Non-ribosomal peptide synthetase engineering focusing on the condensation domain and the condensation/adenylation domain interface

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften (15)

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Janik Kranz

aus Frankfurt am Main

Frankfurt am Main 2022

D30

vom Fachbereich Biowissenschaften (15) der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel

Gutachter: Prof. Dr. Helge B. Bode

Zweitgutachter: Prof. Dr. Martin Grininger

Datum der Disputation:

#IchbinHanna

Table of Contents

Table of abbreviations

4'-PPant	4'-phosphopantetheine
A	adenylation
AA	amino acids
A_{core}	*N*-terminal adenylation core domain
ACN	acetonitrile
ACP(s)	acetyl-carrier-protein(s)
acvA	δ-(*L*-α-aminoadipyl)-cysteinyl-*D*-valine A
Adm	andrimid
AMP	adenosine monophosphate
ANL	acyl-CoA synthetases, NRPS adenylation domains, and luciferase enzymes
A_{sub}	*C*-terminal adenylation subdomain
ATP	adenosine triphosphate
BGC(s)	biosynthetic gene cluster(s)
C	condensation
C_{ASub}	*C*-terminal acceptor condensation subdomain
CAT	chloramphenicol acyltransferase
CDA	calcium-dependent antibiotic synthetase
C_{DSub}	*N*-terminal donor condensation subdomain
CE	dual condensation-epimerisation
CoA	coenzyme A
C_T	terminal condensation
Cy	cyclization
$^{D}C_{L}$	condensation of a *C*-terminal the donor *D*-substrate and acceptor *L*-substrate
CV	column volumes
DhbE	2,3-dihydroxybenzoic acid acitvating bacillibactin synthetase E
DNA	deoxyribonucleic acid
Dpt	daptomycin synthetase
E	epimerization

E2p	dihydrolipoamide acyltransferase
F	formylation
GrsA	gramicidin S synthetase A
GxpS	GameXPeptide synthetase
HPLC/MS	High-performance liquid chromatography - mass spectrometry
HPLC/HRMS	High-performance liquid chromatography – high-resolution mass spectrometry
IP	isoelectric point
KolS	kolossin synthetase
$^{L}C_{L}$	condensation of a *C*-terminal the donor *L*-substrate and acceptor *L*-substrate
LgrA	linear gramicidin synthetase A
mRNA	messenger RNA
MT	methyltransferase
MTSL	S-methyl-thiomethanesulfonate
NP(s)	natural product(s)
NRP(s)	non-ribosomal peptide(s)
NRPS(s)	non-ribosomal peptide synthetase(s)
Ox, Red	redox-active
PCP	peptidyl carrier protein
PDB	precursor-directed biosynthesis
PDB ID	Protein Data Bank identification code
PKS(s)	polyketide synthase(s)
PP_i	pyrophosphate
PPTase	phosphopantheteine transferase
Pvd	pyoverdine
RMSD	root-mean-square deviation
SAM	S-adenosylmethionine
SNACs	*N*-acetylcysteamine thioester
SrfA	surfactin synthetase A
T	thiolation
TE	thioesterase

Tyc	tyrocidine synthetase
WT	wild type
XU	eXchange Unit
XUC	eXchange Unit Condensation domain
SM(s)	specialized metabolite(s)

Summary

Non-ribosomal peptide synthetases (NRPSs) are modular biosynthetic megaenzymes producing many important natural products and refer to a specific set of peptides in bacteria's and fungi's secondary metabolism. With the actual purpose of providing advantages within their respective ecological niche, the bioactivity of the structurally highly diverse products ranges from, e.g., antibiotic (e.g., vancomycin) to immunosuppressive (e.g., cyclosporin A) to cytostatic (e.g., echinomycin or thiocoralin) activity.

An NRPS module consists of at least three core domains that are essential for the incorporation of specific substrates with the 'multiple carrier thiotemplate mechanism' into a growing peptide chain: an adenylation (A) domain selects and activates a cognate amino acid; a thiolation (T) domain shuffles the activated amino acid and the growing peptide chain, which are attached at its post-translationally 4'-phosphopantetheine (4'-PPant) group, between the active sites; a condensation (C) domain links the upstream and downstream substrates. NRPS synthesis is finished with the transfer of the assembled peptide to the *C*-terminal chain-terminating domain. Accordingly, the intermediate is either released by hydrolysis as a linear peptide chain or by an intramolecular nucleophilic attack as a cyclic peptide.

The NRPS's modular character seems to imply straightforward engineering to take advantage of their features but appears to be more challenging. Since the pioneering NRPS engineering approaches focused on the reprogramming and replacement of A domains, several working groups developed advanced methods to perform a complete replacement of subdomains or single or multiple catalytic domains.

The first part of this work focusses parts of the publication with the title '*De novo* design and engineering of non-ribosomal peptide synthetases', which follows up assembly line engineering with the development of a new guideline. Thereby, the pseudodimeric V-shaped structure of the C domain is exploited to separate the *N*-terminal (C_{DSub}) and *C*-terminal (C_{ASub}) subdomains alongside a four-AA-long

linker. This results in the creation of self-contained, catalytically active C_{ASub}-A-T-C_{DSub} (XUC) building blocks. As an advantage over the previous XU concept, the characteristics (substrate- and stereoselectivity) assigned to the C domain subunits are likewise exchanged, and thus, no longer represent a barrier. Furthermore, with the XUC concept, no important interdomain interfaces are disrupted during the catalytic cycle of NRPS, allow to expect much higher production titers. Moreover, the XUC concept shows a more flexible application within its genus origin of building blocks to create peptide libraries. Additionally, with this concept only 80 different XUC building blocks are needed to cover the entire proteinogenic amino acid spectrum.

The second part of this work addresses the influence of the C domain on activity and specificity of A domains. In a comprehensive analysis, a clear influence of different C domains on the *in vitro* activation rate and the *in vivo* substrate spectrum could be observed. Further *in situ* and *in silico* characterizations indicate that these influences are neither the result of the respective A domains promiscuity nor the C domain’s proofreading, but due to an 'extended gatekeeping' function of the C domain. This novel term of an 'extended gatekeeping' function describes the very nature of interfaces that C domains can form with an A domain of interest. Therefore, the C-A interface is assumed to have a more significant contribution to a selectivity filter function.

The third part of this work combines the NRPS engineering with phylogenetic/evolutionary perspectives. At first, the C-A interface could be precisely defined and further identified to encode equivalent information corresponding to the complete C-A didomain. Moreover, the comparison of NRPSs topology reveals hints for a co-evolutionary relatedness of the C-A didomain and could be shown to reassemble even after separation. In this regard, based on a designed CAopt.py algorithm, the reassembling-compatibility of hybrid interfaces could be determined by scoring of the co-expressed NRPS hybrids. This algorithm also enables the randomization of the interface sequences, thus, leading to the identification of more functional interface variant, which cause significantly higher peptide production and could even be applied to other native and hybrid interfaces.

Zusammenfassung

Nicht-Ribosomale Peptidsynthetasen (NRPS) sind modular aufgebaute biosynthetische Megaenzyme, die Peptide mit hoher struktureller und bioaktiver Diversität produzieren. Dabei bilden sie eine spezifische Gruppe von Peptiden im Sekundärstoffwechsel von Bakterien und Pilzen. Mit dem eigentlichen Zweck, Vorteile innerhalb ihrer jeweiligen ökologischen Nische zu ermöglichen, erstreckt sich die Bioaktivität der vielfältigen Produkte von z.B. antibiotischer (z.B. Vancomycin) über immunsuppressiver (z.B. Cyclosporin A) bis hin zu zytostatischer (z.B. Echinomycin oder Thiocoralin) Aktivität. Nicht-ribosomale Peptide (NRPs) oder die von NRPs abgeleiteten Peptide stellen daher eine wichtige Klasse pharmazeutisch relevanter Wirkstoffe dar, die über die letzten Jahrzehnte in der klinischen Anwendung etabliert werden konnten.

Die Synthese von nicht-ribosomalen Peptide ist in einer modularen fließbandähnlichen Logik organisiert, bei dem jedes Modul für den Einbau einer bestimmten Aminosäure verantwortlich ist. Damit ist die Synthese nicht von der mRNA sondern von der Aminosäurespezifität, Anzahl und Organisation der Module anhängig. Trotz determinierter Modulabfolge lassen sich NRPS nach ihrer biosynthetischen Logik in lineare NRPS (Typ A), iterative NRPS (Typ B) und nichtlineare NRPS (Typ C) einordnen. Gemeinsam ist allen jedoch, dass sie den Einbau spezifischer Substrate in eine wachsende Peptidkette durch den „*multiple carrier thiotemplate mechanism*“ katalysieren. Ein NRPS Modul besteht aus mindestens drei Kern-Domänen: Einer Adenylierungs (A) Domäne, die ein passendes Substrat erkennt und aktiviert; einer Thiolierungs (T) Domäne die aktivierten Aminosäuren bzw. die wachsende Peptidkette, die an ihrer posttranslationalen 4'-PPant-Gruppe gebunden ist, zwischen den aktiven Stellen hin und her befördert; und einer Kondensations (C) Domäne, die stromaufwärts und stromabwärts gelegenen Substrate verbindet und eine Kontrollfunktion der Substrat- und Stereoselektivität einnimmt. Im Anschluss an diese Kern-Domänen wird das assemblierte Peptid auf eine endständige Terminationseinheit übertragen und kann entweder durch Hydrolyse als lineare Peptidkette oder durch ein intramolekulares Nukleophil als zyklisches Peptid freigesetzt werden.

Der modulare Charakter der NRPS scheint eine simple Veränderung der Enzyme zu erlauben. Diese Vorteile für das rational Design neuartiger Naturstoffe zu nutzen, hat sich jedoch als eine größere Herausforderung herausgestellt. Seit den ersten NRPS-Engineering Ansätzen, die sich auf die Umprogrammierung und den Untereinheitenaustausch von A Domänen konzentrierten, haben mehrere Arbeitsgruppen fortgeschrittene Methoden entwickelt, um einen vollständigen Austausch von einzelnen Untereinheiten oder Domänen und sogar mehreren katalytischen Domänen durchzuführen.

Der erste Teil beschäftigt sich mit Teilen der Publikation mit dem Titel „*De novo* design and engineering of non-ribosomal peptide synthetases“ und knüpft an die Neukombination von NRPS mit der Entwicklung einer neuartigen NRPS-Engineering Strategie an. Dabei wird die pseudeodimere, V-artige Struktur der C Domäne ausgenutzt, um diese entlang eines vier-Aminosäure-langen Linkers zwischen den *N*-terminalen (C_{DSub}) und *C*-terminalen (C_{ASub}) Untereinheit zu teilen. Dadurch entstehen, entsprechend der Schnittstellenposition in der C Domäne, s.g. *eXchange Unit Condensation domain* (XUC) Bausteine. XUC Bausteine sind katalytisch eigenständige C_{ASub}-A-T-C_{DSub} Einheiten, die die Eigenschaften der flankierenden C Domänen Untereinheiten mit austauschen. Dadurch wird ein flexibles System geschaffen, das die C Domänen Spezifitäten integriert und die Zahl der NRPS-Bausteine, die zur Herstellung oder Veränderung bestimmter Peptide erforderlich sind, drastisch reduziert. Es sind mit dem XUC Konzept nur 80 verschiedene XUC Bausteine (unter Berücksichtigung der *N*- und *C*-terminaler C- oder dualer C/E Domänen) benötigt, um das gesamte Spektrum aller 20 proteinogene Aminosäuren abzudecken.

Ein großer Vorteil dieses Konzepts ist, dass keine Interdomänengrenzflächen verändert werden, was höhere Produktionstiter der generierten Hybride erwarten lässt. Dies ermöglicht die effiziente Herstellung von Peptiden, die auch nicht natürliche Aminosäuren enthalten, mit Ausbeuten von bis zu 280 mg/l. In diesem Zusammenhang konnte gezeigt werden, dass sich sogar die relaxierte Substratspezifität des GameXPeptide (GxpS) Moduls 3 als XUC Baustein auf das

Xenotetrapeptid (XtpS) übertragen lassen. Somit konnten durch Fütterung unnatürlicher Aminosäuren wie *O*-Propargyl-*L*-Tyr, *p*-Bromo-*L*-Phe oder *p*-Azido-*L*-Phe eine Produktion der jeweiligen XtpS Derivate beobachtet werden. Weiterhin ermöglicht die einmalige, auf Hefe basierende transformationsassoziierte Rekombination (TAR) die Erstellung umfassende Peptid Peptidbibliotheken.

Der zweite Teil dieser Arbeit beleuchtet die Rolle der Kondensations (C) Domänen in der NRP Synthese näher. Dabei wurde untersucht, ob C Domänen, die für die stereochemische Filterung und Verknüpfung kovalent gebundener Substrate verantwortlich sind, auch eine intrinsische Substratspezifität besitzen.

In einer Reihe von *multiplexed hydroxamate assay* (HAMA) *in vitro* Versuchen konnte ein deutlicher Einfluss der C Domänen und ihrer Untereinheiten auf die Aktivität der bearbeiteten GxpS_A3 Domäne nachgewiesen werden. In Kombination von *in vivo* Versuchsreihen zeigten sich, von den mitunter erstellten hybriden Interfaces, eine erweiterte Kontrollfunktion C Domänen auf die Feinabstimmung der Aminosäureinkorporation. C Domänen, die dabei der natürlichen C Domäne homolog waren, wiesen ähnliche Fähigkeiten auf die entsprechenden Substrate zu erkennen und zu aktivieren, während fremdartigere C Domänen deutliche Einbußen oder Anstiege der Produktionstiter zur Folge hatten.

Der Einfluss der ausgebildeten C-A Grenzfläche konnte am Beispiel des Fabclavins (Fcl) durch *in situ* Substitutionen weiter verdeutlicht werden. Im Vergleich der sehr homologen Fabclavine BGCs aus *Xenorhabdus budapestensis* (*Xbud*), *Xenorhabdus hominickii* (*Xhom*), und *Xenorhabdus szentirmaii* (*Xsze*), zeigt die A Domäne im sechsten NRPS-Teil von *X. szentirmaii* als Einzige der Drei eine relaxierte Substratspezifität neben Pro, auch für Thr und Val. Durch das Erstellen von sechs chimären NRPS, welche alle möglichen C-A Kombinationen aus der *Xbud*, *Xhom* und *Xsze* an dieser BGC Position repräsentieren, konnte die Promiskuität als ein Effekt veränderter C-A Wechselwirkungen nachgewiesen werden. Diese Eigenschaft konnte ebenfalls durch die Substitution dieser *Xsze*

C Domäne auf die A Domänen auf *Xbud*, und *Xhom* übertragen werden, die anschließend in der Lage waren neben Pro, auch Thr und Val zu akzeptieren.

In einem *in silico* Ansatz war es möglich die erweiterte Kontrollfunktion von C Domänen, zumindest in erster Näherung, anhand von Homologiemodellen zu charakterisieren. Dadurch konnte die erweiterte Kontrollfunktion als die Eigenschaft der C-A Grenzflächeninteraktion in den jeweils katalytischen Phasen der NRPS Biosynthese manifestiert werden. Die Kontrollfunktion der C Domänen ist somit nicht auf eine die der C_{ASub} zugeschriebenen intrinsische Spezifität zurückzuführen, sondern wurde als die Weise, mit der die C Domäne eine die Grenzfläche mit einer A Domäne bildet, definiert.

Der dritte Teil dieser Arbeit befasst sich mit der Grenzfläche zwischen der C und A Domäne und ihre gezielte Veränderung. Die Definition der C-A Grenzfläche umfasst alle Aminosäuren, die sich in unmittelbarer räumlicher Nähe zur Kontaktfläche, zwischen der C und der A Domäne befinden oder eine ausgeprägte Seitenkettenorientierung zu dieser aufweisen.

Im Zuge dessen wurde die C-A Grenzfläche für alle bis dato bekannten Kristallstrukturen bestimmt. Diese dienten als Vorlage um alignmentbasierend die Grenzflächen von C-A Didomänen unbekannter Struktur vorherzusagen. Eine umfassende phylogenetische Analyse zeigte, dass die C-A Grenzfläche allein die gleichen Informationen über Gattungsherkunft, die Subtypen der C Domäne und die Spezifität der A Domäne kodiert, wie die komplette C-A Didomäne. Weiterhin konnten in einer Topologieanalyse etablierter C-A-T Phylogenien wichtige Hinweise auf eine Coevolution der C-A Didomäne durch die geringe topologische Distanz der C-A grenzflächenbildenden Bereiche gegeben werden. Diese topologische Beziehung erwies sich in Co-Expressionsexperimenten von verschiedenen C/A-getrennten NRPS Modulen als ausreichende Affinität, um sich auch ohne kovalente Verbindung wieder zu funktionalen NRPS zu assemblieren. Diese Passfähigkeit hybrider C-A Didomänen konnte weiterhin mit dem maßgeschneiderten *CAopt.py*-Algorithmus erfasst werden, der über die physikochemischen

Eigenschaften der C-A Grenzfläche eine Konvergenz durch ein Scoring-System bewertet und vorhersagt. Dabei konnte eine Korrelation von C-A Grenzflächen mit gutem Scoring und funktionalen NRPS mit guten Produktionsraten identifiziert werden.

Eine in den *CAopt.py*-Algorithmus integrierte Randomisierungsfunktion ermöglicht darüber hinaus die systematische Änderung der grenzflächenbildenden Reste, mit dem unter Anderem, Möglichkeiten für systematische Änderungen der C-A Grenzfläche ausgelotet werden können. Auf diese Weise konnte eine C-A Grenzflächenvariante identifiziert werden, die eine signifikant höhere Produktionsrate der hybriden NRPS zu Folge hatte. Weiterhin konnte diese C-A Grenzflächenvariante auf andere hybride NRPS übertragen werden, was ebenfalls zur Produktionssteigerung führte. Es zeigt sich ein großes Potential in dem *CAopt.py*-Algorithmus zur Identifizierung „besserer“ oder „schlechterer“ Domäneninteraktionen, und um weitere einzigartige C-A Grenzflächenvarianten aufzudecken.

1. Introduction

More than 30,000 distinct natural products (NPs) have been isolated to date. Since 1981, 20 % of these have been approved as new drugs and account for 50 % of the antibacterial drugs in use[1,2]. The most prominent natural products are antibiotics[3]. More than 16,800 have been isolated to date[3], and more than 100 are in clinical use[1], with 43 new antibiotics recently in clinical development[4].

Bioactive substances from nature have been used by humans for thousands of years. The 'Ebers Papyrus' from 1550 BC from ancient Egypt is the oldest medical document describing a variety of applications for plants and even the exact dosage[5]. Moreover, many other ancient records document the natural substance applications in antique Serbia, China, Greece, and Egypt of, e.g., heart-active foxglove glycosides or alkaloid toxin coniin extracted from hemlock plants[2,5]. However, the development of modern medicinal treatment is widely paved by pioneers such as Paul Ehrlich, who discovered 1909 salvarsan to treat the causative agent *Treponema pallidum* of syphilis[6], and Alexander Fleming, who discovered 1928 penicillin from *Penicillium notatum* with antibacterial properties to treat so far life-threatening infections[7].

The subsequent identification of microorganisms as prolific producers of natural products (NPs) or specialized metabolites (SMs)[8] was crucial for the beginning of the 'Golden Age of Antibiotic Discovery'[9]. SMs form a diverse chemical interaction network within microorganisms and create access for chemical properties beyond the Lipinski’s rule of five[10,11]. The biosynthesis of such molecules is usually characteristic for a species. These molecules serve a role in adapting their environment to have advantages over conspecifics or within their respective ecological niche[12].

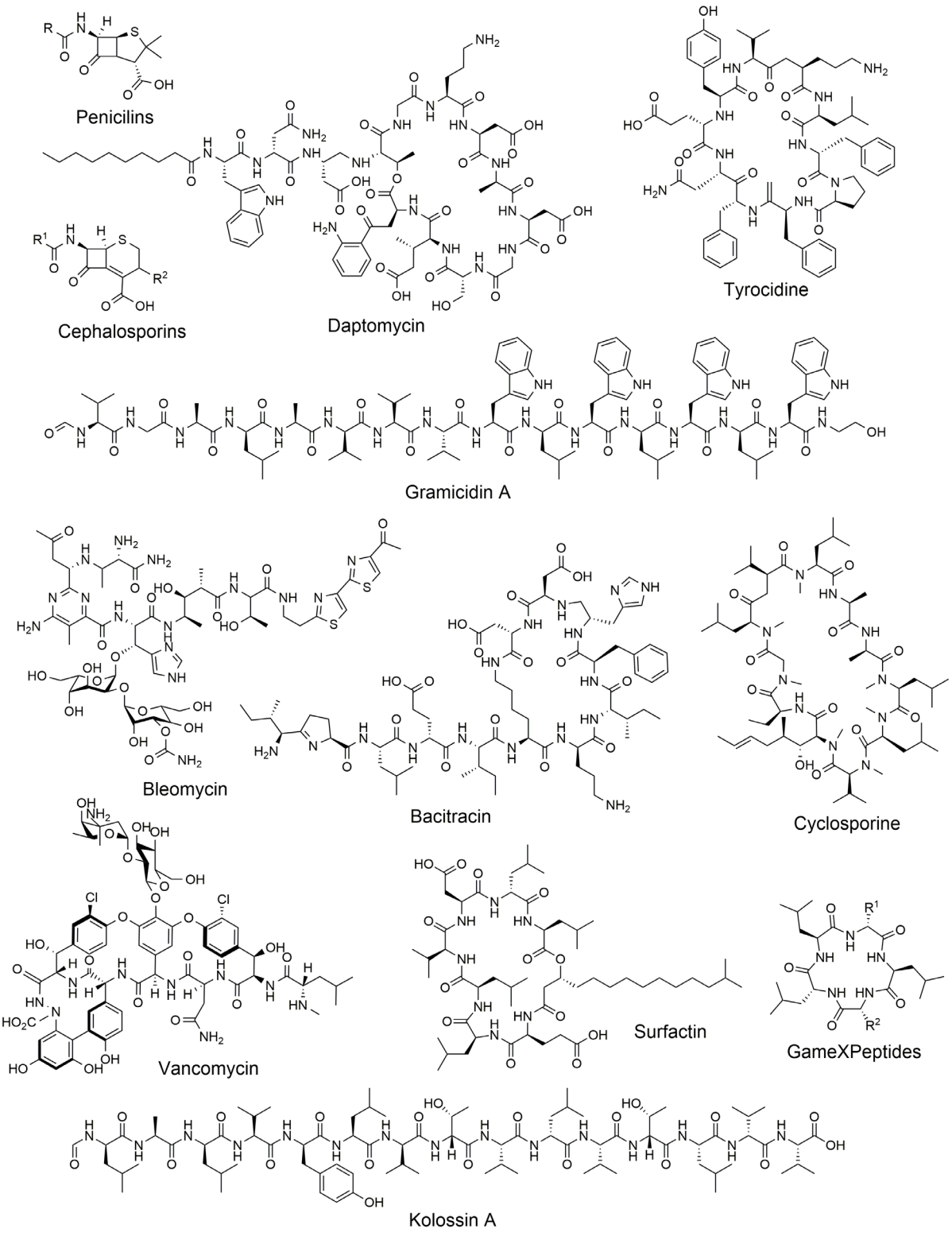


Figure 1: Selection of representative NRP structures.

NPs are often valued for their great variety of functionalities[13]. The established NP main actors of that era like beta-lactam penicillins and cephalosporins[14], macrocyclic tyrocidines and gramicidins[15], or tridecapeptide daptomycins[16] were also identified to have peculiarities of their biosynthesis in common to achieve this diversity. They are all produced by multifunctional enzymes, namely: non-ribosomal peptide synthetases (NRPSs)[9].

NRPSs are a member of the megasynthetase family, which is widespread in various species of archaea, fungi, and bacteria[17]. NRPSs are organized in multiple modules comprising a set of domains which catalyse the incorporation and programmed chemical modifications of an extender unit into the growing peptide chain[18]. This assembly-like concept dependent on its function-imparting modules is, unlike the ribosomal peptides, not limited to the 20 proteinogenic amino acids (AA) and includes a repertoire of over 500 different *L*-amino acid, *D*-amino acid, carboxylic acid, fatty acid, hydroxy acid, and keto acid building blocks[19]. The thereby attained high degree of structural diversity have become a promising scaffold for investigating new drug leads[20]. Thus, NRPS could be key to successfully address the lack of newly developed leads and market-ready drugs to effectively challenge the emerging multidrug resistances[5].

1.1. Historical review of NRPS

The story of NRPS starts in the 1940s with the isolation of tyrocidine from cultures of *Bacillus brevis*[15] without being aware of its non-ribosomal biosynthesis. Only when Mach *et al.* in the 1960s distinguished the biosynthesis of tyrocidine from the biosynthesis of other proteins, they concluded a ribosomal-independent biosynthesis of peptides[21]. Further, the resistance to ribosomal inhibitors like chloramphenicol, puromycine or the inhibition of RNA to the synthesis of gramicidine S[22], edeine[23], tyrocidine[24] and polymyxin[25] lead to the assumption of ribosomal-independence of this mechanism. To describe this new mechanism, scientists introduced the term 'sluggish'-ribosomes, because the substantial biosynthesis of their products were observed only at later exponential growth phases[26].

Berg *et al.* indicated the requirement of ATP or phosphoenolpyruvate and pyruvate kinase as an energy source in their biosynthesis[22]. These findings were complemented by *in vitro* assays showing an exchange between ATP and AMP+PP_i[27] and the dependence of magnesium[28]. In addition, the amino acids were reversibly bound as aminoacyl adenylates with their acyl group to an enzyme-bound 4'-phosphopantetheine (4'-PPant). Leading to the assumption of the presence of a 'carrier-protein' and an unique arrangement of the active sites of the participating enzymes[27–29].

The progress in synthesis was shown by using ^{14}C labelling experiments and thin-layer chromatography. Amino acids were step by step incorporated into a growing peptide chain, linked to the enzymes by acid-stable bonds[29–31]. Gilhuus-Moe *et al.* assumed only one molecule of 4'-phosphopantetheinly attached to the larger enzymes, a view which turned out to be wrong[30]. They thought the central thiol-carrying arm charged with a growing peptide chain would be passed by transthiolation in the reaction cycle – postulated as the 'thiotemplate mechanism'[32]. However, Lipmann and co-workers shaped our understanding as pioneers of NRPS until today. They first associated a modular character of the multifunctional enzymes based on the protein’s size and the number of produced peptides[33,34]. Additionally, they postulated a two-step reaction for the activation of amino acids: The

magnesium-dependent charge by ATP with the enzyme corresponding amino acids and the thioesterification of amino acids to the enzymes[34].

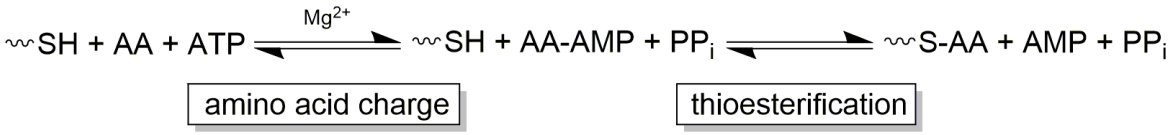


Figure 2: Scheme of NRPS amino acid activation. Adopted from[34].

Up to this point, many NRPs were known, and their biosynthesis became increasingly well described[35]. DNA sequencing became more experimentally accessible and enabled the sequencing of biosynthetic gene clusters as tyrocidine synthetases 1&2 and gramicidin S, leading to the confirmation of the modularity of NRPS[34,36]. Additionally, the assumption that there is only one 4'-PPant cofactor per NRPS got replaced by the 'multiple carrier thiotemplate mechanism'[32,37].

Pioneering work in the 1990s significantly shaped the image of NRPSs: NRPSs are arranged in modules being in charge for the incorporation of an amino acid and can be further subdivided in domains. A minimal module contains: an adenylation domain that recognizes and activates the corresponding amino acid; a peptidyl carrier protein domain with 4'-PPant cofactor at a conserved serine attached for transfer activated amino acids; and a condensation domain between two consecutive modules catalyse the peptide bond formation[38].

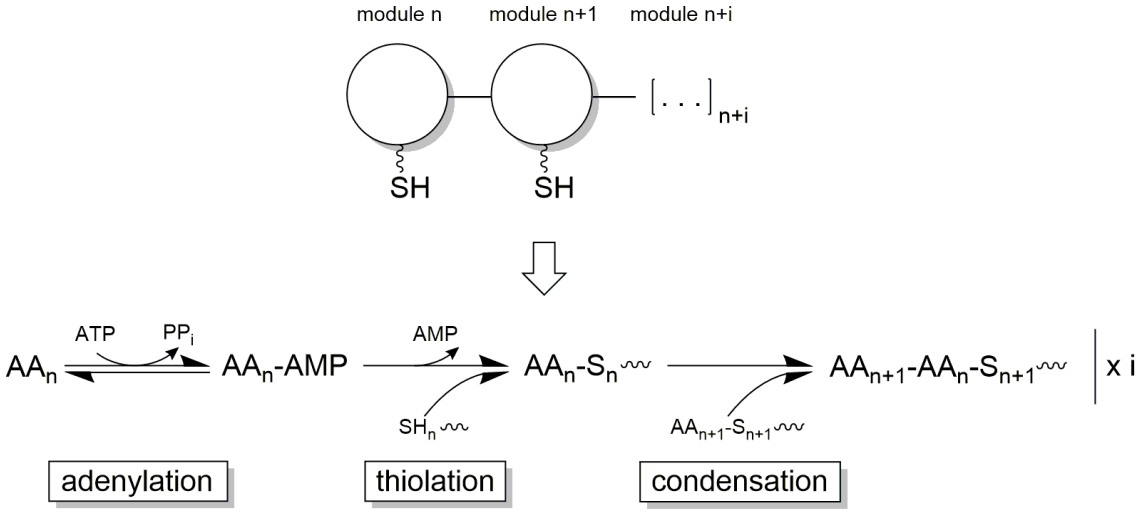


Figure 3: Simplified scheme of thiotemplate-directed NRPS synthesis. Representation of the NRPS module organisation, whereby a module (symbolized by a circle) at position n catalyses the adenylation, thiolation and condensation reaction for the respective amino acid incorporation into the growing peptide chain of length i. Adapted from[39].

Still, after further progress in characterization and future advanced engineering experiments, this model of NRPS has established the foundation for the fundamental NRPS conception as described in the following sections.

1.2. Architecture of non-ribosomal peptide synthetases

NRPSs are large, multifunctional (mega)enzymes in which multiple, repeating modules of enzymatic domains catalyse the incorporation of selected extender units into the growing peptide chain[40]. The corresponding higher-order architecture in NRPSs is thus capable to organize the biosynthesis of highly complex peptide structures due to the underlying 'multiple carrier thiotemplate mechanism'[39] (see Chapter 1.1). Therefore, a set of domains, termed as module, dictates the processing of the thioester-linked building blocks and, as such, determine the sequence of growing NRPs through the module order of NRPS[41]. In consequence, the organization of NRPS into modules, each responsible for the incorporation of one amino acid, makes the structure of the synthesized peptides dependent not on the mRNA but the amino acid specificity, number, and organization of the modules[41].

Three core domains are essential for maintaining the catalytic activity of a module. A specific amino acid is recognized and activated by the adenylation (A) domain, transferred to the post-translationally modified thiolation (T) domain’s 4'-PPant group, and subsequently after shuffling the activated amino acids, the formation of a peptide bond between two T domain bound amino acids is catalysed by the condensation (C) domain[20]. The obtained structural diversity is extended by further editing domains, such as the configuration by epimerization (E) domains, *C*- or *N*-methylation by methyltransferase (MT) domains, heterocyclization through cyclization (CY) domains, or the redox state through redox-active (Ox, Red) domains[42]. Furthermore, the release of the matured peptide is often catalysed by *C*-terminally present thioesterase (TE) (commonly in bacteria NRPS) or terminal C (C_T) (commonly in fungi NRPS) domains, which is also able to associate different structural classes through cleavage, e.g., as a linear product or as head-to-tail-cyclized/side-chain-to-tail peptides of various ring sizes[40].

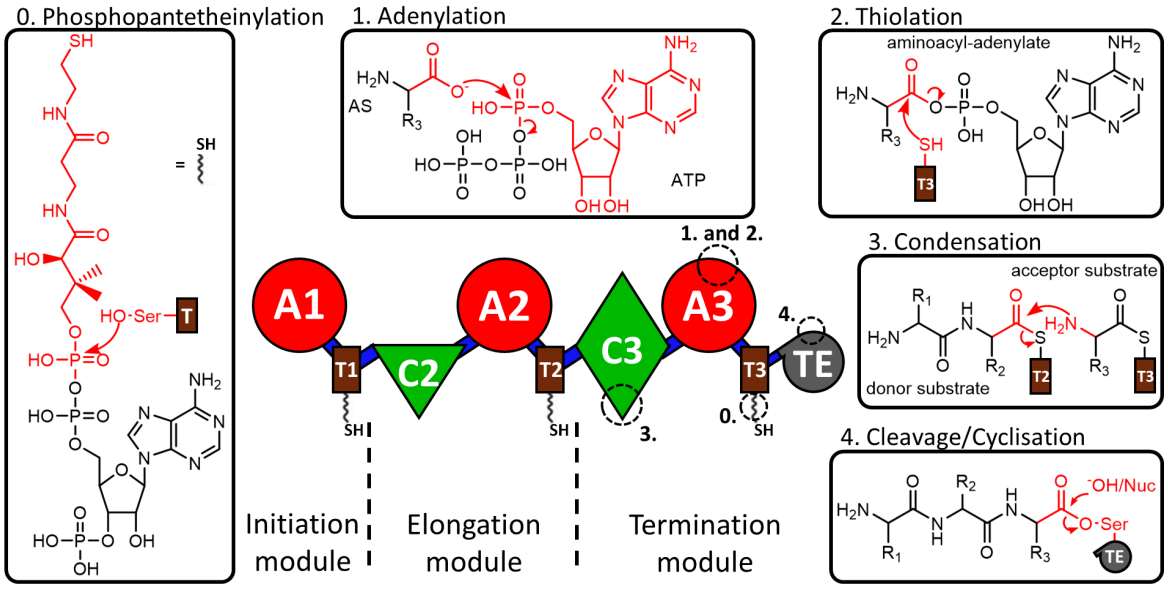


Figure 4: Architecture and the catalysed reactions of NRPS domains predominantly found in bacteria. Schematic representation of a fictive NRPS consisting of an initiation, elongation, and termination module. The domains are illustrated by the following symbols: adenylation (A) domain, large red circle; thiolation (T) domain, small brown rectangle; condensation (C) domain, green triangle; dual condensation/epimerization (C/E) domain, green diamond; thioesterase (TE) domain, small grey circle; and interdomain linkers, bold blue lines. The posttranslational attached 4'-PPant arm at the T domain is represented by a curled line and its terminal thiol group. The catalysed reactions are highlighted with dashed circles at its proper location at the respective domain and the reaction mechanism is depicted in the corresponding boxes (0.-4.). Adapted from[40].

Distinguishing the assembly-line logic of each NRPS, they can be categorized into three groups[43]: (1) In linear (type A) NRPS, each NRPS module incorporates one specific building block, and therefore, the number of modules in linear NRPS is directly dependent on the number of amino acids in their products[40]. The variation of type A NRPS ranges from, e.g., GameXPeptide synthetase (GxpS)[44,45] producing cyclic pentapeptides to tyrocidine synthetase (Tyc)[46], distributed on three separate proteins TycABC, producing cyclic decamers to kolossin synthetase (KolS)[47] producing linear pentadecamers; (2) Iterative (type B) NRPSs use certain modules multiple times during the completion of a catalytic cycle, which can often lead to the reoccurrence of congruent structural elements of the products. The repeated use of modules requires a so-called 'holding position' whereby the stalled reaction intermediates undergo 2-5 elongation cycles before further processing[40]. In congocidine[48] the iteration is achieved by storage on a loading T domain, whereas in enniatin[49] the intermediates are stored on a terminal T domain, and in

gramicidin S (GrsA)[50] on a TE domain; (3) In contrast to type B, the nonlinear (type C) NRPSs use only one domain multiple times in a catalytic cycle[40]. A feature of Type C NRPS is, therefore, the decomposition of the assembly lines, which e.g., results in a repetition of the enduracididine residue in mannopeptimycin[51] or yields the terminal bisthiazole fragment of bleomycin[52].

The highly versatile architecture of the NRPS draws a complex picture of the synthesis of natural products by megasynthetases. However, the individual biosynthetic steps result from the domain-inherent activity. These main catalytically active core domains will be discussed in detail in the next sections.

1.2.1. Adenylation (A) domain

The primary action in non-ribosomal peptide synthesis is the selection and activation of a cognate amino acid[53]. These reactions are carried out by adenylation domains within NRPS. Adenylation domains (~60 kDa) belong to the ANL (**A**cyl-CoA synthetases, **N**RPS adenylation domains, and **L**uciferase enzymes) superfamily and exhibit a characteristic adenylate (AMP) binding motif for adenylate-forming enzymes[20,40]. Several crystal structures of members of this family were solved and reveal an almost identical fold[54,55,56]. A domain consists of a large ~50 kDa *N*-terminal core domain (A_{core}) and the small ~10 kDa *C*-terminal subdomain (A_{sub}) coupled by a flexible linker region of about five amino acids[54]. The active site is located close to the A_{core}-A_{sub} interface within the A_{core} domain. Therein, ten core motifs (A1-A10) serve as structural anchors to determine the substrate specificity and impart the 'gatekeeper' function[53].

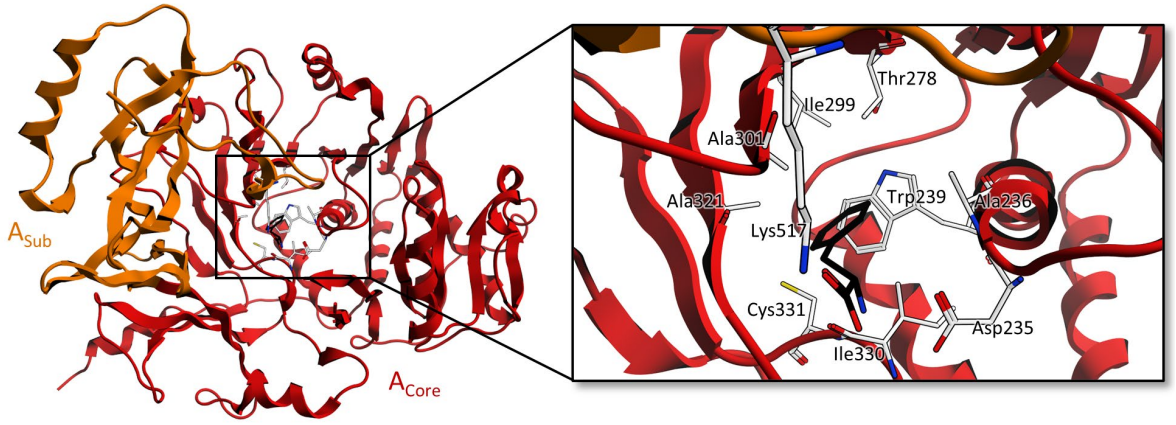


Figure 5: Structure of an A domain and its binding pocket. Ribbon representation of an A domain illustrated by the PheA (PDB ID: 1AMU)[54]. The larger *N*-terminal A_{Core} (red) and smaller *C*-terminal A_{Sub} (orange) domain are highlighted. The binding pocket's core residues (white) and bound substrate (black) are depicted in stick representation.

The positioning of α-amino acids in the active site is accomplished by stabilizing the amino group via the aspartic acid residue 1 (A1) and the carboxy group via the lysine residue 10 (A10)[40]. The A10 residue is essential for the adenylation reaction and is therefore strictly conserved. The A1 residue is also conserved in α-amino acid activating A domains. However, it can also be located in different positions in the binding pocket in other A domains in order to stabilize β-amino acids, α-hydroxy acids, or α-keto acids[40,57]. In addition, there are eight more amino acid (A2-A9) in

the binding pocket, which are responsible for recognizing the correct side chain of the substrate and can be divided into two classes. The highly variable amino acid A3, A4, A5, A7, and A9 show high flexibility between different A domains and exert the most significant influence on the substrate specificity. The moderately variable amino acid A2, A6, and A8 vary only slightly between different binding pockets and mostly have aliphatic side chains. It is assumed that the moderately variable amino acids are responsible for the A domain's catalytic activity and for fine-tuning the substrate specificity[40].

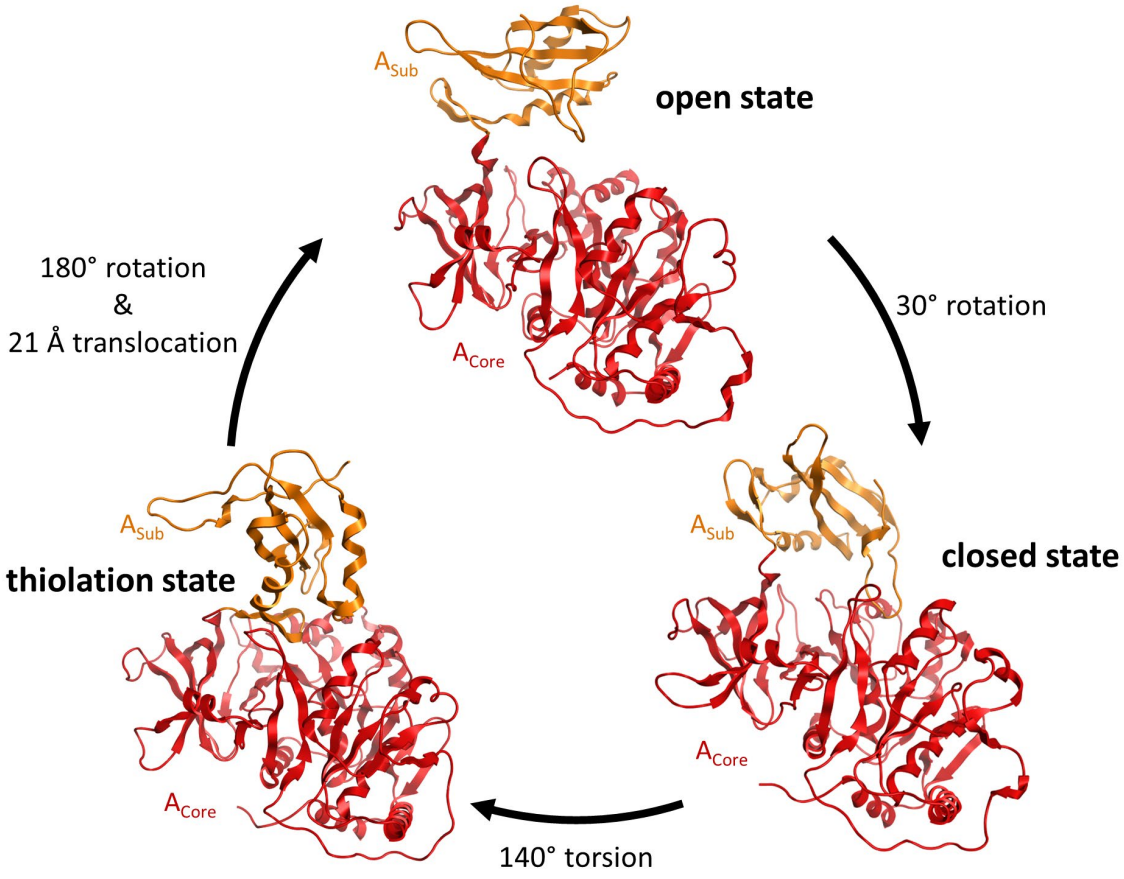


Figure 6: Conformational states of the A domain during catalytic cycle of NRPS. Ribbon representation of an A domain illustrated by the A domain excised from SrfA-C (PDB ID: 2SVQ)[58] in the open state, LgrA; (PDB ID: 6MFX)[59] in the closed state, and EntF (PDB ID: 5T3D)[60] in the thiolation state. The larger *N*-terminal A_{Core} (red) and smaller *C*-terminal A_{Sub} (orange) domain are highlighted. The A_{Sub} rearrangement is specified in respect to the A_{core}.

The activation of the amino acid requires two catalytical steps and an extensive domain rearrangement[60]. An open A domain state – where the A_{sub} is bent away from the A_{core} – promoting amino acid and Mg^{2+} ATP entry[58]. On substrate binding,

the A domain closes, rotating the A_{sub} by ~30° with respect to the A_{core}, to catalyse adenylation reaction[61]. With the formation of the amino acyl adenylate and the release of pyrophosphate (PP_i) the A_{sub} performs a rigid-body torsion of ~140° into the thiolation state[57,60]. Hence, the high-energy intermediate, and therefore prone to non-productive hydrolysis, is protected from bulk solvent exposure. This reaction and transition into thioester-forming conformation triggers the rearrangement of the *holo*-T domain and enables the 4'-PPant arm to reach the aminoacyl adenylate intermediate in the A_{core} binding pocket[40]. Likewise, within the binding pocket, the A domain catalyses the formation of the aminoacyl-thioesters between the carboxylic acid residue of the aminoacyl adenylate and the free thiol group of the 4'-PPant arm[40]. Subsequently, the A_{sub} domain makes a full 180° rotation and 21 Å translocation (as an extension of the linker[62]); accordingly, the aminoacylated *holo*-T domain undergoes a ~75° rotation and 61 Å translocation, allowing the substrate to bridge the 50 Å span between the active sites of the A domain and the adjacent catalytical domain[61]. The A domain is now back in the open state and can activate a second amino acid to prime the system for another cycle[60].

1.2.2. Thiolation (T) domain / Peptidyl Carrier Protein (PCP)

The thiolation domain mainly facilitates the extensive flexibility of NRPS. Essential for the catalytical activity is the transfer from the *apo*-form into the active *holo*-form by posttranslational modification by a phosphopantetheine transferase (PPTase) with an ~20 Å long 4'-PPant cofactor[19].

T domains (10 kDa) consist of a four-helix (α1-4) bundle. The first and second helices are distorted, harbouring the cofactor binding site covalently attached to a conserved serine, at the *N*-terminus of α2[63]. Helices two and three are connected by a short loop in antiparallel orientation to helix two. All helices have an amphiphile property forming the hydrophobic core in the inner molecule[63].

An essential motif is further described for the equivalent polyketide synthases (PKSs) acetyl-carrier-proteins (ACPs) - the α2 with its extensive loop to α1, has been called the 'universal recognition helix'[64]. The highly conserved serine located therein is the anchor point for the domain-activating posttranslational phosphopantetheine modification by a PPTase[65]. This essential motifs serine, also prevalent in NRPS PCPs, catalyses the nucleophilic attack of its beta-hydroxy side chain on the pyrophosphate bond of CoA, resulting in the transfer of the 4'-PPant moiety to the attacking serine[39].

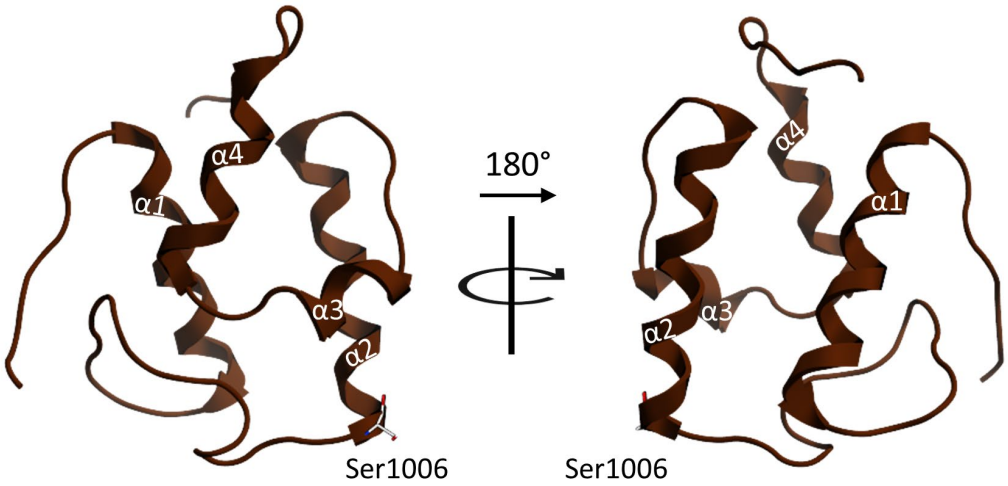


Figure 7: Structure of a T domain. Ribbon representation of a T domain illustrated by the T domain excised from AB3404 (PDB ID: 4ZXH)[60]. Helices (α1-4) of the four-helix bundle are labelled and the conserved serine 4'-PPant binding site is highlighted.

The newly introduced thiol group of the 4'-PPant now acts as a nucleophile for acylation by a substrate[65]. S-methyl-thiomethanesulfonate (MTSL) spin label experiments showed no particular interactions of the PPant-arm and the T domain[66]. Behaving like a swinging arm, the covalent bound 4'-PPant facilitates the pass of substrates to the adjacent reaction centres[67]. However, the helices α2, α3 and their connection loop exhibit a distinct (mainly hydrophobic) interaction with the A, C and TE domain[67,68] and likewise allow a structural mechanism to guide the PCP between active sites[60].

Hitherto, extensive changes in the tertiary structure of PCP depending on its loading state were assumed to influence and regulate PCP interactions – based on the findings of the tyrocidine synthetase PCP domain to maintain in three different conformations (*apo* (A), the A/H, and the *holo* (H) state)[19]. However, this turned out to be a crystallographic artefact, as all other PCP structures studied to date have the A/H conformation[60,64,66,69,70]. The conformational change from A to H state requires a 75° rotation and 61 Å translocation - illustrating the remarkable flexibility of the PCP domain[61].

1.2.3. Condensation (C) domain

The key coupling reaction of the NRPS/PCP-bound upstream amino acyl to the NRPS/PCP-bound intramodule α-amino group is catalysed by condensation domains[40].

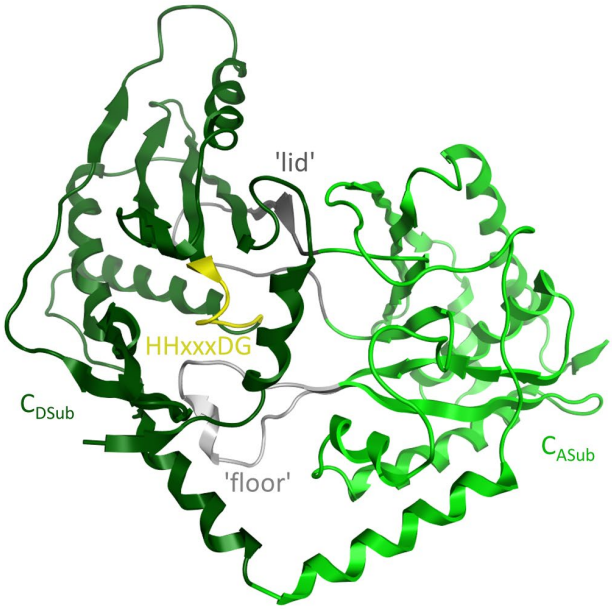


Figure 8: Structure of a C Domain. Ribbon representation of a C domain illustrated by the C domain excised from EntF (PDB ID: 5T3D)[60]. The C domain's *N*-terminal donor C_{DSub} (dark green) and *C*-terminal acceptor C_{ASub} (light green) subunit, active site motif (yellow), 'lid' region (dark grey), and 'floor' loop (light grey) are highlighted.

C domains (~50 kDa) possess a pseudodimeric V-shaped structure, containing an *N*-terminal donor (C_{DSub}) and *C*-terminal acceptor (C_{ASub}) subdomain[71]. The C_{DSub} binds the growing peptide with its electrophilic acyl group, and the C_{ASub} binds the activated amino acid with its nucleophilic α-amino group to enable their subsequent condensation[72,73].

Furthermore, C domains are specific in their stereoselectivity to condense two *L*-amino acids as $^{L}C_{L}$ domain or an upstream *D*- and a downstream *L*-amino acid as $^{D}C_{L}$ domain[40,74]. Like the structurally related chloramphenicol acyltransferase (CAT) and dihydrolipoamide acyltransferase (E2p) of pyruvate dehydrogenase the two subdomains form a cleft at their interfaces accommodating the active site motif HHxxxDG[40,71,75]. The substrate channel in the cleft is covered by a flexible 'lid' region on top and a hydrophobic 'floor loop' covering the active site at the bottom[76].

Access of the PPant arm into this substrate channel appears to be controlled by the entrance located side chains[71,76]. For $^{L}C_{L}$, a conserved Arg is assumed to prevent the undesired 'passing' of unelongated donor substrates in favour of the aminoacyl PPant. However, at this position in $^{D}C_{L}$ domains, there are predominantly small sidechain residues found, believed to permit the unmodified PPant access into the C domain's active site[77].

The second histidine of the active site motive was proposed to be essential for catalysis. Point mutations of the second histidine abolished the peptide cyclization or thioester hydrolysis, displaying its major contribution to catalysis[76]. This histidine residue is seen to act as the general base, mediating the nucleophilic attack of the downstream PCP-bound aminoacyl adenylate α-amino group to the upstream PCP-bound growing peptide chains acyl group[78]. To enable the role as a catalytic base of the second histidine, a reorientation of subdomains hinge region or the closure of the active site by the bridging lid region is deemed to be necessary[76]. Likewise, the 'floor loop' implies its direct involvement in peptide bond formation due to sequence conservation and the formation of hydrogen bonds to the HHxxxDG aspartate residue[71].

The catalytical influence of the histidine residue varies particularly for different C domains, highlighting the persistent debate about the exact enzymatic mechanism of C domains[40,71,79,80].

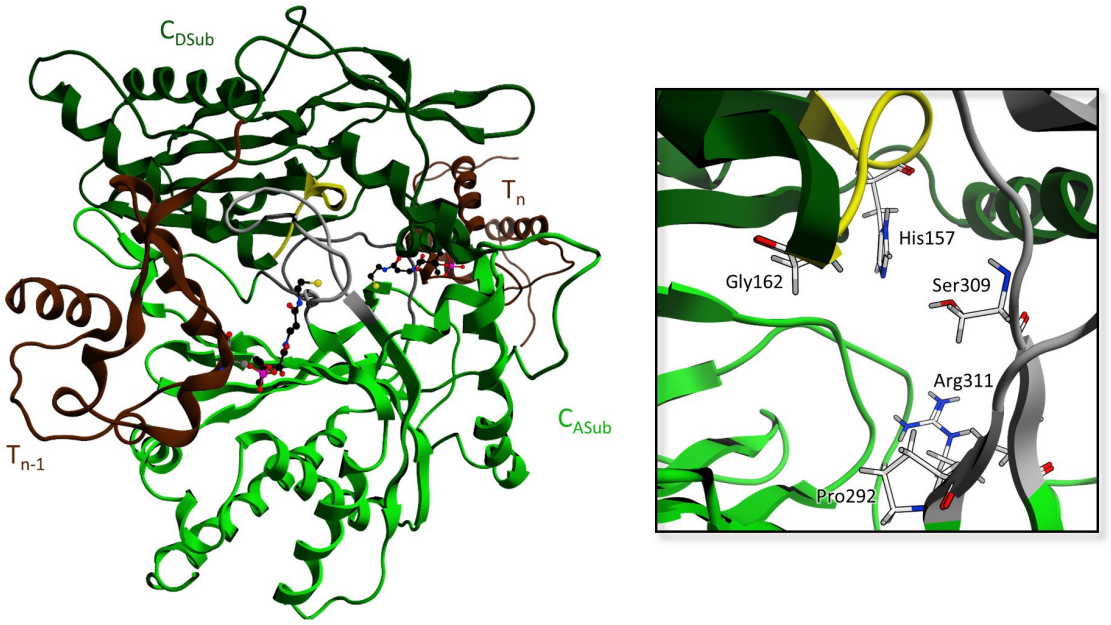


Figure 9: Structure of a C domain and (up-/downstream) binding T domains. Ribbon representation of a C domain-binding illustrated by the manually arranged PCP-E GrsA (PDB ID: 5ISX)[80] and PCP AB3404 (PDB ID: 4ZXH)[60]. The binding pocket is illustrated by the RzmA-Cs (PDB ID: 5DU9)[81] and binding pockets' core residues (white) are depicted in stick representation. The C domain's colouring is according to Fig. 8. The T domains are marked in brown.

Bloudoff *et al.* proposed an alternative mechanism based on the crystal structures of the calcium-dependent antibiotic synthetase CDA-C1 mutants[82], and this view gained support with the solved starter C domain RzmA (rhizomide biosynthetic enzyme) structures by Zhong *et al.*[81]. These structures revealed the HHxxxDG second histidine's protonated δ-nitrogen to interacts with neighbouring backbone carbonyls, enabling its deprotonated ε-nitrogen to accept hydrogen bonds from the α-amino nucleophile of the CoA portion[81,82]. Thereupon, the ε-nitrogen is also in close proximity to the acyl group of the donor substrate, which is presumed to have sufficient space made by the short-chain glycine of this site for proper binding and positioning[81]. They concluded a positioning role of the substrate not only for the acceptor substrate[82] but also for the donor substrate[81] of the second histidine and solvent-mediated proton extraction[83] as the catalytic mechanism of the condensation reaction.

The solved RzmA structures show the active site's compression in the cleft between the subdomains mainly determined by the bridging-lid and the floor-loop. Upon the condensation reaction, the *C*- and *N*-lobe of the condensation domain first open

slightly and then widely, while the floor-loop, under dramatic conformational changes, tosses the substrate out[81].

Moreover, the substrates in the growing peptide chain must undergo a second proofreading step at the C domain, which reduces the error rate of an NRPS[40,73,84–86]. Thus, C domains have been attributed to co-determine substrate specificity of NRPS through a gatekeeping function. These observations based on *in vitro* studies that observed a bypass of editing functions of adenylation domains[87] and numerous unsuccessful engineering efforts[42,88]. However, a recently initiated debate strongly question the C domains' gatekeeping role for their condensed substrates like the A domains do[89]. The underlying data demonstrate that the generation of new NRPS by substituting only the A domain is possible[90], and numerous productive recombination events were observed without the accompanying exchange of C domains[89]. Therefore, the so far assumed dogma of the C domains' second proofreading is under strong criticism[89]. However, the possibility is conceded that other subtypes of C domains exist[90], which possess a stricter acceptor side specificity or that C domains apply substrate filtering regardless of an definite intrinsic specificity[91,92].

Curiously, the C domains' scaffold, core motifs, and protein surface show up in nature in structural homologues with diverse NRPS relevant functions[73]. These are embodied, for example, by epimerization[93], cyclization[94], β-lactam formation or recruitment of auxiliary enzymes[66].

1.2.4. Peptide Release

The NRPS synthesis is finished with the transfer of the assembled peptide to the *C*-terminal chain-terminating domain. The basic task of this termination module is to reactivate the multienzyme by off-loading the tethered linear peptide molecule. Depending on the type of terminal domains, they make use of different chain cleavage mechanisms promoting structural features like macrolactones or macrolactams[95], macrocyclizations[96], aldehydes or alcohols[97], Dieckmann-type cyclizations, amides[98], and *C*-terminal amino sugar or polyamine attachments[99] to the matured peptide due to their mode of action[20]. In bacterial NRPSs, it is most commonly a TE domain[100], which will be highlighted in the next section.

1.2.4.1. Thioesterase (TE) domain

Most commonly the NRP cleavage is mediated by thioesterase (TE) domains. They are named after their homologous fatty acyl thioesterases first discovered in PKS and then characterized in the hydrolysis of acyl-S-ACP[101].

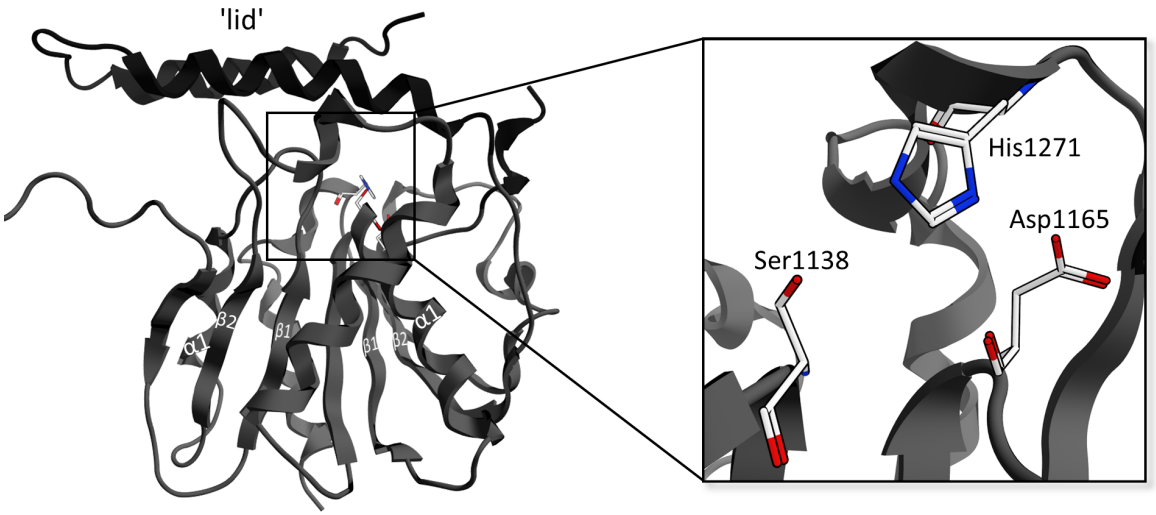


Figure 10: Structure of a TE domain and its binding pocket. Ribbon representation of an TE domain illustrated by the TE domain excised from EntF (PDB ID: 3TEJ)[70]. The TE's core (dark grey) and 'lid' (black) is highlighted. The mixed β/α/β motifs (β1-2, α1) of the four-helix bundle are labelled and binding pocket's core residues (white) are depicted in stick representation.

TE domains (30 kDa) are composed of mixed β/α/β motifs forming a HotDog[102] or, more commonly in bacterial NRPS, an α/β-Hydrolase fold[103]. Unlike PKS TEs, NRPS TEs are monomers and adopt distinct bowl-shaped hydrophobic cavity[104] of an alternating two helix dimerization region and seven- to eight-layered parallel β-sheet with a counter clockwise helical twist[105]. The exact location of the active site is dependent on the TE family[103] but is covered by two α-helices, often referred to as the 'lid' region[106], framing the shape of the substrate channel[105].

The active site is arranged in a very conserved nucleophile-histidine-acid triad[107]. In the first position, a serine, cysteine, or aspartate serve as a nucleophile – the second histidine is always conserved, and the third position must be acidic[103]. When the matured peptide reaches the end of the assembly line, the acid stabilizes the histidine that act as a catalytical base to accept a proton from the nucleophile, which subsequently performs a nucleophilic attack with its hydroxyl oxygen to the carbonyl carbon of the (peptidyl)acyl-PCP-domain bound thioester to generate a (peptidyl)acyl-O-TE oxoester[104]. Peptide release dependents on the TE's nature either by an external nucleophilic attack leading to a linear hydrolyzed or aminolyzed product, or by an internal nucleophile, typically an hydroxyl-group (-OH) or amino group (-NH_2), liberating cyclic peptides as macrolactons or macrolactams[40,103,104].

In order to achieve an accurate substrate loading, the TE domain acts in concert with the mobile tethered upstream PCP as it would otherwise be blocked by the adjacent C domain[100]. In this conformation the active site is able to reach the matured peptide. A second conformation was determined[58,60], which differs in the positioning of the 40 amino acid lid region[104]. The lid region initially attracted the peptidyl-PPant and is now responsible for the transition in the closed state to prevent unwanted access by exogenous nucleophiles in the binding pocket[40].

Furthermore, some TE domains indicate a degree of tolerance for substrate binding[103], but also imply substantial strictness in the recognition of the final products of the assembly line[40]. Therefore, a certain proofreading for processed substrate and its cleavage can be attributed to the TE[105].

Misprimed PCPs within NRPS can be repaired by hydrolysis of the undesired (peptidyl)acyl group through type II TE[40]. These *trans*-acting TE are shallow-shaped and thus allow easy accessibility to the active site that provides a rather unspecific editing/repair role in the NRPS biosynthesis[104]. Therefore, type II TE fulfil an indispensable function to prevent NRPS assembly line stalling[108].

1.3. Strategies for re-engineering NRPS systems

1.3.1. Precursor defined modifications

The most obvious approach to directed biosynthesis is the supplementation of alternative precursor molecules. Precursor-directed biosynthesis (PDB) does not rely on direct engineering of the NRPS protein architecture but feeding rather unusual or inaccessible substrates in nature[44,109]. This phenomenon results from relaxed substrate specificity of certain A domains and facilitates the first diversification of NRPS products of, e.g., novel cyclosporins[110] or lipopeptide iturins[111], by feeding alternative substrate analogues. However, this strategy is strongly limited by the promiscuity of the A domains as well as by the amount of the supplied analogues actually being able to reach their cellular target depending on the cellular processes.

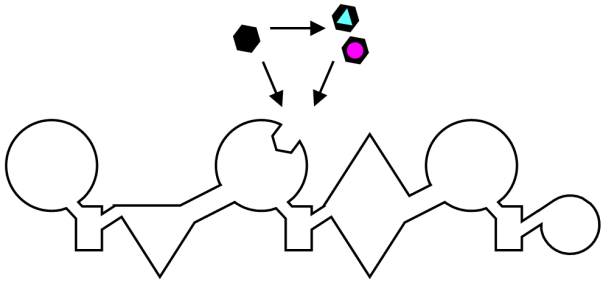


Figure 11: Precursors defined modifications. The supplemented substrates are symbolled by hexagons with coloured shaped cores resembling the analogues. Schematic representation of the NRPS domains by symbols is according to Fig. 4.

The continuation of this concept, the mutasynthesis, additionally manipulates the biosynthesis pathway of the component. In this process, the formation of the building block in the biosynthesis gets disturbed, and the desired chemical alternative is supplemented[112]. This made it possible to, e.g., generate several new vancomycin-type by blocking β-hydroxytyrosine biosynthesis or S-adenosylmethionine (SAM) modified fluoro derivatives by blocking SAM chlorination[113].

Although these approaches relies on the uptake of the substrates and also requires substrate promiscuity of the gatekeeping domains to accept the non-natural

substrate, the expansion of the concept by designing multiblocked mutant strains still finds its application in recent relevant studies of e.g. biosynthetically advanced macrocyclic NRPs[114].

1.3.2. A domain reprogramming and replacement

The investigation of the of the A domain as the main substrate specificity filter permitted initial approaches in the reprogrammed NRPSs biosynthesis with altering amino acid specificities. Pioneering work of Mohamed A. Marahiel and co-workers opened the field of rational NRPs design with the first generation of hybrid NRPS genes by exchanging domain-coding regions within *srfA* of the surfactin synthetase A[115]. However, the production yields in this process were observed to be very low, so an application with appropriate quantities couldn't be reached[116].

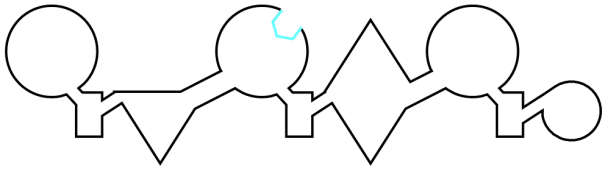


Figure 12: A domain binding pocket modification. The modified binding pocket is represented by its cyan frame. Schematic representation of the NRPS domains by symbols is according to Fig. 4.

The first solved A domain structure of the gramicidin synthetase PheA[56] and the identification of an eight key residue containing specificity-conferring code[53] granted more leeway for rational design. Further progress was driven by revealing a predictability of the substrate binding by previously uncharacterized A domains[53]. With this 'change-of-selectivity mutagenesis', it was possible to manipulate NRPS with minimal structure interferences demonstrated by several minor substitutions around the binding pocket of GrsA and SrfA[56,117].

In the case of SrfA modification, novel surfactin derivatives were obtained containing *L*-Glu instead of *L*-Gln and *L*-Asp instead of *L*-Asn, but for the latter still proportions of wild-type surfactin was observed[117]. In an alternative application, mutant strains of the initially promiscuous third fusaricidin module restricted the product spectra to

the *L*-Phe-incorporation fusaricidin derivative, which exhibits the highest antimicrobial activity[118]. Moreover, even a single mutation is demonstrated to have drastic effects to. On the one hand, is was demonstrated to increase the amount of minor products in luminmide/GameXPeptide biosynthesis[45,119] or, on the other hand, allow the incorporation of non-natural amido acids like *p*-azido-*L*-Phe and *o*-propargyl-*L*-Tyr in GrsA[120], and therefore implying the use of NRPS products for 'click chemistry'[121].

First efforts for large-scale screen methods were achieved by saturation mutagenesis. In this process the A domain binding pocket residues were randomized, and the generated library were screened in LC-MS for catalytically efficient variants. In course of this application, the screening of the *L*-Val specific andrimid (AdmK) revealed *L*-Ile or *L*-Leu incorporation analogues[122]. Furthermore, this allowed successful modifications of the stand-alone A domain 2,3-dihydroxybenzoic acid acitvating bacillibactin synthetase E (DhbE) , and the screening for non-natural 3-hydroxybenzoic acid and 2-aminobenzoic acid activating variants instead of the natural 2,3-dihydroxybenzoic acid[123].

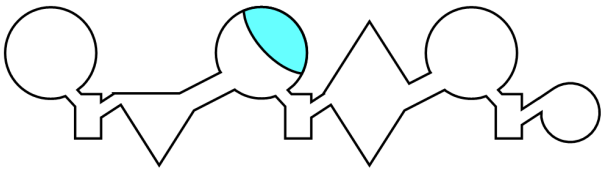


Figure 13: A subdomain swapping. The swapped A domain subunit is coloured in cyan at its proper location at the respective domain. Schematic representation of the NRPS domains by symbols is according to Fig. 4.

An approach that keeps the A domains more in its biosynthetic context is the 'subdomain swapping'. This exchange strategy is based on evolutionary events in NRPS directed by duplication and recombination of compact substructures[124], and therefore meant to be exchanged within these subdomain boundaries. The initial work on hormaomycin[125], followed by work on gramicidin[126], demonstrated the successful partial A domain substitutions transferability, including their substrate preference.

1.3.3. Exchange of single or multiple catalytic domains

Due to the modular nature of the NRPSs, engineering appears to be straightforward. Unfortunately, the exchanges of single or multiple catalytic domains turned out to be more challenging. Again, the pioneering work of Mohamed A. Marahiel and co-workers on *srfA* allowed the construction of hybrid peptide synthetases with altered amino acid specificities incorporating Phe-, Orn-, Leu-, Cys-, and Val-activating domains of *grs* and δ-(*L*-α-aminoadipyl)-cysteinyl-*D*-valine A (*acvA*)[115]. As a consequence of the observed low product yields in this set of engineering efforts, attention was drawn to the influences of the inter-domain interfaces, in particular that with the preceding C domain. The C domains assayed with synthetically loaded aminoacyl-S-coenzyme A molecules[87] and aminoacyl-*N*-acetylcysteamine thioesters (SNACs)[127] to bypass the A domains were found to exhibit strong acceptor site stereo- and side-chain specificity, and donor site stereo specificity. As evidence for the so far low success rates of well-yielding NRPS engineering attempts, further multi-module exchanges focused on maintaining C-A didomains as an inseparable unit.

Therefore, in a first whole-module fusion approach, several tyrocidine synthetases modules could be assembled *in vitro* and the expected tripeptides could be detected[128]. Here, the T-C linker between the domains was used as a reassembly point due to its highly variability.

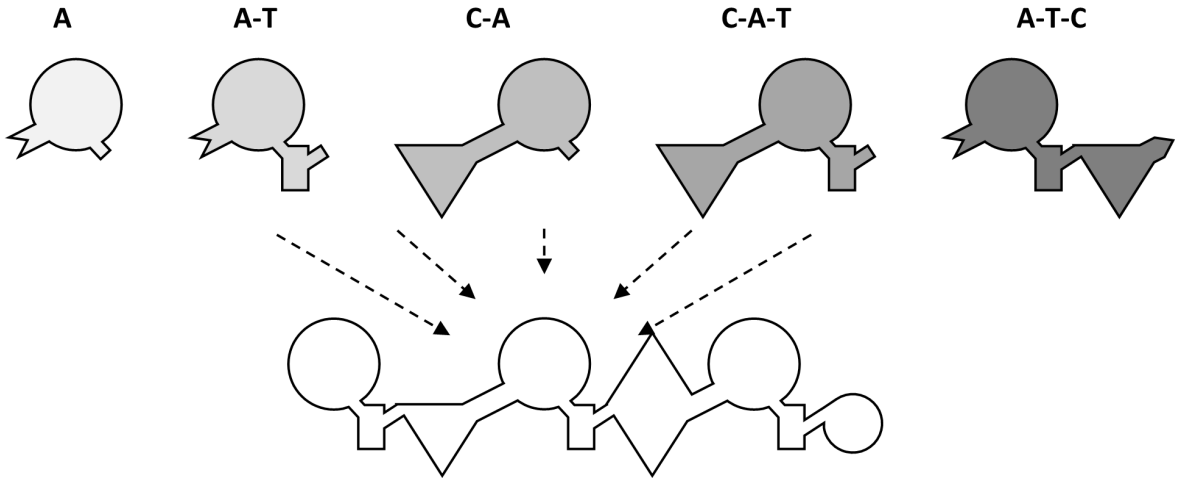


Figure 14: Exchange of single or multiple catalytic domains. Strategies that use the substitution of domain building-blocks comprising A, A-T, C-A, C-A-T, and A-T-C units. Schematic representation of the NRPS domains by symbols is according to Fig. 4.

The further development in NRPSs engineering suggested various strategies for domain assembling. Module fusions with diverse assembly points in, e.g., the C-A, T-C, A-T, or T-TE linker were tested[128–130]. In a comprehensive comparison, these strategies were evaluated by creating several recombinant Asp-Phe NRPSs, which encouraged the conclusion – through fairly variable enzyme activity – that the T-C fusion point should preferably be chosen and the C-A interface should be preserved[131]. The most extensive results were generated by Baltz and co-workers who showed successful *in vivo* exchanges. They replaced single or multiple modules in the daptomycin (*dpt*) BGC with modules from daptomycin and A54145 NRPSs, generating over 40 novel lipopeptide antibiotics in sufficient quantities[132]. Thus, they were able to identify the important antibiotic daptomycin for clinical use[133].

The direction of the engineering attempts was further guided by gained knowledge about the substrate-determination of the C domains. It was possible to demonstrate a strong exhibited stereoselectivity for the donor-T domain (C_{DSub}-) bound substrate and a significant side-chain selectivity for the acceptor-T domain (C_{ASub}-) bound substrate[87,134]. In this context, C domains were identified to cluster according to their stereoselectivity[73]. Moreover, biochemical characterization has assigned a second proofreading role to the C domains that reduces the error rate of an

NRPS[72,135], suggesting C domains as additional substrate gatekeepers[87] and the C-A interface as a stable workbench[58].

The following series of engineering attempts used C-A domains and whole module substitution[133,136], including the renowned studies by Calcott and Ackerley on pyoverdine (Pvd) synthesis in *Pseudomonas aeruginosa*[137,138]. They substituted the second Thr-specific module of PvdD with alternative substrate-specific A, C-A or T-C-A domains and observed only traces of pyoverdine by non-Thr-specific A domains. Two of these C-A domain substitutions synthesised pyoverdine relatively effectively and therefore they concluded that the adjacent C domain has a stronger acceptor-substrate selectivity than the introduced A-domain[137]. However, the extension of the exchanged didomain to T-C-A did not lead to higher yields either[138].

Nevertheless, all of the mentioned NRPS engineering attempts worked with substitution of very homologues BGC, while at the same time struggling with low production titers, and reproducibility. A breakthrough for efficient rational NRPS design was achieved by Bode and co-workers, who investigated that the inter-domain C-A linker can be separated into two structurally unrelated parts[139]. Based on these findings, they defined a new eXchange Unit (XU) concept, in which substitutions are performed along a fusion point in the C-A linker region as A-T-C XUs – leading to several multi-module *Xenorhabdus*-, *Photorhabdus*- and *Bacillus*-NRPS hybrids with considerable production titers *in vivo*[139]. Thereby, they demonstrated the importance to respect the C domain's specificity in its natural context for the successful generation of artificial NRPSs. Thereupon, and unison with the previously solved termination module AB3403 and EntF crystal structures[60], the understanding of a completely stable platform in the C-A interface was displaced by a highly dynamic C-A interface conformation during the catalytical stages.

1.4. Interdomain dynamics and their interfaces

A peculiarity of multi-modular peptide synthesis by NRPS is that the incorporation of each amino acid into the growing peptide chain is catalysed by delocalized active sites in a set of enzymatic domains[100].

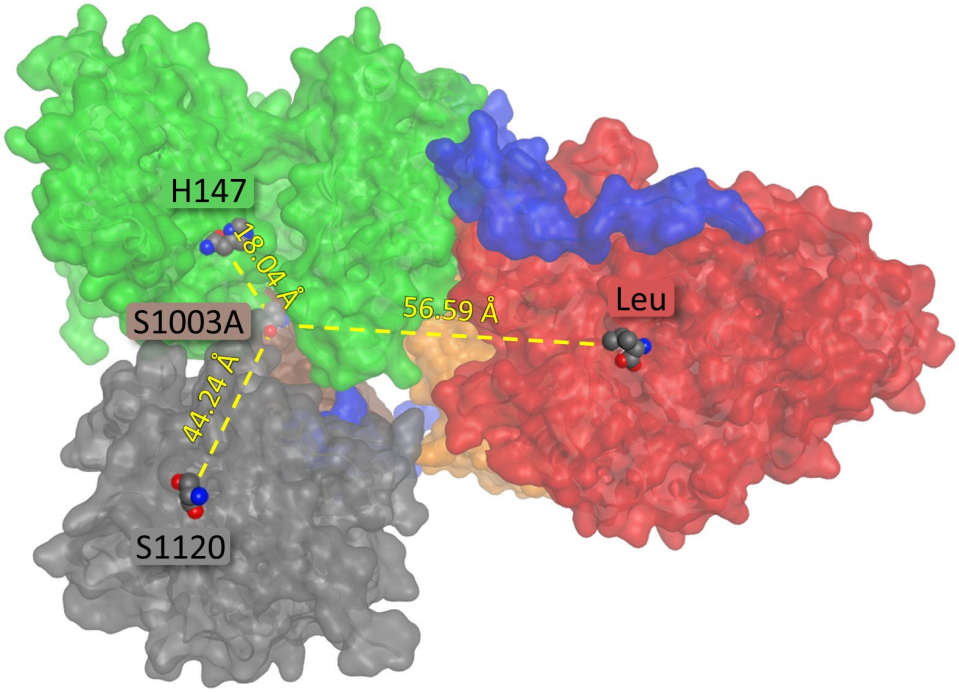


Figure 15: Distances of active sites in a NRPS module. Representation of a C-A-T-TE module with its calculated protein surfaces illustrated by SrfA-C (PDB ID: 2VSQ)[58]. The active site serine (S1003A) of the T domain (brown), which must reach the C domain (green) histidine H147, A domain (red) bound substrate (Leu), and the TE domain (dark grey) serine (S1120) is marked with the respective distances (dashed yellow lines) to these.

Since very large distances of >50 Å between active sites have to be overcome in this process, extensive domain rearrangements are required to complete the NRPS catalytic cycle[60,100]. Accordingly, the organisation of domains of NRPS elongation modules is facilitated by interdomain linker regions, allowing the movement of the domains relative to each other, and adding conformational flexibility. Linker regions are of variable length of up to 32 AA across NRPSs[58] and correspond to short non-functional loops predominantly containing small and hydrophilic residues[128]. The linkers play a crucial role not only in stabilizing the conformational state and intermediates[40] but also contribute greatly in enabling the formation of numerous domain-domain interactions during the dynamic catalytic cycle[58,60]. Thereby formed

inter-domain interfaces are dominated by hydrophobic interactions with few hydrogen bonds and salt bridges[77]. The hydrophobic interaction pattern of the various NRPSs domain interfaces exhibits rapid structural changes that provide sufficient freedom for shuffling of the activated amino acids or the growing peptide chain between the widely separated active sites. Although domain movement is rather undirected, the appropriate topologically guided interaction affinity with unidirectionality of (condensation-) reactions results in progression of NRPS synthesis[100].

1.4.1. The C-A interface

The C-A interface is the most constant interaction surface of any between the two canonical domains[100], which varies in size depending on the catalytic state of the module comprising the interacting *C*-terminal C_{Sub} and *N*-terminal A_{Core} with is smaller *C*-terminal A_{Sub} unit[40]. While in the open conformation the A_{Sub} is packed against the C domain[85] forming a 1,097 $Å^2$ interface[60]. The transition into the adenylate-forming conformation accompanied by a 30° A_{Sub} rotation displays a comparable interface in size with 1,023 $Å^2$[60]. Thereby the interacting *C*-terminal C_{Sub} surface only differs slightly (~25° rotation relative to the adenylation domain)[60]. Furthermore, in the thioester-forming conformation the C domain is shifted closer (~15 Å) towards the A domain, with which, after 140° rotation of the A_{Sub} and folding over the A_{Core}, forms an interface of 780 $Å^2$ with predominantly the A_{Core}[60,85,100].

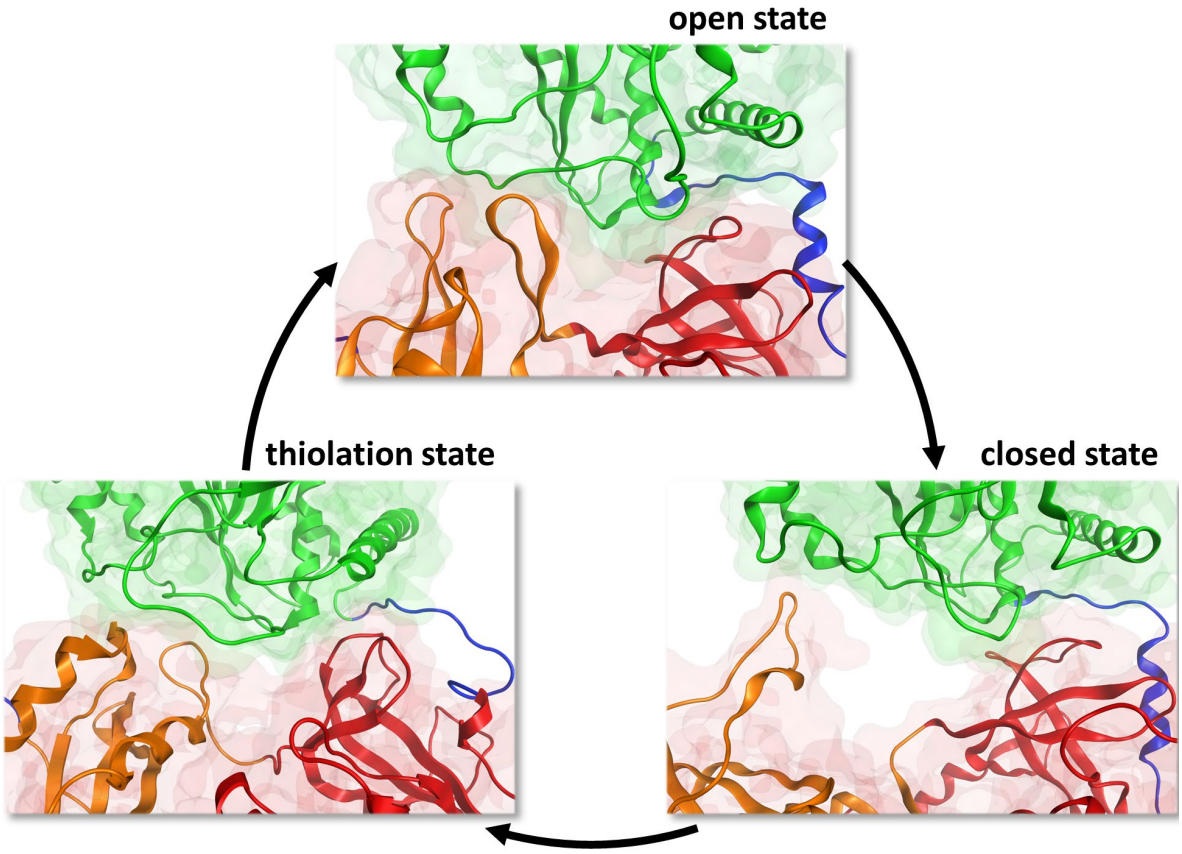


Figure 16: Structure and conformational states of the C-A Interfaces during catalytic NRPS cycle. Ribbon representation of the C-A interface with shaded protein surfaces illustrated by the C-A interface excised from SrfA-C (PDB ID: 2VSQ)[58] in the open state, AB3404 (PDB ID: 4ZXH)[60] in the closed state, EntF (PDB ID: 5T3D)[60] in the thiolation state. The *C*-terminal acceptor C_{ASub} (light green) subunit interacting with the larger *N*-terminal A_{Core} (red) and smaller *C*-terminal A_{Sub} (orange) domain are highlighted.

The L-shaped C-A linker is closely correlated with both the C and the A domain in the SrfA-C crystal structure (open conformation)[58]. Its *N*-terminal part mainly associated with the C domain seems to stabilize the efficient C-A domain interactions by running along with the C-A interface before turning from a helical segment into a structurally unrelated segment only associated with the surface of the A domain[139].

Yet, the interaction-mediating capability of the linker as well as the domain surfaces during the catalytic cycle is formed by 'weak' hydrophobic interactions with several hydrogen bonds, reported in the SrfA-C[58,139]. Therefore, the multitude of interactions between the C and A domains are meant to occlude solvent access

from high-energy intermediates or reaction centres during the catalytical stages paired with extended structural flexibility[40,58,139].

It is difficult to draw a conclusive picture of the interaction network between these two important domains with only four structures available (SrfA-C PDB ID: 2VSQ[58], EntF PDB ID: 5T3D[60], AB3403 PDB ID: 4ZXH[60], and LgrA PDB ID: 6MFX[59,100]), thus, further structural and biochemical data are required to make a comprehensive characterization.

1.4.2. The A-PCP(T) interface

To allow the thiolation reaction and shuffling between active sites to happen, the PCP needs to be guided by interacting with the adenylation domain[60]. The general A-PCP interaction involves PCP helix 2 (that carries the PPant attachment site) and the preceding and subsequent loops, which interact concurrently via hydrophobic as well as ionic interactions with helix 11 of the A_{Core}. In addition, the loop between helices 1 and 2 of the PCP forms a network of charged interactions with the last structural motif (loop + strand) of the A_{Sub} domain[85].

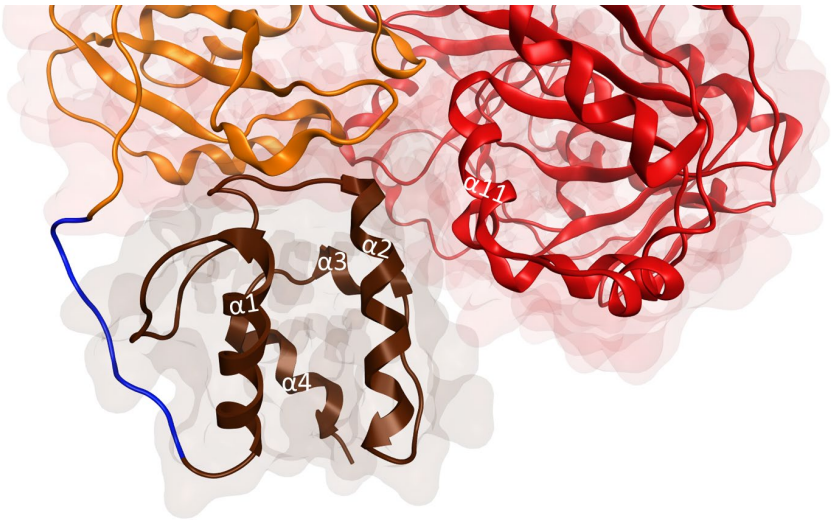


Figure 17: Structure of the A-PCP interface. Ribbon representation of the A-PCP interface with shaded protein surfaces illustrated by the A-PCP interface excised from EntF (PDB ID: 5T3D)[60]. The larger *N*-terminal A_{Core} (red) and smaller *C*-terminal A_{Sub} (orange) domain interacting with the PCP domain (brown) are highlighted. Helices (α1-4) of the four-helix bundle of the PCP and the mainly interacting A domain helix (α11) are labelled.

In contrast to the C-A interface, the linker in the A-PCP interface barely makes any contact with the catalytic platform, being only half the size of the C-A linker[58]. Therefore, the inherent free rotation and translocation of the A_{Sub} serves as a linker extension[100] and additionally creates an electrostatic surface that participates in positioning the PCP along the acceptor side of the C domain after successful thioester-formation[58,85]. This guidance seems necessary since, in the conserved LPxP linker motif, the leucine was revealed as the only one docking to the conserved hydrophobic pocket formed by residues of the β-sheet in the C_{ASub}[85], therefore, forming a quite small PCP-C(/F)-interface (~220 $Å^2$)[100].

1.4.3. The PCP(T)-C interface

The transfer of substrates to the catalytic site of the C domain requires the correct attachment of donor and acceptor PCPs to the C domain's surface[85]. Since no structure of a PCP positioned to the downstream C_{DSub} donor-PCP binding has been solved so far, the primary evidence for the mode of interaction is forwarded by available PCP and C domain homologues structures of fungal TqaA PCP-C_T[140] and GrsA PCP-E[80].

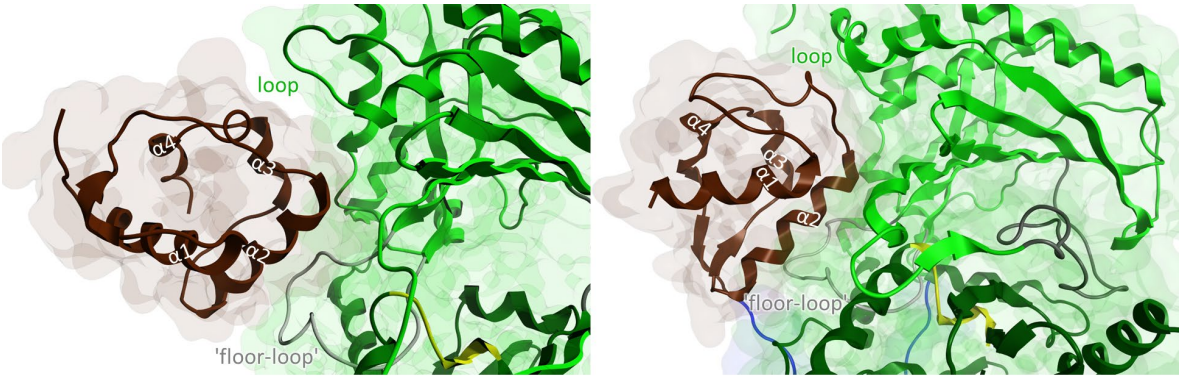


Figure 18: Structure of the PCP(T)-C interface. Ribbon representation of the A-PCP interface with shaded protein surfaces illustrated by the A-PCP interface excised from TqaA PCP-C_T (PDB ID: 5EJD)[140] and GrsA PCP-E (PDB ID: 5ISX)[80]. The C domain's *N*-terminal donor C_{DSub} (dark green) and *C*-terminal acceptor C_{ASub} (light green) subunit, interacting with the PCP domain (brown) are highlighted. Helices (α1-4) of the four-helix bundle of the PCP and active site motif (yellow), 'lid' region (dark grey), and 'floor'(light grey) loop of the C domain are labelled.

Both demonstrates direct (predominantly hydrophobic) interactions between the PCP helices 2 and 3 residues with the C domain-mimics floor-loop, stabilised by four hydrogen bonds[77,85]. This floor-loop, burring approximately one third of the ~898 Å^2 interface area (for the PCP-E interface)[80], is suggested to participate in the correct positioning of the PCP and its helix 2 attached PPant moiety in the C domain's *N*-terminal donor site depression[40].

These are preceded by a 30° rotation of the PCP conformation[100] and additionally directed by a rather ordered and less flexible 9-11 AA long linker between the two domains[76,80]. This mainly charged/polar interaction of the linker along the domain's interface act as electrostatic 'fixture' for correct recognition and binding relative to the C domain active site tunnel[85].

1.4.4. The PCP(T)-TE interface

In order to not stall NRPS catalysis during multiple cycles, recognition of the donor PCP by the TE domain is crucial for the function of the assembly line process[85]. Despite the two types of thioesterase (type I and II) found in NRPS, they indicate a very similar mode of interaction with the PCP[58,60].

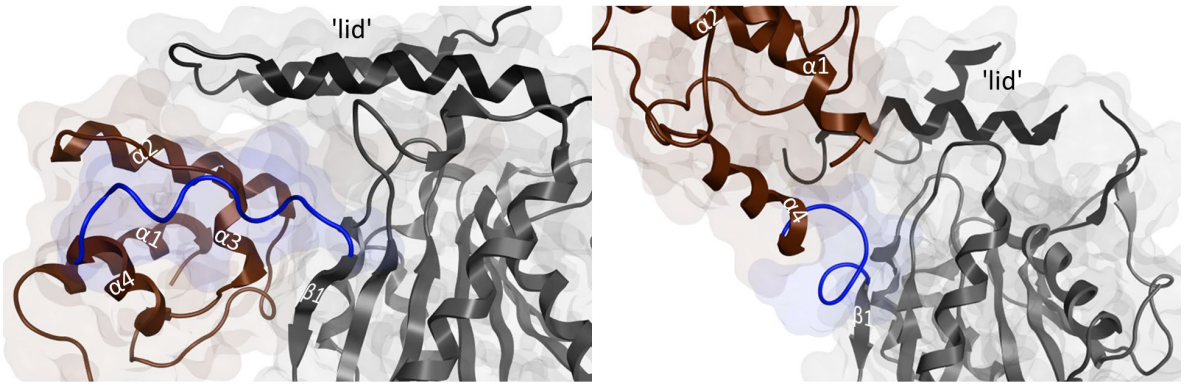


Figure 19: Structure of the PCP-TE interface. Ribbon representation of the PCP-TE interface with shaded protein surfaces illustrated by the PCP-TE interface excised from EntF (PDB ID: 3TEJ)[60] (Typ I TE) and SrfA-C (PDB ID: 2VSQ)[58] (Typ II TE). The PCP domain (brown) interacting with the TE's core (dark grey) and 'lid' (black) are highlighted. Helices (α1-4) of the four-helix bundle of the PCP and the beta-sheet (β1) of the TE domain are labelled.

This interplay predominantly involves helix 2 and the loop between helices 1 and 2 of the PCP and a small notch created by the TE's lid and core region[85]. Moreover, the extending lid of the TE gathers both of the active sites[100]. Forming a mainly hydrophobic domain-domain interface, comprising a large ~1300 $Å^2$ surface, this interaction network is proposed to maintain the PCP-4 helix bundle fold and its hydrophobic embrace of the first beta strand of the TE core[85]. The short (~9 AA) PCP-TE linker lacks any mentionable contacts[58].

Whereas the loaded PCP is suggested to deliver the maturing peptide to the downstream TE domain through a simple rotation[60], the extensive interactions still allow the PCP-TE didomain configuration great conformational mobility for the TE domain involving the lid's tilt-motion to access the PCP PPant moiety to the catalytic centre[60,85,100].

1.5. Evolution of NRPS

The beginning of NRPS's evolutionary history is subjected to the same general mechanisms that act on the evolution of any protein. The evolution of protein domains is the result of a series of random mutations and the selective imprinting of a function[141]. Accordingly, the interplay between protein conformation and enzymatic function with its interaction partner directly affects the protein configuration[142]. This interaction can be considered as a force of the protein's energy landscape, which brings specific conformational changes in the protein structure. Thus, the binding energies must be adjusted in the case of a desired binding, whereby the amino acids account for the largest part of the change in free energy that occurs upon binding[141].

The main evolutionary mechanisms believed to affect protein chemistry are mutational forces and selection pressures[143]. Random mutation in the nucleic acid sequence arise through the incorporation of nucleotide analogues, radiation effects, or low fidelity of polymerase amplification, and also by cellular repair processes meant to counter mutations[141]. The selection pressures determine the positive or negative influences of these mutations and sanction whether they will be distributed or eliminated in following generations[141].

The accommodation of related primary structure similarities are classified into families and by tertiary structure into clans or superfamilies[103]. The arise of domains and their combinations may have diverged beyond recognition by cryptic genetic variations but can thereby be traced back within lineage through combinations of pre-existing elements to their enzyme families or superfamilies[103,142,144]. However, the comparison of solved protein structures show that protein tertiary organizations are much more conserved than their primary sequence, therefore, provide significantly more insight into the relations of protein domains and phylogeny of their biosynthetic gene clusters (BGCs)[141].

BGCs are defined as physically clustered groups of genes that encode the enzymatic pathways necessary to construct specific chemicals[145], and are regarded to be key for driving chemical diversity[141]. However, even if the BGC and

its component enzymes belong to a common biosynthetic or chemical family, that can be highly heterogeneous, holding their own evolutionary histories[142].

Nature appears to have been successful at engineering biosynthetic pathways through the process of gene cluster evolution[145]. Furthermore, the rates of evolutionary events within BGCs of the secondary metabolism are much higher than those seen in comparable gene clusters involved in primary metabolism[146]. Arguably the most intensively studied group of enzymes from secondary metabolism are the PKS class of multi-modular megaenzymes[142], whereas investigations of NRPS only just started[89]. Phylogenetic analyses of the diverse and widespread PKS multi-modular megasynthase family have established an evolutionary framework as a paradigmatic example of biosynthetic assembly lines[142,146].

The modularity of PKS and NRPS is central for current evolution models of these BGCs and led to the proposal of primary forces on the modular biosynthetic assembly mechanisms by which (multi)modular enzymes evolve: (i) duplication and/or a prominent influence of concerted evolution of modules, (ii) iterative duplication by the addition of a new starter unit, and (iii) domain swapping with related gene clusters[142,145]. These mechanisms can be combined, but for the latter, the recruitment of domains over longer evolutionary distances seems to be a feature of NRPS[145] (more detailed in Fig. 20).

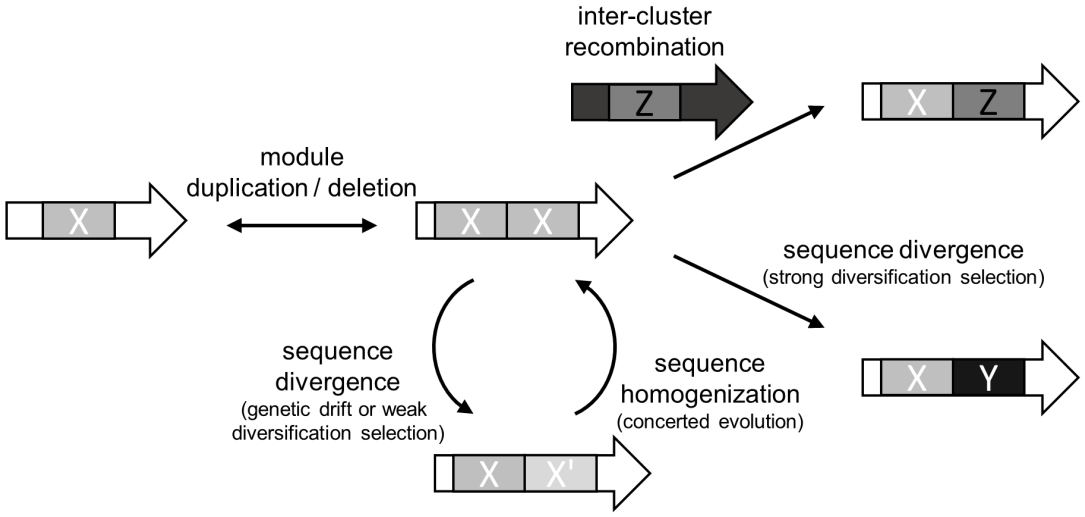


Figure 20: Evolutionary model for modular NRPS/PKS clusters. The initial evolutionary mechanism for adding or eliminating modules (grey, X) in a BGC (white arrow) are duplication or deletion. Afterwards they may pass indefinite cycles of sequence divergences forced by weak diversification selection and sequence homogenization of small sequence segments (light grey, X'). They may escape this cycle through either inter-cluster (anthracite arrow) recombination of domains (dark grey, Z) or sequence divergence forced by strong diversification selection (black, Y). Adapted from[145].

With the observation of so many different evolutionary mechanisms, protein domains have been the tools of evolution to create an enormous and diverse assembly of proteins from likely an initially relatively limited set of domains[141,145]. However, evolutionary mechanisms to achieve the exchange of individual amino acids in NRP scaffolds, or rather driving their remarkable chemical diversity, are still poorly understood[89]. Nevertheless, vigorous effort in generating functional recombinant NRPS modules still exhibits few successful approaches, but provides great contribution to unravel their evolutionary history[42,88].

Phylogenetic analyses focusing on C domains and homologues, have reconstructed an evolutionary history[142] to mainly form clades according to their functional categories in their $^{L}C_{L}$, $^{D}C_{L}$, CY, Starter, E, and dual E/C domain's subtrees[73,146]. C domains divide into a *N*-terminal donor substrate-binding (C_{DSub}) and a *C*-terminal acceptor substrate-binding (C_{ASub}) subunits by their pseudo-dimeric structure (as already described in more detail in Chapter 1.2.3), of which the donor side predominantly exhibits a strongly pronounced stereoselectivity[73]. The assigned role of C domains as stereoselective gatekeepers is supported by phylogenetic events in which clustering of C domains could be observed according to their position in

NRPS rather than their similarity to each other[146]. In addition, recombination events that integrate an E domain into a pathway demonstrate the obligatory exchange of the neighbouring $^{L}C_{L}$ domain with a $^{D}C_{L}$ domain[40,89]. Apart from the C domains' biochemical characterization (as further described in Chapter 1.2.3), the aggregation of productive recombination events within the A domain without the simultaneous exchange of its adjacent C domain increasingly challenges this role as a selectivity gatekeeper[89]. This is supported by recent studies on A domain reengineering, in which novel NRPs could be generated by substituting only the A domains[90] – indicate the C domains as a lower barrier to acceptor-substrate specificity.

Remarkably, recombination boundaries of the NRPS exchanged units can predominantly be identified within the A_{Core} domain[89,125]. Segregation analysis identified these region in the A_{Core} to exhibit the most significant phylogenetic incompatibility between A domains and the surrounding domains as clustering strongly by substrate specificity[90].

Moreover, it became apparent that the A_{Sub} domain also appears to address more complex recombination scenarios in which multiple recombination events contributed to the diversification of NRPS genes. The harbouring of a well-conserved A8 hinge motif[39] in A_{Core}-A_{Sub} interface as a likely recognition site for the C domain[89] and the preference of the A_{Sub} to be exchanged with the adjacent A_{Core} explicates the importance of maintaining the interfaces in functional architecture. As long as the C-A domain junction is intact and the dynamic A_{Sub}-T domain core is unaffected, the overall NRPS topology is less affected. In this context, the existence of modifiers that reside within the A domains as methyltransferases (Mt-A), ketoreductases (Kr-A), oxidases (Ox-A), and monooxygenases (MOx-A) is considered as evidence for nature's 'effort' to keep the native C-A, A_{Sub}-T, and T-C domain interfaces intact while implementing new enzyme functionalities[89].

Whereas the partial or complete A domain substitution is identified as the primary and best documented site for recombination events, the exchange of C-A didomain and A-T-C-A multidomains could also be observed as, however, only the

exception[89,125]. In addition, phylogenetic profiling (presence of a homolog to a given protein in a BGC based on a correlated pattern of inheritance to others), including many well-known and widely conserved motifs belonging to modular BGC architectures of NRPSs and PKSs, showed a significant co-migration of, e.g., C-A, A-T, T-C, C-A-T, and C-A-T-C[145]. These observations suggest dynamic processes by which NRPS parts are exchanged unexpectedly with varying boundaries for exchange, allowing small fragments within a module, entire domains, and even didomains up to multiple domains to recombine[146].

However, it is striking that despite these diverse recombination events, no unified boundaries of NRPS evolution, except for the A_{Core} domain, have yet been established. The reason for this might be that PKSs and NRPSs can evolve through concerted evolution where extremely similar DNA recombines through internal homologous recombination, accompanied by gene conversion[142].

Concerted evolution is a pattern of homology which is consistent with the homogenization of DNA sequences within a given repetitive family caused by high rates of internal recombination[145]. This shows that evolution has solved how to recombine NRPS genes effectively, and explains the natural occurrence of related NRPs on countless occasions[89]. Consequently, concerted evolution serves to increase similarity, but in turn, can confound phylogenetic analyses and mask early evolutionary events[142], removing almost all traces of independent evolution[145].

With the continued growth of available genomes and extensive community efforts to develop computational processing-tools, this starts to change. Numerous genomic data can be identified for their Biosynthetic Gene Clusters (BGCs) using computational tools such as antiSMASH[147], collected and catalogued by MIBiG[148], filtered and visualized due to their chemical properties with 'The Natural Products Atlas'[149], and underpinned by linking MS/MS data and whole-genome or metagenome data thought PoDP[150]. With these maturing methods at hand to navigate through genomic databases it is believed to close gaps in the historical literature of NRPS/PKS for a successful computational redesign[42].

2. Materials and Methods

2.1. Topic A: Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains

The following part contains all materials and methods used for the experiments covered in topic A (Chapter 3.1). All the herein described methods (Chapter 2.1.1.-2.1.9.) are part of the publication '*Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains*' by Kenan A. J. Bozhüyük[#], Annabell Linck[#], Andreas Tietze[#], Janik Kranz[#] (#equally shared first authors), Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-Ni Shi, Peter Grün and Helge B. Bode* (*corresponding author). This publication is attached in Chapter 7.

2.1.1. Cultivation of strains

All *E. coli* strains were cultivated in liquid or on solid LB-medium (pH 7.5, 10 g/l tryptone, 5 g/l yeast extract and 5 g/l NaCl). Solid media contained 1.5 % (w/v) agar. *S. cerevisiae* strain CEN.PK 2-1C and derivatives were grown in liquid and on solid YPD-medium (10 g/l yeast extract, 20 g/l peptone and 20 g/l glucose). Agar plates contained 1.5 % (w/v) agar. Kanamycin (50 µg/ml) and G418 (200 µg/ml) were used as selection markers. For auxotrophic selection cell were supplemented with 1120 µg/ml uracil. All *E. coli* cultures were cultivated at 37 °C, 22 °C or 16 °C for peptide or protein production purposes. *S. cerevisiae* strains were grown at 30 °C.

2.1.2. Microorganisms

All microorganisms used for topic A, including their genotype and reference are listed in the following table.

Table 1. Microorganisms used for topic A.

Strain	Genotype	Reference
E. coli DH10B	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ- *rps*L(Str^R) *nup*G/ -	Invitrogen
E. coli DH10B::*mtaA*	*E. coli* DH10B with Δ*entD*::*mtaA*	[151]
S. cerevisiae CEN. PK 2-1C	*MATa*; *his3D1*; *leu2-3_112*; *ura3-52*; *trp1-289*; *MAL2-8c*; *SUC2*	Euroscarf
P. luminescens TTO1	wildtype	NCBI:txid243265
X. nematophila ATCC 19061	wildtype	NCBI:txid406817
X. budapestensis DSM16342	wildtype	NCBI:txid290110
B. licheniformis ATCC 10716	wildtype	M. A. Marahiel / ATCC

2.1.3. Plasmids

All plasmids used for topic A, including their genotype and reference are listed in the following table.

Table 2. Plasmids used for topic A.

Plasmid	Genotype	Reference
pAT41	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	gifted by Andreas Tietze, AK Bode Frankfurt, Germany
pFF1_*gxpS*	2µ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*	[139]

pFF1_NRPS_1	2µ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1C2-*bicA*_A2T2C_{Dsub}3-*gxpS*_C_{Asub}3A3T3C4A4T4C5A5T5TE	gifted by Kenan A. J. Bozhüyük, AK Bode Frankfurt, Germany
pFF1_NRPS_2	2µ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1C2A2T2C_{Dsub}3-*bicA*_C_{Asub}3A3T3C_{DSub}4-*gxpS*_C_{Asub}4A4T4C5A5T5TE	gifted by Kenan A. J. Bozhüyük, AK Bode Frankfurt, Germany
pFF1_NRPS_3	2µ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*_A1T1C2A2T2C_{Dsub}3-*gxpS*_C_{Asub}3A3T3C_{Dsub}4-*xtpS*_C_{Asub}4A4T4TE	gifted by Kenan A. J. Bozhüyük, AK Bode Frankfurt, Germany
pAT41_NRPS_4	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bacA*_A1T1Cy2A2T2C_{Dsub}3-*gxpS*_C_{Asub}3A3T3CE4A4T4CE5A5T5TE	this work

2.1.4. Oligonucleotides

All oligonucleotides used for topic A are listed in the following table.

Table 3. Oligonucleotides used for topic A. For all constructs the digested pFF1 with EcoRI/SgsI was used for TAR cloning.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pAT41_NRPS_4	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCC GTTTTTTTGGGCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAA AATG	*B. licheniformis* ATCC 10716
	ALAT_1	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAATTAAAT GTATGATTCCATTCCACATAATC	*B. licheniformis* ATCC 10716
	AT_99	TCAGGTGAAGGAGTACAGGCAC	pFF1_*gxpS*
	ABCR14_neu	GAAACGGGTATGTTCAGCCTGAC	pFF1_*gxpS*
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	pFF1_*gxpS*
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCG CCTCCGCTTCAC	pFF1_*gxpS*

2.1.5. Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus*, *Photorhabdus* and *Bacillus* strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. All PCRs were performed with oligonucleotides obtained from Eurofins Genomics (Tab. 3). NRPS fragments for HiFi cloning (NEB) or TAR cloning[151] were amplified with primers coding for the respective homology arms (40-80 bp) in a two-step PCR program. Polymerases Phusion High-Fidelity DNA polymerase (Thermo Scientific), Q5 High-Fidelity DNA polymerase (New England BioLabs), and Velocity DNA polymerase (Bioline) were used according to the manufacturers' instructions. All generated plasmids (Tab. 2) were introduced into *E. coli* DH10B::*mtaA*[151] by electroporation. Each NRPS (subunit) was under the control of a P_{BAD} promotor for peptide production. Plasmid isolation from *E. coli* was achieved by alkaline lysis or with the Invisorb Spin Plasmid Mini Two Kit (stratec molecular).

2.1.6. Transformation-associated recombination (TAR) cloning

Transformation of yeast cells was done according to the protocols from Gietz and Schiestl[152]. Therefore 100-2,000 ng of each fragment was used for transformation. All generated plasmids were isolated from yeast transformants and introduced into *E. coli* DH10B::*mtaA* by electroporation. Successfully transformed plasmids were isolated from *E. coli* transformants and verified by restriction digest.

2.1.7. Heterologous expression of NRPS templates and HPLC/MS analysis

Constructed plasmids were transformed into *E. coli* DH10B::*mtaA*[151]. Cells were grown overnight in LB-medium containing the 50 µg/ml kanamycin. 100 µl of an overnight culture were used for inoculation of 10 ml LB-cultures supplemented with

the respective antibiotics as selection markers and additionally containing 0.002 mg/ml *L*-arabinose and 2 % (v/v) XAD-16. For the feeding experiments 2 mM of the respective AA was added into the growing *E. coli* culture. After incubation for 72 h at 22 °C the XAD-16 was harvested. One culture volume methanol was added and incubated for 30 min at RT. The organic phase was filtrated and evaporated to dryness under reduced pressure. The extract was diluted in 1 ml methanol and a 1:10 dilution was centrifugation (17,000 x *g*, 20 min) and used for HPLC/MS analysis. All measurements were performed by using an Ultimate 3000 LC system (Dionex) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml/min using acetonitrile (ACN) and water containing 0.1 % formic acid (v/v) in a gradient ranging from 5-95 % of ACN over 16 min (40 °C) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. High-resolution mass spectra were obtained on an Ultimate 3000 RSLC (Dionex) coupled to an Impact II qTOF (Bruker) equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

2.1.8. Peptide quantification

The absolute production titers of selected peptides were calculated with calibration curves based on pure synthetic **8**, (for quantification of **8**) or purified **2** (for quantification of **2**), **3** (for quantification of **3**), **4** (for quantification of **4**, **5** and **9**), **7** (for quantification of **6**, **7** and **10**).

Therefore, the pure compounds were prepared at different concentrations (100, 50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, and 0.195 µg/ml) and measured by HPLC/MS measurements as described above. 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 origin. Triplicates of all *in vivo* experiments were measured. The pure peptide standards **8** was synthesized in-house[139].

2.1.9. Peptide isolation and structure elucidation

Seven peptides (**1**, **2**, **3**, **4**, **5**, **6**, **7**) were isolated from *E. coli* DH10B::*mtaA* (see Chapter 7). The strains were cultivated, and the extracts were generated as described above, from 1 l cultures. Compounds were isolated in a first chromatography using either Sephadex LH-20 (MeOH, 25-100 µm, Pharmacia Fine Chemical Co. Ltd.) or a 1260 Infinity II LC system coupled to a G6125B LC/MSD ESI-MS (Agilent). A 25-55 % water/acetonitrile gradient was applied over 25 min on an Agilent Eclipse XDB-C18, 7 µm, 21.2 x 250 mm column using a flow rate of 20 ml/min. Subsequently, **4**, **5**, **6**, **7** were purified in an additional chromatographic step using a 1260 Semiprep LC system coupled to a G6125B LC/MSD ESI-MS (Agilent). A 35-65 % water/acetonitrile gradient was applied over 25 min on an Eclipse Plus Phenyl-Hexyl, 5 µm, 9.4 x 250 mm column using a flow rate of 3 ml/min. If required an additional semipreparative HPLC was performed on an Agilent 1260 Infinity II LCMS Systems with a Cholester column (1OID × 250 mm, COSMOSIL). The structures of all isolated compounds were elucidated by detailed 1D and 2D NMR experiments (see Chapter 7).^{1}H, ^{13}C, HSQC, HMBC, ^{1}H-^{1}H COSY, and ROESY spectra were measured on Bruker AV500 and AV600 spectrometers, using DMSO as solvent. Coupling constants are expressed in Hz and chemical shifts are given on a ppm scale. High-resolution MS analysis was performed as described above.

2.2. Topic B: Influence of condensation domains on activity and specificity of adenylation domains

The following part contains all materials and methods used for the experiments covered in topic B (Chapter 3.2). All the herein described methods (Chapter 2.2.1.-2.2.12.) are part of the manuscript in preparation '*Influence of condensation domains on activity and specificity of adenylation domains*' by Janik Kranz[#] (#first author), Sebastian L. Wenski, Alexander A. Dichter, Helge B. Bode* and Kenan A. J. Bozhüyük* (*corresponding authors). A preprinted (non-peer-reviewed) version of the manuscript is available on *bioRxiv* (doi: 10.1101/2021.08.23.457306).

2.2.1. Cultivation of strains

All *E. coli* and *Xenorhabdus* strains were cultivated in liquid or on solid LB-medium (pH 7.5, 10 g/l tryptone, 5 g/l yeast extract and 5 g/l NaCl). Solid media contained 1 % (w/v) agar. Kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml) were used as selection markers. All *E. coli* cultures were cultivated at 37 °C, 22 °C, or 16 °C for peptide or protein production purposes. *Xenorhabdus* strains were grown at 30 °C.

2.2.2. Microorganisms

All microorganisms used for topic B, including their genotype and reference are listed in the following table.

Table 4. Microorganisms used for topic B.

Strain	Genotype	Reference
E. coli DH10B	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ$^-$ *rps*L(Str^R) *nup*G/ -	Invitrogen

E. coli DH10B::*mtaA*	*E. coli* DH10B with Δ*entD*::*mtaA*	[151]
E. coli BL21 (DE3) Gold	*E. coli* B F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte	Invitrogen
E. coli S17-1 λpir	Tp^r Sm^r *rec*A, *thi*, *pro*, *hsd*R-M+RP4: 2- Tc:Mu:Km Tn7 λpir	Invitrogen
P. luminescens TTO1	wildtype	NCBI:txid243265
X. nematophila ATCC 19061	wildtype	NCBI:txid406817
X. budapestensis DSM16342	wildtype	NCBI:txid290110
X. miraniensis DSM 17902	wildtype	NCBI:txid351674
X. indica DSM 17382	wildtype	NCBI:txid333964
B. licheniformis ATCC 10716	wildtype	M. A. Marahiel / ATCC
X. szentirmaii DSM 16338	wildtype	NCBI:txid1427518
X. budapestensis DSM 16342	wildtype	NCBI:txid290110
X. hominickii DSM 17903	wildtype	NJAI00000000_NCBI
X. szentirmaii Δ*fclIJ*	wildtype *ΔfclI*, *ΔfclJ*	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany

2.2.3. Plasmids

All plasmids used for topic B, including their genotype and reference are listed in the following table.

Table 5. Plasmids used for topic B.

Plasmid	Genotype	Reference
pTF16	Chaperone tig, *L*-arabinose inducible Promotor *araB*, cm^R	TaKaRa Bio Inc., Singapore
pCOLA_ara/tacI	ori ColA, kan^R, *araC*-P_{BAD} and *tac*I	[153]
pCOLA_GxpS A3-T3	ori ColA, kan^R, *tac*I, HIS-*gxpS*_A3T3	gifted by Sarah Nowak, AK Bode Frankfurt, Germany
pAD_GxpS C_A3-A3-T3	ori ColA, kan^R, *tac*I, *gxpS*_CA3A3T3	gifted by Alexander Dichter, AK Bode Frankfurt, Germany

pAD_GxpS C3-A3-T3	ori ColA, kan^R, *tac*I, HIS-*gxpS*_C3A3T3	gifted by Alexander Dichter, AK Bode Frankfurt, Germany
pAD_XtpS C3 GxpS A3-T3	ori ColA, kan^R, *tac*I, HIS-*xtpS*_C3 *gxpS*_A3T3	gifted by Alexander Dichter, AK Bode Frankfurt, Germany
pJK_BacA C3 GxpS A3-T3	ori ColA, kan^R, *tac*I, HIS-*bacAS*_C3 *gxpS*_A3T3	this work
pJK11.2	ori ColA, kan^R, *araC*-P_{BAD} *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK17	ori ColA, kan^R, *araC*-P_{BAD} *gxpS*_C3A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK17.2	ori ColA, kan^R, *araC*-P_{BAD} *gxpS*_C_{ASub}3A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK12	ori ColA, kan^R, *araC*-P_{BAD} *xtpS*_C3 *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK13	ori ColA, kan^R, *araC*-P_{BAD} *kolS*_C5 *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK14	ori ColA, kan^R, *araC*-P_{BAD} *bicA*_C3 *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK15	ori ColA, kan^R, *araC*-P_{BAD} *X. mira. AmbS*_C5 *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pJK16	ori ColA, kan^R, *araC*-P_{BAD} *X. indi. AmbS*_C5 *gxpS*_A3T3CE4A4T4C5A5T5TE and *tac*I	this work
pEB17	R6Kγ ori, oriT, *araC*, *ara*BAD promoter, Km^r	[154]
pEB17_Δ*fclIJ* *X. szentirmaii*	pEB17 with *Xsze fclI, Xsze fclJ*	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pCOLA_ara_tacI_*fclJ Xsze*	ori ColA, kan^R, *araC*-P_{BAD} *Xsze fclj*_C5A5T6C6A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany

pCOLA_ara_tacI_*fclJ Xbud*	ori ColA, kan^R, *araC-P$_{BAD}$ Xbud fclj*_C5A5T6C6A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pCOLA_ara_tacI_*fclJ Xhom*	ori ColA, kan^R, *araC-P$_{BAD}$ Xhom fclj*_C5A5T6C6A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pDD3	ori ColA, kan^R, *araC-P$_{BAD}$ Xbud fclj*_C5A5T6C6 *Xsze. fclj*_A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pDD4	ori ColA, kan^R, *araC-P$_{BAD}$ Xhom fclj*_C5A5T6C6 *Xsze fclj*_A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pDD5	ori ColA, kan^R, *araC-P$_{BAD}$ Xsze fclj*_C5A5T6C6 *Xbud fclj*_A6T6 and *tac*I	this work
pDD6	ori ColA, kan^R, *araC-P$_{BAD}$ Xhom fclj*_C5A5T6C6 *Xbud fclj*_A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pDD7	ori ColA, kan^R, *araC-P$_{BAD}$ Xsze fclj*_C5A5T6C6 *Xhom fclj*_A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany
pDD8	ori ColA, kan^R, *araC-P$_{BAD}$ Xbud fclj*_C5A5T6C6 *Xhom fclj*_A6T6 and *tac*I	gifted by Sebastian Wenski, AK Bode Frankfurt, Germany

2.2.4. Oligonucleotides

All oligonucleotides used for topic B are listed in the following table.

Table 6. Oligonucleotides used for topic B.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pJK_BacA C3 GxpS A3-T3	JK-P194	CGAGAACCTGTACTTCCAATCCATGAGAGAAGAAAGCGATATGGAAC	*B. licheniformis* ATCC 10716
	JK-P195	GATACGGTTCTTCGGTCGCATTCCAGGACTCTAAAAGTGTCCGT	*B. licheniformis* ATCC 10716
	JK-P132	TGGAATGCGACCGAAGAACC	*P. laumondii* TTO1
	AD14	CTTTACCAGACTCGATTACGGTACTTGCTCAACAATACG	*P. laumondii* TTO1
	DUET_Gib_FW	CAGCTTAATTAACCTAGGCTG	pCOLADuet_ara/tacI
	JK-P184	CATGGATTGGAAGTACAGGTTCTCGGTACCCAGATCTACACCAGAAGAATG ATGATGATGATGGTGCATGGTATATCTCCTTATTAAAGTT	pCOLADuet_ara/tacI
pJK11.2	jw0142_for	GGGCTAACAGGAGGAATTCCATGGTATGTGTCCATCAGTTGTTTGAAC	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATAC GAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK17	ALAT_6	TCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGA ATTCCATGTGTGCACAACGTAATACGGGC	*P. laumondii* TTO1

	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATAC GAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK17.2	ALAT_5	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGA GGAATTCCATGTCAGGTGAAGGAGTACAGGCAC	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATAC GAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK12	JK-P93	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTT TTTTGGGCTAACAGGAGGAATTCCATGCGTCATGCGAACAGTGATC	*X. nematophila* ATCC 19061
	AD74	GGTTCTTCGGTCGCATTCCAGGTTTTTAACAACAATGTGC	*X. nematophila* ATCC 19061
	AD75	GCACATTGTTGTTAAAAACCTGGAATGCGACCGAAG	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATAC GAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK13	JK-P95	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTT TTTTGGGCTAACAGGAGGAATTCCATGTTTGATGAGTCAGGCGAGG	*P. laumondii* TTO1
	JK-P67	AGTAGGATACGGTTCTTCGGTCGCATTCCAGGTTTCCAATAACAGCTTACG T	*P. laumondii* TTO1
	JK-P68	GTAGCGGAACGTAAGCTGTTATTGGAAACCTGGAATGCGACCGAAGAAC	*P. laumondii* TTO1

	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK14	JK-P99	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGGCAGTGCAGGCCCGGATTGATAAAC	*X. budapestensis* DSM16342
	JK-P97	AGTAGGATACGGTTCTTCGGTCGCATTCCAAGTTTTCAGCAACAACTGGCG	*X. budapestensis* DSM16342
	JK-P98	TCGGCAGAGCGCCAGTTGTTGCTGAAAACTTGGAATGCGACCGAAGAAC	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pJK15	JK-P103	ACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGCTTAATAAAAAAGAAAGTATACAGCCTGTTATCGTACCGATATCC	*X. miraniensis* DSM 17902
	JK-P101	AGTAGGATACGGTTCTTCGGTCGCATTCCACGTTTCCAGCAATAACCGG	*X. miraniensis* DSM 17902
	JK-P102	CCGGCAGAGCGCCGGTTATTGCTGGAAACGTGGAATGCGACCGAAGAAC	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI

pJK16	JK-P107	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGTTGGGTGAATATAGCAATAAACTTCCTGCTATTTTGCC	*X. indica* DSM 17382
	JK-P105	AGTAGGATACGGTTCTTCGGTCGCATTCCACGTTTCCAGCAATAATTTACG	*X. indica* DSM 17382
	JK-P106	CCAGTGGAACGTAAATTATTGCTGGAAACGTGGAATGCGACCGAAGAAC	*P. laumondii* TTO1
	JK-P92	TGCTAGCCTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCACAATTCATTGCC	*P. laumondii* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pCOLADuet_ara/tacI
	jw0064	CATGGAATTCCTCCTGTTAGCC	pCOLADuet_ara/tacI
pDD5	SW497	CCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGCAGCAAGATACATTTAAATTTAAAGCATCCC	*X. szentirmaii* DSM 16338
	SW498	CCCGGCGCTCCCTGATTGAAGGCTTCAATTTTCCGCTGCTGGG	*X. szentirmaii* DSM 16338
	SW499*	AGCAGCGGAAAATTGAAGCCTTCAATCAGGGAGCGCCG	*X. budapestensis* DSM 16342

2.2.5. Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus* and *Photorhabdus* strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. All PCRs were performed with oligonucleotides obtained from Eurofins Genomics (Tab. 6). NRPS fragments for HiFi cloning (NEB) were amplified with primers coding for the respective homology arms (20-30 bp) in a two-step PCR program. The coding sequences for the His6-Tag were amplified with the pCOLADUET^{TM}-1 (Merck/Millipore) plasmid backbone. Polymerases Phusion High-Fidelity DNA polymerase (Thermo Scientific), Q5 High-Fidelity DNA polymerase (New England BioLabs), and Velocity DNA polymerase (Bioline) were used according to the manufacturers' instructions. DNA purification was performed using Invisorb Fragment CleanUp or MSB Spin PCRapace Kits (stratec molecular). All generated plasmids (Tab. 5) were introduced into *E. coli* DH10B::*mtaA*[151] by electroporation. Each NRPS (subunit) was under the control of a P_{BAD} promotor for peptide production or a *tac*I promotor for protein expression. Plasmid isolation from *E. coli* was achieved with the Invisorb Spin Plasmid Mini Two Kit (stratec molecular).

2.2.6. Generation of deletion mutants

The generation of deletion mutants was performed as described previously[155,156]: The upstream and downstream flanking regions of the corresponding gene (approximately 1000 bp) were amplified and cloned into the either PCR-amplified or digested vector pEB17 to generate deletion vectors[10]. After the Hot Fusion Assembly *E. coli* S17 were transformed with the vectors, followed by conjugation with the corresponding *Xenorhabdus* strain as described previously[157].

2.2.7. Transformation of *Xenorhabdus szentirmaii*

Hetero- and homologous complementation as well as NRPS-engineering plasmids were transformed into the corresponding *X. szentirmaii* strain by heatshock transformation by an adapted protocol of Xu *et al*. as described previously[155,158].

2.2.8. Heterologous expression of NRPS templates and LC-MS analysis

Constructed plasmids were transformed into *E. coli* DH10B::*mtaA*[151]. Cells were grown overnight in LB-medium containing the necessary antibiotics (50 µg/ml kanamycin, or 34 µg/ml chloramphenicol). 100 µl of an overnight culture were used for inoculation of 10 ml LB-cultures supplemented with the respective antibiotics as selection markers and additionally containing 0.002 mg/ml *L*-arabinose and 2 % (v/v) XAD-16. After incubation for 72 h at 22 °C the XAD-16 was harvested. One culture volume methanol was added and incubated for 30 min at RT. The organic phase was filtrated, and a sample was taken of the cleared extract. After centrifugation (17,000 x *g*, 20 min) the methanol extracts were used for HPLC/MS analysis. All measurements were performed by using an Ultimate 3000 LC system (Dionex) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml/min using acetonitrile (ACN) and water containing 0.1 % formic acid (v/v) in a gradient ranging from 5-95 % of ACN over 16 min (40 °C) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. High-resolution mass spectra were obtained on an Ultimate 3000 RSLC (Dionex) coupled to an Impact II qTOF (Bruker) equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

2.2.9. Expression and purification of His6-tagged proteins

For overproduction and purification of the His6-tagged ~72 kDa GxpS_A3-T3, ~98 kDA GxpS_C3$_{ASub}$-A3-T3, ~122 kDa GxpS_C3-A3-T3 and ~122 kDa XtpS_C3-GxpS_A3-T3 a 5 ml overnight culture in LB-medium of *E. coli* BL21 (DE3) Gold cells harbouring the corresponding expression plasmid and the TaKaRa chaperone-plasmid pTF16 (TAKARA BIO INC.) were used to inoculate 500 mL of autoinduction medium (464 mL LB-medium, 500 µl 1M MgSO4, 10 ml 50x5052, 25 ml 20xNPS) containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. The cells were grown at 37 °C up to an OD_{600} of 0.6. Thereupon, the cultures were cultivated for additional 72 h at 16 °C. The cells were pelleted (10 min, 4,000 rpm, 4 °C) and stored overnight at -20 °C.

For protein purification the cell pellets were resuspended in binding buffer (500 mM NaCl, 20 mM imidazol, 50 mM HEPES, 10 % (w/v) glycerol, pH 8.0). For cell lysis benzonase (500 U, Fermentas), protease inhibitor (Complete EDTA-free, Roche), 0.1 % Triton-X and lysozym (0.5 mg/mL, ~20,000 U/mg, Roth) were added and the cells were incubated rotating for 30 min at 4 °C. After this the cells were placed on ice and lysed by ultra-sonication. Subsequently, the lysed cells were centrifuged (25,000 rpm, 30 min, 4 °C).

The yielded supernatant was passed through a 0.2 µm filter and loaded with a flow rate of 0.5 ml/min on a 5 ml HisTrap HP column (GE Healthcare) equilibrated with 12 CV binding buffer. Unbound protein was washed off with 8 CV with 4 % and 8 CV with 8 % elution buffer (500 mM NaCl, 500 mM imidazol, 50 mM HEPES, 10 % (w/v) glycerol, pH 8.0). The purified protein of interest was eluted with 35 % elution buffer. Following, the purified protein containing fraction was concentrated (Centriprep® Centrifugal Filters Ultacel® YM - 50, Merck Millipore).

2.2.10. Multiplexed hydroxamate assay (HAMA)

The hydroxamate formation assay was performed as published previously[159]. The 100 µl reaction volume containing 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 150 mM hydroxylamine (pH 7.5-8.0, adjusted with HCl), 5 mM ATP, 1 mM TCEP and 2 µM of purified enzyme were stared by adding 1 mM proteinogenic amino acid mix (in 100 mM TRIS, pH 8) and incubated for 30 min at RT. The reactions were stopped by diluting in 10-fold 95 % acetonitrile (ACN) and water containing 0.1 % formic acid. All measurements were performed by using an Ultimate 3000 RSLC (Dionex) with an ACQUITY UPLC BEH Amide Column, 130 Å, 1.7 µm, 2.1 mm X 50 mm, 1/pkg coupled to an Impact II qTOF (Bruker) equipped with an ESI Source set to positive ionization mode. UPLC conditions were performed as published previously[159]. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

2.2.11. Homology modelling and interface identification

The homology-modelling was performed with the homology modelling algorithm within MOE (Molecular Operating Environment). This process undergoes an (I) initial partial geometry, where all coordinates are copied if residue identity is conserved. Next, a (II) Boltzmann-weighted randomized sampling, which (IIa) consists of a backbone fragments collection from a high-resolution structural database, and alternative side chain conformations assembly from an extensive rotamer library for non-identical residues and (IIb) a creation of independent number models based upon loop and side chain placements scored by a contact energy function[160].

For homology modelling, the C-A didomains within the 2.70 Å crystal structure of AB3403 (PDB-ID: 4ZXH)[60], 2.80 Å crystal structure of EntF (PDB-ID: 5T3D)[60] and 2.60 Å crystal structure of SrfA-C (PDB-ID: 2SVQ)[58] were selected as homologous-protein-templates.

With the homology models build, HSPred[161], a support vector machine(SVM)-based method, was used to predict the critical interaction partners across the

interface. This approach systematically mutated individual amino acids (excluding Pro and Gly) to alanine and calculates the changes in free energy of binding (*ΔΔG*). 'Critical Interaction Partners' or 'Hot Spots Residues' are defined as those residues for which *ΔΔG* ≥ 2 kcal/mol. The HSPred output score (its exact calculation can be read here[161]) predicts mutated residues with a score greater than zero as hot spots (*ΔΔG* ≥ 2 kcal/mol) and negative scores (*ΔΔG* < 2 kcal/mol) as non-hot spots. Others are not involved in the interface.

2.2.12. Peptide quantification

The absolute production titers of selected peptides were calculated with calibration curves based on pure synthetic **19** (for quantification of **19**), **20** (for quantification of **20**), **21** (for quantification of **21** and **23**), **22** (for quantification of **22**, **24** and **28**), **25** (for quantification of **25** and **26**), and **27** (for quantification of **27**). Therefore, the pure compounds were prepared at different concentrations (100, 50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, and 0.195 µg/ml) and measured by HPLC/MS measurements as described above. 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 origin. Triplicates of all *in vivo* experiments were measured. The pure peptide standards **19**, **20**, **21**, **22**, **25** and **27** were synthesized in-house[139].

2.3. Topic C: Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering

The following part contains all materials and methods used for the experiments covered in topic C (Chapter 3.3). All the herein described methods (Chapter 2.3.1.-2.3.11.) are part of the manuscript in preparation '*Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering*' by Janik Kranz[#], Nadya Abbood[#] (#equally shared first authors), Mohammad Alanjary, Helge B. Bode* and Kenan A. J. Bozhüyük* (*corresponding authors).

2.3.1. Cultivation of strains

All *E. coli* strains were cultivated in liquid or on solid LB-medium (pH 7.5, 10 g/l tryptone, 5 g/l yeast extract and 5 g/l NaCl). Solid media contained 1 % (w/v) agar. Kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml) were used as selection markers. All *E. coli* cultures were cultivated at 37 °C.

2.3.2. Microorganisms

All microorganisms used for topic C, including their genotype and reference are listed in the following table.

Table 7. Microorganisms used for topic C.

Strain	Genotype	Reference
E. coli DH10B	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ- *rps*L(Str^R) *nup*G/ -	Invitrogen
E. coli DH10B::*mtaA*	*E. coli* DH10B with Δ*entD*::*mtaA*	[151]
P. luminescens TTO1	wildtype	NCBI:txid243265
X. budapestensis DSM 16342	wildtype	NCBI:txid290110
X. nematophila ATCC 19061	wildtype	NCBI:txid406817
X. miraniensis DSM 17902	wildtype	NCBI:txid351674
B. licheniformis ATCC 10716	wildtype	M. A. Marahiel / ATCC

2.3.3. Plasmids

All plasmids used for topic C, including their genotype and reference are listed in the following table.

Table 8. Plasmids used for topic C.

Plasmid	Genotype	Reference
pFF1	ori 2µ, kanMX4, *araC-P$_{BAD}$*, ori ColA, Ypet-Flag, kan^R, MCS	[139]
pFF1_gxpS_bicA-A1T2C2	ori 2µ, kanMX4, *araC-P$_{BAD}$*, ori ColA, Ypet-Flag, kan^R, *bicA*_A1T1CE2 *gxpS*_A2T2C3A3T3CE4A4T4CE5A5T5TE	[139]
pJW82	ori p15A, cm^R, *araC-P$_{BAD}$* *gxpS*_A1T1CE2A2T2C3 and *tac*I-*araE*	gifted by Jonas Watzel, AK Bode Frankfurt, Germany
pJW63	ori p15A, cm^R, *araC-P$_{BAD}$* *xtpS*_A1T1CE2A2T2C3 and *tac*I-*araE*	[162]

pNA93	ori p15A, cm^R, *araC-*P_{BAD} *X. mira. ambS*_C1A1T1CE2A2T2C3 and *tac*I-*araE*	gifted by Nadya Abbood, AK Bode Frankfurt, Germany
pNA142	ori p15A, cm^R, *araC-*P_{BAD} *bicA*_A1T1CE2A2T2C3 and *tac*I-*araE*	gifted by Nadya Abbood, AK Bode Frankfurt, Germany
pJW115	ori p15A, cm^R, *araC-*P_{BAD} *bacA*_A1T1Cy2A2T2C3 and *tac*I-*araE*	gifted by Jonas Watzel, AK Bode Frankfurt, Germany
pJW83	ori ColA, kan^R, *araC-*P_{BAD} *gxpS*_A3T3CE4A4T4CE5A5T5TE and *tac*I	gifted by Jonas Watzel, AK Bode Frankfurt, Germany
pJW83_A7	pJW83 with inserted gene fragment A7	this work
pJW83_A8	pJW83 with inserted gene fragment A8	this work
pJW83_A9	pJW83 with inserted gene fragment A9	this work
pJW83_A10	pJW83 with inserted gene fragment A10	this work
pJW83_A11	pJW83 with inserted gene fragment A11	this work
pJW83_A12	pJW83 with inserted gene fragment A12	this work
pJW83_B2	pJW83 with inserted gene fragment B2	this work
pJW83_B3	pJW83 with inserted gene fragment B3	this work
pJW83_B4	pJW83 with inserted gene fragment B4	this work
pJW83_B5	pJW83 with inserted gene fragment B5	this work
pJW83_B6	pJW83 with inserted gene fragment B6	this work
pJW83_B7	pJW83 with inserted gene fragment B7	this work
pJW83_B8	pJW83 with inserted gene fragment B8	this work
pJW83_B9	pJW83 with inserted gene fragment B9	this work
pJW83_B10	pJW83 with inserted gene fragment B10	this work
pJW83_B11	pJW83 with inserted gene fragment B11	this work
pJW83_B12	pJW83 with inserted gene fragment B12	this work
pJW83_C1	pJW83 with inserted gene fragment C1	this work
pJW83_C3	pJW83 with inserted gene fragment C3	this work
pJW83_E9	pJW83 with inserted gene fragment E9	this work
pJW83_E10	pJW83 with inserted gene fragment E10	this work
pJW83_E11	pJW83 with inserted gene fragment E11	this work

pFF1_gxpS_bicA-A1T2C2_modi	pFF1_gxpS_bicA-A1T2C2 with T5,236C, T5,237G and A5,238T	this work
pJK10.2	ori 2µ, kanMX4, *araC*-P_{BAD}, ori ColA, kan^R, MCS	this work
pJK23	ori 2µ, kanMX4, *araC*-P_{BAD}, ori ColA, kan^R, *bicA*_A1T1CE2A2T2C3 *gxpS*_A3T3CE4A4T4CE5A5T5TE	this work
pJK23_modi	pJK23 with T8,377C and T8,378G	this work
pJK29	ori 2µ, kanMX4, *araC*-P_{BAD}, ori ColA, kan^R, *hcta*_A1T1CE2A2T2C3 *gxpS*_A3T3CE4A4T4CE5A5T5TE	this work
pJK29_modi	pJK29 with T8,464C and T8,465G	this work
pJK30	ori 2µ, kanMX4, *araC*-P_{BAD}, ori ColA, kan^R, *hcta*_A1T1CE2A2T2C3A3T3CE4 *gxpS*_A3T3CE4A4T4CE5A5T5TE	this work
pJK30_modi	pJK30 with T11,707C and T11,708G	this work

2.3.4. Oligonucleotides

All oligonucleotides used for topic C are listed in the following table.

Table 9. Oligonucleotides used for topic C. Primer inserted point-mutations are marked with bold nucleotide positions. For all constructs marked with an asterisk the digested pFF1 with EcoRI/SgsI was used for TAR cloning.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pFF1_gxpS_bicA-A1T2C2_modi*	JK-P1	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCC GTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTG G	pFF1_gxpS_bicA-A1T2C2
	JK-P174	GATTGCCGTCCGG**ACG**GTAGCGCACC	pFF1_gxpS_bicA-A1T2C2
	JK-P173	GGTGCGCTAC**CGT**CCGGACGGCAATC	pFF1_gxpS_bicA-A1T2C2
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
pJK23	JK-P1	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCC GTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTG G	*X. budapestensis* DSM 16342
	JK-P97	AGTAGGATACGGTTCTTCGGTCGCATTCCAAGTTTTCAGCAACAACTGGCG	*X. budapestensis* DSM 16342
	JK-P145	GCAGAGCGCCAGTTGTTGCTGAAAACTTGGAATGCGACCGAAGAACCGTAT CCTACTCAGGTATGTGTCCATCAGTTGTTTGAAC	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1

	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2
pJK23_modi	JK-P1	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCC GTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTG G	*X. budapestensis* DSM 16342
	JK-P97	AGTAGGATACGGTTCTTCGGTCGCATTCCAAGTTTTCAGCAACAACTGGCG	*X. budapestensis* DSM 16342
	JK-P145	GCAGAGCGCCAGTTGTTGCTGAAAACTTGGAATGCGACCGAAGAACCGTAT CCTACTCAGGTATGTGTCCATCAGTTGTTTGAAC	*P. luminescens* TTO1
	JK-P113	CCGG**CCG**ATAGCGAGCCAAATCTCCTGCCCGGTACATTCGGGCATTCG	*P. luminescens* TTO1
	JK-P114	CAGGAGATTTGGCTCGCTAT**CGG**CCGGACGGCAATCTGGTGTTTGTCGG	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2
pJK29	JK-P160	GGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGCGGGAAATA TTCTTAATGG	*X. miraniensis* DSM 17902
	JK-P161	GGTTCTTCGGTCGCATTCCAGGTTTTCAGCAACAACGTGCATTCTGC	*X. miraniensis* DSM 17902
	JK-P132	TGGAATGCGACCGAAGAACC	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1

	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2
pJK29_modi	JK-P160	GGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGCGGGAAATA TTCTTAATGG	*X. miraniensis* DSM 17902
	JK-P161	GGTTCTTCGGTCGCATTCCAGGTTTTCAGCAACAACGTGCATTCTGC	*X. miraniensis* DSM 17902
	JK-P132	TGGAATGCGACCGAAGAACC	*P. luminescens* TTO1
	JK-P113	CCGG**CCG**ATAGCGAGCCAAATCTCCTGCCCGGTACATTCGGGCATTCG	*P. luminescens* TTO1
	JK-P114	CAGGAGATTTGGCTCGCTAT**CGG**CCGGACGGCAATCTGGTGTTTGTCGG	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2
pJK30	JK-P160	GGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGCGGGAAATA TTCTTAATGG	*X. miraniensis* DSM 17902
	JK-P162	CGGTTCTTCGGTCGCATTCCAGATATTTAACAATAAGGTACGTTCGGCTTCC G	*X. miraniensis* DSM 17902
	JK-P132	TGGAATGCGACCGAAGAACC	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1

	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2
pJK30_modi	JK-P160	GGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGCGGGAAATA TTCTTAATGG	*X. miraniensis* DSM 17902
	JK-P162	CGGTTCTTCGGTCGCATTCCAGATATTTAACAATAAGGTACGTTCGGCTTCC G	*X. miraniensis* DSM 17902
	JK-P132	TGGAATGCGACCGAAGAACC	*P. luminescens* TTO1
	JK-P113	CCGG**CCG**ATAGCGAGCCAAATCTCCTGCCCGGTACATTCGGGCATTCG	*P. luminescens* TTO1
	JK-P114	CAGGAGATTTGGCTCGCTAT**CGG**CCGGACGGCAATCTGGTGTTTGTCGG	*P. luminescens* TTO1
	JK-P6	CCAAGTTTTTAACAACAATGTGC	*P. luminescens* TTO1
	JK-P7	CCATTTGTGTTATCGGTAGAGGATAGTGG	*P. luminescens* TTO1
	JK-P8	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCG CCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
	jw0061	TGACAATTAATCATCGGCTCG	pJK10.2
	jw0064	CATGGAATTCCTCCTGTTAGCC	pJK10.2

2.3.5. Gene fragments

All gene fragments used for topic C are listed in the following table.

Table 10. Gene fragments used for topic C. For all constructs the digested pJW83 with BpmI/SacII was used for HIFI cloning.

Plasmid	#Gene Fragments		
pJW83_A7	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCGGCTGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TATACCGGCGGTGTTATCCCGATCGGCTCTCCG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GTAGCAAACAAACGTCTCTACCTCCTGGATAACC	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	ACGGTCAGCCGCAGCCGCTCGGCGCCGTGGGC	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	GAACTGTATATCGGCGGTGTTGGCGTTGCACGT	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GGCTACCTGAACCGTCCGGAACTGACCGCTGAA	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	CGTTTTCTGAATGATCCGTTCTCTGATGAAACTAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCACGTATGTACCGTGCCGGCGATCTGGCACG	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	CTACCTGCCTGACGGTAACCTGGTTTTTGTTGGC	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGCAACGATCAACAGCTGAAAATCCGCGGTTTCC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	GCATTGAACTCGGTGAGATCGAAGCCCGCCTGG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TTGAACACTCTGAAGTATCTGAAGCACTGGTACT	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	GGCACTGGGCGACGGTCAGGATAAACGCCTGGT	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	AGCCTACGTGGTTGCGCTGGCAGATGACGGCCT	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	GGCGACCAAACTGCGCGAACACCTGTCTGACGT	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	ACTGCCAGATTATATGGTGCCGGCAGCTTTTGTC	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	CGTCTGGATGCTTTTCCGCTGACCCCGAACGGT	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA	AAGCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
	ACCACTTGGGATTGC		
pJW83_A8	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCTGCTGATT	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	ACACTGGTGGCGTTATCCCGATTGGTTCTCCGGT	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GGCGAACAAACGTCTGTACCTCCTGGATGACCAT	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	GGCCAGCCGGTGCCTTTCGGTGCAGTTGGTGAA	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	CTGTATATCGGCGGTGTAGGTGTTGCGCGTGGT	CATTAGCCACTATTTGGCGCG

	ACGCTGACACAAATTACCCAGTTTGATGT	TATCTGAATCGTCCGGAACTGACCGCTGAACGTT	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	TTCTGAACGACCCGTTCAGCGACGAAACGAACG	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CTCGTATGTACCGTGCGGGCAACCTCGCACGTT	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	ACCTGCCAGATGGTAACCTGGTCTTTGTGGGTC	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	GCAACGACCAGCAAGTTATTATCCGCGGTTTCCG	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	TATCGAACTGGGCGAAATTGAAGCACGTCTGGTT	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	GAACACTCTGAAGTATCCGAGGCTCTGGTTCTGG	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	CGCTGGGTAACGGTCAAGACAAACGTCTGGTCG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	CGTACGTAGTTGCTCTGGCAGATGACGGTCTGG	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	CGACCAAACTGCGCGAACATCTGTCTGACGTGC	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	TCCCAGACTACATGATTCCGGCAGCGTTCGTGC	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	GTCTGGACGCCTTCCCTCTGAGCCCGAACGGTA	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA	AGCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
	ACCACTTGGGATTGC		
pJW83_A9	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TACACGGGTGGCGTGATCCCGATCGGTTCCCCA	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GTGGCTAACAAACGCCTGTACCTCCTGGAAAACC	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	ATGGTCACCCTGTGCCACTGGGTGAAGTGGGTG	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	AACTGTACATCGGTGGCGTTGGTGTCGCGCGTG	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GTTACCTGAACCGCCCGGAACTGACGGCAGAAC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GCTTCCTGAACGACCCGTTCTCCGATGAGACCAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCGCGTATGTACCGTGCGGGCCACCTGGCACG	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	CTACCTCCCGGACGGCAACCTGGTTTTCGTTGGT	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGTAACGATCAACAGGTTAAGGTCCGCGGCTTTC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	GTATTGAAGTTGGTGAAATCGAAGCTCGCCTGGT	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TGAACACTCTGAAGTATCTGAAGCACTGGTACTC	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	GCGCTGGGTGACGGCCAGAACAAACGCCTGGTC	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	GCTTACGTAGTTGCCCTGGCTGACGATGGCCTG	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	GCTACCAAACTGCGTGAACACCTGTCCGATATCC	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	TGCCGGACTATATGGTGCCGGCCGCTTTCGTTC	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	GTCTGGATGCTTTTCCACTGACTCCTGATGGCAA	AGCAATGTTGCCGCTGATTGA
		ACTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT

	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC		
pJW83_A10	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAT TACACGGGCGGTGTCATCCCTATTGGTAGCCCG GTGGCAAACAAGCGTCTGTATCTCCTGGATGAG CACGGCGGTCCGGACCCGACCTGTACGGTGGG CGAACTGTATATTGGCGGTGTTGGCGTCGCACG TGGCTACCTGAATCGCCCGGAACTGACGGCAGA GCGTTTTCTGAACGACCCGTTCTCTGACGAAACT AACGCTCGCATGTACCGTGCCGGCCTCCTGTCT CGTTACGACCCGGACGGTAACCTGGTGTTCGTT GGCCGTAACGATCAACAGGTGAAAATGGACGGT TTCCGTCTGAACTCCTGGGAGATCGAAGCGCGT CTGGTGGAACACTCTGAGGTCAGCGAAGCGCTG GTGCTGGCTCTGGGTATCTGGCAATCCAAACGT CTGGTTGCGTACGTAGTTGCACTGGCGGATGAC GGCCTGGCTACCAAACTGCGCGAACACCTGTCT GATATCATTCCGAACCTCCTGATCCCGGCGGCTT TTGTGCGCCTGGACGCATTCCCGCTGACGTACC ATGGCAAACTGGATCGCCGGTCGCTGCCAGCAC CG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_A11	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT	GTGTGTGCAACCACTTGGGATTGCCCGGCTGAT TACACCGGCGGTGTGATTCCTATCGGTTCTCCAG TTGCTAATAAGCGTCTCTATCTCCTGGATGCCGC AAACACTCCGGTGGTTCTCGGCACTGTTGGTGAA CTGTACATCGGCGGTGTCGGCGTAGCTCGCGGT TATCTGAACCGCCCGGAGCTGACCGCGGAACGC TTTCTGAACGACCCGTTTTCTGACGAGACCAACG CACGTATGTACCGTGCTGGTGACTCTGGTGAGTA CCTGCCGGATGGTAACCTCGTGTTCGTTGGTCG CAATGATCAGCAACAGAAAATCTGCGGTTTCTGG TACACTTTCGGTGAGATCGAAGCACGCCTCGTC	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC

	TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GAACACAGCGAAGTGTCCGAAGCTCTGGTACTC GCTCTGGGTCCAACTTCTGACAAACGTCTGGTTG CGTACGTTGTGGCGCTGGCGGACGATGGTCTGG CTACCAAGCTCCGTGAACACCTCTCTGACATCTA CAAAGACAAGAAAATCCCTGCGGCATTCGTACGT CTGGACGCGTTCCCGCTGGGTCCGAACACGAAA CTGGATCGCCGGTCGCTGCCAGCACCG	TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_A12	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCGGCGGAT TATACCGGTGGCGTAATCCCGATCGGTTCTCCG GTAGCGAATAAACGCCTCTACCTCCTGATCGAAG GCAAAATCCCTGTGCCGTACTTCACTGTTGGCGA ACTGTACATCGGCGGTGTAGGCGTTGCACGTGG TTATCTGAACCGTCCAGAACTCACTGCTGAACGT TTCCTGAACGATCCGTTCTCTGATGAAACCAACG CGCGTATGTACCGTGCTGGCGATGAACAGGACT ATCTGCCGGATGGTAATCTGGTTTTCGTAGGTCG CAACGATCAACAGGTTAACATCCACGGTTTTCGT CAAGAACTGGGCGAGATCGAAGCGCGTCTCGTG GAGCACTCCGAGGTAAGCGAAGCACTCGTTCTG GCTCTGGGCGACCGTCAGCTCAAATGTCTGGTC GCGTACGTGGTCGCTCTCGCGGATGACGGTCTG GCGACTAAACTGCGCGAACACCTGTCCGATATTC GTCCGGATGCGGATATGCCGGCCGCGTTTGTCC GCCTGGACGCTTTCCCGCTGACCCCTAAAGGTC ACCTGGATCGCCGGTCGCTGCCAGCACCG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B2	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAC TATACCGGTGGCGTAATCCCGATCGGCTCTCCG GTGGCCAACAAACGTCTGTATCTCCTGGTTGAAC ACCGTTGGGTTGTGCGTTGCCGCACTGTGGGTG AACTGTACATCGGCGGTGTAGGTGTAGCGCGTG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG

	ACGCTGACACAAATTACCCAGTTTGATGT	GTTACCTGAATCGCCCGGAACTGACGGCTGAGC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GTTTCCTGAACGACCCGTTCTCTGATGAAACGAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	TGCACGCATGTACCGTGCGGGCGACTGTGCGCC	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	GTACTGTCCGGATGGCAATCTGGTTTTCGTGGG	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CCGCAACGATCAGCAAGTTTGTATCATGAAATTT	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	TGCATCGAACCGCGTCCTATTGAGGCACGTCTG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	GTTGAGCATTCTGAGGTGAGCGAAGCTCTGGTT	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	CTGGCTCTGGGCGACCGTCAGGACTGTTGCCTC	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	GTAGCATACGTGGTAGCCCTGGCGGATGACGGT	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	CTGGCGACCAAACTGCGCGAACACCTCTCTGAT	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	ATCTGTTGCGACGCTAAGATCCCGGCTGCATTCG	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	TACGCCTGGATGCTTTCCCGCTCACTCCGAACC	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA	GTTGCCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
	ACCACTTGGGATTGC		
pJW83_B3	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TACACCGGTGGCGTCATCCCGATTGGCAGCCCG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GTGGCTAACAAACGTCTGTACCTCCTGGACGAG	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	CACCGTTGCTGTCCATGTCGCGAAACGGTTGGT	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	GAACTGTACATCGGTGGCGTTGGCGTTGCGCGT	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GGCTATCTGAATCGCCCGGAACTGACTGCGGAA	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	CGTTTCCTGAACGATCCGTTCTCTGATGAAACCA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	ACGCACGCATGTATCGTGCAGGTGATTGCGCCG	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	GCTATTGTCCTGATGGCAACCTGGTGTTCGTAGG	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	TCGCAACGATCAACAGGTTTGCCCTTGCCGCTTT	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	CGTATTCCTCCGTGGGAAATTGAGGCGCGTCTG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	GTAGAACACTCTGAGGTAAGCGAAGCACTGGTC	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	CTGGCACTGGGTGACCGCCAGGATTGCCCGCTG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	GTTGCGTACGTGGTTGCGCTCGCTGATGACGGT	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	CTGGCCACCAAGCTGCGTGAACATCTGAGCGAC	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	ATCTGTTGCCGCTACCCGATCCCGGCCGCATTC	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	GTTCGTCTGGACGCATTCCCGCTGACTTGTAACC	AGCAATGTTGCCGCTGATTGA
		GTAAACTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT

	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC		
pJW83_B4	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCTGCTGATT	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	ACACTGGTGGCGTGATCCCTATCGGTAGCCCGG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	TTGCTAACAAACGTCTCTATCTCCTGCAGGTGCA	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	GCTGCAGCTGTGCCCGCTGGGTACGGTGGGCG	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	AACTGTATATCGGTGGCGTAGGTGTAGCTCGTG	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GTTACCTGAACCGTCCGGAACTGACCGCTGAAC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GTTTCCTGAACGACCCTTTTTCTGACGAAACGAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCGCGTATGTATCGTGCAGGTGATCTGGCGAA	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	AATGCTGCCGGACGGTAATCTGGTTTTCGTTGGT	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGTAACGATCAACAGGTTAAAATCCGTGGTTATC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	GTGCGGAACCTGGCGAAATTGAAGCTCGTCTGG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TTGAACACAGCGAGGTGAGCGAAGCTCTGGTTC	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	TGGCTCTGGGTGACGGTGATGACGATCCGCTGG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	TAGCTTACGTAGTGGCTCTGGCCGACGATGGTC	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	TGGCTACTAAACTGCGTGAACATCTGTCTGACTC	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	CCTGCCGGACTACATGATCCCGGCCGCGTTCGT	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	GCGTCTGGACGCTTTCCCGCTGACCGTAAACGG	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	TAAGCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
pJW83_B5	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCGGCGGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TATACCGGTGGCGTTATCCCGATCGGTAGCCCT	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GTTGCGAACAAGCGCCTGTATCTCCTGGACAAG	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	GACATCCAGCTGTACCCGCTGGGTACCGTTGGT	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	GAACTGTACATCGGCGGTGTAGGTGTAGCACGT	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GGCTATCTGAACCGCCCGGAACTGACCGCTGAG	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	CGCTTTCTGAACGACCCTTTCAGCGACGAAACCA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	ACGCCCGTATGTACCGTGCAGGCAACCTGGCTA	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	AAATGCTGCCGGACGGCAATCTGGTCTTTGTGG	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	GTCGCAACGATCAACAGGTATATGTTCGTGGTTA	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	CCGCATCGAACCTGGTGAAATCGAAGCACGTCT	TGGGCCTGTCGGTGCAAACGC

	TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GGTGGAGCACTCTGAAGTGAGCGAAGCGCTGGT ACTGGCGCTGGGCGACGGTCAACACGATCCACT GGTAGCCTATGTCGTTGCACTGGCGGACGATGG TCTGGCCACCAAACTGCGCGAACACCTGTCTGA CTCCCTGCCGAACTACATGATCCCGGCTGCGTTT GTACGTCTGGATGCATTCCCTCTGACTGTTCAGG GTAAACTCGATCGCCGGTCGCTGCCAGCACCG	TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B6	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCGGCTGAC TATACTGGCGGTGTAATCCCGATTGGTAGCCCTG TGGCGAACAAACGTCTGTATCTCCTGTCCGAGCA CGCGCAGATCTGTCCGCTGGGCGCTGTAGGTGA ACTGTATATTGGCGGTGTAGGCGTAGCACGCGG CTACCTGAACCGCCCGGAACTGACCGCAGAACG TTTCCTGAACGATCCATTTAGCGATGAAACTAAC GCGCGCATGTATCGCGCGGGTGATCTGGCGAAG CATCTGCCTGATGGTAACCTGGTATTCGTTGGTC GTAACGACCAACAGGTTAAAATCCGTGGTATTAA AATCGAACCGGGCGAAATCGAAGCCCGCCTGGT AGAACATAGCGAGGTAAGCGAAGCGCTGGTGCT GGCGCTGGGTGACGGCAACGATGACCCTCTGGT TGCTTACGTGGTAGCGCTGGCTGATGACGGTCT GGCGACCAAGCTGCGTGAGCATCTGTCTGATAA CCTGCCGGACTATATGATTCCGGCTGCGTTCGT GCGCCTGGATGCCTTTCCGCTGACCGTACACGG CAAACTGGATCGCCGGTCGCTGCCAGCACCG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B7	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG	GTGTGTGCAACCACTTGGGATTGCCCGGCTGAT TACACTGGCGGTGTAATCCCGATCGGTTCCCCG GTTGCGAACAAGCGCCTGTACCTCCTGTACGAAT ACGCTCAGGGCGAACCGCTGGTCACCGTAGGTG AGCTGTACATCGGTGGCGTGGGTGTTGCTCGTG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG

	ACGCTGACACAAATTACCCAGTTTGATGT	GTTACCTGAACCGTCCAGAACTGACCGCAGAAC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GTTTTCTGAATGACCCGTTCTCTGATGAGACGAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCCCGTATGTATCGTGCAGGTGACCCGGCTTG	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	CATCCTGCCAGATGGCAACCTGGTTTTTGTGGGT	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGCAATGATCAACAGGTTAAAGCGCGTGGCTTTG	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	AACCGTTTATCTACGAGATTGAGGCGCGCCTGGT	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TGAACACAGCGAGGTGAGCGAAGCTCTGGTGCT	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	GGCCCTGGGTCACGGCCAACAGAAACCGCTGGT	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	AGCATATGTTGTCGCGCTGGCCGACGATGGTCT	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	GGCCACCAAACTGCGTGAACACCTGTCCGATCA	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	TAGCCCGGATATGATGCACCCGGCGGCTTTTGT	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	GCGTCTGGACGCCTTCCCGCTGACCATCAACGG	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA	CTGCCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
	ACCACTTGGGATTGC		
pJW83_B8	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCAGCGGAT	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TACACCGGTGGCGTGATCCCTATCGGTTCCCCG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GTTGCAAACAAACGTCTGTACCTCCTGGACACCG	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	GCCGTCAGGTGGTTCCTAGCTATACCGTTGGCG	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	AGCTGTACATTGGCGGTGTTGGCGTGGCGCGTG	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	GCTATCTGAATCGTCCGGAACTGACCGCGGAAC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GCTTTCTGAACGACCCGTTCAGCGACGAAACCAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCGCGTATGTACCGCGCCGGTGACCTCTCTAA	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	ATACGGTCCGGATGGCAACCTGGTATTCGTGGG	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CCGTAACGACCAACAGGTATGTATCCGCGGTTTC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	CGCGCGGAAATCTGGTACATCGAAGCGCGTCTG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	GTCGAACATTCTGAGGTTTCTGAAGCGCTGGTCC	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	TGGCTCTGGGCCAGGGCCAGGACGGCAACCTG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	GTGGCGTACGTAGTTGCGCTGGCGGATGACGGC	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	CTGGCTACCAAACTGCGTGAACACCTGAGCGAT	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	ATCGTATTTGACACCATGGTCCCGGCTGCCTTCG	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	TACGTCTGGACGCCTTTCCGCTGAACCCGCCAG	AGCAATGTTGCCGCTGATTGA
		GCAAACTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT

	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC		
pJW83_B9	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCAGCAGAC TACACCGGTGGCGTTATTCCTATCGGTTCTCCGG TCGCGAACAAGCGTCTGTACCTGCTCGACGAGG TGGGTAGCCCGTTCCCGCTGAAGTTTGTAGGTG AACTGTATATCGGTGGCGTTGGCGTTGCCCGTG GTTATCTCAACCGTCCAGAACTGACCGCTGAACG TTTCCTGAACGATCCGTTTAGCGACGAAACCAAC GCACGCATGTACCGTGCTGGTGATATGGGTGCA TACCTGCCGGATGGTAACCTGGTTTTCGTGGGTC GCAACGACCAACAGGTTCGTATCTACGGTTTCTG GGAGGAATGGGGCGAAATCGAGGCGCGCCTGG TTGAACACAGCGAAGTTTCTGAAGCCCTGGTGCT GGCCCTGGGCGTGGGCCAGATGAAATGGCTGGT TGCATACGTAGTGGCACTGGCGGACGATGGTCT GGCGACTAAACTGCGTGAGCACCTGTCTGATAA CAATCCTTGGTATTCCATCCCGGCTGCGTTTGTT CGTCTGGATGCATTCCCTCTGACTCGTAACCATA AACTGGATCGCCGGTCGCTGCCAGCACCG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B10	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT	GTGTGTGCAACCACTTGGGATTGCCCGGCCGAT TATACTGGCGGTGTGATCCCAATCGGTTCCCCG GTTGCTAACAAACGCCTGTATCTCCTGAAAGAAG TCCGTTGTCCGGTAGCATGCAAAACGGTGGGCG AGCTGTATATCGGTGGCGTAGGCGTCGCTCGTG GCTATCTCAACCGCCCGGAACTGACCGCCGAGC GTTTCCTGAACGATCCGTTCTCTGACGAGACTAA CGCTCGCATGTACCGTGCAGGTATTTGTGCTCGT TATTGCCCAGACGGCAACCTGGTTTTTGTCGGCC GTAACGATCAACAGGTATGTATCCGTCGCTTCCG CATCTGCCGCAAGGGCATCGAAGCTCGTCTGGT	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC

	TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	TGAACACTCTGAAGTTAGCGAAGCCCTGGTACTG GCACTGGGCATTCGTCAACCGAAACGTCTCGTG GCTTATGTCGTTGCTCTGGCGGACGATGGCCTG GCAACCAAGCTGCGCGAACATCTGTCCGATATCT GTTGCGTGGAGATGATCCCGGCCGCGTTTGTAC GTCTGGACGCCTTCCCTCTGACTCCTAACGAATG TCTCGATCGCCGGTCGCTGCCAGCACCG	TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B11	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAT TACACTGGTGGCGTGATCCCGATTGGTTCTCCG GTTGCGAACAAGCGTCTGTATCTCCTGTGGGAAC CGCGCTGTCCGGTGGTATGCGAAACCGTAGGTG AACTGTATATCGGCGGTGTCGGTGTTGCGCGCG GTTACCTGAACCGTCCGGAGCTCACGGCTGAAC GTTTTCTGAACGACCCGTTCTCTGATGAGACCAA CGCACGTATGTACCGCGCGGGTGATTGTCGCCG TTACTGCCCGGATGGTAACCTGGTATTTGTTGGT CGCAACGACCAACAGGTATGCATTTGCGGTTTCG CTATTTGTATGTGGGAAATCGAGGCGCGCCTGG TTGAGCACTCCGAAGTTTCCGAAGCACTGGTTCT GGCGCTGGGTGTTCGTGTAGATAAGCGTCTGGT AGCCTATGTGGTAGCCCTGGCGGACGATGGCCT GGCGACCAAGCTGCGTGAACACCTGTCTGATATT TGTTGCGATGCGGGCATCCCGGCGGCATTCGTG CGCCTGGACGCCTTCCCGCTGACCCCGAACAAG TGCCTGGATCGCCGGTCGCTGCCAGCACCG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_B12	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG	GTGTGTGCAACCACTTGGGATTGCCCGGCTGAC TACACCGGCGGTGTAATCCCAATCGGTTCTCCAG TAGCGAACAAGCGTCTGTACCTCCTGGAAGAGC ATCGCCCACCGAAAGAATGCCGTACCGTTGGCG AACTGTACATCGGCGGTGTTGGCGTGGCTCGTG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG

	ACGCTGACACAAATTACCCAGTTTGATGT	GTTACCTGAACCGTCCGGAACTGACCGCGGAGC	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	GTTTCCTGAACGACCCGTTTTCTGACGAAACGAA	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCTCGTATGTATCGTGCCGGTGATTGTTGCCG	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	CTACTGTCCGGATGGTAACCTGGTGTTCGTTGGT	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGTAACGACCAGCAAGTTAAAATCCCGTGGTTCC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	CTATTGAATGCCGTGAACGTGAGGCGCGTCTGG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TAGAGCACAGCGAAGTAAGCGAGGCCCTGGTCC	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	TGGCGCTGGGTGTTCGTCAGCCGAAACGTCTGG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	TAGCATATGTAGTGGCACTGGCGGATGACGGCC	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	TGGCGACGAAACTGCGCGAACACCTGAGCGATA	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	TCTGCTGTGACGAGCGTATCCCGGCGGCATTCG	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	TACGTCTGGATGCGTTCCCGCTGACCCCGAACC	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA	GCTGCCTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
	ACCACTTGGGATTGC		
pJW83_C1	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCAGCGGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TATACCGGCGGTGTTATCCCGATCGGTTCTCCTG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	TGGCAAACAAACGCCTGTACCTCCTGGATGCGTT	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	TGGTCAGCCGGTACCGCTGGGTACCGTGGGTGA	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	GCTGTATATCGGTGGCGTTGGTGTGGCACGCGG	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	CTATCTGAACCGTCCTGAACTGACCGCGGAGCG	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	TTTCCTGAACGATCCGTTCTCTGACGAAACTAAC	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	GCGCGCATGTATCGTGCGGGTGACCTGGCGCGT	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	TATCGTCCGGACGGTAACCTGGTATTCGTGGGC	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	CGTAACGACCAACAGGTTAAAATCCGTGGCATCC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	GTATCGAACCGGGTGAAATCGAAGCGCGCCTGG	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	TGGAACATTCCGAGGTTTCCGAAGCCCTGGTACT	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	CGCTCTCGGCGATGGCCAGGATAAACGTCTGGT	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	TGCATACGTTGTGGCTCTGGCCGACGATGGTCT	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	GGCTACCAAACTGCGTGAGCACCTGTCTGACAT	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	CCTGCCGGACTACATGGTGCCGGCAGCGTTTGT	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	CCGTCTGGACGCGTTCCCGCTGACCCCAAACGG	AGCAATGTTGCCGCTGATTGA
		TAAACTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT

	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC		
pJW83_C3	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCAGCAGATT	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	ATACCGGCGGTGTAATCCCGATCGGTTCTCCAGT	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	GGCGAACAAACGTCTGTATCTCCTGGATGCTCGT	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	GGTCAGCCAGTACCGCTGGGTAAAGTTGGTGAA	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	CTGTACATCGGTGGCGTAGGCGTGGCCCGTGGT	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	TATCTGAACCGTCCTGAGCTGACCGCAGAGCGT	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	TTCCTGAACGACCCGTTCTCTGATGAAACCAATG	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	CGCGTATGTACCGCGCGGGTCACCTGGCGCGTT	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	ATCGCCCGGATGGCAACCTGGTATTCGTTGGTC	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	GCAACGACCAGCAAGTGAAGATCCGTGGTTTCC	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	GTATTTTCCCGGGTGAAATCGAAGCTCGCCTGGT	TGGGCCTGTCGGTGCAAACGC
	TACCTGGAAGAGCAGGCGATAACCCATG	CGAACACAGCGAAGTGTCCGAAGCGCTGGTGCT	TATTCCAGCATCCGACATTAAG
	CCTGTTTGACACCGGCTATGTTTCATGAC	GGCGCTGGGCGACGGCCAGGACAAACGCCTGG	TGTATTGGCCCAATCTTTGGT
	GGCACCGATTTACCGGCGATAGCGATAA	TTGCCTACGTAGTCGCTCTGGCAGATGACGGCC	GCCGCATCGTGAAATTAGCGT
	AACCAACCTTAATATTTGCAGGAGAAGCG	TGGCGACCAAACTGCGCGAACACCTGTCCGACA	GCCTGATAACGGTATTACCGC
	CCGAGTCCCGCGCTATTTCAGGCACTGT	TCCTGCCGGATTACATGGTGCCAGCTGCCTTCGT	GGATACCACTGTGCTGACGCC
	GCAGCCGGGCCGATTTGTTTAACGCGTA	TCGCCTGGATGCTTTCCCTCTGACTCCGGATGGT	AGCAATGTTGCCGCTGATTGA
	TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	AAACTGGATCGCCGGTCGCTGCCAGCACCG	TCTGACTCAAGCT
pJW83_E9	GAATACCTTGCTGGAGCAGCCAGACCAT	GTGTGTGCAACCACTTGGGATTGCCCGGCGGAC	GATCGCCGGTCGCTGCCAGC
	AATCCGCAAGTCTCTGGACTGACCCCAC	TACACTGGCGGTGTTATCCCTATTGGTTCTCCGG	ACCGGGGGAGGATGCCTTTG
	AGCATCTGGCCTATGTGATTTACACTTCC	TAGCAAACAAACGTCTGTATCTCCTGGACACCCA	CTCGCCAAGCTTACCAAGCGC
	GGCTCAACCGGTAGACCAAAAGGTGTGA	CGGTCAGCCGAAACCACTGGGTAGCGTTGGTGA	CACAAGGGGAAATTGAGATAG
	TGATTGAACATCGTAGTGTGGTTAATTTG	ACTGTACATCGGCGGTGTCGGTGTAGCGCGTGG	CATTAGCCACTATTTGGCGCG
	ACGCTGACACAAATTACCCAGTTTGATGT	TTACCTGAACCGTCCTGAGCTGACCGCAGAACG	AATTGCTGAATGTTGAACAGG
	TTGCGCCACCAGCCGAATGTTGCAGTTT	TTTCCTGAACGATCCTTTCTCCGATGAGACTAAC	TTGGCCGACATGACAGTTTCT
	GCCTCATTTGGGTTCGATGCCAGTGTCTG	GCACGCATGTATCGTGCGGGCGATATGGCCCGT	TTGCCTTGGGCGGTCACTCGC
	GGAAATCATGATGGCGCTGAGTTGTGGG	TATCGCCCAGACGGTAATCTGGTTTTTGTCGGTC	TGTTGGCAGTCAGGATGATCG
	GCTATGCTGGTTATTCCTACAGAAACCGT	GTAACGACCAACAGGTTAAAATCCGTGGCTTCCG	AACGTTTACGGCGTATAGGAT
	GCGGCAAGATCCACAGCGCCTTTGGCGT	CTACGAGGCCGGTGAAATCGAAGCTCGCCTGGT	TGGGCCTGTCGGTGCAAACGC

	TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	TGAACATTCTGAAGTTTCTGAAGCGCTGGTGCTG GCGCTGGGTGATGGTCAGGATAAACGTCTGGTG GCGTACGTCGTAGCTCTGGCGGATGACGGCCTG GCGACCAAGCTGCGTGAACACCTGTCTGATGTC CTGCCGGACTACATGGTGCCAGCCGCGTTCGTT CGCCTGGATGCTTTCCCGCTGACCCCAAACGGC AAACTGGATCGCCGGTCGCTGCCAGCACCG	TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_E10	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA TGATTGAACATCGTAGTGTGGTTAATTTG ACGCTGACACAAATTACCCAGTTTGATGT TTGCGCCACCAGCCGAATGTTGCAGTTT GCCTCATTTGGGTTCGATGCCAGTGTCTG GGAAATCATGATGGCGCTGAGTTGTGGG GCTATGCTGGTTATTCCTACAGAAACCGT GCGGCAAGATCCACAGCGCCTTTGGCGT TACCTGGAAGAGCAGGCGATAACCCATG CCTGTTTGACACCGGCTATGTTTCATGAC GGCACCGATTTACCGGCGATAGCGATAA AACCAACCTTAATATTTGCAGGAGAAGCG CCGAGTCCCGCGCTATTTCAGGCACTGT GCAGCCGGGCCGATTTGTTTAACGCGTA TGGGCCGACGGAAATCACCGTGTGTGCA ACCACTTGGGATTGC	GTGTGTGCAACCACTTGGGATTGCCCAGCTGAC TACACTGGTGGCGTAATCCCGATTGGCTCCCCG GTGGCCAATAAACGTCTGTATCTCCTGGATTCTT GCGGCCGTCCGGTTCCGCTGGGTTCCGTCGGC GAACTGTATATCGGCGGTGTTGGCGTTGCACGT GGTTATCTCAACCGTCCGGAACTCACCGCTGAAC GCTTTCTGAATGATCCGTTTAGCGATGAAACGAA CGCTCGCATGTACCGCGCGGGCGACCACGCCC GCTACCGTCCGGACGGCAACCTGGTCTTCGTTG GCCGTAACGATCAGCAAGTTAAAATCCGTGGTTT CCGCGTTGAAAAATCCGAGATTGAAGCCCGTCT GGTTGAGCATTCCGAAGTTTCTGAAGCACTGGTC CTGGCGCTGGGCAACGGTCAGGATAAACGTCTG GTCGCTTACGTTGTCGCTCTGGCTGATGACGGT CTGGCGACTAAACTGCGTGAGCATCTCTCTGAC GTGCTGCCGGATTACATGATTCCGGCAGCTTTTG TTCGCCTGGACGCATTCCCGCTGTCCCCGAACG GCAAGCTGGATCGCCGGTCGCTGCCAGCACCG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG CATTAGCCACTATTTGGCGCG AATTGCTGAATGTTGAACAGG TTGGCCGACATGACAGTTTCT TTGCCTTGGGCGGTCACTCGC TGTTGGCAGTCAGGATGATCG AACGTTTACGGCGTATAGGAT TGGGCCTGTCGGTGCAAACGC TATTCCAGCATCCGACATTAAG TGTATTGGCCCAATCTTTGGT GCCGCATCGTGAAATTAGCGT GCCTGATAACGGTATTACCGC GGATACCACTGTGCTGACGCC AGCAATGTTGCCGCTGATTGA TCTGACTCAAGCT
pJW83_E11	GAATACCTTGCTGGAGCAGCCAGACCAT AATCCGCAAGTCTCTGGACTGACCCCAC AGCATCTGGCCTATGTGATTTACACTTCC GGCTCAACCGGTAGACCAAAAGGTGTGA	GTGTGTGCAACCACTTGGGATTGCCCGGCAGAT TACACCGGCGGTGTAATTCCAATTGGCTCTCCGG TGGCCAATAAACGTCTGTACCTCCTGGACAAACA CGGTCAGGGCGTCCCGCTGGGCAGCGTAGGTG	GATCGCCGGTCGCTGCCAGC ACCGGGGGAGGATGCCTTTG CTCGCCAAGCTTACCAAGCGC CACAAGGGGAAATTGAGATAG

TGATTGAACATCGTAGTGTGGTTAATTTG
ACGCTGACACAAATTACCCAGTTTGATGT
TTGCGCCACCAGCCGAATGTTGCAGTTT
GCCTCATTTGGGTTCGATGCCAGTGTCTG
GGAAATCATGATGGCGCTGAGTTGTGGG
GCTATGCTGGTTATTCCTACAGAAACCGT
GCGGCAAGATCCACAGCGCCTTTGGCGT
TACCTGGAAGAGCAGGCGATAACCCATG
CCTGTTTGACACCGGCTATGTTTCATGAC
GGCACCGATTTACCGGCGATAGCGATAA
AACCAACCTTAATATTTGCAGGAGAAGCG
CCGAGTCCCGCGCTATTTCAGGCACTGT
GCAGCCGGGCCGATTTGTTTAACGCGTA
TGGGCCGACGGAAATCACCGTGTGTGCA
ACCACTTGGGATTGC

AACTGTATATCGGTGGCGTGGGTGTCGCACGTG
GTTACCTCAACCGCCCGGAGCTGACCGCTGAAC
GTTTCCTGAACGACCCGTTCTCTGACGAGACCAA
CGCACGCATGTACCGTGCCGGTCATATGGCGCG
TTACCGTCCGGACGGTAACCTGGTGTTTGTTGGT
CGCAATGACCAACAGGTTAAAACCCGCGGCTTC
CGCATTGAACTGTCTGAGATCGAAGCGCGCCTG
GTAGAACACTCCGAGGTCAGCGAAGCGCTGGTG
CTGGCACTGGGCGATGGCCAAAACAAACGTCTG
GTGGCGTACGTAGTTGCTCTGGCCGATGACGGT
CTGGCAACCAAACTCCGTGAACATCTGTCTGATA
TCCTCCCGGACTACATGGTGCCGGCTGCGTTCG
TGCGCCTGGATGCCTTCCCGCTGACCCCGGATG
GTAAGCTGGATCGCCGGTCGCTGCCAGCACCG

CATTAGCCACTATTTGGCGCG
AATTGCTGAATGTTGAACAGG
TTGGCCGACATGACAGTTTCT
TTGCCTTGGGCGGTCACTCGC
TGTTGGCAGTCAGGATGATCG
AACGTTTACGGCGTATAGGAT
TGGGCCTGTCGGTGCAAACGC
TATTCCAGCATCCGACATTAAG
TGTATTGGCCCAATCTTTGGT
GCCGCATCGTGAAATTAGCGT
GCCTGATAACGGTATTACCGC
GGATACCACTGTGCTGACGCC
AGCAATGTTGCCGCTGATTGA
TCTGACTCAAGCT

2.3.6. Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus* and *Photorhabdus* strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. All PCRs were performed with oligonucleotides obtained from Eurofins Genomics (Tab. 9). NRPS fragments for HiFi cloning (NEB) or TAR cloning[151] were amplified with primers coding for the respective homology arms (20-120 bp) in a two-step PCR program or Gene Fragments were obtained from Integrated DNA Technologies, Inc. (IDT) (Tab. 10). Plasmid backbones were generated by digest with EcoRI/SgsI (pFF1), digest with BpmI/SacII (pJW83) or amplified using the primer jw0061/jw0064 (Tab. 9-10). Polymerases Phusion High-Fidelity DNA polymerase (Thermo Scientific), Q5 High-Fidelity DNA polymerase (New England BioLabs), and Velocity DNA polymerase (Bioline) were used according to the manufacturers' instructions. DNA purification was performed using Invisorb Fragment CleanUp or MSB Spin PCRapace Kits (stratec molecular). All generated plasmids (Tab. 8) were introduced into *E. coli* DH10B::*mtaA*[151] by either electroporation or heat shock. Each NRPS (subunit) was under the control of a P_{BAD} promotor for peptide production. Plasmid isolation from *E. coli* was achieved with the Invisorb Spin Plasmid Mini Two Kit (stratec molecular).

2.3.7. Heterologous expression of NRPS templates and LC-MS analysis

Constructed plasmids were transformed into *E. coli* DH10B::*mtaA*[151]. Cells were grown overnight in LB-medium containing the necessary antibiotics (50 µg/ml kanamycin, or 34 µg/ml chloramphenicol). 100 µl of an overnight culture were used for inoculation of 10 ml LB-cultures supplemented with the respective antibiotics as selection markers and additionally containing 0.002 mg/ml *L*-arabinose and 2 % (v/v) XAD-16. After incubation for 72 h at 22 °C the XAD-16 was harvested. One culture volume methanol was added and incubated for 30 min at RT. The organic phase was filtrated, and a sample was taken of the cleared extract. After centrifugation (17,000 x *g*, 20 min) the methanol extracts were used for HPLC/MS

analysis. All measurements were performed by using an Ultimate 3000 LC system (Dionex) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml/min using acetonitrile (ACN) and water containing 0.1 % formic acid (v/v) in a gradient ranging from 5-95 % of ACN over 16 min (40 °C) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. High-resolution mass spectra were obtained on an Ultimate 3000 RSLC (Dionex) coupled to an Impact II qTOF (Bruker) equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

2.3.8. High throughput cloning and heterologous expression

The generated primary sequences with CAopt.py were back translated using codon usage table 'Escherichia coli K12 (high)' (EMBOSS package), respective homology arms (20-30 bp) were added, and nucleotide sequences were purchased from Twist Bioscience (Tab. 10). Plasmid assembly with HiFi cloning (NEB) was performed according to the manufacturers' instructions into pJW83 (Tab. 8-9) in a 96well plate. 40 µl of competent *E. coli* DH10B::*mtaA*[151] cells were added in each well and heatshock was performed. Clones were grown overnight in 2 ml LB-medium containing the necessary antibiotics (50 µg/ml kanamycin, or 34 µg/ml chloramphenicol) in a 96-deep-well plate. Plasmid isolation from *E. coli* was achieved with the PureYield™ Plasmid Miniprep System (Promega). Desired plasmid combinations were introduced into *E. coli* DH10B::*mtaA*[151] by electroporation. 40 µl of an overnight culture were used for inoculation of 4 ml LB-cultures in a 24-deep-well plate supplemented with the respective antibiotics as selection markers and additionally containing 0.002 mg/ml *L*-arabinose and 2 % (v/v) XAD-16. After incubation for 72 h at 22 °C the XAD-16 was harvested. One culture volume methanol was added and incubated for 30 min at RT. The organic phase was filtrated, and a sample was taken of the cleared extract. After centrifugation (17,000 x *g*, 20 min) the methanol extracts were used for HPLC/MS

analysis. All measurements were performed by using an Ultimate 3000 LC system (Dionex) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml/min using acetonitrile (ACN) and water containing 0.1 % formic acid (v/v) in a gradient ranging from 5-95 % of ACN over 16 min (40 °C) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. High-resolution mass spectra were obtained on an Ultimate 3000 RSLC (Dionex) coupled to an Impact II qTOF (Bruker) equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

2.3.9. Homology modelling

The homology-modelling was performed with the homology modelling algorithm within MOE (Molecular Operating Environment). This process undergoes an (I) initial partial geometry, where all coordinates are copied if residue identity is conserved. Next, a (II) Boltzmann-weighted randomized sampling, which (IIa) consists of a backbone fragments collection from a high-resolution structural database, and alternative side chain conformations assembly from an extensive rotamer library for non-identical residues and (IIb) a creation of independent number models based upon loop and side chain placements scored by a contact energy function[160].

For homology modelling, the C-A didomains within the 2.60 Å crystal structure of SrfA-C (PDB-ID: 2SVQ)[58] was selected as homologous-protein-templates. The identity of the primary sequence of the bicornutine (BicA) CE2 from *Xenorhabdus budapestensis* DSM 16342[163] and GameXPeptide (GxpS) A2 from *Photorhabdus laumondii subsp. laumondii* TTO1[45] (BicA_CE2-GxpS_A2) to the SrfA-C template was 28.8 %, and the final model a *root-mean-square deviation* (RMSD) of 2.1 Å respectively, in comparison to the template structures.

2.3.10. C-A interface optimizer – CAopt.py algorithm

Variant sequences for C-A interface optimization were generated using a custom python script that automated identification of reference specified interface regions, randomization of amino acid sequence at these positions, and scoring based on physiochemical similarity to the native interface of the natural cluster. Default refence interface sequences and regions are provided along with the CAopt.py script at the git repository (https://github.com/malanjary-wur/NRPS-CAopt; accessed January 31, 2022). Alternate reference and regions are also supported through the script options. The default reference file includes C-A domain interface sequences form four varying specie's clusters from which corresponding crystal structures have been solved: EntF (PDB ID: 5T3D)[60], AB3404 (PDB ID: 4ZXH)[60], SrfA-C (PDB ID: 2VSQ)[58], in addition to the build BicA_CE2-GxpS_A2 homology model. Query sequences are then mapped to these reference locations using MAFFT with local pair and maximum 1000 iterations. The input for the script assumes three sequences in FASTA format comprised of natural cluster interfaces from the two source clusters (cluster A/B) and the engineered cluster interface which corresponds to these regions (cluster C). Amino acid sequences on these interface locations, and specified by user input, are then mutated using a sequential replacement at random sites. This results in saving of single to multi-site replacements of varying sequences. The pool of sequences is then scored using an estimation of physiochemical similarity to the corresponding locations of the natural interface partner. This is rapidly done through a sequence embedding using the principle component-based *E*-descriptors as described in Venkatarajan and Braun *et al.*[164].The Euclidian distance of the corresponding points in this vector space are then used to calculate native similarity.

2.3.11. Peptide quantification

The absolute production titers of selected peptides were calculated with calibration curves based on pure synthetic **32**, (for quantification of **32** and **33**), **34** (for quantification of **34** and **35**), **36** (for quantification of **36**), **37** (for quantification of **37**),

38 (for quantification of **38** and **39**). Therefore, the pure compounds were prepared at different concentrations (100, 50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, and 0.195 µg/ml) and measured by LC-MS using HPLC/MS measurements as described above. 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 origin. Triplicates of all *in vivo* experiments were measured. The pure peptide standards **32**, **34**, **36** and **38** were synthesized in-house[139] and the pure synthetic **37** was produced by WuXi AppTec (Wuhan) Co., Ltd..

3. Results

3.1. Topic A: Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains

The following chapter cover parts of the publication '*Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains*' by Kenan A. J. Bozhüyük[#], Annabell Linck[#], Andreas Tietze[#], Janik Kranz[#] (#equally shared first authors), Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-Ni Shi, Peter Grün and Helge B. Bode* (*corresponding author). All results shown here are also included in this publication. The complete publication is attached in Chapter 7. When results from other co-authors are described, these authors are mentioned.

The following chapter presents direct applications for a new NRPS re-engineering concept, which is a continuation of the XU concept and targets the condensation domains as a novel engineering site. In this approach, the structural and biochemical features of the C domain are used to determine a suitable fusion point. C domains have a pseudo-dimeric V-shaped structure with a *N*- and *C*-terminal subdomain[76], which together are forming two opposite tunnels that lead from the donor-T and acceptor-T domain binding sites to the conserved key catalytic-residues containing active site motif HHxxxDG[40,71,75,77] (see Chapter 1.2.3.). Accordingly, the crystal structure of the TycC5-6 T-C didomain (PDB ID: 2JGP)[76] as well as sequence alignments of targeted *Photorhabdus* and *Xenorhabdus* C domains led to the assumption, that the four-AA-long linker region (Gln267-Ala270) between the two subdomains might be an ideal fusion site. Thereupon, the e**X**change **U**nit **C**ondensation domain (XUC) concept was postulated, where self-contained catalytically active C_{ASub}-A-T-C_{DSub} units are substituted to create

chimeric C domains and subsequently modulate C domain specificities (see attached publication in Chapter 7).

3.1.1. Extraction and purification of *de novo* designed NRPS in large-scale production cultures

The quantification of production yields for microbial synthesis is an important value for determining production quality. For this purpose, chemically synthesized standards can be used for HPLC/MS-based quantification of production culture extracts (as described previously[139]). However, to determine the exact amount of a compound that can be obtained from a heterologous expression culture, a selection of seven peptides were isolated as part of this work from large-scale *E. coli* DH10B::*mtaA* cultures.

In the following, large-scale cultivation in shaking flasks and isolation of the produced peptides via HPLC/MS are presented (see attached publication in Chapter 7; the final structure elucidation of the isolated compounds via NMR was performed by Yan-Ni Shi, AK Bode, Frankfurt, Germany)

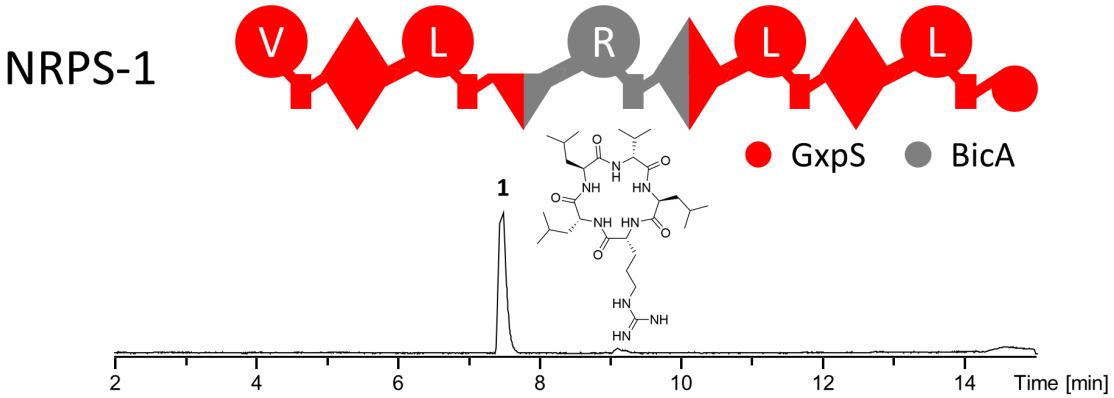


Figure 21. HPLC/HRMS chromatogram of compound 1 produced by NRPS-1 – after purification. Schematic representation of a fictive NRPS with its domains illustrated by the following symbols: adenylation (A) domain, large circle; thiolation (T) domain, small rectangle; condensation (C) domain, triangle; dual condensation/epimerization (C/E) domain, diamond; thioesterase (TE) domain, small circle. The base peak chromatogram of the purified compound **1** is shown and measured via HPLC/HRMS. Additionally, the structure of compound **1** is depicted.

NRPS-1 produced compound **1**, yielding a total amount of 10.73 mg from 1 l *E. coli* DH10B::*mtaA* culture after purification.

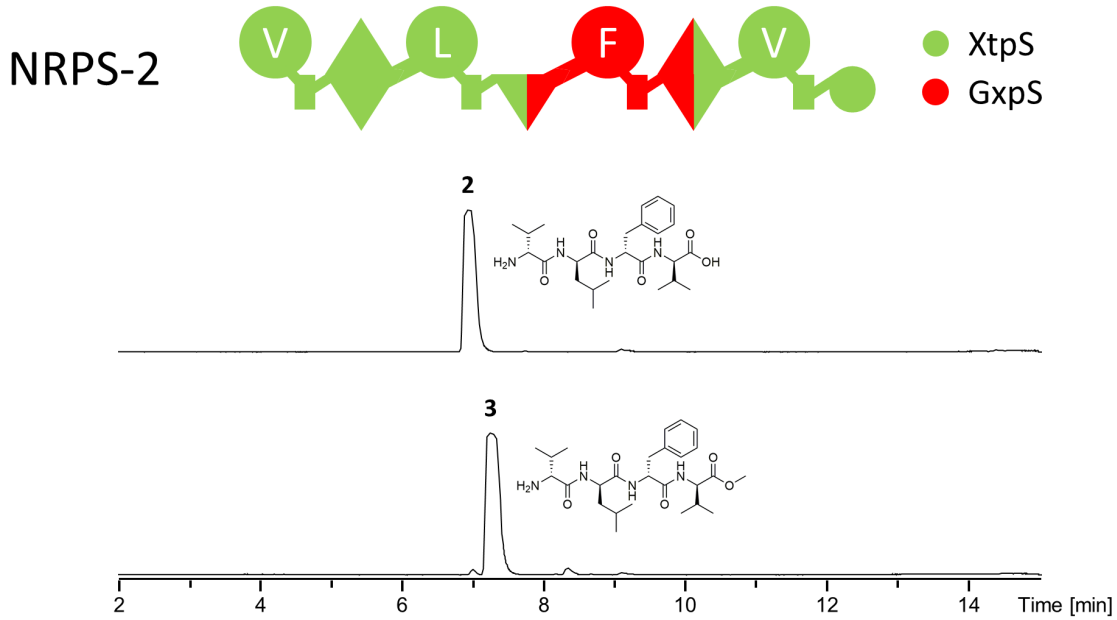


Figure 22 HPLC/HRMS chromatogram of compound 2 and 3 produced by NRPS-2 – after purification. Schematic representation of NRPS-2 domains by symbols is according to Fig. 21. The base peak chromatograms of the purified compounds **2** and **3** are shown and measured via HPLC/HRMS. Additionally, the structure of compounds **2** and **3** are depicted.

NRPS-2 produced compound **2** and **3**, yielding a total amount of 52.6 mg and 47.3 mg from 1 l *E. coli* DH10B::*mtaA* culture after purification.

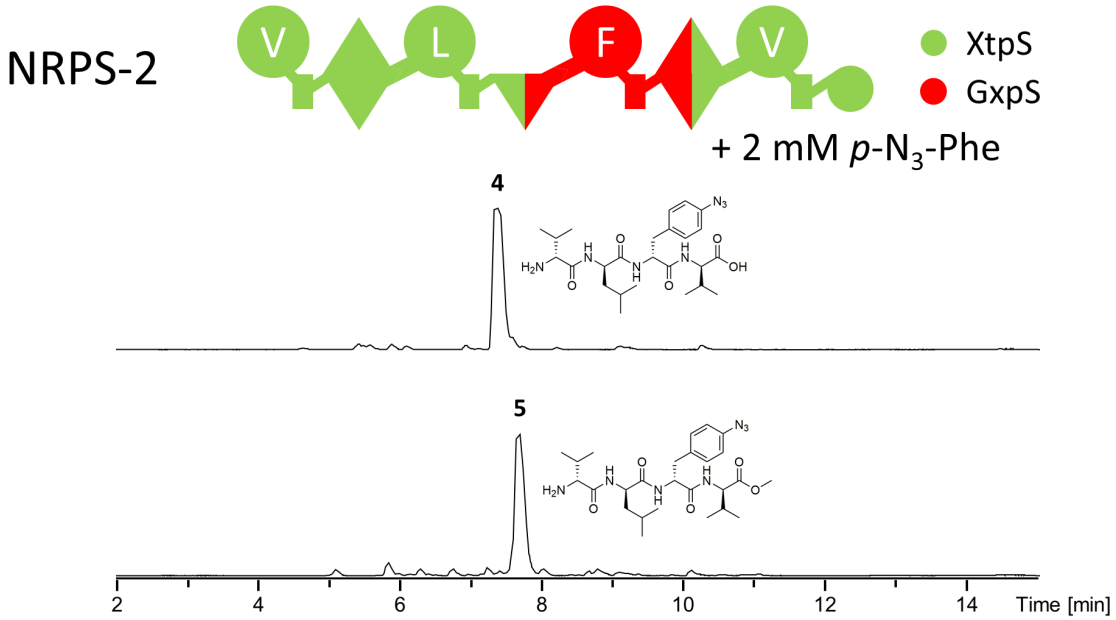


Figure 23. HPLC/HRMS chromatogram of compound 4 and 5 produced by NRPS-2 fed with *p*-N_3-Phe – after purification. Schematic representation of NRPS-2 domains by symbols is according to Fig. 21. The base peak chromatograms of the purified compounds **4** and **5** are shown and measured via HPLC/HRMS. Additionally, the structure of compounds **4** and **5** are depicted.

NRPS-2 supplemented with 2 mM *p*-N_3-Phe produced compound **4** and **5**, yielding a total amount of 6.67 mg and 4.83 mg from 1 l *E. coli* DH10B::*mtaA* culture after purification.

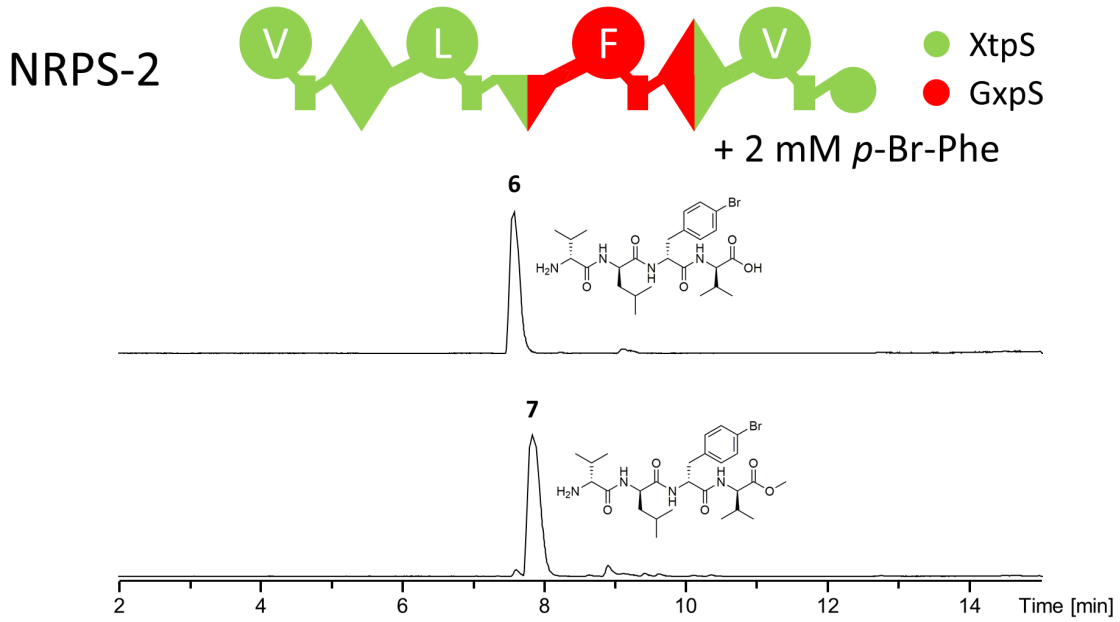


Figure 24. HPLC/HRMS chromatogram of compound 6 and 7 produced by NRPS-2 fed with *p*-Br-Phe – after purification. Schematic representation of NRPS-2 domains by symbols is according to Fig. 21. The base peak chromatograms of the purified compounds **6** and **7** are shown and measured via HPLC/HRMS. Additionally, the structure of compounds **6** and **7** are depicted.

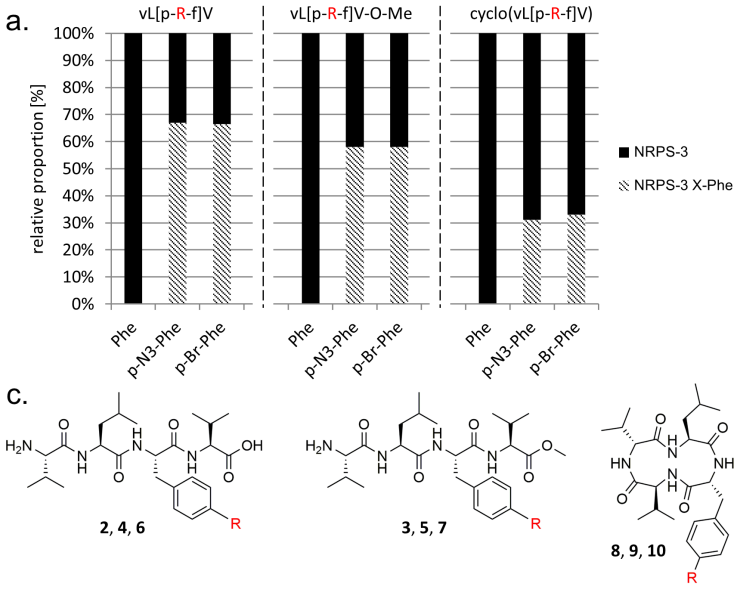
NRPS-2 supplemented with 2 mM *p*-Br-Phe produced compound **6** and **7**, yielding a total amount of 6.5 mg and 2 mg from 1 l *E. coli* DH10B::*mtaA* culture after purification.

Compared to the small-scale cultivations, different ratios of linear to cyclic peptides were observed for some peptides resulting in the linear forms as the main derivatives, e.g., NRPS-1 produced the almost twice amount of the linear product **1** in the large-scale production (*cf.* Chapter 7). Additionally, the observed methyl ester derivatives **3**, **5**, and **7** were derived from the use of MeOH as solvent during the work-up procedure as confirmed by NMR (see attached publication in Chapter 7).

3.1.2. *In vivo* characterization of a *de novo* designed NRPS in *E. coli* DH10B::*mtaA*

As part of this work, the generation of a new variant of the xenotetrapeptide synthetase (XtpS) from *Xenorhabdus nematophila* ATCC 19061[165] by exchanging

XUC3 against GameXPeptide producing Synthetase's (GxpS) XUC3 (NRPS-2) allowed the *in vivo* characterization of the GxpS_XUC3 (cloning of NRPS-2 was performed by Kenan A. J. Bozhüyük, AK Bode, Frankfurt, Germany). GxpS from *Photorhabdus laumondii subsp. laumondii* TTO1[44,45] is known to produce a small library of peptide derivatives due to the relaxed selectivity of the A domains from modules 1 (A1: leucine & valine) and 3 (A3: *p*-NH_2-phenylalanine, phenylalanine, leucine). In addition, the latter has already been characterized *in vitro* and *in vivo* in previous work[44,119,166].



b.

Compound		R
vLfV	**2**	H
vL[p-N_3-f]V	**4**	N_3
vL[p-Br-f]V	**6**	Br
vLfV-OMe	**3**	H
vL[p-N_3-f]V-OMe	**5**	N_3
vL[p-Br-f]V-OMe	**7**	Br
cyclo(vLfV)	**8**	H
cyclo(vL[p-N_3-f]V)	**9**	N_3
cyclo(vL[p-Br-f]V)	**10**	Br

Figure 25. *In vivo* characterization of NRPS-2 in *E. coli* DH10B::*mtaA*. a. Stacked bar chart showing the relative proportion of produced compounds by NRPS-2 in *E. coli* DH10B::*mtaA* supplemented with Phe derivatives (R = H, N_3, Br) and the extracts were measured via HPLC/MS. **b.** Produced compounds (**2**-**10**) with the respective Phe derivative residue R (red) highlighted. **c.** Structure of compounds **2**-**10** produced from NRPS-2 with the highlighted residue R (red). Adapted from[166].

Growing NRPS-2 in *E. coli* DH10B::*mtaA* cultures were fed with 2 mM *p*-N_3-Phe or *p*-Br-Phe resulting in the production of six functionalized XtpS peptides (**4**-**7**, **9**-**10**). The production level of the respective derivative was compared in its relative proportion to the Phe derivative. The observed methyl ester derivatives **3**, **5** and **7** in all these experiments (Fig. 22-24) were derived from the use of MeOH as solvent during the peptide-extraction procedure as confirmed by NMR (see attached publication in Chapter 7).

3.1.3. Generation of a chimeric semi-functional *Bacillus*/*Photorhabdus* NRPS

In order to show the ability of the XUC approach when combining NRPS systems of different origins, the chimeric NRPS-4 was generated (Fig. 26a) consisting of the more distant related bacitracin A (BacA) XUC1-2 from *Bacillus licheniformis* ATCC 10716[167] and the terminal GameXPeptide (GxpS) XUC3-5 from *Photorhabdus laumondii subsp. laumondii* TTO1[45] (Fig. 26a).

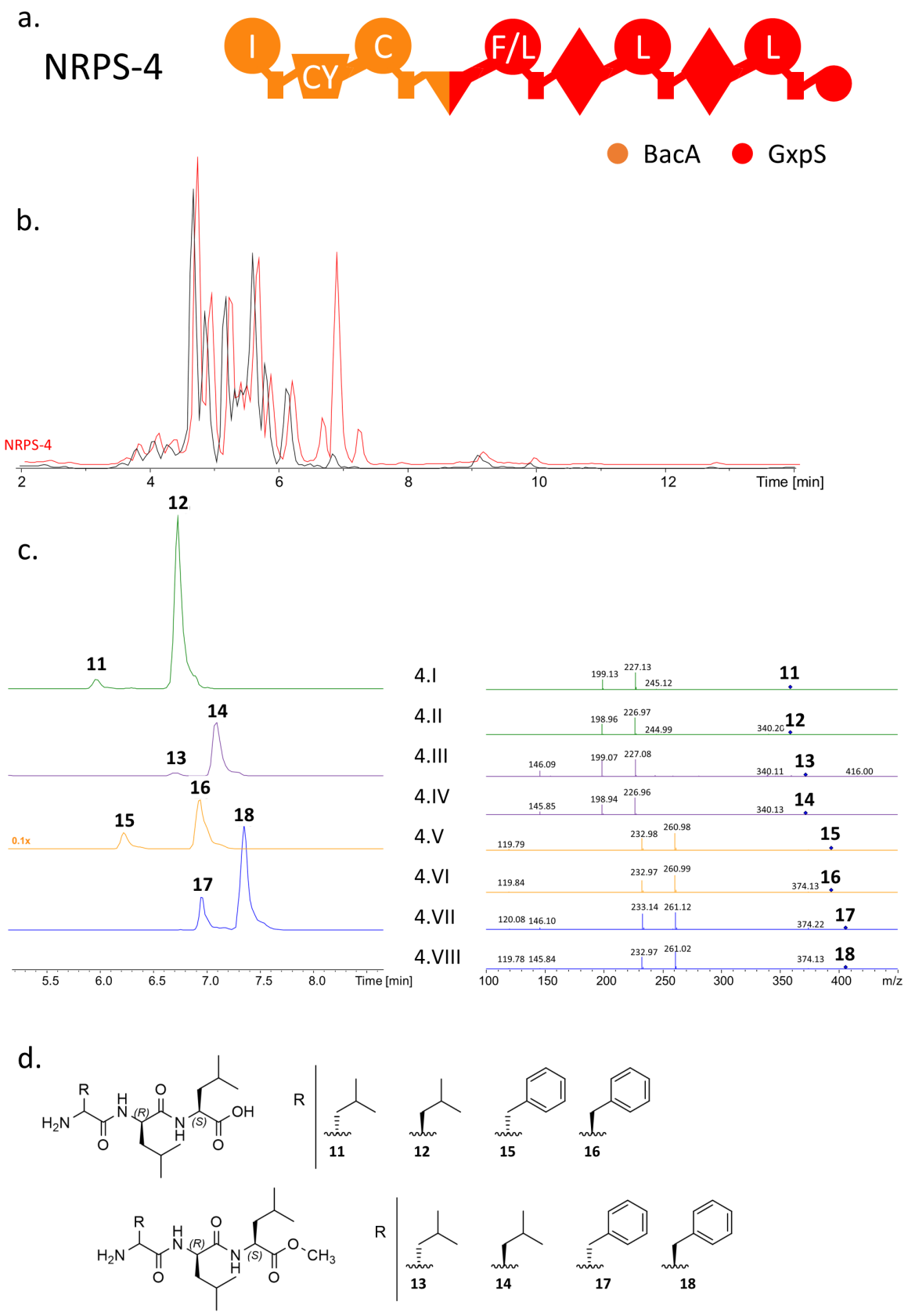


Figure 26: Design and HPLC/MS data of NRPS-4. a. Schematic representation of NRPS-4 domains by symbols is according to Fig. 21 in addition to the heterocyclization (CY)

domain, trapezium. **b.** Base peak chromatogram of extracts from productions cultures of NRPS-4 in *E. coli* DH10B::*mtaA* (red: induced, black non-induced). **c.** Extracted ion chromatogram (left) and HPLC/MS^2 data (right) of **11** (m/z [M+H]+ = 358.22), **12** (m/z [M+H]+ = 358.22), **13** (m/z [M+H]+ = 372.22), **14** (m/z [M+H]+ = 372.22), **15** (m/z [M+H]+ = 392.22), **16** (m/z [M+H]+ = 392.22), **17** (m/z [M+H]+ = 406.22) and **18** (m/z [M+H]+ = 406.22). The shift of the retention time of **11**, **13**, **15** and **17** compared to **12**, **14**, **16** and **18** respectively is supposed to be due to partial epimerization by NRPS-4. **d.** Structure of compounds **11-18** produced from NRPS-4 expressed in *E. coli* DH10B::*mtaA* and the extracts were measured via HPLC/MS. Adapted from[166].

Surprisingly, NRPS-4 only produced truncated peptides **11-18**, which exclusively relates to the expected activity of the GxpS portion (Fig. 26b-d). Thereupon, SDS-PAGE was conducted to test whether NRPS-4 is expressed as a full-length protein or not.

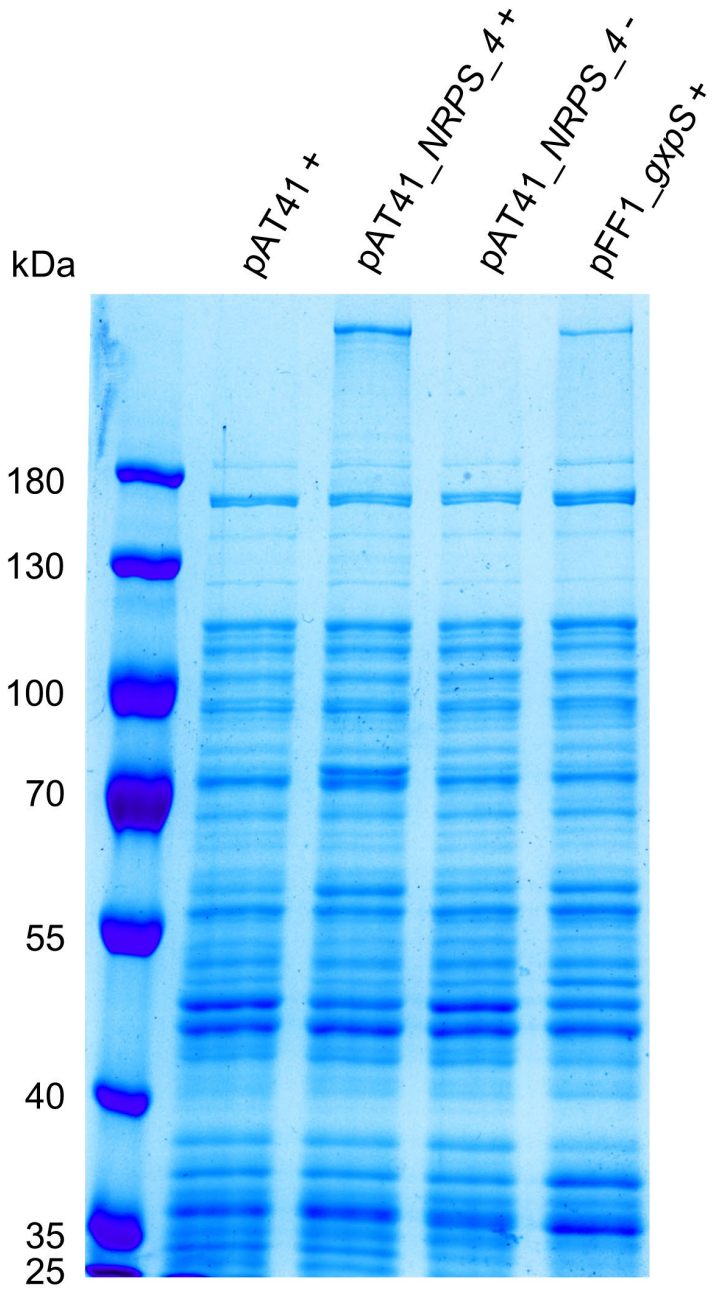


Figure 27. SDS-PAGE of protein extracts of *E. coli* DH10B::*mtaA* Cultivated *E. coli* cultures harboured pAT41 with no insert (control), pAT41_NRPS_4 and pFF1_*gxpS*. All samples marked with '+' were induced with 0.02 % arabinose and '-' without arabinose. The expected molecular weight for NRPS-4 and GxpS is 572 kDa and 514 kDa, respectively. Adapted from[166].

Protein extracts of *E. coli* production cultures of NRPS-4 (induced/non-induced) and, for comparison, the approximately equally sized GxpS were plotted in an SDS-PAGE (Fig. 27). NRPS-4 and as well as GxpS showed a clear band (Fig. 27, line three and five) and could therefore confirmed to be produced as full-length proteins.

3.2. Topic B: Influence of condensation domains on activity and specificity of adenylation domains

The following chapter is based on the manuscript in preparation '*Influence of condensation domains on activity and specificity of adenylation domains*' by Janik Kranz[#] (#first author), Sebastian L. Wenski, Alexander A. Dichter, Helge B. Bode* and Kenan A. J. Bozhüyük* (*corresponding authors). All the herein shown results are part of this manuscript in preparation. A preprinted (non-peer-reviewed) version of the manuscript is available on *bioRxiv* (doi: 10.1101/2021.08.23.457306). When results from other co-authors are described, these authors are mentioned.

The following chapter contributes to the controversial role of C domain specificity and whether C domain selectivity is indeed just a presumption that has unnecessarily complicated rational NRPS redesign – as suggested by the latest studies[77,89,90,162,168]. With this in mind, the recombinant NRPSs created in our lab were reviewed to identify functional artificial BGCs showing an unexpected behavior, like C domains accepting noncognate substrates or altered A domain activation profiles not matching the profiles observed in the natural context[139,162,166]. To shed further light on the role of C domains for NRP synthesis, the effect of C domains onto A domains were systematically analyzed via series of *in vitro*, *in vivo*, *in situ*, and *in silico* characterizations.

3.2.1. *In vitro* characterization of C domains influence on GxpS_A3's activity and selectivity

To quickly grasp the influence of C domains on the activity and selectivity of A domains, initial advantage was taken of the GameXPeptide A-F producing Synthetase (GxpS) from *Photorhabdus laumondii subsp. laumondii* TTO1[44,45]. GxpS, besides being the most widespread BGC in *Photorhabdus* and *Xenorhabdus* strains[13], is one of our best studied, most engineered, and most promiscuous model

systems[119,139,162,166] – producing a library of cyclic penta-peptides[44]. This library of peptide derivatives is synthesized due to the relaxed selectivity of the A domains from modules 1 (A1: leucine & valine) and 3 (A3: *p*-NH_2-phenylalanine, phenylalanine, leucine). In addition, the latter has already been characterized *in vitro* and *in vivo* in previous work[119,166]. In the course of these characterizations, it was even possible to determine that the GxpS_A3, in addition to a broad variety of proteinogenic amino acids (*in vitro*), recognizes and activates non-natural *para*- (*p*), *meta*- (*m*), and *ortho*- (*o*) substituted amino acids (*in vitro* and *in vivo*), such as *m*/*o*/*p*-Cl-Phe, *m*/*o*/*p*-F-Phe, *m*/*p*-Br-Phe, and *p*-O(C_3H_3)-Phe[166]. Since this promiscuity is the ideal prerequisite to analyse the influence of C domains on the activity profile of A domains, GxpS_A3 was selected as a first framework for further experiments.

To get first biochemical evidence of the hypothesized influence of C domains on A domain selectivity the cloning, heterologously production (in *E. coli* BL21 (DE3) Gold), purification (via His6-Tag affinity chromatography), and *in vitro* assays were performed of three GxpS derived proteins (P1: GxpS_A3-T3; P2: GxpS_C_{ASub}-A3-T3; and P3: GxpS_C3-A3-T3) (cloning of P1 and P2-3 was performed by Sarah Nowak and Alexander A. Dichter, AK Bode, Frankfurt, Germany) for adenylation activity against all 20 proteinogenic amino acids in the presence or absence of an upstream domain (C or C_{Asub}) (Fig. 28). To this end the recently introduced multiplexed hydroxamate assay (HAMA) was chosen (Fig. 28a-b)[159]. The HAMA targets the second half-reaction of amino acid activation, quenching the formed aminoacyl adenylate by adding hydroxylamine and analysing the resulting amino-hydroxamates by tandem mass spectrometry (MS/MS) analysis[159]. In contrast to the 'traditionally' applied *in vitro* assays, like the radioactive ATP-^{32}PPi isotope exchange assay[169,170], γ-[$^{18}O_4$]-ATP assay[171], and colorimetric assays[172], which all only allow testing of one substrate per reaction – HAMA allows the parallel testing of dozens of competing amino acid substrates. Therefore, the novel HAMA is supposed to mimic the natural conditions in the cell much better, as all substrates are present at the same time.

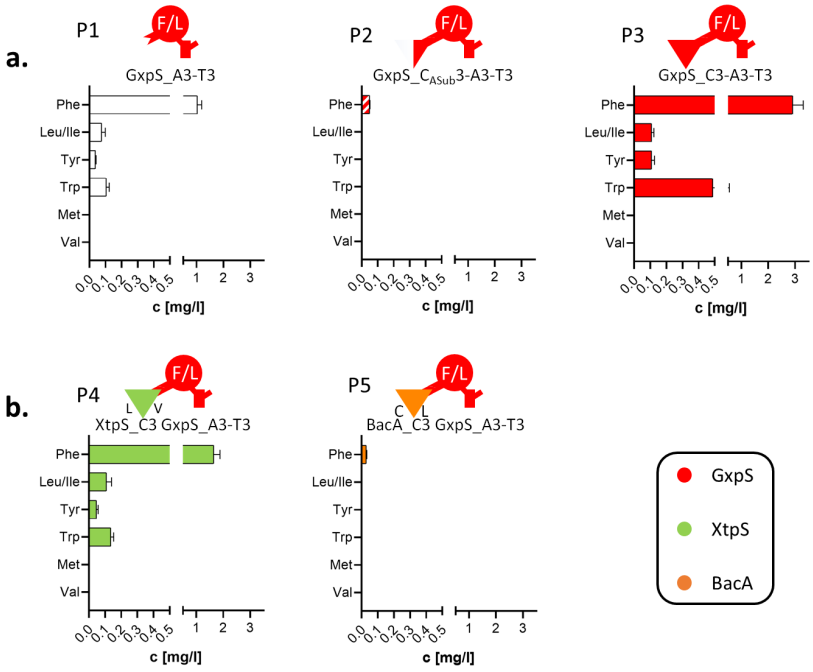


Figure 28. *In vitro* characterization of the GxpS_A3. a. GxpS_A3-T3 with no preceding C domain, GxpS_C3$_{ASub}$, or GxpS_C3 domain tested in a HAMA for produced peptide yields measured with HPLC/HRMS. **b.** GxpS_A3-T3 with XtpS_C3, or BacA_C3 tested in a HAMA for produced peptide yields measured with HPLC/HRMS. Schematic representation of the NRPS domains by symbols is according to Fig. 21. The C_{DSub} and C_{ASub} are labelled corresponding to the preferred up- and down-stream A domain substrate in their WT NRPS.

As a result of this first *in vitro* characterization of P1-P3 via HAMA, it can be stated that the presence and absence of any domain (GxpS_C$_{ASub}$3, GxpS_C3, XtpS_C3) upstream of GxpS_A3 showed a great influence on adenylation activities and substrate recognition profiles (Fig. 28a). In general, the heterologously produced proteins (P1-P3) only showed adenylation activities for hydrophobic polar (Phe, Leu/Ile, Trp) and non-polar aromatic amino acids (Tyr) (Tab. S2), closely resembling the A domain's *in vivo* behaviour, but with notable differences[44,119,139,162,166]. As expected, however, P1 showed highest specificity for phenylalanine (1.03 mg/l), followed by tryptophan (0.11 mg/l), leucine/isoleucine (0.08 mg/l), and tyrosine (0.04 mg/l) (Tab. S2). P2, carrying the *C*-terminal subdomain of GxpS_C3 upstream of GxpS_A3, showed impaired catalytic potential to activate the offered substrates, leaving only a detectable signal for phenylalanine (0.05 mg/l) (Tab. S2). In contrast, for P3, carrying the full length GxpS_C3 domain, an improved catalytic efficiency to activate the offered substrates was observed. P3 revealed an almost identical

activation profile as P1, but with almost three-fold increased turnover rates (Fig. 28a; Tab. S2).

Next, and to better understand the influence of C-A interfaces on substrate recognition and activation of adjacent A domains, P1 was targeted by creating two chimeric proteins with three domains each (P4 & P5), which were again analysed using HAMA (Fig. 28b). While P4 was generated by fusion of the C3 domain of the xenotetrapeptide-producing synthetase (XtpS) from the Gram-negative *Xenorhabdus nematophila* ATCC 19061[165], which is very similar to the originally present GxpS_C3 (~86 % sequence identity) (Tab. S4), P5 was generated by fusion of the C2 domain of the peptide antibiotic-producing bacitracin synthetase (BacA) from the unrelated Gram-positive *Bacillus licheniformis* ATCC 10716[167] (~41 % sequence identity) (Tab. S4). Both hybrid proteins – as well as all hybrid constructs described below – were created according to the splicing position established within the XU concept[139]. As expected, P4 showed a very similar activity and amino acid recognition profile to P3, with phenylalanine being the preferred substrate (Fig. 28a-b). In contrast, P5 almost completely lost its catalytic activity, with a barely detectable signal for phenylalanine left (Fig. 28b; Tab. S2), confirming that altered C-A interactions do have a great impact on the A domains capacity to recognise and activate respective substrates, indeed.

3.2.2. *In vivo* characterization of C domains influence on GxpS_A3 using truncated GxpS versions

As *in vitro* experiments sometimes can lead to results not reflecting the enzymes true *in vivo* behaviour, for example, as experienced with results from biochemical characterizations of C domains[73,87,134,173,174], a series of *in vivo* experiments with truncated chimeric GxpS versions were also performed. GxpS has the rare potential to *in vivo* initiate peptide synthesis even after deletion of the initiation module[162] – as was also recently reported for the teicoplanin-producing NRPS in an *in vitro*

study[92]. However, for the experimental setup, the deletion of the first two modules (A1 to C3) of GxpS, inserted either no preceding C domain (NRPS-5), GxpS_C_{ASub}3 (NRPS-6), GxpS_C3 (NRPS-7) (Fig. 29a) or related C domains (63 to 69 % sequence identity) (Tab. S4) of various origins with different ascribed acceptor site specificities (NRPS-8 to -12) (Fig. 29a-b) were performed, produced the resulting NRPSs heterologously in *E. coli* DH10B::*mtaA*[151], and analysed the culture extracts by HPLC/MS. Throughout the present work, the resulting peptides and yields were confirmed by HPLC/MS (Tab. S3) and in comparison of retention times with synthetic standards.

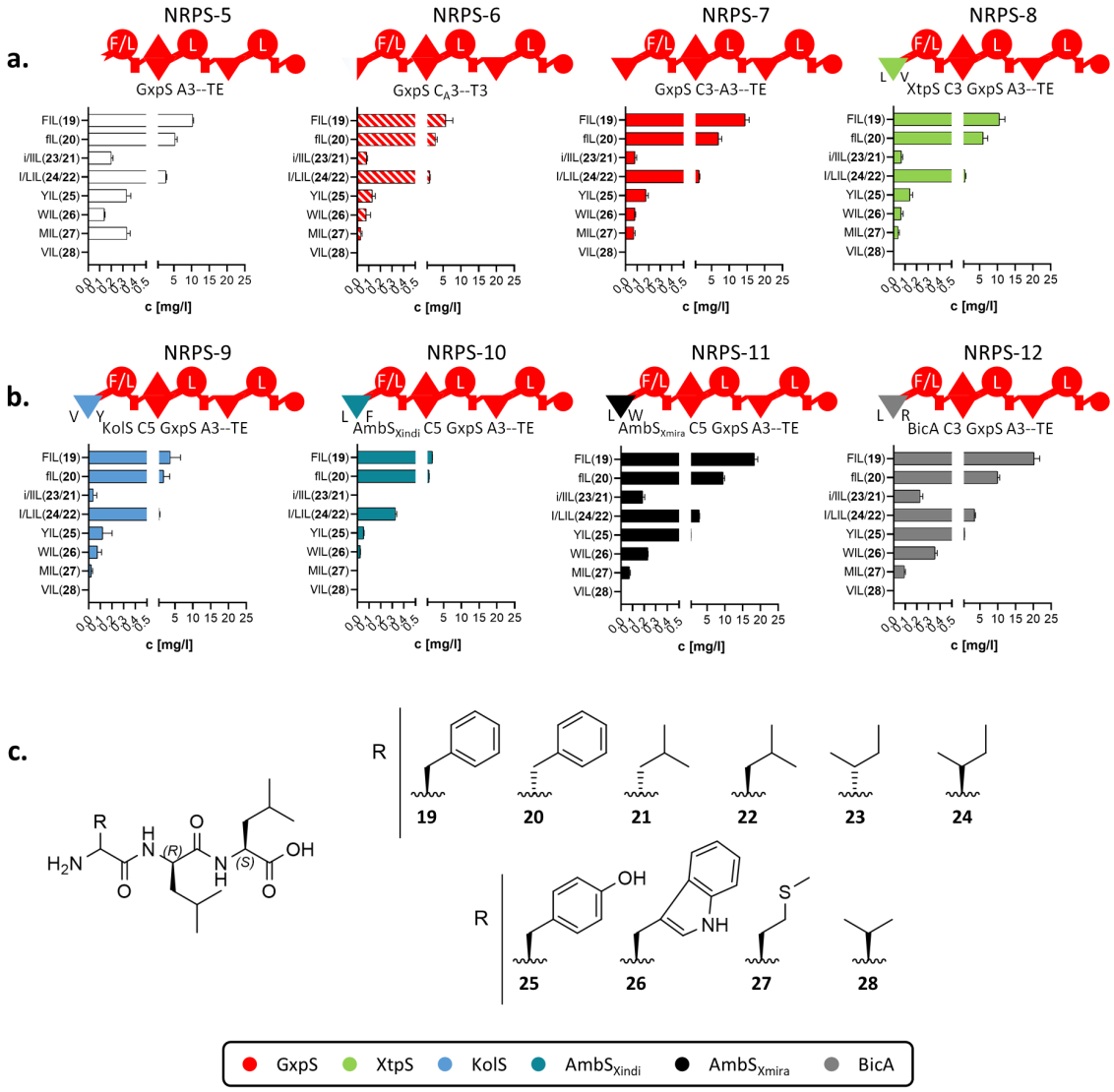


Figure 29. *In vivo* characterization of the GxpS_A3 with varying C domains using truncated GxpS versions. a. terminal GxpS_A3--TE with no preceding C domain, GxpS_C3$_{ASub}$, GxpS_C3 or XtpS_C3 domain heterologous expressed in *E. coli* DH10B::*mtaA* and the extracts were measured via HPLC/MS. **b.** terminal GxpS A3--TE with KolS_C5, BicA_C3, AmbS$_{Xmira}$_C5 or AmbS$_{Xindi}$_C5 domain heterologous expressed in *E. coli* DH10B::*mtaA* and the extracts were measured via HPLC/MS. Schematic representation of the NRPS domains by symbols is according to Fig. 21, and C$_{DSub}$ and C$_{ASub}$ are labelled corresponding to the preferred up- and down-stream A domain substrate in WT NRPS. **c.** Structure of compounds **19-28** produced from NRPS-5 to -12 expressed in *E. coli* DH10B::*mtaA* and the extracts were measured via HPLC/MS.

Briefly, all of them, NRPS-5 to -12, were catalytically active showing biosynthesis of the same range of tripeptides (**19-28**), due to the promiscuity of GxpS_A3 – with FIL (**19**) having highest titer followed by flL (**20**) and I/LIL (**21** & **22**) (Fig. 29a-b). Interestingly, despite the C/E domain downstream of GxpS_A3, all peptides are

produced with higher titers towards the *L*-configuration (**19**, **22**, and **24**). As the epimerization reaction is reversible and finds its end in the adjustment of an equilibrium between both isomers[175], this might indicate that the downstream C/E domain, which usually expects a peptidyl chain, is unable to make sufficient contact with the activated amino acid, resulting in delayed condensation followed by late thiolation. Hence, this change in the reaction velocity caused by the length of the substrate[176] might lead to the observed diastereomer with a trend towards the *L*-isomer and not the expected *D*-isomer. However, titers of NRPS-6 are slightly lower but NRPS-7's are ~30 % higher compared to NRPS-5 (Fig. 29a; Tab. S3) – confirming the *in vitro* observed influence of the C-A interface on general biocatalytic activity of A domains (*cf.* Fig. 28a-b). Remarkably, GxpS_A3 as part of NRPS-5, -6, and -7 showed a different substrate activation profile than as part of P1, P2, and P3, respectively. NRPS-5, -6, and -7 mainly synthesised the tripeptides **19**-**24**, known and expected from WT behaviour, illustrating why biochemical *in vitro* characterizations must always be treated with the utmost caution, especially with regard to multi-modular assembly lines.

Compared to NRPS-5, the chimeric proteins NRPS-8 to -12 showed no difference in the number of substrates activated by GxpS_A3, but the overall peptide production rates of NRPS-9 and -10 were ~50 % lower and of NRPS-11 and -12 ~200 % higher (Fig. 29b; Tab. S3). For example, the latter produces **25** with 2-fold higher yield than NRPS-5 and 4-fold higher than NRPS-7. Consequently, NRPS-5 to -12 are supporting the *in vitro* observed extended gatekeeping function of C domains (Fig. 28) by fine tuning the A domains' substrate selectivity.

3.2.3. *In situ* characterization of fabclavines through plasmid-based XU complementation of *fclJ* in *Xenorhabdus szentirmaii*

To further investigate the influence of altered C-A interactions on the product spectra of NRPSs, homologous BGCs present in several bacterial strains producing the same peptide scaffold but resulting in different derivatives were studied next. The great advantage of such highly homologous systems is the possibility to study the effect of altered C-A interactions without having to consider further potential incompatibilities that could inhibit synthesis. Therefore, the fabclavine-producing BGCs (*fcl*; Fig. 30a) present in several *Xenorhabdus* strains[155], including *Xenorhabdus budapestensis* DSM 16342 (*Xbud*), *Xenorhabdus hominickii* 17903 (*Xhom*), and *Xenorhabdus szentirmaii* DSM 16338 (*Xsze*), were studied in this work.

Fabclavines are bioactive peptide-polyketide-polyamine hybrids with broad-spectrum activity against bacteria, fungi, and other eukaryotic cells[155,177]. In previous work, the deletion of *fclI* of the NRPS encoding genes *fclIJ* led to shortened polyamine carrying fabclavine derivatives (**29**-**31**), and thus to the assumption that FclJ can also be used as a starting unit without FclI[155]. Accordingly, and due to *fclJ*'s rather small size, encoding two NRPS elongation modules (~7 kbp), FclJ was chosen as promising starting point to investigate C-A interface substitutions *in situ*. FclJ, however, bears another advantage necessary to study the impact of an altered C-A interface on the A domain's substrate recognition profile – namely FclJ_A6. While this particular A domain recognizes and activates proline in *X. budapestensis* and *X. hominickii*, it also activates threonine and valine in *X. szentirmaii*[178]. As the promiscuity of the latter neither can be explained by differences within the respective proteins' primary sequence (~89 % similarity) (Tab. S4) nor with changes within the substrate specificity conferring amino acid residues within the A domains' active site (Tab. S5), led to the hypothesis that the respective upstream C domain (FclJ_C6) may be the reason for the product diversification in *X. szentirmaii* (Fig. 30).

To explore the substrate promiscuity of FclJ_A6 in *X. szentirmaii*, a *X. szentirmaii ΔfclI ΔfclJ* double knockout mutant was generated and a library of plasmids were

prepared encoding WT FclJ from *X. szentirmaii* (NRPS-13), *X. budapestensis* (NRPS-14) and *X. hominickii* (NRPS-15), as well as six chimeric FclJ combinations (Fig. 30b, NRPS-16 to -21) for plasmid-based gene complementation experiments (Fig. 30b) (deletion mutants and cloning of plasmids – with the exception of the plasmid for NRPS-18 – were performed by Sebastian Wenski, AK Bode, Frankfurt, Germany). These six chimeric NRPSs represent all possible XU-C-A interface combinations from the chosen set of target-BGCs and therefore allows to investigate whether the observed promiscuity of FclL_A6 is due to intrinsic proofreading of the given C domain or rather an effect of altered C-A interactions.

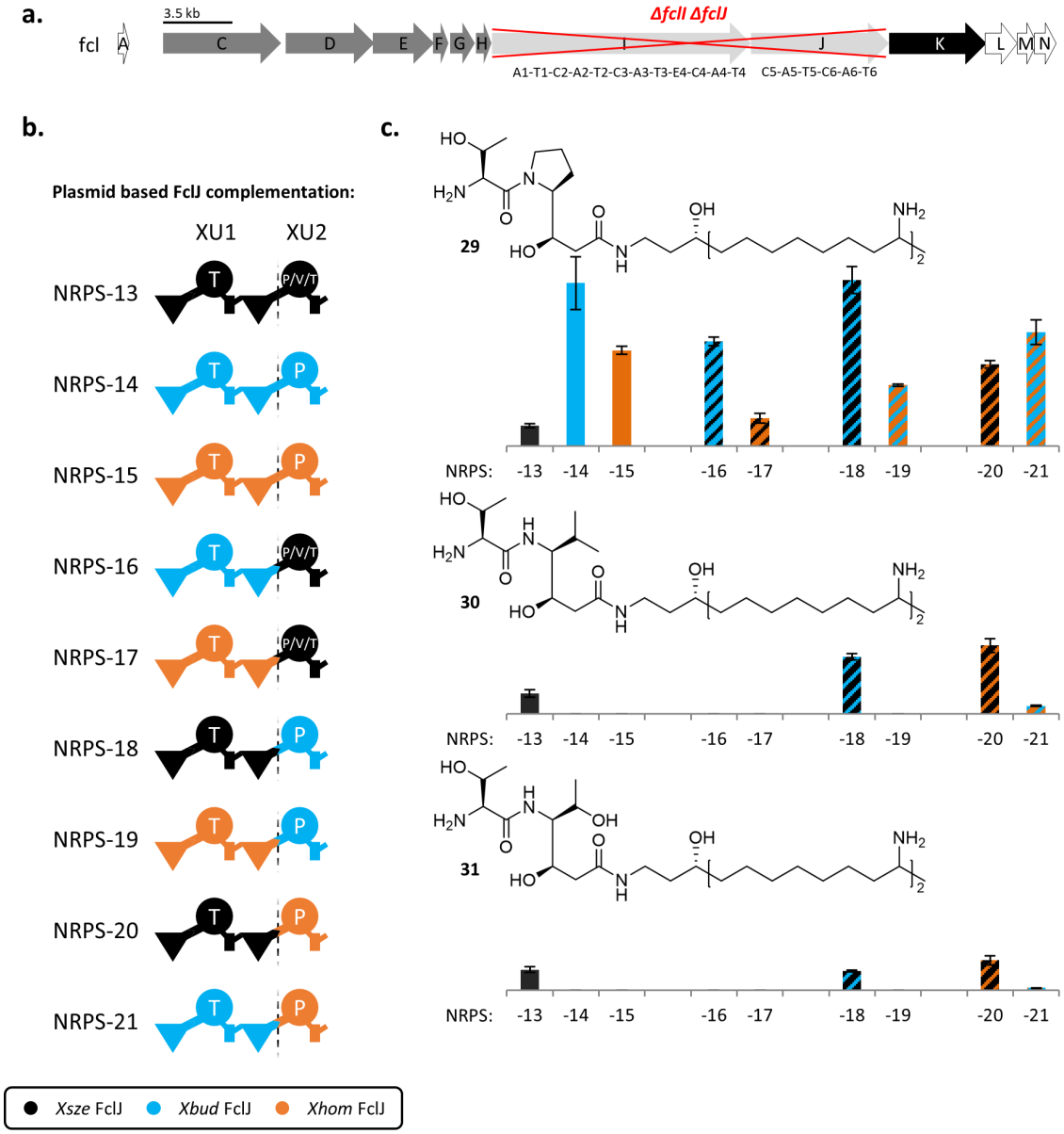


Figure 30. Fabclavines and the plasmid-based XU combinations for *fclJ* complementation in *X. szentirmaii ΔfclIJ*. a. Fabclavine biosynthesis gene cluster (BGC) with the polyamine (dark grey), NRPS (light grey), PKS (black), other (white) genes, and the *fclI fclJ* deletion (*Δ*) highlighted in red. Adapted from[155]. **b.** Schematic representation of the carried-out XU combinations of the *Xsze* FclJ_C5--C6/A6-T6 (black), *Xbud* FclJ C5--C6/A6-T6 (light blue), and *Xhom* FclJ C5--C6/A6-T6 (orange). Schematic representation of the NRPS domains by symbols is according to Fig. 21. **c.** Hatched bar charts with corresponding colour code of the plasmid based FclJ insertions of the Pro derivative (Top), Val derivative (Middle), and the Thr derivative (Bottom). The produced quantity of each product **29**, **30** or **31** was compared in percentage relative to the amount of produced **29**, **30** or **31** by *Xsze* FclJ_C5--C6/A6-T6 (set as 100 %) (Fig. S1-3).

As intended, plasmid based-complementation and production of WT FclJs (NRPS-13 to -15) in *X. szentirmaii ΔfclI ΔfclJ* led to the production of the expected

range of shortened fabclavines (**29**-**31**) – with NRPS-13 synthesizing peptides **29**-**31** and NRPS-14 and -15 only the proline derivative **29**. For the chimeric NRPS-16 and -17, both carrying the putative promiscuous XU2 of FclJ (A6T6) from *X. szentirmaii* (Fig. 30b-c), only synthesis of **29** could be detected, and thus FclJ_A6's substrate promiscuity could not be transferred – indicating that production of derivatives other than **29** is not the sole result of the A domain's substrate specificity. This indication is further supported by NRPS-18 and -20, both carrying XU1 (C5A5T5C6) of FclJ from *X. szentirmaii* and XU2 from *X. budapestensis* (NRPS-18) and *X. hominickii* (NRPS-20), respectively, now capable to biosynthesize peptides **29**-**31**. Interestingly, production of NRPS-19 only led to detectable amounts of **29**, while NRPS-21 led to the synthesis of **29**-**31**, but to a much lesser extent than NRPS-18 and -20 (Fig. 30b-c).

3.2.4. *In silico* mapping of hot-spot areas in the C-A interface to determine crucial C-A didomain interactions

To reveal the very nature of the C domains' influence on downstream A domains, a series of WT and chimeric C-A interfaces were investigated on a structural level.

Since the structural elucidation of the targeted C-A interface forming proteins of NRPS-15 and NRPS-17 was intended but failed, an *in silico* approach was chosen to characterize the extended gatekeeping function of C domains, at least in first approximation. Therefore, protein homology modelling[160] by using the Molecular Operating Environment (MOE 2019.01[179]) was combined along with HSPred[161]. The latter is a support-vector-machine-based method to predict critical interaction partners across protein-protein interfaces. HSPred systematically mutates *in silico* individual amino acids (excluding Pro and Gly) to alanine and calculates the changes in free energy of binding ($\Delta\Delta G$). 'Critical Interaction Partners' or 'Hot Spots Residues' are defined as those residues for which $\Delta\Delta G \geq 2$ kcal/mol. The HSPred output score predicts mutated residues with a score greater than zero as hot spots ($\Delta\Delta G \geq 2$ kcal/mol) and negative scores ($\Delta\Delta G < 2$ kcal/mol) as non-hot spots. Others are not involved in interface formation (Fig. 31; Fig. S4).

For comparative structural *in silico* analysis, the *in vitro* assayed GxpS WT interface of P3 (GxpS_C3-A3) and hybrid interfaces of P4 (XtpS_C3-GxpS_A3) and P5 (BacA_C3-GxpS_A3) (*cf*. Fig. 28) were chosen. In terms of sequence homology and catalytic activity, P4 and P5 were chosen to represent the two extremes, with P4 being almost WT-like and P5 completely unrelated (Tab. S4). Additionally, the *in situ* investigated FclJ_C6-A6 WT and hybrid interfaces presented above (Fig. 30b, NRPS-13 to -21) was chosen. For homology modelling, three different crystal structure templates of NRPS proteins with multiple domains: AB3403 (PDB ID: 4ZXH)[60], EntF (PDB ID: 5T3D)[60], and SrfA-C (PDB ID: 2SVQ)[58] were selected. This template diversity is intended to cover the majority of C-A interface-forming regions, such as the adenylate-forming conformation of AB3403, the thioester-forming conformation of EntF and the open conformation of SrfA-C (Fig. 31a). Based on these templates, a total of 30 models of the selected C-A interfaces (Tab. S6) were created, and then analyzed the protein-protein interactions of the C and A domains via HSPred (Fig. 31; Fig. S4).

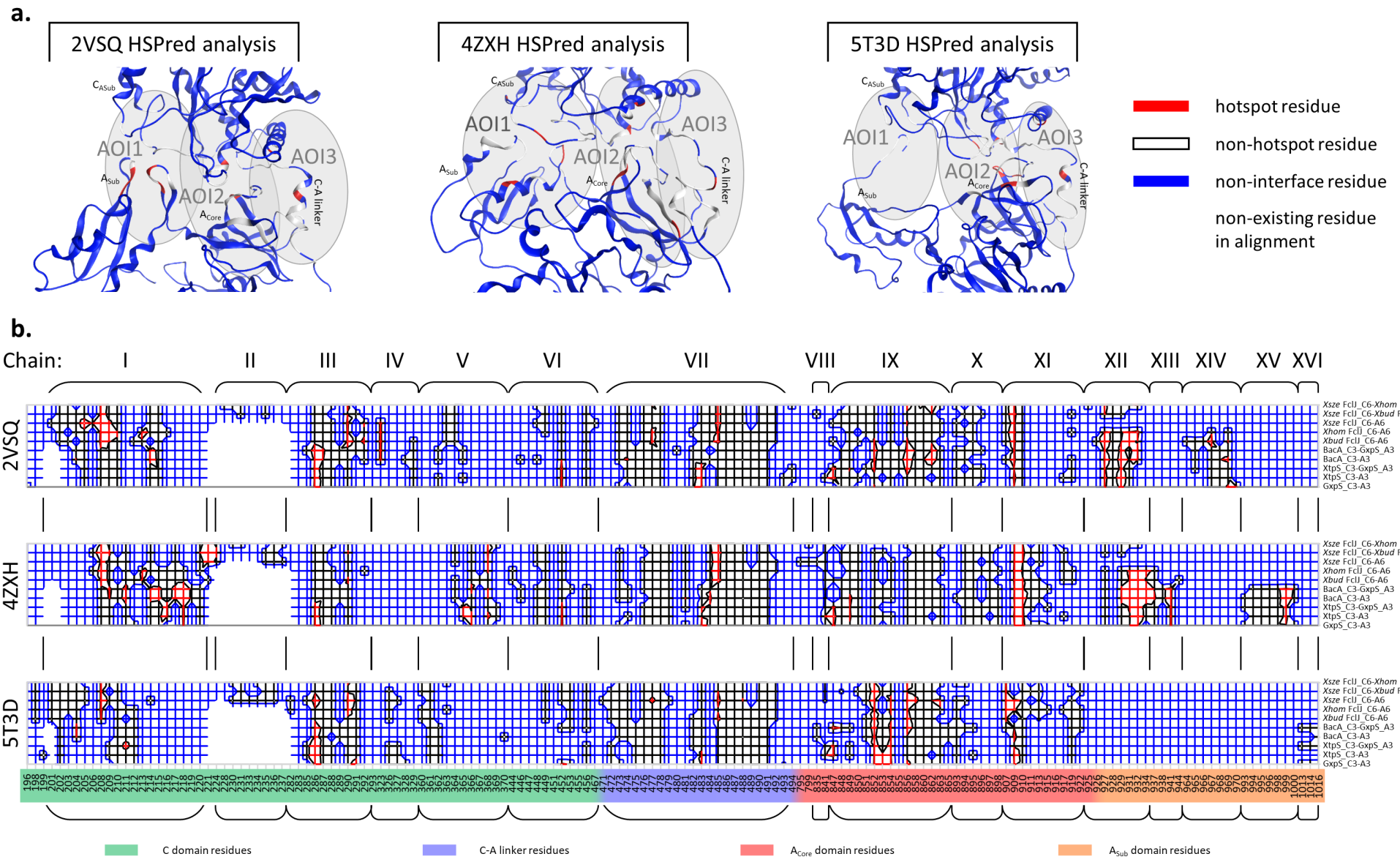

a.
2VSQ HSPred analysis
4ZXH HSPred analysis
5T3D HSPred analysis
AOI1
AOI2
AOI3
C-A linker
hotspot residue
non-hotspot residue
non-interface residue
non-existing residue in alignment
b.
Chain: I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
2VSQ
4ZXH
5T3D
C domain residues
C-A linker residues
A domain residues

Figure 31. HSPred interface prediction for the created homology models. a. Exemplary MOE models of the GxpS_C3-A3 homology model calculated with SrfA-C (PDB ID: 2SVQ)[58], AB3403 (PDB ID: 4ZXH)[60], and the EntF (PDB ID: 5T3D)[60] as templates representative for all models from the HSPred analysis with highlighted hotspot residues (red), non-hotspot residues (white), non-interface residues (blue), and non-existing residues (colourless) in the reference alignment. The positions of the Area-Of-Interactions (AOI) 1-3 are greyed out in the structures. **b.** Contour wireframe model only showing the interface forming residues of the HSPred interface prediction of the ten *Xsze* FclJ_C6-*Xhom* FclJ_A6, *Xsze* FclJ_C6-*Xbud* FclJ_A6, *Xsze* FclJ_C6-A6, *Xhom* FclJ_C6-A6, *Xbud* FclJ_C6-A6, BacA_C3-GxpS_A3, BacA_C3-A3, XtpS_C3-GxpS_A3, XtpS_C3-A3, and GxpS_C3-A3 models build in reference to the 4ZXH, 5T3D, and 2SVQ templates. The interacting Chains I-XVI are indicated with black frames in the contour wireframe models. Colouring of the residue positions in the reference alignment domains (Tab. S7) is according to the C domain (green), C-A linker (blue), A_{Core} (red), and A_{Sub} (orange).

The resulting interface plots in the contour wire model of the HSPred prediction (Fig. 31) show at a glance the distinct conformational changes of the different interfaces formed. In brief, numerous interactions can be ascribed to sixteen different Chains I-XVI that contribute to interface formation (Fig. 31b). These Chains, of which the C domain has six, the C-A linker one, and the A domain nine, interact in the so-called Area-Of-Interaction (AOI) 1-3 (Fig. 31a), which show highly dynamic conformational changes in the course of the catalytic NRP synthesis cycle. However, based on the chosen crystal structure templates, a comprehensive description of the structural C-A interface differences and the changes that occur during the transition of the individual catalytic states into each other can be found in the supplementary information (Explan. S1). In the following, the most important differences between the WT and hybrid C-A interfaces modelled and analyzed in this chapter are highlighted. It should be noted that the amino acid numbering used below is based on the residue position in the protein sequence alignment of calculated models (Tab. S7).

When the WT C-A interface of P3 is compared to the hybrid interfaces of both, the C-A interface of P4 and P5, differences mainly are present in the AOI1 A_{Sub} area of all catalytic states (Fig. 31). P4 introduces additional hot spot residues in the adenylate forming conformation (Fig. 31, $P4_{4ZXH}$) via Chain I (R216) & V (R365), loosened A_{Core}/A_{Sub} transition of Chain III (D291), and a tightening to the C-A linker in AOI2 of Chain IX (R841, R847, Y849) in the thioester forming conformation (Fig. 31, $P4_{5T3D}$; Tab. S7). Although the exchange of GxpS_C3 for XtpS_C3 in P4 leads to a slightly closer interaction of the C domain with the A_{Sub} domain during adenylate and thioester formation as well as to a slight relaxation of the A_{Core}/A_{Sub} hinge region (AA907 to AA944; Tab. S7), the hybrid interface of P4 appears to be very similar to the wild type one of P3 – as could be expected from their high sequence similarity (Tab. S4). In turn, when the interface of P5 is compared to the WT P3 interface, multiple additional hot residues within Chain I (R204, K209, D211, Y214, D215, K217, R218; Tab. S7), and Chain V (R323) in all three models (Fig. 31, $P5_{2VSQ}$, $P5_{4ZXH}$, and $P4_{5T3D}$) could be observed, indicating a much stronger association around the otherwise flexible A_{Sub} domain in AOI1.

Compared to all other interfaces, investigated within present work, the *Xsze* FclJ_C6-A6 WT interface (NRPS-13) as well as the recombinant interfaces from NRPS-16 to -21 (*cf*. Fig. 30) show a novel interaction site with a significant impact on

especially the substrate binding/release conformation (Fig. 31, models based on 2VSQ). All these constructs have the *Xsze* FclJ_C6 domain in common – introducing the unique Chain II (Fig. S5-7) that highly contributes to the C-A interface formation by tightly interacting with the respective A_{Sub} domains in AOI1. Chain II shows up in all conformations but with the highest abundance of hot residues in the adenylate forming state (models based on 4ZXH) at S220, H221, and E224 (Tab. S7). Chain XIV, which regularly participates in interface formation in substrate binding/release (Fig. 31, models based on 2VSQ models), disappears entirely in the constructs containing the respective *Xsze* C domain. Further, all of the studied fabclavine interfaces lack contribution of Chain XV and XVI, which have been involved in the adenylate-forming and thioester-forming conformation in the GxpS, XtpS and BacA interfaces (Fig. 31b). Consequently, the pronounced conformational changes of the A_{Sub} domain observed *in silico* are less determined by the opposite C domain, suggesting dynamic detachment.

3.3. Topic C: Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering

The following chapter is based on the manuscript in preparation '*Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering*' by Janik Kranz[#], Nadya Abbood[#] (#equally shared first authors), Mohammad Alanjary, Helge B. Bode* and Kenan A. J. Bozhüyük* (*corresponding authors). All the herein shown results are part of this manuscript in preparation. When results from other co-authors are described, these authors are mentioned.

Parts of the results in 3.3.1. are already mentioned in the master's thesis *Kranz J. 2018. Verbesserung der Produktion nichtribosomal hergestellter Peptide durch Optimierung der Interaktionsfläche zwischen der C- und A-Domäne*[180].

A correct C-A communication is of great importance for maintaining NRPS functionality during peptide biosynthesis. The C-A interface thus illustrates the highly dynamic NRPS domain architecture by adapting different interface-conformations during the catalytic cycle[100]. For example the smaller *C*-terminal A_{Sub} domain's interaction surface ranges from a tight interaction in the open conformation[58] to barely making any contact in the thioester-forming conformation[60] with the opposite C domain. These observations along with decades of unsuccessful NRPS engineering attempts[42,88] resulted in a persistent image of the C-A didomain as an unit not to be separated[58]. However, we got quite experienced by breaking this dogma with various developed engineering strategies[139,162,168]. With the XU concept, substitutions are performed along a fusion point in the C-A linker region as A-T-C eXchange Units (XUs), allowing the generation of several multi-module NRPS hybrids[139]. Furthermore, the potential of this fusion site was recently highlighted for the application of synthetic zippers (SZs)[162] to generate peptide libraries quickly and further increase the bio-combinatoric potential of NRPSs[168].

The following chapter focuses the targeted altering of the C-A interface for more effective NRPS engineering. Therefore, the interface of the C and A domain was precisely defined, aiming to change it towards an improved version to increase peptide production. For this purpose, it was also made use of phylogenetic analyses to investigate how C-A interfaces behave on an evolutionary level and to deduce its importance for NRPS engineering.

3.3.1. Manual assignment of the interface forming regions between the C and A domain

To gain initial access to the C-A domain, the interface-forming regions were identified. Since the C-A contact area was already outlined in a previous study[139] based on the earlier identified hydrogen bonds in the SrfA-C termination module[58], here, the C-A interface-forming regions sought to be preciously defined with a manual assignment approach.

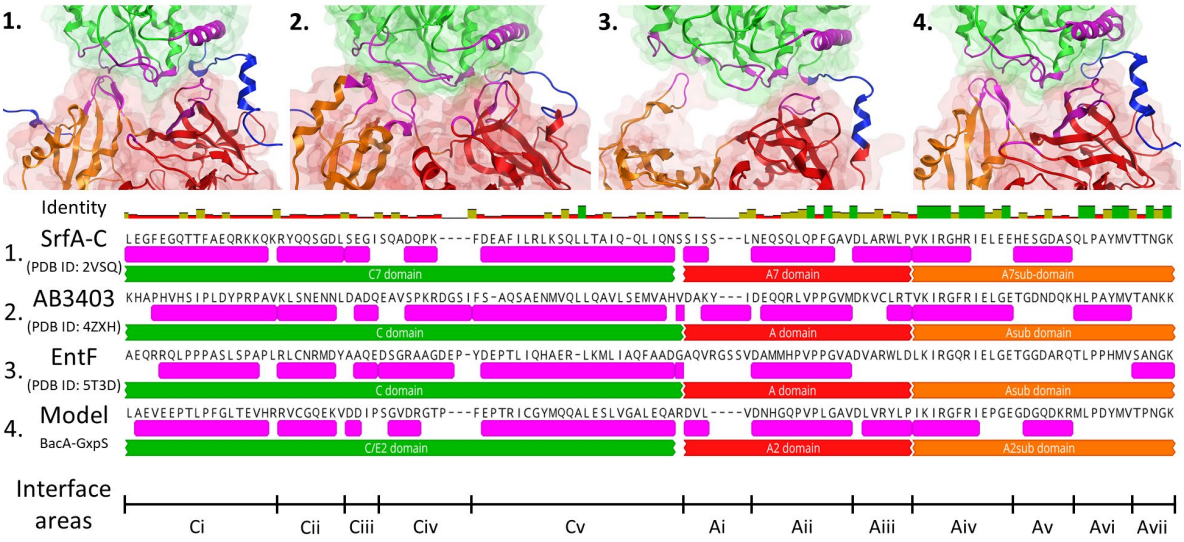


Figure 32. Identification of the interface forming regions. Illustration of the C-A interface by crystal structures excised from 1. SrfA-C (PDB ID: 2SVQ)[58], 2. AB3403 (PDB ID: 4ZXH)[60], 3. EntF (PDB ID: 5T3D)[60], and the calculated 4. Homology model of the BicA_CE2-GxpS_A2 (based on SrfA-C template) with highlighted C(/E) domain (green), A_{Core} domains (red), A_{Sub} domain (orange) and interface forming regions (purple). Additionally, the primary sequences of only the interface forming regions for all structures are shown in an underneath Clustal Omega alignment (identification of the interfaces was also part of the master's thesis by Kranz J. 2018[180]). The average interface areas for the C domain (Ci-v) and the A domain (Ai-vii) are marked in the scale below.

Therefore, from the available C-A didomain crystal structures comprising the open conformation of SrfA-C (PDB ID: 2VSQ)[58] with a 1,097 Å^2 interface (Fig. 32, no. 1.), the adenylate-forming conformation in AB3403 (PDB ID: 4ZXH)[60] with a 1,023 Å^2 interface (Fig. 32, no. 2.), and the thioester forming conformation of EntF (PDB ID: 5T3D)[60] with a 780 Å^2 interface (Fig. 32, no. 3.), all interface-forming areas were identified. These areas were selected based on their spatial proximity of AAs with a distance less than 5 Å to AAs of the opposite domain and, thus, defined as interface-forming AAs (Fig. 32) (identification of the interfaces was also part of the master's thesis by Kranz J. 2018[180]). Thereupon, 12 crucial contact areas were identified from which 5 can be assigned to the C domain (Ci-v) and 7 to the A domain (Ai-vii) (Fig. 32).

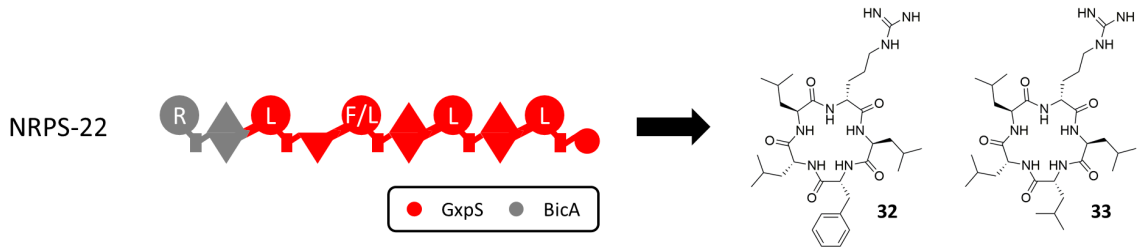


Figure 33. NRPS-22 and its produced compounds. Schematic representation of the NRPS domains by symbols is according to Fig. 21. The structure of the produced compounds **32** and **33** by NRPS-22 are depicted.

Next aim was to determine the contact areas Ci-v and Ai-vii as well for one of our previously constructed hybrid NRPS-22[139] composed of bicornutine (BicA) A1--C/E2 from *Xenorhabdus budapestensis* DSM 16342[163] and the terminal GameXPeptide (GxpS) modules A2--TE from *Photorhabdus laumondii subsp. laumondii* TTO1[45] (Fig. 33). Therefore, it was intensively attempted to generate protein crystals of the hybrid BicA_C/E2-GxpS_A2 but with no success. However, again based on the SrfA-C crystal structure harbouring the largest C-A interface and probably spanning most of the interface-forming residues, the protein structure was created *in silico* by homology modelling (Fig. 32, no. 4) with a final model *root-mean-square deviation* (RMSD) of 4.70 Å.

3.3.2. Phylogenetic analyses of selected NRPS C-A didomains and their interfaces

To categorize the importance of the C-A interface in an evolutionary context, but also to better understand reduced production titers of some previously engineered NRPSs[139,162], several phylogenetic trees were created by using RAxML v4.0[181] (gathering of data for the tree calculation was done in cooperation with Nadya Abbood, AK Bode, Frankfurt, Germany). In total, 20 of our well-established NRPSs' from *Xenorhabdus*, *Photorhabdus*, and *Bacillus* were included, comprising a sum of 174 A domains, 164 C domains, 164 C-A didomains with consequently 164 C-A interfaces to construct phylogenetic trees of: (1) C domains (Fig. S8), (2) A domains (Fig. S9), (3) C-A didomains (Fig. 34) and (4) extracted C-A interfaces (Fig. 34).

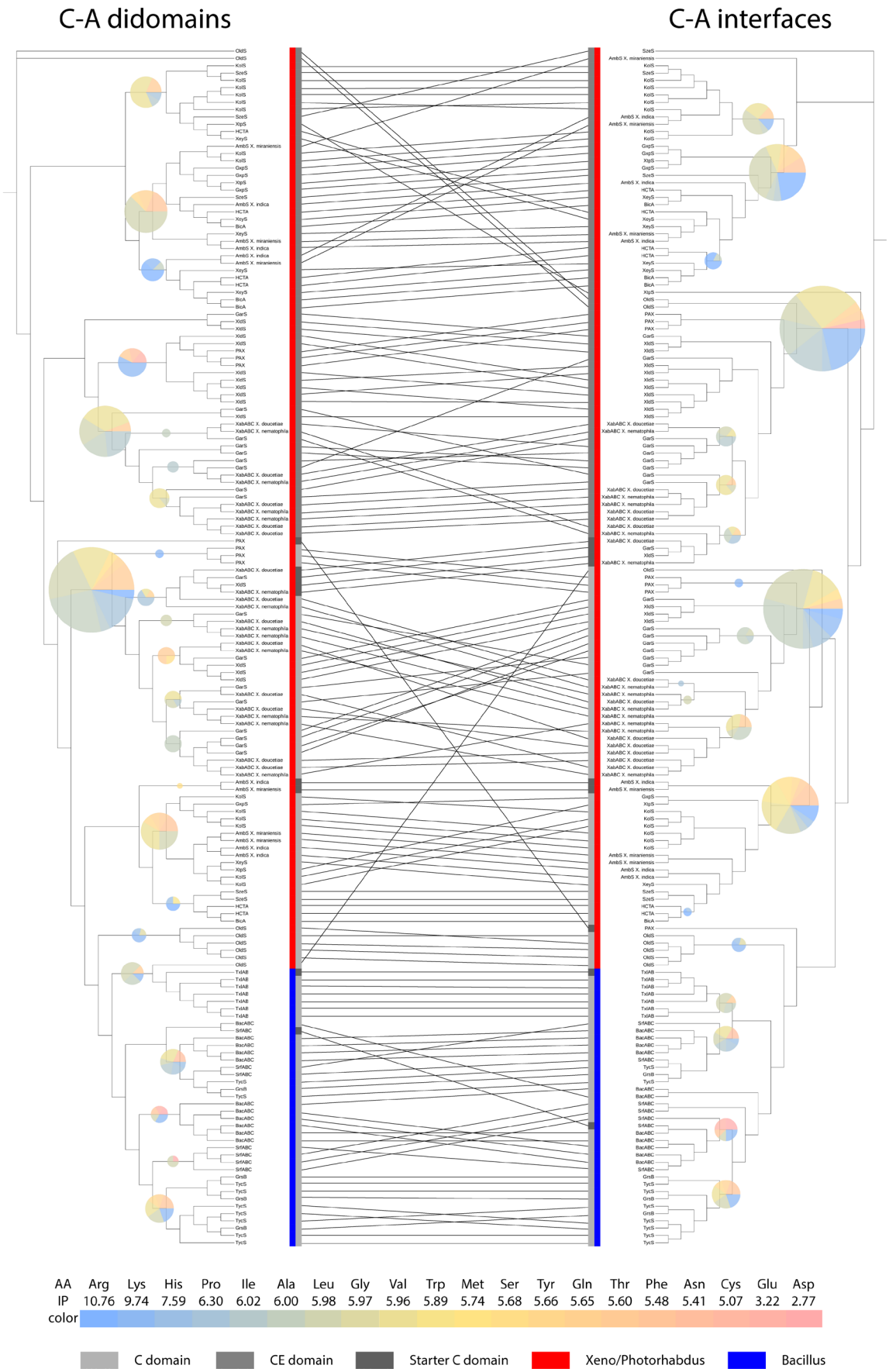


Figure 34. Phylogenetic tree of selected NRPS C-A domains compared to C-A interfaces. Tree was build using RAxML of Clustal Omega alignment from selected NRPS primary sequences (gathering of data for the tree calculation was done in cooperation with Nadya

Abbood, AK Bode, Frankfurt, Germany). The A domains' specificities are displayed as a pie chart (relative pie chart radius corresponds to number of included branch leaves) directly on the node branch and are color-coded according to their isoelectric point (IP)[182] as shown in the scale below for all branch leaves. Genus's affiliations and C domain types are resembled by colored strips. The C-A didomain of the others tree corresponding C-A interface is connected by a straight line. An interactive representation of the dendrogram can be accessed here: https://itol.embl.de/tree/14121241184572116074220 26 (C-A didomains dendrogram), https://itol.embl.de/tree/14121241184573116074220 27 (C-A interfaces dendrogram) (accessed January 31, 2022).

In brief, phylogenetic reconstructions of the (1) C domains and (2) A domains, results showed that both the A and C domain tree cluster according to their genus in the first place. In second place, the C domains cluster by their stereospecific subclass ($^{L}C_{L}$, $^{D}C_{L}$, $C_{Starter}$, C/E), while A domains clustered by their substrate specificity (Fig. S8-9). These observations are consistent with previous studies[18,53,73,90] and demonstrate the strict hierarchy by which the individual domain-phylogenies evolve.

The calculated phylogenetic tree of the (3) C-A didomains (Fig. 34) revealed a strict hierarchy clustering by genus (highest priority), then by C domain subtypes (second priority), and lastly by A domain specificity (lowest priority). Compared to the C domain (Fig. S8) or A domain (Fig. S9) phylogenies, that of the C-A didomains is much more complex. While A domains alone form clearer clades according to their specificity, the specificity plays a subordinated role in C-A didomains as the type of C domain dominates the phylogenetic branching.

Lastly, a phylogenetic tree (4) including only the interface-forming regions of the C and A domains (Fig. 32) was constructed. Interestingly, the tree depicts a replica of the complete C-A didomain phylogeny (Fig. 34), since the interface branches similar to the full-length C-A phylogeny, although the sequences of the entire C_{DSub} and large parts of the C_{ASub} from the C domain and all putative binding-pocket constituents[53] from the A domain were removed. Accordingly, tree (4) demonstrates that even the truncated 'only interface-forming regions' seems to encode the didomain specific characteristics, allowing its branching into genus, C domain subtypes, and A domain specificity (Fig. 34).

3.3.3. Topological evaluation of the C-A interface

Next, it was investigated whether or not there are hints for evolutionary events occurring within the C-A interface. Therefore, phylogenetic subsection-trees within NRPS modules were generated and their topological distance compared. For this purpose, a total of 15 in our laboratory well-established NRPSs were used from different *Xenorhabdus*/*Photorhabdus* species and the first C-A-T elongation module of each NRPS was extracted as a representative of the entire NRPS system. All modules were divided into 21 subsections of 50-100 bp in length, and the phylogenetic trees of each section were constructed using RAxML. With the use of TreeCmp[183] it was finally possible to compare generated phylogenetic trees and calculate topologic distances between the 21 subsections.

TreeCmp comprehensively compares each of these segments, represented by one calculated tree, with all other segments within the considered C-A-T module by a selected metric – of which the weighted and rooted *Cophenetic Metric with L^2 norm* (*CML*)[183] was chosen. *CML* measures quantitatively the difference between a pair of phylogenetic trees by first encoding them by means of their half-matrices of cophenetic values, and then comparing these matrices with the addition to be safely used on weighted phylogenetic trees with nested taxa[184]. Therefore, a pair of trees with smallest non-negative values in a plot can be understood as 'adjacent' in the corresponding metric space at minimum distance[184]. Accordingly, the compared sub-trees of a segment in a plot are considered to be evolutionarily correlated if the calculated *CML* value is approximately zero (Tab. S8).

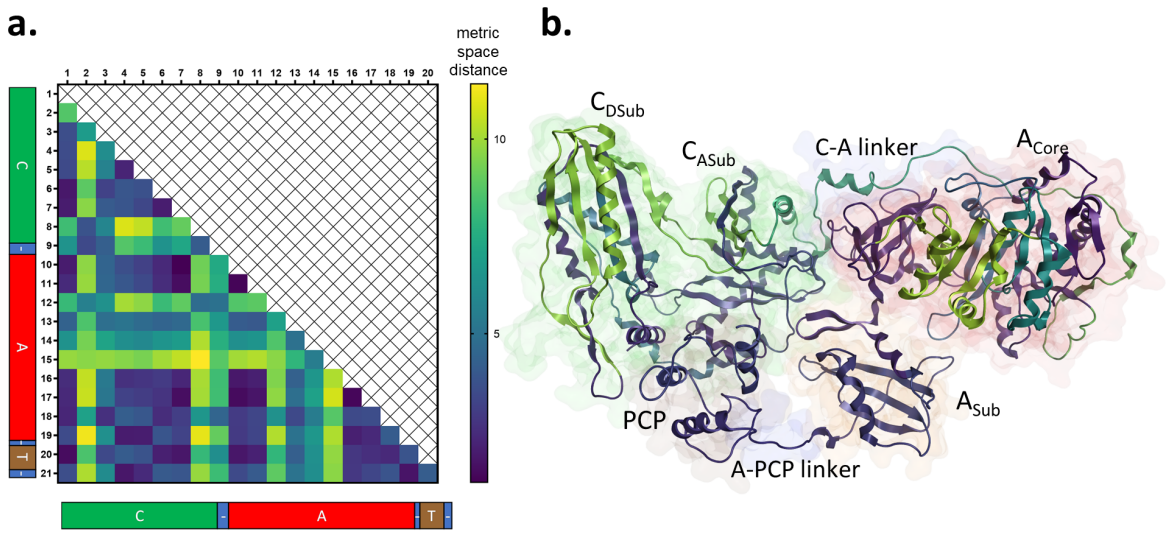


Figure 35. Topological evaluation of the C-A interface. a. Topological heatmap calculated *CML* values of corresponding metric space of segmented NRPS sub-trees is coloured according to their metric space distance, ranging from yellow to dark blue with metric space at minimum distance. On the axes, C-A-T scales are denoted representative for the position of the NRPS domain segments in the considered module with the respective domains colouring according to Fig. 32. **b.** Excised C-A-T from the SrfA-C (PDB ID: 2SVQ)[58] crystal structure with highlighted ribbons according to the averaged (a.) segments 16-20 heatmap colouring and the calculated protein surface are colored according to the module-scales in a..

At a glance, segments 2, 8-9, and 12-15 (Fig. 35a-b) represent areas with large topological distance. Moreover, the previously assigned interface forming regions (*cf.* Fig. 32) largely correspond to segments 4-6 and 17-19, which show a close topological proximity (Fig. 35a-b). In detail, segments 4-6 correspond to C_{ASub} areas Ci-v and segment 16 to A_{Core} areas Ai-ii, which is connected via the hinge region in 17 with those 18-19 to the A_{Sub} areas Aiii-vii (*cf.* Fig. 32). Additionally, the complete the T domain segments 20-21 are also show close topological proximity. According to this, the interface forming areas reflect a surface with a low variation in their phylogenetic branching.

3.3.4. Creation of an automated computer-aided algorithm for C-A interface annotation and scoring

To correlate C-A identities, a scoring algorithm was designed to describe the quality of formed chimeric interfaces compared to WT interfaces (Fig. 36). The designed customized scoring algorithm – namely CAopt.py – automates the very complex (*cf.* Fig. 32; Chapter 3.3.1.) manual interface identification and characterization (coding of

the CAopt.py python-script was performed by Mohammad Alanjary, Wageningen University, Wageningen, The Netherlands). Therefore, the algorithm uses the interface forming regions of 2VSQ, 4ZXH, 5T3D, and BicA_CE2-GxpS_A2 (Fig. 36) as reference and aligns them with the C-A didomain to be investigated to propose the interface forming region within the boundaries of the reference interfaces (*cf*. Fig. 32, no 1-4).

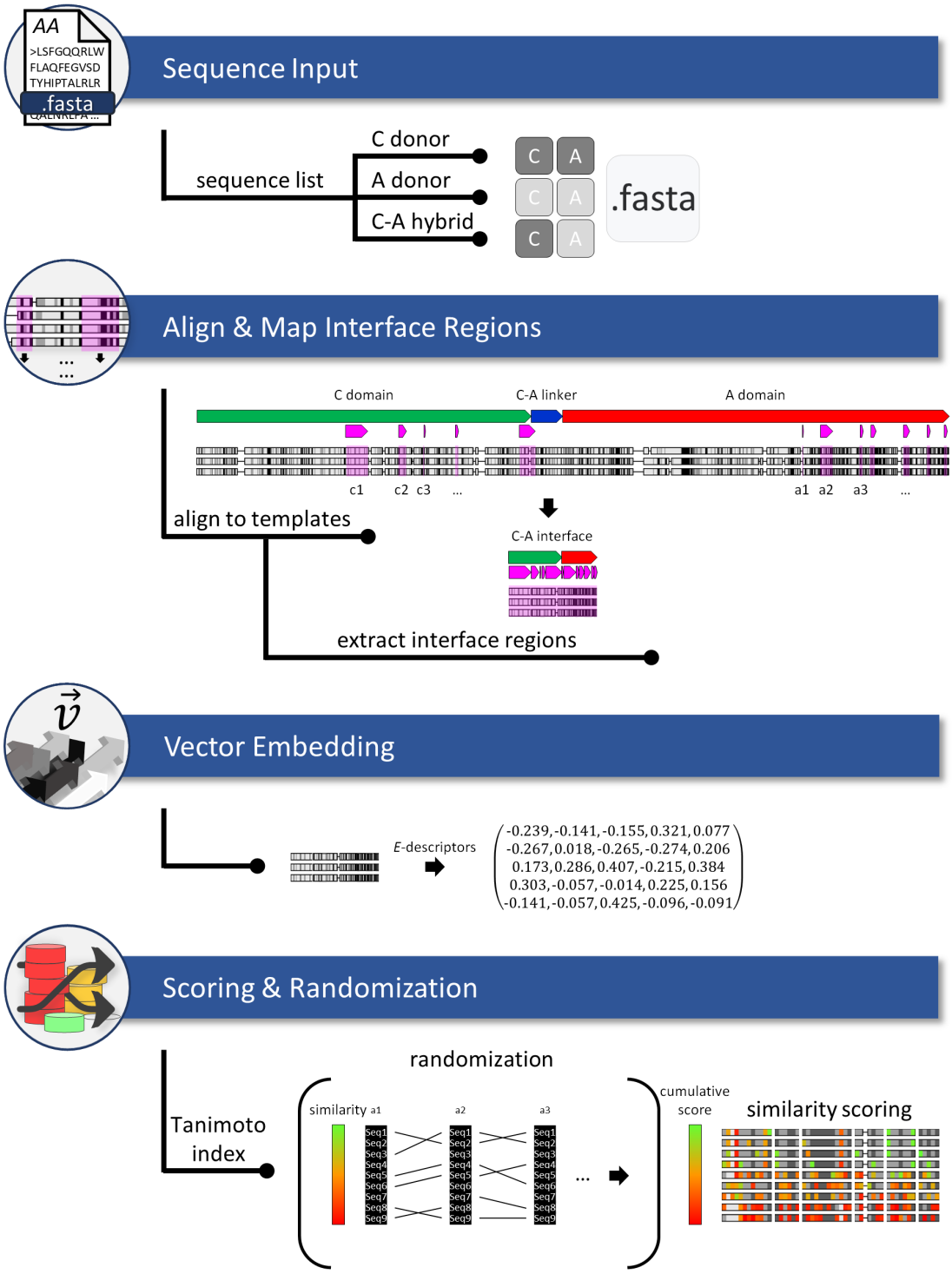


Figure 36. CAopt.py algorithm overview. Flowchart of the CAopt.py algorithm (coding of the CAopt.py python-script was performed by Mohammad Alanjary, Wageningen University, Wageningen, The Netherlands) showing the query C-A didomains and their hybrid input sequences as .fasta; sequence alignment to a template using MAFFT for mapping the interface regions (domain colouring is according to Fig. 32); embedding only the interface regions by *E*-descriptor values in a vector; optional single to multi-site randomization of varying sequences in the respective interface regions and scoring of an estimated physiochemical high (green)/low (red) similarity (Tab. S9) in this vector space to the corresponding locations of the natural interface partner to calculate native similarity.

For the interface characterization, the CAopt.py algorithm interprets the interfaces primary sequence to its physiochemical properties as *E*-descriptor values (Fig. 36). Subsequently, each amino acid is translated into a vector whose coordinates represent the valence of the *E*-descriptor hydrophobicity, size, helix-forming properties, relative abundance, and beta-sheet-forming properties[164]. Moreover, the vectors are transformed into binary fingerprints and compared in terms of Tanimoto coefficients[185]. In this case, the Tanimoto coefficient specifies the similarity measure between the list of properties of the compared amino acids as their distance. The thereby established scoring determines its limits with the lowest distance of 0.19 by exchanging Ser with Thr – whereas of course, the exchange of an amino acid against itself always results in a score of 0 – and the highest possible distance of 1.18 by exchanging Pro with Cys (Tab. S9). In addition, the complete interface can be rendered as the sum of all differences of the artificial to the natural interface by cumulative scores (Fig. 36).

3.3.5. Computer-aided evaluation of interface-combinability by co-expressing unlinked NRPS units

For a direct application of the CAopt.py algorithm, and in reference to the work of Kaniusaite et al.[92], who demonstrated productive NRPS reconstitution of non-covalently linked building blocks at their C-A position, non-covalently linked NRPS of several previously engineered constructs[139,162] were created. Therefore, cloning of a in module three separated GameXPeptide (GxpS) version (NRPS-23) and several hybrids with starting modules of xenotetrapeptide (XtpS) from *Xenorhabdus nematophila* (NRPS-24), ambactin (AmbS) from *Xenorhabdus miraniensis* (NRPS-25), bicornutine (BicA) from *Xenorhabdus budapestensis* (NRPS-26) or bacitracin (BacA) from *Bacillus licenformis* (NRPS-27) was performed (cloning experiments were performed by Nadya Abbood, AK Bode, Frankfurt, Germany) to exclusively analyse the interface effects of intermodular communication (Fig. 37). Throughout the present work, the resulting peptides and yields were confirmed by HPLC/MS (Fig. S10-32) and in comparison of retention times with synthetic standards.

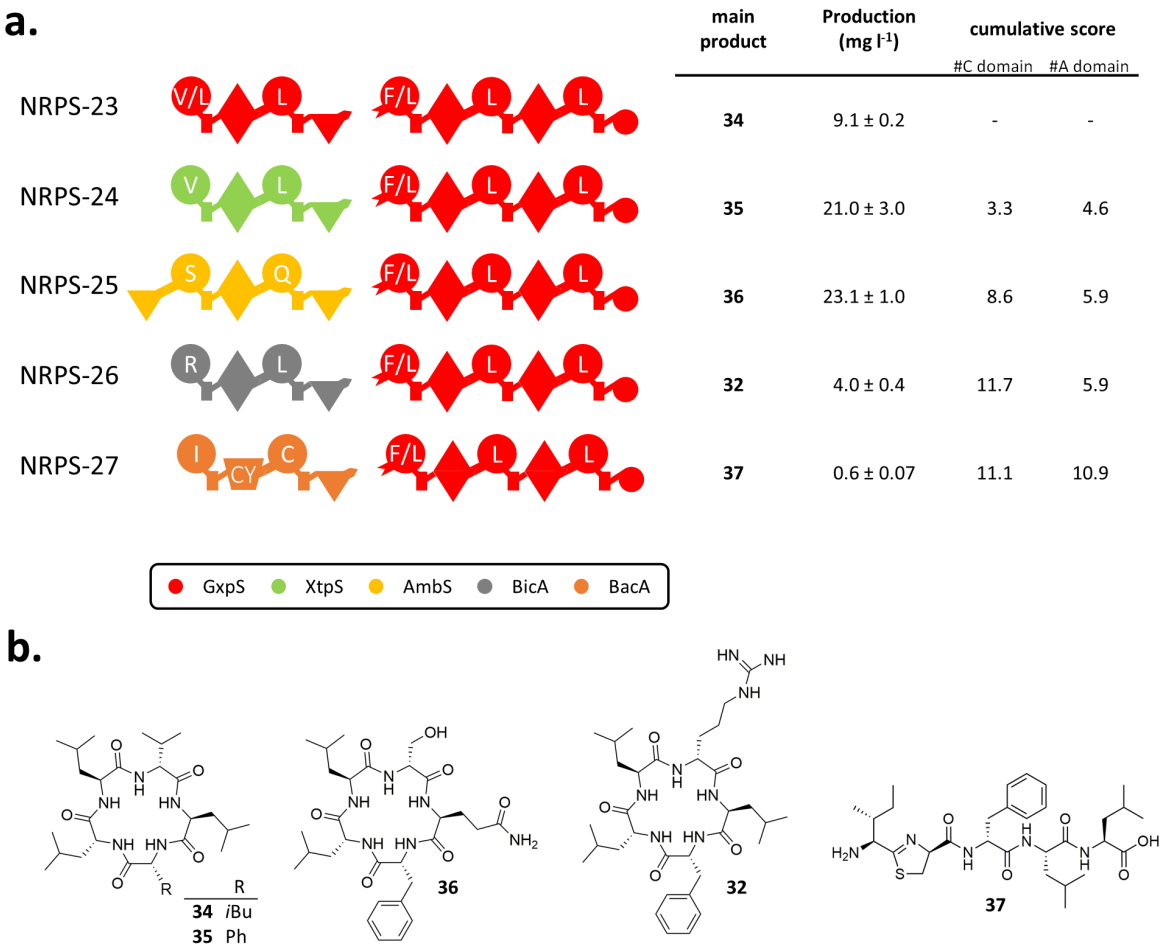


Figure 37. Evaluation of interface-combinability by co-expressing non-covalently linked NRPS units. a. Schematic representation of the NRPS domains by symbols is according to Fig. 21. The recombinant NRPSs were co-expressed with two individual subunits (NRPS-23 to -27) using GxpS, XtpS, AmbS, BicA, and BacA NRPS building blocks (cloning experiments were performed by Nadya Abbood, AK Bode, Frankfurt, Germany). The corresponding peptide yields were obtained from triplicate experiments and cumulative scores are calculated with CAop.py. **b.** Compounds **32** and **34-37** produced from NRPS-23 to -27 expressed in *E. coli* DH10B::*mtaA* and the extracts were measured via HPLC/MS.

Therefore, all co-expressed constructs NRPS-23 to -27 were analysed and a clear tendency was observed accordingly to that high production titers correlate to a low cumulative score of the C and A domain interfaces and vice versa. For instance, while NRPS-24 has a low and NRPS-25 a decent score for both the A (4.6/5.9) and C (3.3/8.6) domain, a similar high production amount of >20 mg/l are observed (Fig. 37). In comparison, NRPS-26 shows a significant higher C domain score (11.7) and therefore suffer losses in production – despite comparable score of the A domain (5.9) to NRPS-24 and NRPS-25 – leading to significantly lower titers of 4 mg/l. Surprisingly, even the combination of very high cumulative scores for both the C and A domain (11.1/10.9) for NRPS-27 is still accompanied by marginally product yields (Fig. 37).

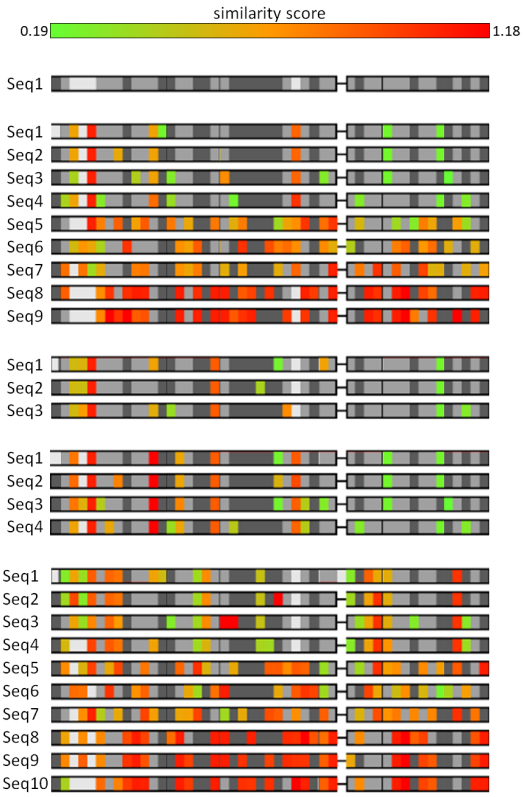
This indicates that despite a low compatibility of the C-A interaction, a certain degree of minimal communication can still be expected.

3.3.6. Systematically alternation of the interface forming region and the identification of a key position

In a further approach, the value of the CAopt.py algorithm was addressed, on the one hand, to test whether the interface can be systematically altered and, on the other hand, to apply synthetic *in silico* calculated interface versions to understand what makes a 'good' interface. Unlike the more conventional alanine scanning used to identify key positions[186], it was possible to implement a randomizer-tool into the CAopt.py script (coding of the CAopt.py python-script was performed by Mohammad Alanjary, Wageningen University, Wageningen, The Netherlands). This enables the mutation of interface residues using a sequential replacement at random sites and generates a pool of sequences, which are then scored using an estimation of physicochemical similarity to the corresponding locations of the natural interface partner.

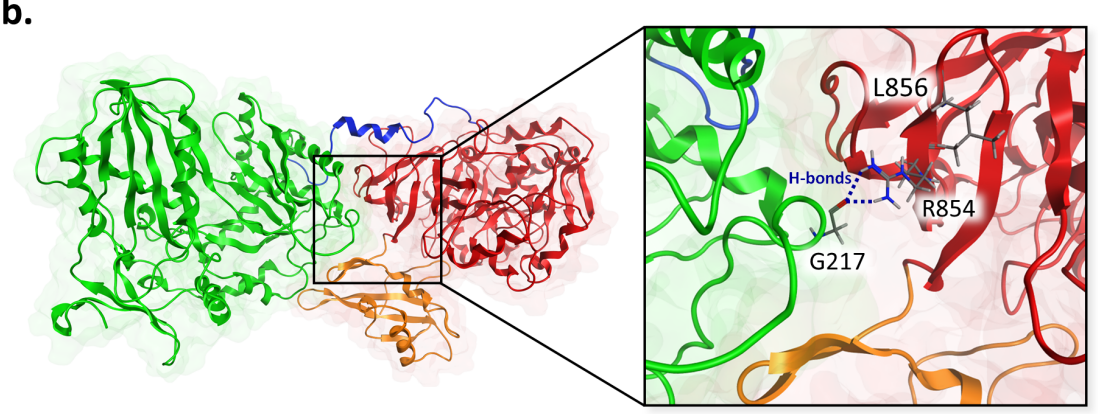
In doing so, the recombinant NRPS-24 to -27 interfaces (Fig. 37a) served as query sequences to calculate an individual set of 10e6 A domain interface variants of the GxpS_A3 for each NRPS. Accordingly, a set of 22 truncated GxpS variant with altered GxpS_A3 interface sequences were generated and cloned for heterologous expression in *E. coli* DH10B::*mtaA*. These selected clones contained GxpS_A3 interfaces with low (<5), medium (10-20), or high (>30) native similarity scores compared to the native NRPS-3 and were analysed for their production yields of the respective produced product (Fig. 38a; Tab. S10) (HPLC/MS analysis for the production cultured of the generated *E. coli* clones was performed by Nadya Abbood, AK Bode, Frankfurt, Germany).

a.

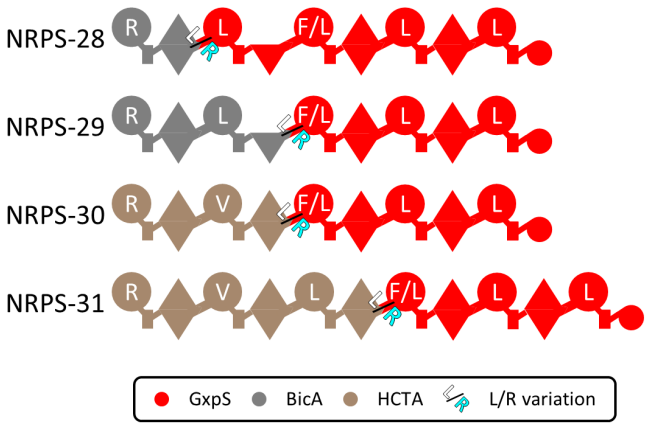


	main product	cumulative scoring	relative production [%]
NRPS-23	**34**	-	
NRPS-24	**35**	4.6	100.0 ± 3.3
		1.2	34.1 ± 5.0
		2.5	24.0 ± 2.4
		3.2	25.6 ± 1.1
		17.0	45.5 ± 2.0
		17.5	41.1 ± 3.1
		17.8	31.7 ± 1.6
		28.5	42.4 ± 2.3
		28.8	32.1 ± 1.6
NRPS-25	**36**	5.9	100.0 ± 13.7
		1.8	558.4 ± 13.7
		3.6	-
NRPS-26	**32**	5.9	100.0 ± 1.9
		2.1	-
		3.3	-
		3.6	-
NRPS-27	**37**	10.9	100.0 ± 6.5
		3.4	23.2 ± 6.3
		4.4	20.9 ± 11.3
		4.8	29.2 ± 9.4
		19.6	18.5 ± 3.8
		20.7	21.6 ± 10.3
		21.5	34.4 ± 10.6
		32.7	30.2 ± 10.7
		33.0	26.6 ± 3.1
		33.3	22.3 ± 6.8

b.



c.



	main product	WT level (**36**) (mol%)	production (mg l^{-1})
NRPS-28	cyclo(rLflL)	100.0 ± 3.2	58.3 ± 1.8
	32	↗ 122.3 ± 10.0	71.3 ± 5.8
NRPS-29	cyclo(rLflL)	126.7 ± 12.6	73.9 ± 7.3
	32	↗ 132.6 ± 11.3	77.3 ± 6.6
NRPS-30	rLflL	11.6 ± 0.7	6.7 ± 0.7
	38	↗ 14.2 ± 0.2	8.3 ± 0.5
NRPS-31	rvlflL	0.7 ± 0.13	0.5 ± 0.09
	39	↗ 0.9 ± 0.09	0.7 ± 0.06

Figure 38. Computer aided interface alternation and the identification of a key residue for optimization. a. Interface sequence alternation of the GxpS_A3 domain part of NRPS-23 to -27 using CAopt.py randomising function. The depicted sequence representations of the generated variants (Seq1 corresponds in each case to the unmodified interfaces) coloured according to the similarity scoring colour-scale above. in comparison to NRPS-23's interface. The cumulative scores are calculated with CAop.py and relative peptide yields were obtained from triplicate experiments via HPLC (HPLC/MS analysis for the production cultured of the generated *E. coli* clones was performed by Nadya Abbood, AK Bode, Frankfurt, Germany). **b.** Illustration of the C-A interface by the calculated homology model of the BicA_CE2-GxpS_A2 (based on SrfA-C template) and the close-up view of the identified L/R variation in the interface with the adjacent H-bond (dark blue) builders. **c.** Schematic representation of the NRPS domains by symbols is according to Fig. 21. The recombinant GxpS (NRPS-28 to -31) using BicA and HCTA NRPS building blocks are depicted with the implementation of the identified L/R variation at its respective position. The corresponding peptide yields were obtained from triplicate experiments.

Unlike the expectations from the previously assessed C-A reassembling-compatibilities (*cf.* Fig. 37), most of the carried-out interface variants resulted in lower than initial production or even unfunctional NRPS (Fig. 38a) – even with good scorings. In detail, alternations in conserved regions nearby the hydrogen bonds (Aiii-v, *cf.* Fig. 32) appear to have a particularly negative effect. Especially for all constructs of which no production could be detected (Fig. 38a, NRPS-25 Seq3/NRPS-26 Seq2-4), area Aiii in the A_{Core}/A_{Sub} hinge region (*cf.* Fig. 32) shows significant changes in polarity by adding, e.g., a histidine close or at the Asp82 position (Tab. S10). In addition, the scoring system incorporates amino acids that would have a low abundance in the natural interface region like Cysteine (0 % to 6 %), Tryptophan (0 % to 2.2 %), or Methionine (1.1 % to 2.2 %) (Tab. S11) – which makes the interface even more artificial.

Nevertheless, the CAopt.py demonstrates its use to generate entirely novel interfaces based only on the introduced scoring system. Therefore, it is still remarkable that almost all interfaces are functional. However, even if the low-scoring/highly-similar variants show significantly fewer site-replacements, they are thus assumed to have changed some critical positions. The CAopt.py algorithm is therefore required of further adjustments with the aim of creating better producing NRPS.

One low-scoring/highly-similar variant was generated for the NRPS-25, which was able to achieve a 550 % higher production rate (Fig. 38a, NRPS-25 Seq2). Upon closer examination of the altered positions – notwithstanding some minor changes – the L to

R (score = 0.8939) exchange, located exactly in the A_{Core}/A_{Sub} hinge region (corresponding to L856R in BacA_CE2-GxpS_A3 homology model, Fig 39b), stands out. This region is part of the highly conserved A8 motif[39] and ensures the rigid body torsion of the A_{Sub} domain of 140° after the adenylation reaction and preserves the domain flexibility for the catalytic machinery[100]. Furthermore, a hydrogen bond in the SrfA-C from R847 to E217 was already identified in direct proximity[58], suggesting the role of L856R (corresponds to the position 849 in SrfA-C) in the BacA_CE2-GxpS_A3 homology model (Fig. 38b) as a potential additional hydrogen bond former in this area. Maintaining this region is essential for a successful thiolation reaction[187] and is assumed to lead to significantly higher production yields (*cf*. Chapter 3.2.4).

To narrow down the exact origin of the production-increasing effects and a transferability on additional interfaces, this interface L/R variation was applied to other NRPS hybrids (Fig. 38c). In addition to the hybrid NRPS used for homology modelling BacA_CE2-GxpS_A2 interface (NRPS-28), the adjustment was transferred to the related interfaces having an arginine at the exact same position in their A domains, in BacA (NRPS-29) module three, as well as module three HeptaCycloTetraArginine (HCTA) from *Xenorhabdus miraniensis* (NRPS-30), and module four of HCTA (NRPS-31) (Fig. 38c).

With the L/R variation the yields of **32** for NRPS-28 were raised by 22.3 % (Fig. 38c). For NRPS-29 the production rate of **32** with a production level compared to NRPS-28 of 126.7 % can be further increased slightly to 132.6 %. In case of the HCTA containing NRPSs, only linear products are produced with significant losses in production rates to 60.0 % for NRPS-30 and 0.7 % for NRPS-31. Nevertheless, an increase in the production with the L/R variation can be achieved, for both, of **38** by 14 % and of **39** by almost one third (Fig. 38c, Fig. S33-36).

Thus, the optimizing properties of the L/R variation can be confirmed since the production rate of all tested NRPS-28 to -31 were increased. Furthermore, this property was transferred to another interface in BicA and even to another NRPS – HCTA. BicA from *X. budapestensis* and HCTA from *X. miraniensis* both produces arginine-rich products and are related in their domain phylogeny and product spectra. It is assumed that the C domains in the respective NRPS anticipate an interaction of R856 (in the

BacA_CE2-GxpS_A2 model) with the A domain. For these NRPS, the L/R variation seems to represent a more functional interface variant.

4. Discussion

NRPSs produce highly functionalized natural products with bioactivities, important for many pharmaceutical and agrochemical applications[19]. For the successful synthesis of these highly diverse compounds, it is essential that the correct substrates are activated and that these substrates are then properly linked together to form the specific product. The central importance of these reactions for the understanding of NRPS biosynthesis is evident in the efforts that have been made in recent decades to decipher especially the adenylation and condensation domains inherent in these reactions[88].

The focus of this work has been to shed light on the particular role of C domains for NRP biosynthesis and how to correctly apply them to achieve successful NRPS reprogramming.

4.1. Topic A: Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains

4.1.1. Placement of the XUC concept between its predecessor and successor concepts

A concept in which C domains are actively used as the subject of NRPS engineering has been developed with the e**X**change **U**nit **C**ondensation domain (XUC) concept[166]. In this concept, self-contained catalytically active C_{ASub}-A-T-C_{DSub} (XUCs) units are used as building blocks. Thereby, and due to its pseudodimeric 'V' structure, the C domains are cut in half in its *N*-terminal C_{ASub} and *C*-terminal C_{DSub} subunits at a four amino acid long conformationally flexible loop/linker region.

Strikingly, the fusion site is located right beneath the binding pocket, thus partitioning it. Although the HHxxxDG core motif on the C_{DSub} and the substrate tunnel-forming lid

and floor loop elements on the C_{ASub} are separated, they appear to be structurally conserved enough to restore catalytic activity upon replacement. However, this massive intervention in the C domains' nature demonstrates the limitations of XUCs in a cross-species application. The exchange of *Xenorhabdus*/*Photorhabdus* with *Bacillus* XUCs could not produce functional NRPS[166]. Nevertheless, the XUC concept shows essential advantages compared to its preceding XU concept.

The XU concept uses a conserved fusion point in the C-A linker at the termed 'WNATE' amino acid motif to exchange A-T-C(/E) (XU) building blocks[139]. However, this always results in hybrid interdomain interfaces between the main catalytically active C and the A domains. Thus, the XU concept seems to find more universal applications but is paired with significantly reduced production titers[139]. The XUC concept, on the other hand, does not interfere with any important interdomain interfaces during the catalytic cycle of NRPS and therefore permits substantially higher production titers. This is also reflected in its use to increase product specificity and reduce by-product formation[166].

Although there are some barriers to accessing the C-A interface, the XU concept cannot be neglected as a potent engineering strategy. For instance, the insertion of a certain C domain bears the opportunity to influences the substrate recognition capacity of the respective downstream A domain (*cf*. Chapter 3.2). Whereas the inserted C domain used to be restricted by matching former downstream C domain specificities[139], it turned out that the C-A interface is more flexible for those C domains that do not directly conform to this rule (*cf*. Chapter 3.2). Moreover, the C domain interface forming properties, namely 'extended gatekeeping' function (*cf*. Chapter 3.2), contribute to the A domain's substrate recognition capacity and can be used to fine-tune substrate acceptance. Furthermore, with the XU concept synthetic zippers (SZs) can be applied at its respective fusion point[162]. In this process, the NRPS parts are separated on individual plasmids and can reassemble post-transcriptionally *in vivo*. With SZs at hand, peptide libraries can be created quickly and with a high success rate, even if the NRPSs are from different strains or species.

However, the XUC concept has an enormous advantage over the XU concept. The XUC concept only needs 80 building blocks (considering distinctive *N*- and *C*-terminal C or dual C/E domains) to cover the entire proteinogenic amino acid spectrum[166]. In

contrast, the XUs, despite its 'extended gatekeeping' (*cf.* Chapter 3.2) and SZs[162] application, up to 800 building blocks (each with matching downstream specificity and matching C domain type) still have to be made available to keep up with the XUC concept[166].

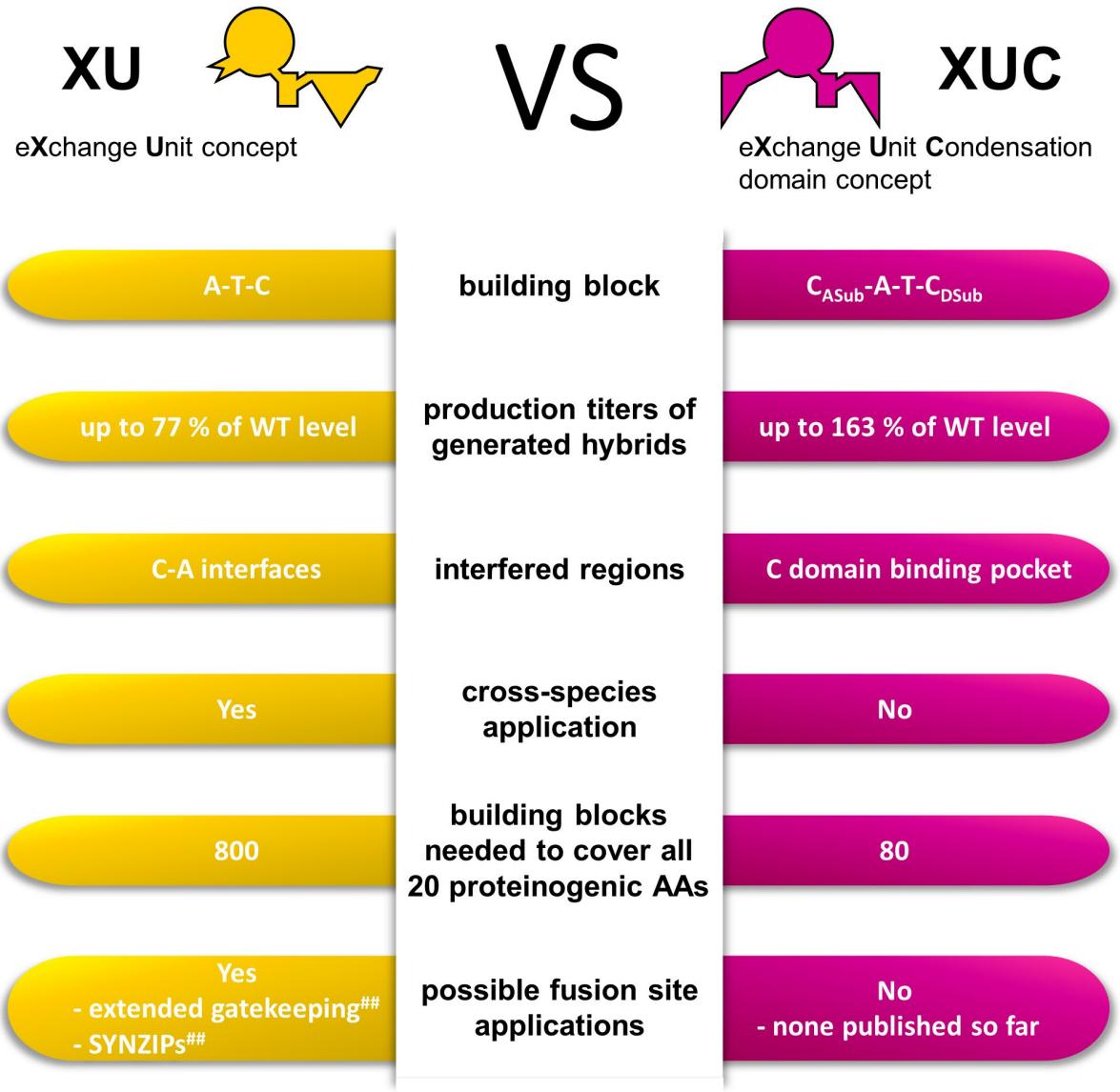


Figure 39: Comparison of XU and XUC concept. The eXchange Unit (XU) concept (yellow) on the left side is compared with the eXchange Unit Condensation domain concept (purple) on the right side by the arguments in the middle. The illustration of the building blocks on the top is depicted by the following domain symbols: adenylation (A) domain, large circle; thiolation (T) domain, small rectangle; condensation (C) domain, triangle; *N*-terminal donor condensation (C_{DSub}) domain subunit; right half triangle; *C*-terminal acceptor condensation (C_{ASub}) domain subunit; left half triangle.

4.2. Topic B: Influence of condensation domains on activity and specificity of adenylation domains

Even 60 years after the discovery of NRPSs, it is not clear whether condensation domains responsible for the condensation of covalently bound substrates have intrinsic substrate specificities or not - the proofreading role attributed to them is controversial. Since this lack of knowledge is a bottleneck in efficiently engineering, answering the question about the exact C domains' role is of great importance. It is imperative to prevent long standing design paradigms becoming a false dogma that influences future engineering efforts in the wrong way, such as the inseparability of C-A didomains and the C domains gatekeeping role[40,87,89,90,132,135,139,166].

The following chapter focuses on the recently sparked debate about the role of C domains in the non-ribosomal synthesis of peptides[174]. Based on our group's established expertise in engineering NRPSs, I tried to rethink the problem and approach it from different angles, focusing particularly on the changing behaviour of A domains in the context of chimeric biosynthetic pathways. Therefore, comprehensive experimental procedures ranging from *in vitro* and *in vivo* characterizations – targeting in our group preferred model system GxpS – to *in situ* investigations of the fabclavine producing BGC and *in silico* characterization of selected C-A didomain interfaces were created in this study.

4.2.1. Characterization of C domains influence on GxpS_A3

Within this study, the experience was made that *in vitro* results can paint a picture that contradict the results obtained *in vivo*. However, all *in vitro* and *in vivo* results concerning the selected promiscuous GxpS_A3 framework, revealed a significant influence of all C domains on both (Fig. 28-29), the general catalytic activity and the substrate recognition profile within the identified 'substrate group specificity' of the GxpS_A3 domain. The gathered data along with insights from recent literature questioning the C domains' role as selectivity filters[18,77,89,90,173], provide strong

indications for an extended gatekeeping function of C domains upstream of A domains rather than strict intrinsic selectivity.

Interestingly, in terms of phylogenetic distance and sequence homology, less similar C domains (e.g. $AmbS_{Xindi.}$_C5 & BicA_C3) seem to have a more pronounced effect (Fig. 28-29) – in both directions (Fig. 29b, NRPS-10 & -12). On the other hand, and as evident from the *in silico* investigations, it is not surprising that the introduced changes at the interface by exchanging the very similar XtpS_C3 with GxpS_C3 had almost no effect (Fig. 28) on the catalytic activity and substrate activation profile of GxpS_A3, as evidenced in the *in vitro* (Fig. 28) and *in vivo* (Fig. 29) assays. This effect could be described as an extended gatekeeping function of the C-A interface in fine-tuning the A domains selectivity and thus contributing to their role as a primary substrate selectivity filter. Similar results have also been reported previously for non-ribosomal assembly lines producing sulfazecin[188] and microcystin[84]. The latter for example comprises mono- and multi-specific modules that either strictly incorporate leucine or arginine or incorporate chemically diverse amino acids in parallel into microcystin[84]. Interestingly, in that study, the presence of the C domain's *C*-terminal subdomain, including all C-A interface-forming residues, was sufficient to restore, at least in part, the specificities observed *in vivo*, whereas in this work, the presence of the *C*-terminal subdomain (Fig. 28-29) even impaired the A domains capacity to efficiently recognize the substrates presented.

Noteworthy, BacA_C3 as a representative outside the *Photorhabdus* and *Xenorhabdus* genus exerts significant influence on the interface-dynamics which almost abolished the activity *in vitro* (Fig. 28b). In addition, the *in silico* analysis suggests an increased rigidity of the C-A interface, seemingly interfering with the A domain's ability to switch between the different conformations required for proper substrate binding/release and adenylate formation[100], as highlighted by P5's poor catalytic *in vitro* activity (Fig. 28b).

This observation now explains our previous inability to recombine building blocks of Gram-negative and -positive origin with each other by using the XU strategy in most cases[139]. Interestingly, most recently it was possible to functionally apply the very same interface (P5) by introducing synthetic leucin zippers (type S NRPSs) within the C-A linker region[162]. The created type S NRPS not only synthesized a thiazoline containing peptide with high fidelity at high titers, but the GxpS_A3 domain exclusively

activated leucine, thus completely omitting the domains *in vitro* confirmed preferred substrate phenylalanine[162] – representing another example of how altered C-A interactions are capable to contribute to the A domains attributed role as primary selectivity filter.

4.2.2. Characterization of C domains influence on FclJ_A6 in *Xenorhabdus szentirmaii*

By targeting the fabclavine BGCs from *X. szentirmaii*, *X. budapestensis*, and *X. hominicii*, XU substitutions to alter C-A interface interactions could be made that were least out of their natural context (Fig. 30). Here, a more dominant and transferable extended gatekeeping effect of the *Xsze* FclJ_C6 domain could be observed, which, through an additional loop in the interface (Fig. S5-7), mainly interacts with the A_{Sub} domain. Introduction of *Xsze* XU1 (C5A5T5C6) upstream of FclJ_A6 from *X. budapestensis* and *X. hominickii* empowered the formerly proline specific A domains to also activate valine and threonine (Fig. 30b-c).

These obtained *in situ* and *in silico* results suggest that, at least in case of fabclavine biosynthesis, the previously described promiscuity at this position in fabclavine biosynthesis[178] does not seem to be the exclusive result of the FclJ_A6 domains activity, but from the extended gatekeeping function of the FclJ_C6 domain that grants additional spatial flexibility (mainly in AOI1, Fig. 31).

The *Xsze* FclJ_C6 seems to follow its very own path in C-A interface formation with considerable differences, especially in the yet unreported interaction of the C domain's Chain II, and loss of interaction of Chain XIV, XV & XVI with A_{Sub} (Fig. 31), extending its already dynamic 30° rotation from the substrate binding/release to the adenylate forming conformation and subsequent 140° body torsion in the thioester forming conformation[60]. Interestingly, this extended gatekeeping, leading to less spatial restrictions on the A domain's movement, is not only transferred on a structural level when chimeric interfaces are created, but also influences the substrate recognition capacity of the respective downstream A domain.

4.3. Topic C: Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering

The C-A relatedness seems to have particular significance as the centre of the NRPSs' main catalytical activity. Whereas decisive recombination events or co-evolution of the C-A didomain could not be clearly determined so far[89,90,132,135], a dependency of correctly formed interdomain interfaces for accurate substrate incorporation and biosynthesis is observed[91,139]. In this context, on the one hand, a correlation of C-As' primary sequence similarities to NRPS production titers are already described by comparing the artificial and natural interfaces of generated NRPS hybrids[139]. In case of low similarity of the natural compared to the non-natural C-A interface, significant losses in the production rates were observed. On the other hand, an 'extended gatekeeping' function of the C domain (*cf*. Chapter 3.2 and Chapter 4.2) is suggested, not only significantly affecting the interface formation with the directly adjacent downstream A domain but also an influence on fine-tuning the substrate specificity of the A domain[84,91,92].

The following chapter combines an engineering with a phylogenetic/evolutionary perspective to uncover the traces of C-As evolutionary history to access the C-A (and its interface) not as a barrier, as it has been handled so far, but as another resource for effective NRPS reconstitution.

4.3.1. The communication between C and the A domain

Hitherto, the co-evolution of the C and A domains has been questioned, since their phylogenies show divergent evolution[89]. However, here, it was able to demonstrate in a comprehensive phylogenetic analysis that, on the one hand, the C-A interface alone encodes equivalent characteristics as corresponding to the complete C-A didomain (Fig. 34) and, on the other hand, the declared C-A interface acts as a topological-phylogenetic unit (Fig. 35, 37-38).

It turned out that the interface bears characteristics, according to which it can be clearly assigned to its donor strain (Fig. 34), and that recombination events most likely rarely happen within the interface (Fig. 35). Therefore, it is very likely that artificial C-A interfaces, as long as they are not adapted to their donor strain, directly affect productivity of engineered NRPS. Accordingly, the catalytic efficiency of redesigned NRPSs (constructed accordingly to the rules of the XU concept[139]) probably indeed correlate with the C-A identity (Fig. 34, 37), since the C domain subtype and A domain specificity are already specified in their interface. The latter was recently exemplified on bi-partite type S NRPSs[168], where the change of interface type from C/E-A to C-A had a negative effect on production titers.

Accordingly, a previously suspected further substrate-determining C domain subclasses[73,90] and a putative filtering function by a correctly formed C-A interface[91] seem to become more apparent with these interface inherent characteristics. These may be therefore the reason, rather than an intrinsic C_{ASub} substrate specificity, why certain C domains more likely form a functional interface with the subsequent A domain to ensure correct substrate incorporation. The phylogenetic distribution of C-A interfaces (Fig. 34) could thus provide a hint for further C domain subtypes that divide into subclasses based on their interface characteristics.

Furthermore, in line with previous results[92], the constructed NRPS-23 to -27 were functional even though individual building blocks were non-covalently linked (Fig. 37). However, the question arises how building blocks are capable to communicate instead. The C-A interface represents the largest domain-domain interface (in comparison to A-T and T-C[58,60,100]), which makes it conceivable that the sum of non-covalent interactions such as hydrogen bonds or hydrophobic interactions are sufficient to form a stable interface. The strength of interaction probably decreases with the identity of the C-A interface, which is reflected in the product titers (Fig. 37).

Nevertheless, the possibility remains that production titers of artificial NRPS not only depend on the C-A identity, but on many other parameters, e.g., genus, G/C content[189], TE specificity[105], correct linkage of accepted substrates by the C domain[127], or catalytic efficiency of donor NRPS[139,162,166,168]. Still, the C-A identity

is revealed to be a crucial parameter that has not been taken into consideration until now.

It turns out that the contextual consideration of the C and A domain indicates a minimal topological distance of the phylogenetic trees of interface-forming regions for both domains – which could be interpreted as co-evolutionary remnants of the C-A didomain migration. Unlike the A domain or rather the A domain binding pocket, it is conceivable for the C-A commonality to have confounded evolutionary boundaries, which could, however, be deduced from its phylogenetic (interface-)branching and comparison of tree-topologies by a broader perspective on the NRPS. This could be seen as another hint for the suspected effect of gene homogenisation through concerted evolution[145,150] and supports an explanation for how evolution solved the problem of effective NRPS genes recombination and the abundance of natural NRP congeners on countless occasions[190].

4.3.2. Computer-aided evaluation of the compatibility of C and A domains with the CAopt.py algorithm

At this point, the CAopt.py algorithm can be considered as a mimic of the concerted evolution, which is interpreted by a computer-aided scoring system (Fig. 36). This pattern of homology, consistent with the homogenisation of DNA sequences within a particular repetitive family caused by high rates of internal recombination, is meant to be captured with CAopt.py scoring in its properties. Unlike the comparison of primary sequence identity in an alignment, the coverage of biochemical properties can intend as more trustworthy for domain interfaces[191]. Furthermore, by comparing the scorings of the to be combined native interfaces, a prediction can be made for the compatibility of the hybrid interface. It is surprising that none of the carried-out hybrid interface combinations, even with higher scores, resulted in unfunctional NRPSs (Fig. 37). Still, CAopt.py gives reliable predictions to avoid the generation of unproductive NRPS hybrids already during the *in silico* planning phase.

Moreover, a randomization tool implemented in this CAopt.py algorithm enables the systematic alternation of the interface-forming residues (Fig. 36, 39a). The targeted

alternation for 'better' or 'worse' fitting interfaces carried out in this process showed few successful-producing NRPS hybrids, nevertheless, highlighting its huge potential with the NRPS-26 variant's improved score of 5.9 to 1.8 and a massively 5.5 times increased production rate (Fig. 38a). Particularly, the program guided us to the R/L variation, which turned out to be a better interface variant and transferrable to other interfaces, contribute to more productive NRPS biosynthesis (Fig. 38c). The numerous non-functional NRPS generated by the randomization tool raises the conclusion that the interface does not allow any serious changes at crucial positions even with good scoring. Since the C-A interface has to adopt a multitude of interaction surfaces during the catalytic cycle of NRPS in a set of mainly hydrophobic interactions, the rash alternation of interaction residues might result in impairment of C-A communication. Thus, the interface alternation can affect the functionality and production rate in the same way as it has been shown for hybrid interfaces (*cf*. Chapter 3.2 and Chapter 3.3). However, the algorithm is still proved to be suitable for evaluating functional NRPS since these critical regions are expected to show a certain consistency in the compared natural interfaces but is in any case in need of further optimization.

To integrate this 'natural calculation guidance' by homologues regions in predominantly examined *Xenorhabdus*/*Photorhabdus* C-A interfaces and overcome the poor functional NRPS success-rates (Fig. 38a), the CAopt.py script needs to implement a rating for these regions to be alternated so that more crucial areas are less likely changed or have a more significant effect on the final scoring. Noteworthy, evaluating and optimizing a hybrid interface based on its native interface to increase production raises the question of whether the underlying native interface is a 'good' interface variant at all. The term 'good' or 'bad' interface is to be understood as the user's subjective assessment of a better or worse producing NPRS, however, even hybrid NRPS, which always has a worse score than the native interface in the CAopt.py algorithm, allow improved production rates. Accordingly, this holds the potential to discover more unique interface variants in the future with an optimized CAopt.py algorithm to hopefully find a universal interface that guarantees proper production titers for general C-A interfaces even with inappropriate C-A matching.

Interface editing will not overcome the limitation of the introduced NRPS modules but can act as another setscrew in the three-dimensional puzzle of NRPS. Yet, it has been shown that it is worth to risk a different point of view to give a complex picture a different perspective.

4.4. Further prospects

4.4.1. Role of NRPS domains and their interacting interfaces for substrate selection

Protein-protein interactions (PPIs) show enormous importance for the effective performance of protein function[192]. A variety of activities essential for life rely on correct PPIs to perform DNA replication, transcription, protein translation, or signal transduction[131,193]. A correctly formed interaction surface has significant impact not only on the protein itself, but also on its interacting environment[192]. NRPSs showcase the complexity achieved by the multi-modular assembly lines of protein domains acting in *cis* and *trans* through a diverse interaction landscape[40]. A number of studies have shown that the specificity and activity of the A domain function can be influenced by the surrounding domains[84,90,91,131,135,139]. Especially the interaction of the A domain with the neighbouring C domain is shown to be the most extensive protein interaction within NRPSs (see Chapter 1.4.1). While this C-A interface has long been treated as inseparable[58,88], substitutions of the C domain have been able to provide important insights into this domain communication[84,91,162,166,168]. At this point, the current controversy, whether the C domain exhibits intrinsic specificity of the linked substrates and takes on the role as an additional gatekeeper, lines up[89]. The combination of these observations raises the question of the exact nature of the domains' contribution to a functional C-A domain interaction and the resulting impact of the formed C-A interface for non-ribosomal synthesis of peptides.

In the first place, the almost complete deciphering of the A domains' nature allows it to be attributed as the primary contributor to substrate selection and activation (see Chapter 1.2.1). The amino acid composition of an NRP can be predominantly predicted by the A domains and the specificity-conferring codes of their binding pockets[53,147]. In contrast, the C domains show a more incomplete picture in which the origin of the substrate-determining properties still leaves room for speculations[89]. However, an extended regulatory function for fine-tuning substrate selection could be assigned to the C-A interaction[91] (*cf.* Chapter 3.2 and 4.2). This is determined, on the one hand, by the structural configuration of the C domain itself, as evidenced by *in vitro*

experiments where no, half- or full-length preceding C domains results in drastic changes in substrate activation of the adjacent A domain[84]. On the other hand, the contact area of the C_{ASub} formed with the A domain could be attributed an important role in maintaining effective C-A communication in the catalytical stages of biosynthesis[91,92]. Essentially, this, termed as the 'extended gatekeeping' function of the C domain (*cf*. Chapter 3.2 and 4.2), can therefore be interpreted as the contribution of the C domain to the substrate fine-tuning by a functional domain-domain or protein-protein interaction. In this regard, it was possible to identify influences for substrate acceptance by added physicochemical strictness or flexibility to the interface, or even unique structural elements of certain C domains (*cf*. Chapter 3.2). These C domains were identified to confer a differential substrate- or activity-affecting effect on the A domain despite the same stereoselective classification as, e.g., $^{L}C_{L}$ domains[73]. Previously, the influence of the C domains was tried to be described by the mixed success rate of engineering attempts through a substrate specificity assigned to the C_{ASub} and different C subtypes[42,88]. Whereas the C domain specificity could be explained by the 'extended gatekeeping' function, as it is more likely that C domains form functionally similar interfaces with adjacent A domains with similar specificity (Fig. S37), the presents of further C subtypes remain to be proven. However, for PCPs, different subtypes could already be described, which differ in interaction with E or C domains and are therefore termed as PCP^{E} or PCP^{C}[170]. Also, for TEs, it could be shown in phylogenetic studies that they do not group according to the type of offloading chemistry they catalyse but according to their function at the corresponding position in the NRPS[105]. Accordingly, it stands to reason that likewise for C domains, additional subtypes may exist that exert a more pronounced influence through interaction with surrounding environments on specificity and activity.

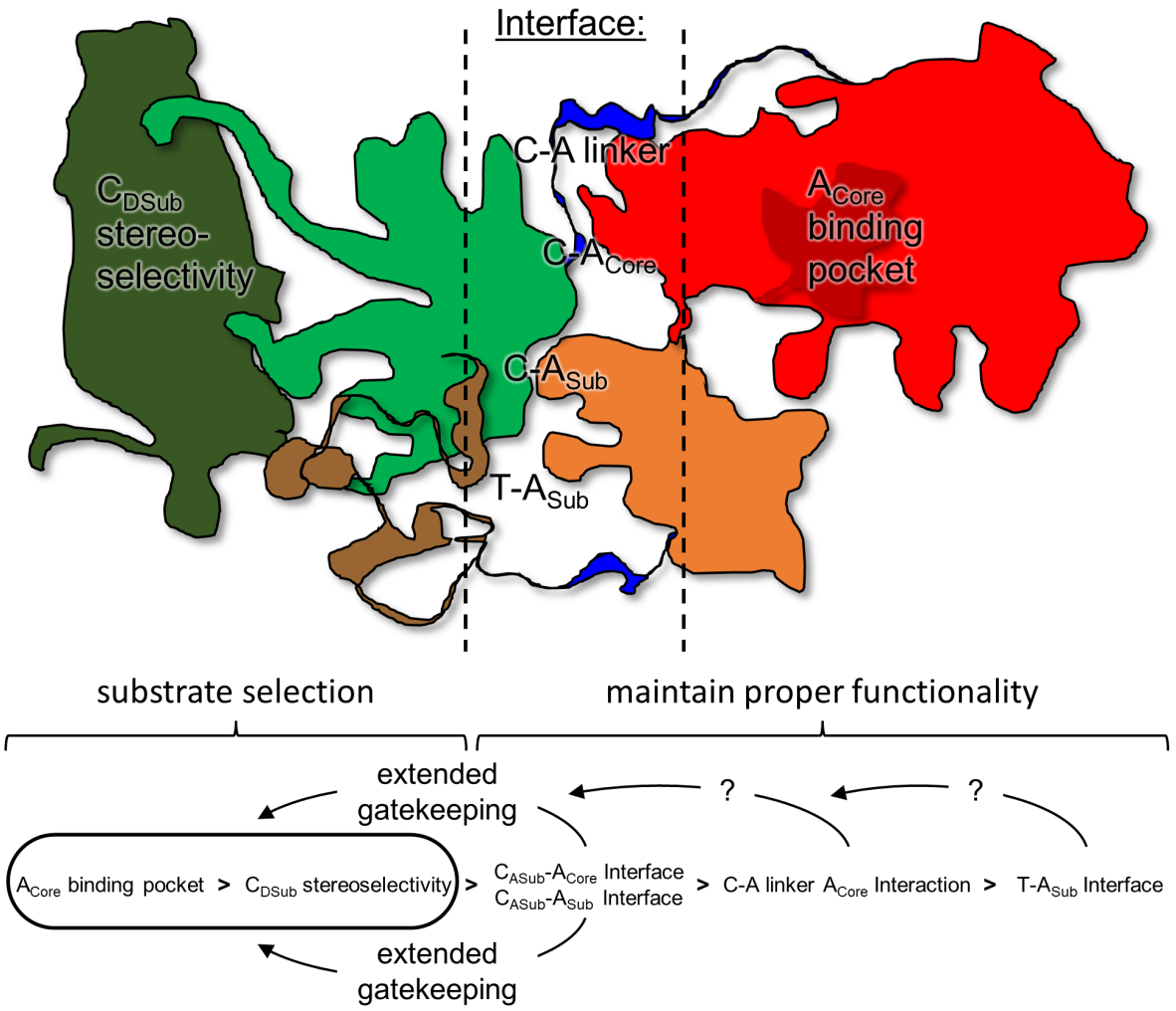


Figure 40: Hierarchy of substrate acceptance in NRPS. Crystal cartoon model of the structure of SrfA-C (PDB ID: 2VSQ)[58]. The C domain's *N*-terminal donor C_{DSub} (dark green) and *C*-terminal acceptor C_{ASub} (light green) subunit, the larger *N*-terminal A_{Core} (red) and smaller *C*-terminal A_{Sub} (orange) domain, the PCP (brown), and linkers (blue) are highlighted. Major substrate incorporating check-points including the A_{Core} binding pocket, C_{DSub} stereoselectivity, and the interacting interface C-A_{Core}, C-A_{Sub}, C-A linker, T-A_{Sub} are labelled. In the interaction hierarchy below, the proposed influences on substrate selection for maintaining proper functionality are shown with their suspected mechanisms.

These multiple sub-pictures allow to sketch a hierarchy of consecutive check-points for proper substrate incorporating (Fig. 40). The A domain is the main gatekeeper, allowing the selection and activation of the appropriate amino acid by the composition of its binding pocket. The C domains' role as an additional gatekeeper through an intrinsic substrate selectivity is unclear, however, it is undoubtedly essential for stereoselective filtering. These two main check-points are further influenced by their interacting domain environment. Next in importance is the C-A interface, to which a direct influence on substrate selectivity could already be assigned by the 'extended gatekeeping' function of the C domain (*cf.* Chapter 3.2 and 4.2). Here, the interface-forming areas with the A_{Core} must be distinguished from the more dynamic A_{Sub}. The

A_{Core} is rather rigid, but depicting substantial interactions with the adjacent C-A linker region, which could be assigned to influence the correct transfer of the A domain function when substitution experiments[90]. The A_{Sub}, on the other hand, mainly shows an influence through the freedom of flexibility left to it, making it responsible for the adaption of the different interface conformations during the catalytic cycle and the recruitment of the PCP[60,100]. Accordingly, a complex picture of substrate activating influences emerges, which requires the correct interaction of the domains involved and, consequently, can also be influenced to a final selection.

4.4.2. Anticipation of evolutionary perspectives on the reengineering of biosynthetic assembly lines

Understanding the evolution of NRPS and PKS biosynthetic pathways is of central interest for the development of new engineering approaches[18]. This offers unique opportunities for effective *de novo* design of (clinically) relevant natural products.

Many NRPS engineering strategies have already been established, which generally showed rather moderate success in universal application[42,88]. These mainly involved substitutions within the same or with related NRPS BGCs. However, the less related the application was from the subject of the original experimental object or genus, the lower success rates were to be expected (see Chapter 1.3). The basic common feature of these methods is that, in addition to modification of the binding pockets, they rely mainly on separation at the domain boundaries of the assignable protein families[128–131,139]. In this regard, the consideration of the evolutionary history of multi-modular assembly line biosynthesis could reveal primal recombination sites and evolutionary events and provide important lessons on NRPS refactoring.

There are several primary evolutionary mechanisms, which offer a potent arsenal for natural NRPS re-engineering strategies (see Chapter 1.5). The main mechanisms to be named are duplications of gene coding regions in a linear or iterative manner, and the exchanges with similar gene clusters[145]. These can occur individually and in combination, and contribute to the divergence and convergence of BGCs[18]. It is important to note that recombination events are bound to both sequence similarity and the function of the molecule on its target. Accordingly, the recombination-tendency is

bound to the molecules function, and loss of a benefit in this function can also lead to the loss of a BGC[142]. In addition, there is a significantly higher occurrence of co-evolving events than in comparable gene clusters involved in primary metabolism[145,146]. Thereby, the boundaries of exchange can vary and thus recombine small fragments within a module, whole domains, and even didomains up to multiple domains[145,146]. This reveals a manifold toolbox of suitable assembly building blocks imposed by NRPSs nature. Accordingly, the sum of moderately successful NRPS engineering attempts can nonetheless be interpreted as a set of already possible fusion points, which, however, essentially apply effectively only to the individually considered or related BGC.

A more extensive elucidation of the evolutionary history of NRPS and the individual NRPS BGC itself may reveal novel and 'already established and proven by nature' assembly strategies. Thus, engineering strategies could be uncovered that rely on the evolutionary/recombination events undergone by the particular NRPS BGC and with which it is confidently compatible. It is not expected that for each NRPS BGC an individual engineering strategy must be deciphered for a successful reassembling, since the occurring evolution/recombination events are limited[142]. However, at least individual clusters could be grouped together based on their related evolutionary histories and addressed by corresponding engineering strategies. For example, *Xenorhabdus*/*Photorhabdus* NRPS clusters can be distinguished from *Bacillus* clusters regarding their occurrence of epimerisation elements. While in *Bacillus* NRPS, the separate E and C domains are most likely present and dual C/E domains are rarely found[15,46,115,167], as it is the case for *Xenorhabdus*/*Photorhabdus* NRPS[45,47,165]. Thus, a different evolutionary history of epimerization installation can be assumed, which should be considered when selecting and applying an appropriate reassembling site.

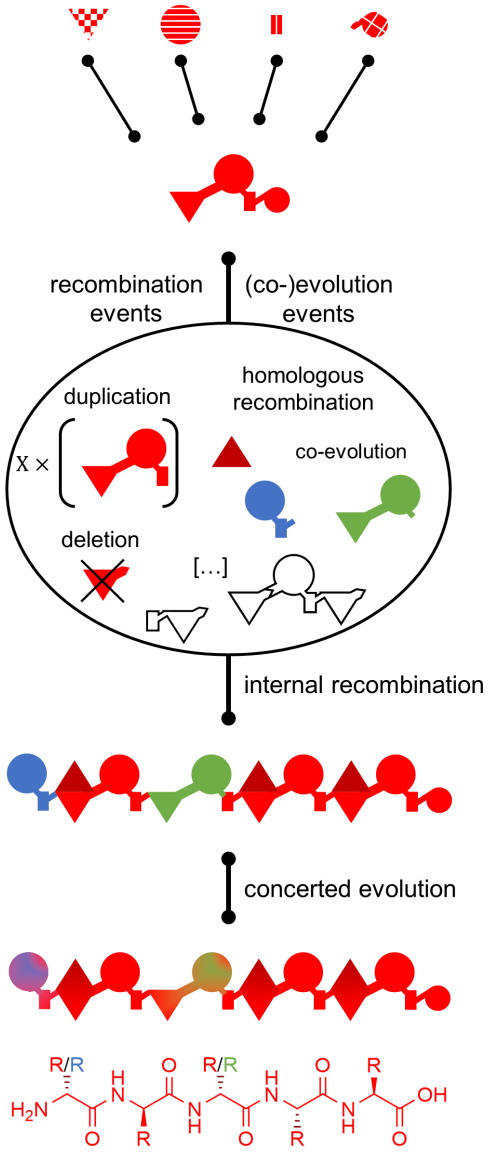


Figure 41: Fictional evolution of an NRPS. Flowchart of the evolution history of a hypothetical NRPS through a series of evolutionary events. The domains are illustrated by the following symbols: adenylation (A) domain, large circle; thiolation (T) domain, small rectangle; condensation (C) domain, triangle; epimerisation domain, upside-down triangle; dual condensation /epimerization (C/E) domain, diamond; thioesterase (TE) domain, small circle. The domain origin from different protein families is represented by the patterns of the domain symbols (red↔white tiled/hatched) and the origin of NRPS building blocks from different NRPS BGCs are represented by their colour (red, dark red, blue, green, colourless). The colouring also applies to the structure of the hypothetically produced peptide and the respective side chains.

In another example of an assumed evolutionary history of an NRPS, the domain composition and nature can be inferred (Fig. 41). In this illustration, the origin of the NRPS is derived from a single multifunctional NRPS module. This undergone in its further development several duplications of this module, the insertion of other co-evolving and functional units. Accordingly, the different specificities of modules 1 and 3 could have arisen in this way in addition to intersected epimerisation units to form dual C/E domains. Concerted evolution would then have increased the homology of the inserted domain to the final NRPS in the process, confounding recombination boundaries. The relaxed flexibility in modules 1 and 3 could therefore be interpreted as relicts of this evolutionary history. Regarding evolution-derived engineering strategies, these events would have to be considered to decipher suitable fusion sites. One reason for an unpromising engineering strategy could be that an 'unnatural' fusion

point within the originally foreign NRPS part is chosen and attempt to normalize it to the NRPS intrinsic elements. Suppose a fusion point is chosen within the boundaries of the natural recombination site, more effective reassembly is expected with other building blocks that were also separated at natural recombination sites.

Nature has been practicing the recombination of NRPS into fantastic natural product producers for several million years and consequently has far greater expertise in natural product research than we can ever achieve. Therefore, elucidating the natural trails of evolutionary history can be a great educator on the path to novel engineering approaches and natural products.

5. References

[1] D. J. Newman, G. M. Cragg, *Journal of natural products* **2020**, *83*, 770–803.

[2] M. I. Hutchings, A. W. Truman, B. Wilkinson, *Current opinion in microbiology* **2019**, *51*, 72–80.

[3] J. Bérdy, *The Journal of antibiotics* **2012**, *65*, 385–395.

[4] PEW Trust, "Antibiotics Currently in Global Clinical Development", to be found under https://www.pewtrusts.org/de/research-and-analysis/data-visualizations/2014/antibiotics-currently-in-clinical-development, **2021**.

[5] G. Kreysa, S. Grabley **2007**.

[6] P. Ehrlich, *Anwendung und Wirkung von Salvarsan*, Frankfurt am Main, **1910**.

[7] A. Fleming, *Br J Exp Pathol* **1929**, 226–236.

[8] S. A. Waksman, A. Schatz, D. M. Reynolds, *Ann N Y Acad Sci* **1946**, 73–86.

[9] L. Katz, R. H. Baltz, *Journal of industrial microbiology & biotechnology* **2016**, *43*, 155–176.

[10] E. Bode, A. K. Heinrich, M. Hirschmann, D. Abebew, Y.-N. Shi, T. D. Vo, F. Wesche, Y.-M. Shi, P. Grün, S. Simonyi et al., *Angewandte Chemie (International ed. in English)* **2019**, *58*, 18957–18963.

[11] K. A. J. Bozhüyük, J. Micklefield, B. Wilkinson, *Current opinion in microbiology* **2019**, *51*, 88–96.

[12] E. A. Felnagle, E. E. Jackson, Y. A. Chan, A. M. Podevels, A. D. Berti, M. D. McMahon, M. G. Thomas, *Mol Pharm* **2008**, *5*, 191–211.

[13] Y.-M. Shi, H. B. Bode, *Nat Prod Rep* **2018**, *35*, 309–335.

[14] H. S. Burton, E. P. Abraham, *Biochem J* **1951**, 168–174.

[15] R. J. Dubos, R. D. Hotchkiss, *J Exp Med* **1941**, *73*, 629–640.

[16] J. R. Woodworth, E. H. Nyhart Jr, G. L. Brier, J. D. Woly, H. R. Black, *Antimicrob Agents Chemother* **1992**, 318–325.

[17] H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen, *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 9259–9264.

[18] T. J. Booth, K. A. J. Bozhüyük, J. D. Liston, E. Lacey, B. Wilkinson **2021**.

[19] M. Strieker, A. Tanović, M. A. Marahiel, *Curr Opin Struct Biol* **2010**, *20*, 234–240.

[20] S. A. Sieber, M. A. Marahiel, *Chem Rev* **2005**, *105*, 715–738.

[21] B. Mach, E. Reich, E. R. Tatum, *Proc Natl Acad Sci U S A* **1963**, *50*, 175–181.

[22] T. L. Berg, L. O. Froholm, S. G. Laland, *Biochem J* **1965**, *96*, 43–52.

[23] Z. K. Borowska, E. L. Tatum, *Biochim Biophys Acta* **1966**, *114*, 206–209.

[24] K. Fujikawa, T. Suzuki, K. Kurahashi, *J Biochem* **1966**, *60*, 216–218.

[25] M. J. Daniels, *Biochim Biophys Acta* **1968**, *156*, 119–127.

[26] U. Spaeren, L. O. Froholm, S. G. Laland, *Biochem J* **1967**, *102*, 586–592.

[27] W. Gevers, H. Kleinkauf, F. Lipmann, *Proc Natl Acad Sci U S A* **1968**, *60*, 269–276.

[28] H. Kleinkauf, W. Gevers, F. Lipmann, *Proc Natl Acad Sci U S A* **1969**, *62*, 226–233.

[29] W. Gevers, H. Kleinkauf, F. Lipmann, *Proc Natl Acad Sci U S A* **1969**, *63*, 1335–1342.

[30] C. C. Gilhuus-Moe, T. Kristensen, J. E. Bredesen, T. L. Zimmer, S. G. Laland, *FEBS Letters* **1970**, *7*.

[31] R. Roskoski, H. Kleinkauf, W. Gevers, F. Lipmann, *Biochemistry* **1970**, *9*, 4846–4851.

[32] F. Lipmann, *Science (New York, N.Y.)* **1971**, *173*, 875–884.

[33] S. G. Lee, R. Roskoski, JR., K. Bauer, F. Lipmann, *Biochemistry* **1973**, *12*, 398–405.

[34] F. Lipmann, *Acta Endocrinol Suppl (Copenh)* **1973**, *180*, 294–300.

[35] K. Kurahashi, *Annu Rev Biochem* **1974**, *43*, 445–459.

[36] a) R. Weckermann, R. Fürbass, M. A. Marahiel, *Nucleic Acids Res* **1988**, 1988; b) G. Mittenhuber, R. Weckermann, M. A. Marahiel, *J Bacteriol* **1989**, *171*, 4881–4887; c) J. Krätzschmar, M. Krause, M. A. Marahiel, *J Bacteriol* **1989**, *171*, 5422–5429;

[37] a) W. Schlumbohm, T. Stein, C. Ullrich, J. Vater, M. Krause, M. A. Marahiel, V. Kruft, B. Wittmann-Liebold, *J Biol Chem* **1991**, *266*, 23135–23141; b) T. Stein, J. Vater, V. Kruft, B. Wittmann-Liebold, Franke. P., M. Panico, R. Mc Dowell, H. R. Morris, *FEBS Letters* **1994**, *340*, 39–44;

[38] H. D. Mootz, M. A. Marahiel, *J Bacteriol* **1997**, *179*, 6843–6850.

[39] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem Rev* **1997**, *97*, 2651–2673.

[40] R. D. Süssmuth, A. Mainz, *Angew Chem Int Ed Engl* **2017**, *56*, 3770–3821.

[41] M. Hahn, T. Stachelhaus, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 15585–15590.

[42] M. Alanjary, C. Cano-Prieto, H. Gross, M. H. Medema, *Nat Prod Rep* **2019**, *36*, 1249–1261.

[43] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Chembiochem* **2002**, *3*, 490 ± 504.

[44] F. I. Nollmann, C. Dauth, G. Mulley, C. Kegler, M. Kaiser, N. R. Waterfield, H. B. Bode, *Chembiochem* **2015**, *16*, 205–208.

[45] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün, *Chemistry* **2012**, *18*, 2342–2348.

[46] F. Lipmann, H. Grevers, H. Kleinkauf, R. Roskoski, JR., *Adv Enzymol Relat Areas Mol Biol* **1971**, 1–34.

[47] H. B. Bode, A. O. Brachmann, K. B. Jadhav, L. Seyfarth, C. Dauth, S. W. Fuchs, M. Kaiser, N. R. Waterfield, H. Sack, S. H. Heinemann et al., *Angewandte Chemie (International ed. in English)* **2015**, *54*, 10352–10355.

[48] M. Juguet, S. Lautru, F.-X. Francou, S. Nezbedová, P. Leblond, M. Gondry, J.-L. Pernodet, *Chemistry & biology* **2009**, *16*, 421–431.

[49] R. Zocher, U. Keller, H. Kleinkauf, *Biochemistry* **1982**, *21*, 43–48.

[50] K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chemistry & biology* **2007**, *14*, 13–22.

[51] N. A. Magarvey, B. Haltli, M. He, M. Greenstein, J. A. Hucul, *Antimicrob Agents Chemother* **2006**, *50*, 2167–2177.

[52] L. Du, C. Sánchez, M. Chen, D. J. Edwards, B. Shen, *Chemistry & biology* **2000**, *7*, 623–642.

[53] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem Biol* **1999**, *6*, 493–505.

[54] E. Conti, N. F. Franks, F. Brick, *Structure (London, England : 1993)* **1996**, *4*, 287–298.

[55] a) A. M. Gulick, V. J. Starai, A. R. Horswill, K. M. Homick, J. C. Escalante-Semerena, *Biochemistry* **2003**, *42*, 2866–2873; b) J. J. May, N. Kessler, M. A. Marahiel, M. T. Stubbs, *Proc Natl Acad Sci U S A* **2002**, *99*, 12120–12125;

[56] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J* **1997**, *16*, 4174–4183.

[57] A. M. Gulick, *ACS Chem Biol* **2009**, *4*, 811–827.

[58] A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science (New York, N.Y.)* **2008**, *321*, 659–663.

[59] J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt, T. M. Schmeing, *Science (New York, N.Y.)* **2019**, *366*.

[60] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235–238.

[61] J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, *Nature* **2016**, *529*, 239–242.

[62] A. M. Gulick, *Current opinion in chemical biology* **2016**, *35*, 89–96.

[63] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure (London, England : 1993)* **2000**, *8*, 407–418.

[64] J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips Jr., B. Shen, *Proteins* **2014**, *82*, 1210–1218.

[65] R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem Biol* **1996**, *3*, 923–936.

[66] K. Haslinger, C. Redfield, M. J. Cryle, *Proteins* **2015**, *83*, 711–721.

[67] T. Kittilä, A. Mollo, L. K. Charkoudian, M. J. Cryle, *Angew. Chem.* **2016**, *128*, 9988–9995.

[68] K. J. Weissman, H. Hong, B. Popovic, F. Meersman, *Chemistry & biology* **2006**, *13*, 625–636.

[69] a) P. Tufar, S. Rahighi, F. I. Kraas, D. K. Kirchner, F. Löhr, E. Henrich, J. Köpke, I. Dikic, P. Güntert, M. A. Marahiel et al., *Chemistry & biology* **2014**, *21*, 552–562; b) S. Zimmermann, S. Pfennig, P. Neumann, H. Yonus, U. Weininger, M. Kovermann, J. Balbach, M. T. Stubbs, *FEBS Lett* **2015**, *589*, 2283–2289; c) A. C. Goodrich, B. J. Harden, D. P. Frueh, *Journal of the American Chemical Society* **2015**, *137*, 12100–12109; d) M. J. Jaremko, D. J. Lee, S. J. Opella, M. D. Burkart, *Journal of the American Chemical Society* **2015**, *137*, 11546–11549; e) J. R. Lai, A. Koglin, C. T. Walsh, *Biochemistry* **2006**, *45*, 14869–14879; f) J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, *Chemistry & biology* **2012**, *19*, 188–198;

[70] Y. Liu, T. Zheng, S. D. Bruner, *Chemistry & biology* **2011**, *18*, 1482–1488.

[71] T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nature structural biology* **2002**, *9*, 522–526.

[72] R. Finking, M. A. Marahiel, *Annu Rev Microbiol* **2004**, *58*, 453–488.

[73] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol Biol* **2007**, *7*.

[74] S. L. Clugston, S. A. Sieber, M. A. Marahiel, C. T. Walsh, *ACS Chem Biol* **2003**, 12095–12104.

[75] V. de De Crécy-Lagard, P. Marlière, W. saurin, *C R Acad Sci III* **1995**, *318*, 927–936.

[76] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L.-O. Essen, *Structure (London, England : 1993)* **2007**, *15*, 781–792.

[77] T. Izoré, Y. T. Candace Ho, J. A. Kaczmarski, A. Gavriilidou, K. H. Chow, D. L. Steer, R. J. A. Goode, R. B. Schittenhelm, J. Tailhades, M. Tosin et al., *Nature communications* **2021**, *12*, 2511.

[78] T. Stachelhaus, H. D. Mootz, V. Bergendah, M. A. Marahiel, *The jounal of biological chemistry* **1998**, *273*, 22773–22781.

[79] K. Bloudoff, D. Rodionov, T. M. Schmeing, *J Mol Biol* **2013**, *425*, 3137–3150.

[80] W.-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, *ACS Chem Biol* **2016**, *11*, 2293–2303.

[81] L. Zhong, X. Diao, N. Zhang, F. Li, H. Zhou, H. Chen, X. Bai, X. Ren, Y. Zhang, D. Wu et al., *Nature communications* **2021**, *12*, 296.

[82] K. Bloudoff, D. A. Alonzo, T. M. Schmeing, *Cell Chem Biol* **2016**, *23*, 331–339.

[83] S. Kuhlenkoetter, W. Wintermeyer, M. V. Rodnina, *Nature* **2011**, *476*, 351–354.

[84] S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth, E. Dittmann, *Cell chemical biology* **2016**, *23*, 462–471.

[85] T. Izoré, M. J. Cryle, *Nat Prod Rep* **2018**, *35*, 1120–1139.

[86] T. Izoré, J. Tailhades, M. H. Hansen, J. A. Kaczmarski, C. J. Jackson, M. J. Cryle, *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 2913–2918.

[87] P. J. Belshaw, C. T. Walsh, T. Stachelhaus, *Science (New York, N.Y.)* **1999**, 486–489.

[88] A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nat Prod Rep* **2018**, *35*, 1210–1228.

[89] M. Baunach, S. Chowdhury, P. Stallforth, E. Dittmann, *Molecular biology and evolution* **2021**.

[90] M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nature communications* **2020**, *11*, 4554.

[91] M. Kaniusaite, J. Tailhades, E. A. Marschall, R. J. A. Goode, R. B. Schittenhelm, M. J. Cryle, *Chemical science* **2019**, *10*, 9466–9482.

[92] M. Kaniusaite, R. J. A. Goode, J. Tailhades, R. B. Schittenhelm, M. J. Cryle, *Chem. Sci.* **2020**, *11*, 9443–9458.

[93] S. A. Samel, P. Czodrowski, L.-O. Essen, *Acta crystallographica. Section D, Biological crystallography* **2014**, *70*, 1442–1452.

[94] D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh, C. L. Drennan, *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 12432–12437.

[95] E. S. Sattely, M. A. Fischbach, C. T. Walsh, *Natural product reports* **2008**, *25*, 757–793.

[96] X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh, Y. Tang, *Nature chemical biology* **2012**, *8*, 823–830.

[97] A. Tietze, Y.-N. Shi, M. Kronenwerth, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2020**, *21*, 2750–2754.

[98] S. Weinig, H.-J. Hecht, T. Mahmud, R. Müller, *Chemistry & biology* **2003**, *10*, 939–952.

[99] J. A. Salas, C. Méndez, *Journal of molecular microbiology and biotechnology* **2005**, *9*, 77–85.

[100] J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, *Curr Opin Struct Biol* **2018**, *49*, 104–113.

[101] a) Lawson, D., M., U. Derewenda, L. Serre, S. Ferri, R. Szittner, Y. Wei, Meighen, E., A., Z. S. Derewenda, *Biochemistry* **1994**, 9382–9388; b) S. D. Bruner, T. Weber, Rahul, M., Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure (London, England : 1993)* **2002**, 301–310;

[102] S. C. Dillon, A. Bateman, *BMC bioinformatics* **2004**, *5*, 109.

[103] D. C. Cantu, Y. Chen, P. J. Reilly, *Protein science : a publication of the Protein Society* **2010**, *19*, 1281–1295.

[104] L. Du, L. Lou, *Natural product reports* **2010**, *27*, 255–278.

[105] M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Natural product reports* **2016**, *33*, 183–202.

[106] a) R. M. Kohli, C. T. Walsh, *Chemical communications (Cambridge, England)* **2003**, 297–307; b) F. Kopp, M. A. Marahiel, *Nat Prod Rep* **2007**, *24*, 735–749;

[107] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **2000**, *407*, 215–218.

[108] K. Buntin, K. J. Weissman, R. Müller, *Chembiochem : a European journal of chemical biology* **2010**, *11*, 1137–1146.

[109] Y. Xie, Q. Cai, H. Ren, L. Wang, H. Xu, B. Hong, L. Wu, R. Chen, *Journal of natural products* **2014**, *77*, 1744–1748.

[110] R. Traber, H. Hofmann, H. Kobel, *J Antibiot (Tokyo)* **1989**, 591–597.

[111] S. Moran, D. K. Rai, B. R. Clark, C. D. Murphy, *Organic & biomolecular chemistry* **2009**, *7*, 644–646.

[112] A. Kirschning, F. Hahn, *Angewandte Chemie (International ed. in English)* **2012**, *51*, 4012–4022.

[113] a) R. D. Süssmuth, W. Wohlleben, *Applied microbiology and biotechnology* **2004**, *63*, 344–350; b) A. S. Eustáquio, B. S. Moore, *Angewandte Chemie (International ed. in English)* **2008**, *47*, 3936–3938;

[114] F. Wesemann, A. Heutling, P. Wienecke, A. Kirschning, *Chembiochem : a European journal of chemical biology* **2020**, *21*, 2927–2930.

[115] T. Stachelhaus, A. Schneider, M. A. Marahiel, *Science (New York, N.Y.)* **1995**, *269*, 69–72.

[116] A. Schneider, T. Stachelhaus, M. A. Marahiel, *Mol Gen Genet* **1998**, 308–318.

[117] K. Eppelmann, T. Stachelhaus, M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718–9726.

[118] J. W. Han, E. Y. Kim, J. M. Lee, Y. S. Kim, E. Bang, B. S. Kim, *Biotechnology letters* **2012**, *34*, 1327–1334.

[119] X. Bian, A. Plaza, F. Yan, Y. Zhang, R. Müller, *Biotechnology and bioengineering* **2015**, *112*, 1343–1353.

[120] H. Kries, R. Wachtel, A. Pabst, B. Wanner, D. Niquille, D. Hilvert, *Angew Chem Int Ed Engl* **2014**, *53*, 10105–10108.

[121] A. J. Pérez, F. Wesche, H. Adihou, H. B. Bode, *Chemistry (Weinheim an der Bergstrasse, Germany)* **2016**, *22*, 639–645.

[122] B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao, N. L. Kelleher, *Chemistry & biology* **2011**, *18*, 601–607.

[123] K. Zhang, K. M. Nelson, K. Bhuripanyo, K. D. Grimes, B. Zhao, C. C. Aldrich, J. Yin, *Chemistry & biology* **2013**, *20*, 92–101.

[124] H. Jenke-Kodama, T. Börner, E. Dittmann, *PLoS computational biology* **2006**, *2*, e132.

[125] M. Crüsemann, C. Kohlhaas, J. Piel, *Chem. Sci.* **2013**, *4*, 1041–1045.

[126] H. Kries, D. L. Niquille, D. Hilvert, *Chemistry & biology* **2015**, *22*, 640–648.

[127] D. E. Ehmann, J. W. Trauger, T. Stachelhaus, C. T. Walsh, *Chemistry & biology* **2000**, *7*, 765–772.

[128] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 5848–5853.

[129] D. Schwarzer, H. D. Mootz, M. A. Marahiel, *Chemistry & biology* **2001**, *8*, 997–1010.

[130] S. Doekel, M. A. Marahiel, *Chemistry & biology* **2000**, *7*, 373–384.

[131] T. Duerfahrt, S. Doekel, T. Sonke, P. J. L. M. Quaedflieg, M. A. Marahiel, *European journal of biochemistry* **2003**, *270*, 4555–4563.

[132] R. H. Baltz, *ACS synthetic biology* **2014**, *3*, 748–758.

[133] K. T. Nguyen, D. Ritz, J.-Q. Gu, D. Alexander, M. Chu, V. Miao, P. Brian, R. H. Baltz, *Proc Natl Acad Sci U S A* **2006**, 17462–17467.

[134] U. Linne, M. A. Marahiel, *Biochemistry* **2000**, *39*, 10439–10447.

[135] S. Lautru, G. L. Challis, *Microbiology (Reading, England)* **2004**, *150*, 1629–1636.

[136] a) M.-F. Coëffet-Le Gal, L. Thurston, P. Rich, V. Miao, R. H. Baltz, *Microbiology (Reading, England)* **2006**, *152*, 2993–3001; b) V. Miao, M.-F. Coëffet-Le Gal, K. Nguyen, P. Brian, J. Penn, A. Whiting, J. Steele, D. Kau, S. Martin, R. Ford et al., *Chemistry & biology* **2006**, *13*, 269–276; c) S. Doekel, M.-F. Coëffet-Le Gal, J.-Q. Gu, M. Chu, R. H. Baltz, P. Brian, *Microbiology (Reading, England)* **2008**, *154*, 2872–2880; d) D. C. Alexander, J. Rock, X. He, P. Brian, V. Miao, R. H. Baltz, *Appl Environ Microbiol* **2010**, *76*, 6877–6887; e) K. T. Nguyen, X. He, D. C. Alexander, C. Li, J.-Q. Gu, C. Mascio, A. van Praagh, L. Mortin, M. Chu, J. A. Silverman et al., *Antimicrob Agents Chemother* **2010**, *54*, 1404–1413; f) D. C. Alexander, J. Rock, J.-Q. Gu, C. Mascio, M. Chu, P. Brian, R. H. Baltz, *The Journal of antibiotics* **2011**, *64*, 79–87;

[137] M. J. Calcott, J. G. Owen, I. L. Lamont, D. F. Ackerley, *Appl Environ Microbiol* **2014**, *80*, 5723–5731.

[138] M. J. Calcott, D. F. Ackerley, *BMC microbiology* **2015**, *15*, 162.

[139] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat Chem* **2018**, *10*, 275–281.

[140] J. Zhang, N. Liu, R. A. Cacho, Z. Gong, Z. Liu, W. Qin, C. Tang, Y. Tang, J. Zhou, *Nature chemical biology* **2016**, *12*, 1001–1003.

[141] C. P. Bagowski, W. Bruins, A. J. W. Te Velthuis, *Current genomics* **2010**, *11*, 368–376.

[142] M. G. Chevrette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson, F. Barona-Gómez, *Natural product reports* **2020**, *37*, 566–599.

[143] A. Ureta-Vidal, L. Ettwiller, E. Birney, *Nature reviews. Genetics* **2003**, *4*, 251–262.

[144] S. K. Forslund, M. Kaduk, E. L. L. Sonnhammer, *Methods in molecular biology (Clifton, N.J.)* **2019**, *1910*, 469–504.

[145] M. H. Medema, P. Cimermancic, A. Sali, E. Takano, M. A. Fischbach, *PLoS computational biology* **2014**, *10*, e1004016.

[146] S. Götze, J. Arp, G. Lackner, S. Zhang, H. Kries, M. Klapper, M. García-Altares, K. Willing, M. Günther, P. Stallforth, *Chemical science* **2019**, *10*, 10979–10990.

[147] K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, T. Weber, *Nucleic acids research* **2021**.

[148] S. A. Kautsar, K. Blin, S. Shaw, J. C. Navarro-Muñoz, B. R. Terlouw, J. J. J. van der Hooft, J. A. van Santen, V. Tracanna, H. G. Suarez Duran, V. Pascal Andreu et al., *Nucleic acids research* **2020**, *48*, D454-D458.

[149] J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castaño-Espriu, C. Chang, T. N. Clark et al., *ACS central science* **2019**, *5*, 1824–1833.

[150] M. A. Schorn, S. Verhoeven, L. Ridder, F. Huber, D. D. Acharya, A. A. Aksenov, G. Aleti, J. A. Moghaddam, A. T. Aron, S. Aziz et al., *Nature chemical biology* **2021**, *17*, 363–368.

[151] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2014**, *15*, 1290–1294.

[152] a) R. D. Gietz, R. H. Schiestl, *Nature protocols* **2007**, *2*, 1–4; b) R. D. Gietz, R. H. Schiestl, *Nature protocols* **2007**, *2*, 31–34;

[153] W. Lorenzen, T. Ahrendt, K. A. J. Bozhüyük, H. B. Bode, *Nature chemical biology* **2014**, *10*, 425–427.

[154] E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, C. Dauth, Q. Zhou, M. Kaiser, P. Klemmt, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2015**, *16*, 1115–1119.

[155] S. L. Wenski, D. Kolbert, G. L. C. Grammbitter, H. B. Bode, *J Ind Microbiol Biotechnol* **2019**, *46*, 565–572.

[156] A. O. Brachmann, S. A. Joyce, H. Jenke-Kodama, G. Schwär, D. J. Clarke, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2007**, *8*, 1721–1728.

[157] a) C. Fu, W. P. Donovan, O. Shikapwashya-Hasser, X. Ye, R. H. Cole, *PloS one* **2014**, *9*, e115318; b) R. Simon, U. Priefer, A. Pühler, *Bio/Technology* **1983**, 784–791; c) S. Thoma, M. Schobert, *FEMS microbiology letters* **2009**, *294*, 127–132; d) N. Philippe, J.-P. Alcaraz, E. Coursange, J. Geiselmann, D. Schneider, *Plasmid* **2004**, *51*, 246–255;

[158] J. Xu, S. Lohrke, I. M. Hurlbert, R. E. Hurlbert, *Appl Environ Microbiol* **1989**, 806–812.

[159] A. Stanišić, A. Hüsken, H. Kries, *Chem Sci* **2019**, *10*, 10395–10399.

[160] A. Nayeem, D. Sitkoff, S. Krystek, *Protein science : a publication of the Protein Society* **2006**, *15*, 808–824.

[161] S. Lise, D. Buchan, M. Pontil, D. T. Jones, *PloS one* **2011**, *6*, e16774.

[162] K. A. J. Bozhüyük, J. Watzel, N. Abbood, H. B. Bode, *Angew Chem Int Ed Engl* **2021**, 17531–17538.

[163] K. Lengyel, E. Lang, A. Fodor, E. Szállás, P. Schumann, E. Stackebrandt, *Systematic and applied microbiology* **2005**, *28*, 115–122.

[164] M. S. Venkatarajan, W. Braun, *Journal of Molecular Modeling* **2001**, *7*, 445–453.

[165] C. Kegler, F. I. Nollmann, T. Ahrendt, F. Fleischhacker, E. Bode, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2014**, *15*, 826–828.

[166] K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F. Fleischhacker, Y.-N. Shi, P. Grün, H. B. Bode, *Nat Chem* **2019**, *11*, 653–661.

[167] D. Konz, A. Klens, K. Schörgendorfer, M. A. Marahiel, *Chem Biol* **1997**, 927–937.

[168] N. Abbood, T. D. Vo, J. Watzel, K. A. J. Bozhüyük, H. B. Bode, *bioRxiv* **2021**.

[169] S. G. Lee, F. Lipmann, *Methods Enzymol* **1975**, 585–602.

[170] U. Linne, M. A. Marahiel, *Methods Enzymol* **2004**, 293–315.

[171] a) V. V. Phelan, Y. Du, J. A. McLean, B. O. Bachmann, *Chem Biol* **2009**, *16*, 473–478; b) D. A. Wirtz, K. C. Ludwig, M. Arts, C. E. Marx, S. Krannich, P. Barac,

S. Kehraus, M. Josten, B. Henrichfreise, A. Müller et al., *Angewandte Chemie (International ed. in English)* **2021**, *60*, 13579–13586;

[172] a) A. Ishida, Y. Yamada, T. Kamidate, *Analytical and bioanalytical chemistry* **2008**, *392*, 987–994; b) N. Kadi, G. L. Challis in *Methods in Enzymology*, Elsevier, **2009**;

[173] A. Stanišić, A. Hüsken, P. Stephan, D. L. Niquille, J. Reinstein, H. Kries, *ACS Catal.* **2021**, 8692–8700.

[174] S. Dekimpe, J. Masschelein, *Nat Prod Rep* **2021**.

[175] T. Stachelhaus, C. T. Walsh, *Biochemistry* **2000**, *39*, 5775–5787.

[176] D. B. Stein, U. Linne, M. A. Marahiel, *The FEBS journal* **2005**, *272*, 4506–4520.

[177] a) S. W. Fuchs, F. Grundmann, M. Kurz, M. Kaiser, H. B. Bode, *Chembiochem : a European journal of chemical biology* **2014**, *15*, 512–516; b) H. Donmez Ozkan, H. Cimen, D. Ulug, S. Wenski, S. Yigit Ozer, M. Telli, N. Aydin, H. B. Bode, S. Hazir, *Front. Microbiol.* **2019**, *10*, 2672;

[178] S. L. Wenski, H. Cimen, N. Berghaus, S. W. Fuchs, S. Hazir, H. B. Bode, *Beilstein journal of organic chemistry* **2020**, *16*, 956–965.

[179] Molecular Operating Environment (MOE), Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2021**.

[180] J. Kranz, Master's thesis, Ernst-Abbe-Hochschule, Jena, **2018**.

[181] A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, *Bioinformatics (Oxford, England)* **2019**, *35*, 4453–4455.

[182] D. R. Lide (Ed.) *Handbook of Chemistry and Physics, 72nd Edition*, Boca Raton, FL, **1991**.

[183] T. Goluch, D. Bogdanowicz, K. Giaro, *Methods Ecol Evol* **2020**, *11*, 494–499.

[184] G. Cardona, A. Mir, F. Rosselló, L. Rotger, D. Sánchez, *BMC bioinformatics* **2013**, *14*, 3.

[185] T. Tanimoto, *IBM Corp.* **1958**.

[186] J. A. Wells, *Methods Enzymol* **1991**, 390–411.

[187] R. Wu, A. S. Reger, X. Lu, A. M. Gulick, D. Dunaway-Mariano, *Biochemistry* **2009**, *48*, 4115–4125.

[188] R. Li, R. A. Oliver, C. A. Townsend, *Cell chemical biology* **2017**, *24*, 24–34.

[189] W. Basile, O. Sachenkova, S. Light, A. Elofsson, *PLoS computational biology* **2017**, *13*, e1005375.

[190] D. F. Ackerley, *Cell chemical biology* **2016**, *23*, 535–537.

[191] I. Dimitrov, L. Naneva, I. Doytchinova, I. Bangov, *Bioinformatics (Oxford, England)* **2014**, *30*, 846–851.

[192] B. Tian, X. Wu, C. Chen, W. Qiu, Q. Ma, B. Yu, *Journal of theoretical biology* **2019**, *462*, 329–346.

[193] a) L. Hu, T. Huang, X. Shi, W.-C. Lu, Y.-D. Cai, K.-C. Chou, *PloS one* **2011**, *6*, e14556; b) Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter et al., *Nature* **2012**, *490*, 556–560; c) Z. Yin, T. Deng, L. E. Peterson, R. Yu, J. Lin, D. J. Hamilton, P. R. Reardon, V. Sherman, G. E. Winnier, M. Zhan et al., *Molecular and cellular endocrinology* **2014**, *394*, 80–87; d) S.-B. Zhang, Q.-R. Tang, *Journal of theoretical biology* **2016**, *401*, 30–37; e) Y. Murakami, L. P. Tripathi, P. Prathipati, K. Mizuguchi, *Current opinion in structural biology* **2017**, *44*, 134–142;

6. Supplementary information

Supplementary Tables

Table S1. ESI-MS data of all produced compounds. The compound amino acid compositions marked with an asterisk are attached with one malonate unit and two polyamine units corresponding to fabclavines biosynthesis[155] to obtain the denoted molecular mass.

Compound	amino acid composition	theoretical mass-to-charge ratio (*m/z*) $[M+H]^+$	Molecular formular	Reference
1	cyclo(vLrlL)	595.4290	$C_{29}H_{54}N_8O_5$	[166]
2	vLfV	477.3071	$C_{25}H_{40}N_4O_5$	[166]
3	vLfV-OMe	491.3228	$C_{26}H_{42}N_4O_5$	[166]
4	vL[*p*-N_3-f]V	518.3085	$C_{25}H_{39}N_7O_5$	[166]
5	vL[*p*-N_3-f]V-OMe	532.3242	$C_{26}H_{41}N_7O_5$	[166]
6	vL[*p*-Br-f]V	555.2177	$C_{25}H_{40}BrN_4O_5$	[166]
7	vL[*p*-Br-f]V-OMe	569.2333	$C_{26}H_{42}BrN_4O_5$	[166]
8	cyclo(vLfV)	459.2966	$C_{25}H_{38}N_4O_4$	[166]
9	cyclo(vL[*p*-N_3-f]V)	500.2979	$C_{25}H_{37}N_7O_4$	[166]
10	cyclo(vL[*p*-Br-f]V)	537.2062	$C_{25}H_{38}BrN_4O_4$	[166]
11	llL	358.2700	$C_{18}H_{36}N_3O_4$	[166]
12	LlL	358.2700	$C_{18}H_{36}N_3O_4$	[166]
13	llL-OMe	372.2857	$C_{19}H_{38}N_3O_4$	[166]
14	LlL-OMe	372.2857	$C_{19}H_{38}N_3O_4$	[166]
15	flL	392.2544	$C_{21}H_{34}N_3O_4$	[166]
16	FlL	392.2544	$C_{21}H_{34}N_3O_4$	[166]
17	flL-OMe	406.2700	$C_{22}H_{36}N_3O_4$	[166]
18	FlL-OMe	406.2700	$C_{22}H_{36}N_3O_4$	[166]
19	FlL	392.2544	$C_{21}H_{33}N_3O_4$	[162]
20	flL	392.2544	$C_{21}H_{33}N_3O_4$	[162]
21	llL	358.2700	$C_{21}H_{35}N_3O_4$	[162]
22	LlL	358.2700	$C_{21}H_{35}N_3O_4$	[162]

23	ilL	358.2700	$C_{21}H_{35}N_3O_4$	[162]
24	IlL	358.2700	$C_{21}H_{35}N_3O_4$	[162]
25	YIL	408.2493	$C_{21}H_{33}N_3O_5$	this work
26	WIL	431.5405	$C_{23}H_{34}N_4O_4$	this work
27	MIL	376.2265	$C_{17}H_{33}N_3O_4S$	this work
28	VIL	344.2544	$C_{17}H_{33}N_3O_4$	this work
29	TP*	713.6263	$C_{39}H_{80}N_6O_5$	[178]
30	TV*	715.6420	$C_{39}H_{82}N_6O_5$	[178]
31	TT*	717.62.12	$C_{38}H_{80}N_6O_6$	[178]
32	cyclo(rLflL)	643.4289	$C_{33}H_{54}N_8O_5$	[139]
33	cyclo(rLllL)	609.4446	$C_{30}H_{56}N_8O_5$	[139]
34	cyclo(vLflL)	586.3963	$C_{32}H_{51}N_5O_5$	[139]
35	cyclo(vLllL)	552.4119	$C_{29}H_{53}N_5O_5$	[139]
36	cyclo(sQflL)	589.3344	$C_{29}H_{44}N_6O_7$	[162]
37	l-Thiazolin-CflL	556.3527	$C_{27}H_{49}N_5O_5S$	[162]
38	rvflL	646.8212	$C_{32}H_{54}N_8O_6$	this work
39	rvlflL	760.5079	$C_{38}H_{65}N_9O_7$	this work

Table S2. Raw data refer to Fig. 28 of the multiplexed hydroxamate *in vitro* assay.

	GxpS_A3-T3		GxpS_C_{ASub}3-A3-T3		GxpS_C3-A3-T3		XtpS_C3-GxpS_A3-T3		BacA_C3-GxpS_A3-T3	
	MEAN mg/l	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**
Ala	-	-	-	-	-	-	-	-	-	-
Val	-	-	-	-	-	-	-	-	-	-
Met	-	-	-	-	-	-	-	-	-	-
Leu	0.0757	0.0229	-	-	0.1080	0.0144	0.1080	0.0323	-	-
Ile	0.0757	0.0229	-	-	0.1080	0.0144	0.1080	0.0323	-	-
Pro	-	-	-	-	-	-	-	-	-	-
Trp	0.1050	0.0185	-	-	0.4880	0.0625	0.1360	0.0175	-	-
Phe	1.0300	0.1597	0.0504	0,0015	2.9100	0.3972	1.6500	0.2294	0.0280	0.0043
Lys	-	-	-	-	-	-	-	-	-	-
Arg	-	-	-	-	-	-	-	-	-	-
His	-	-	-	-	-	-	-	-	-	-
Tyr	0.0382	0.0037	-	-	0.1090	0.0171	0.0466	0.0106	-	-
Thr	-	-	-	-	-	-	-	-	-	-
Gln	-	-	-	-	-	-	-	-	-	-
Gly	-	-	-	-	-	-	-	-	-	-
Ser	-	-	-	-	-	-	-	-	-	-
Cys	-	-	-	-	-	-	-	-	-	-
Asn	-	-	-	-	-	-	-	-	-	-
Glu	-	-	-	-	-	-	-	-	-	-
Asp	-	-	-	-	-	-	-	-	-	-

Table S3. Raw data refer to Fig. 29 of the GxpS with deleted first two modules (A1 to C3) and varying preceding C domains *in vivo* assay.

	GxpS_A3--TE		GxpS_C3--TE		GxpS_C_{ASub}3--TE		XtpS_C3-GxpS_A3--TE		KolS_C5-GxpS_A3--TE		$AmbS_{indica}$_C5-GxpS_A3--TE		$AmbS_{mira}$_C5-GxpS_A3--TE		BicA_C3-GxpSA3--TE	
	MEAN mg/l	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**	**MEAN mg/l**	**SD mg/l**
AIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MIL	0.4510	0.0598	0.0711	0.0105	0.0304	0.0115	0.0412	0.0094	0.0222	0.0113	-	-	0.0740	0.0066	0.0933	0.0077
I/LIL	2.7240	0.3624	1.5264	0.1965	1.2375	0.1136	0.9644	0.1902	0.5896	0.2422	0.3275	0.0102	2.8215	0.0902	3.5202	0.2752
i/IIL	0.1758	0.0247	0.0792	0.0161	0.0824	0.0057	0.0681	0.0120	0.0398	0.0296	-	-	0.1868	0.0175	0.2273	0.0234
PIL	0.0038	0.0085	0.4944	0.0619	0.0919	0.0418	0.4964	0.1275	0.0107	0.0815	0.0467	0.0047	0.1606	0.0382	0.0683	0.0244
WIL	0.1516	0.0126	0.0805	0.0078	0.0772	0.0371	0.0663	0.0164	0.0753	0.0355	0.0255	0.0013	0.2334	0.0010	0.3548	0.0199
FIL	9.6263	0.5551	14.4737	1.1411	5.7133	1.9941	10.5606	1.5582	3.6512	2.8903	1.8598	0.0457	18.3480	0.9115	20.2288	1.5753
fIL	5.3513	0.5291	6.9523	0.9291	2.8380	0.4718	6.0191	1.2855	1.8271	1.6909	0.9494	0.0735	9.5853	0.3810	9.9711	0.6059
KIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YIL	0.3356	0.1154	0.1749	0.0217	0.1289	0.0265	0.1428	0.0230	0.1197	0.0818	0.0543	0.0039	0.5044	0.0175	0.6745	0.0432
TIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
QIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table S4. HeatMap of the Clustal Omega alignment of C domains investigated in this work. Higher similarity is represented by darker plot colours.

Identity [%]	GxpS_C3	XtpS_C3	BacA_C3	KolS_C5	AmbS*X. indi.*_C5	AmbS*X. mira.*_C5	BicA_C3	*Xsze* FclJ_C6	*Xbud* FclJ_C6	*Xhom* FclJ_C6
GxpS_C3		86.33	41.38	67.29	64.43	63.87	62.10	27.52	28.15	28.22
XtpS_C3	86.33		41.38	68.86	64.35	65.05	63.04	28.59	28.33	28.51
BacA_C3	41.38	41.38		41.80	39.96	42.34	42.03	27.69	28.56	29.42
KolS_C5	67.29	68.86	41.80		69.95	71.14	70.72	27.91	28.56	28.45
AmbS*X. indi.*_C5	64.43	64.35	39.96	69.95		67.26	65.10	30.66	30.79	30.45
AmbS*X. mira.*_C5	63.87	65.05	42.34	71.14	67.26		70.40	28.03	29.07	27.82
BicA_C3	62.10	63.04	42.03	70.72	65.10	70.40		27.69	28.56	29.42
***Xsze* FclJ_C6**	27.52	28.59	27.69	27.91	30.66	28.03	27.69		74.59	73.91
***Xbud* FclJ_C6**	28.15	28.33	28.56	28.56	30.79	29.07	28.56	74.59		80.54
***Xhom* FclJ_C6**	28.22	28.51	29.42	28.45	30.45	27.82	29.42	73.91	80.54	

Table S5. The selectivity-conferring code[53] of the FclJ A6 domains and antiSMASH[147] substrate prediction.

FclJ A6	AA1	AA2	AA3	AA4	AA5	AA6	AA7	AA8	AA9	AA10	antiSMASH prediction
Xbud	D	A	W	F	I	G	G	H	-	K	Leu
Xhom	D	A	L	F	I	G	G	Y	-(E)	K	Leu
Xsze	D	A	L	F	I	G	G	Y	-(E)	K	Leu

Table S6. RMSD to mean of the MOE homology models.

model	Template			
	2VSQ	4ZXH	5T3D	
Xsze FclJ_C6-*Xhom.* FclJ_A6	1.26	1.75	1.47	RMSD [Å]
Xsze FclJ_C6-*Xbud.* FclJ_A6	1.27	1.68	1.52	
Xsze. FclJ_C6-A6	1.49	1.86	1.77	
Xhom. FclJ_C6-A6	1.29	1.67	1.50	
Xbud. FclJ_C6-A6	1.30	1.70	1.47	
BacA_C3-GxpS_A3	1.14	1.69	1.57	
BacA_C3-A3	1.31	1.77	1.52	
XtpS_C3-GxpS_A3	1.11	1.67	1.39	
XtpS_C3-A3	1.35	1.68	1.48	
GxpS_C3-A3	1.13	1.60	1.28	

Table S7. Clustal Omega alignment of the C-A didomains used for homology modeling and in HSPred analysis including the AB3403 (PDB-ID: 4ZXH)[60], EntF (PDB-ID: 5T3D)[60] and SrfA-C (PDB-ID: 2SVQ)[58] templates. The interacting Chains I-XVI are highlighted in grey and entitled below. Colouring of the consensus sequence is according to the C domain (green), C-A linker (blue), A_{Core} (red), and A_{Sub} (orange).

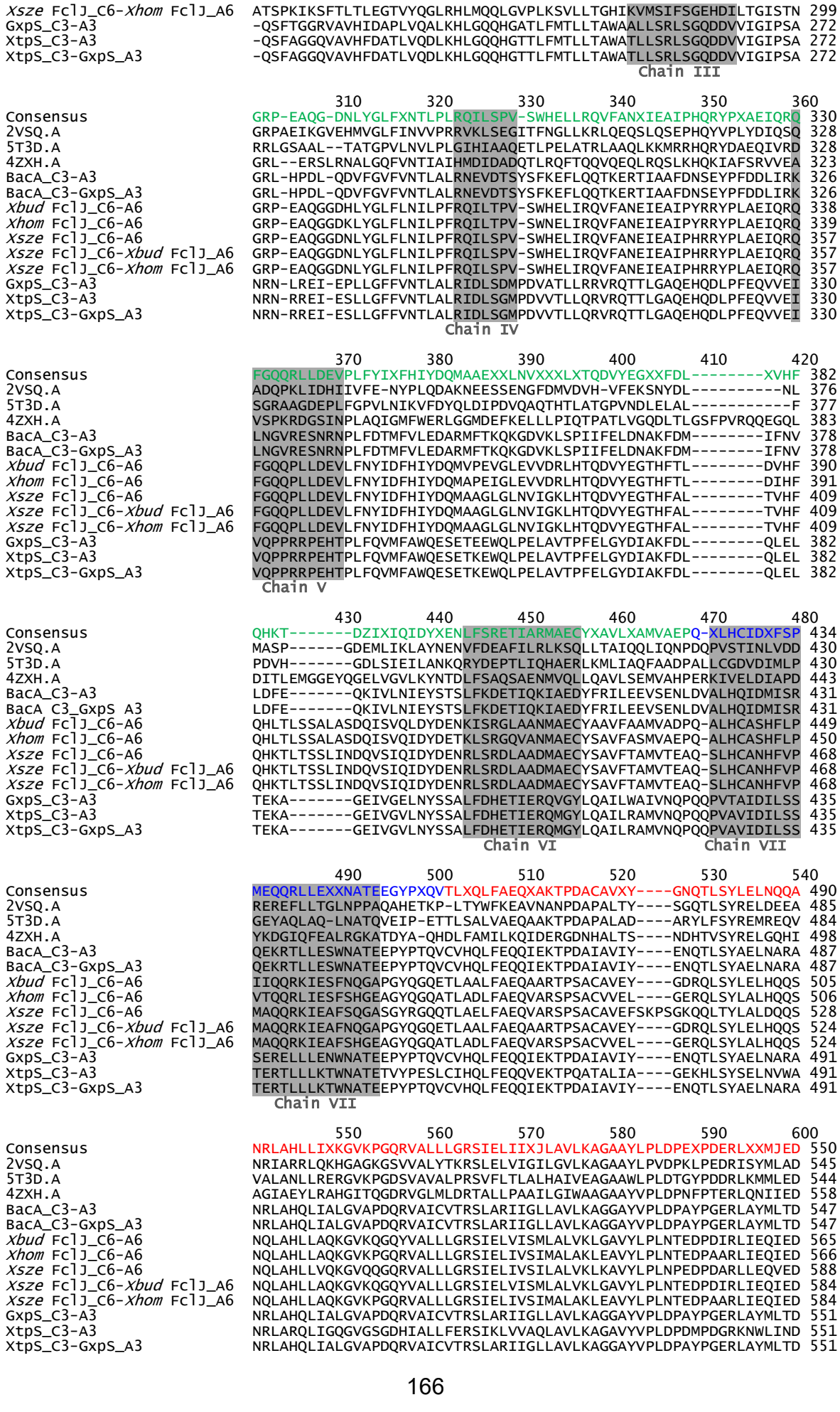
```
                                        10        20        30        40        50        60
Consensus                               LSRMQQGMXYHMNLXPDXNVYHXX--TXLRVNGRXDEXAFREALAQTVARHDVLRTAFDM   58
2VSQ.A                                  LSPMQEGMLFHAILNPGQSFYLEQ--ITMKVKGSLNIKCLEESMNVIMDRYDVFRTVFIH   58
5T3D.A                                  LVAAQPGIWMAEKLSELPSAWSVA--HYVELTGEVDSPLLARAVVAGLAQADTLRMRFTE   58
4ZXH.A                                  ISSEQLGIWYIQRLEPTCSAYNMVVAFDVKVNQSLGNKPI-EILEAVMHDYPLLRVSMPA   59
BacA_C3-A3                              TSPAQQRMYMLSMLENERGAYHIP--MALLVEGRINAMQLENALKTFLQRHEILRTGFEI   58
BacA_C3-GxpS_A3                         TSPAQQRMYMLSMLENERGAYHIP--MALLVEGRINAMQLENALKTFLQRHEILRTGFEI   58
Xbud FclJ_C6-A6                         LSRMQAGMFYHMNMSPDANVYHCTGTSHLRVNGRFDEKAFIEATAQTVAAHDVLRTAFDM   60
Xhom FclJ_C6-A6                         FSRMQEGMFYHMDLSPESNVYHCTGTSHLRINGPFDERAFREATAQTVAAHDVLRTAFDM   60
Xsze FclJ_C6-A6                         LSRMQAGMFYHMNLSPDANVYHCTGSTHIFINGQFDETAFREATAQTVAAHDVLRTAFDM   60
Xsze FclJ_C6-Xbud FclJ_A6               LSRMQAGMFYHMNLSPDANVYHCTGSTHIFINGQFDETAFREATAQTVAAHDVLRTAFDM   60
Xsze FclJ_C6-Xhom FclJ_A6               LSRMQAGMFYHMNLSPDANVYHCTGSTHIFINGQFDETAFREATAQTVAAHDVLRTAFDM   60
GxpS_C3-A3                              LSFGQQRLWFLAQFEGVSDTYHIP--TALRLRGRLNFSAWQQALNRLFARHEALRSVFIT   58
XtpS_C3-A3                              LSFGQQRLWFLAQFEGISDTYHIP--TALRLRGRLDLSAWQQALNCLFARHEALRSVFIT   58
XtpS_C3-GxpS_A3                         LSFGQQRLWFLAQFEGISDTYHIP--TALRLRGRLDLSAWQQALNCLFARHEALRSVFIT   58

                                        70        80        90        100       110       120
Consensus                               EN-SEPLQLVLEFAEFPJVVEDL-RHLSDEEQDEQXKALLEXERVNPFDLLRPPLLRFFJ  116
2VSQ.A                                  EKVKRPVQVVLKKRQFHIEEIDL-THLTGSEQTAKINEYKEQDKIRGFDLTRDIPMRAAI  117
5T3D.A                                  DN-GEVWQWVDDALTFELPEIIDLRTNIDPHGTAQALMQADLQQDLRVDSGKP-LVFHQL  116
4ZXH.A                                  ND--QGIEQLIWDRVYPNIIFSDARHIEAS----DLTQLVEQDTKQPFDLTQPPLWRIHC  113
BacA_C3-A3                              QN-NELIQKIYENVDFRLEYECLDASITDQHALMEITSRYCKESIKPFDLSRPPLMRAKL  117
BacA_C3-GxpS_A3                         QN-NELIQKIYENVDFRLEYECLDASITDQHALMEITSRYCKESIKPFDLSRPPLMRAKL  117
Xbud FclJ_C6-A6                         ENYSEPLQLVYQFAELPIVVEDL-RHLSEEEQDERAKALLEAERVNPFDLLNPTLLRFFI  119
Xhom FclJ_C6-A6                         ENYSEPLQLVHEFAELPIVVEDL-RHLSVEEQDDQAMALLEAERVNPFDLLKPTLLRFFI  119
Xsze FclJ_C6-A6                         ESYSEPLQLVLEFAELPIVVEDL-RHLSDEEQDERVKALLEAERVNPFDLLNPTLLRFFI  119
Xsze FclJ_C6-Xbud FclJ_A6               ESYSEPLQLVLEFAELPIVVEDL-RHLSDEEQDERVKALLEAERVNPFDLLNPTLLRFFI  119
Xsze FclJ_C6-Xhom FclJ_A6               ESYSEPLQLVLEFAELPIVVEDL-RHLSDEEQDERVKALLEAERVNPFDLLNPTLLRFFI  119
GxpS_C3-A3                              VD-GQPQVELLP-AEFGLPVKKYDLRKASN-VDKQFKRLCVQEAKTPFDLARGPLIRCAL  115
XtpS_C3-A3                              VE-GQPQVELLP-AEFALPMQKYDLRKASD-VDEQLAHLCAQEAKTPFDLARGPLIRCAL  115
XtpS_C3-GxpS_A3                         VE-GQPQVELLP-AEFALPMQKYDLRKASD-VDEQLAHLCAQEAKTPFDLARGPLIRCAL  115

                                        130       140       150       160       170       180
Consensus                               XLRXDXSFQFTMTEHHIIXDGWSYGILIVEVFXRYATLTGQXPWPXPXK-SLEYRDYIAW  175
2VSQ.A                                  FKKAEESFEWVWSYHHIILDGWCFGIVVQDLFKVYNALREQKPYSLPPV--KPYKDYIKW  175
5T3D.A                                  IQVADNRWYWYQRYHHLLVDGFSFPAITRQIANIYCTWLRGEPTPASPF--TPFADVVEE  174
4ZXH.A                                  YECGQNHYVIAFVIHHALMDFWSIGLLLRDVSKRFGLVA-----ESDAVNGIEFVQYADK  168
BacA_C3-A3                              IKIDDIRHILVINFHHIISDGVSQGILMNEILELYSNV------PLPEV-NVQYKDYVEW  170
BacA_C3-GxpS_A3                         IKIDDIRHILVINFHHIISDGVSQGILMNEILELYSNV------PLPEV-NVQYKDYVEW  170
Xbud FclJ_C6-A6                         QLRTDNSFQFTMTECHPVFDGWSYNTMIVEVFNRYATLTGQASWQAPDKKSLEYRDFIAL  179
Xhom FclJ_C6-A6                         QLRTDNSFQFTMTECHPVFDGWSYNTMIVEVFNRYATLTGQVNWQAPDKKSLEYRDFIAL  179
Xsze FclJ_C6-A6                         YLRTDHSFQFTMTECHPIFDGWSYNTMIVEVFSRYATLTGLDNWQAPDKKSLEYRDFIAL  179
Xsze FclJ_C6-Xbud FclJ_A6               YLRTDHSFQFTMTECHPIFDGWSYNTMIVEVFSRYATLTGLDNWQAPDKKSLEYRDFIAL  179
Xsze FclJ_C6-Xhom FclJ_A6               YLRTDHSFQFTMTECHPIFDGWSYNTMIVEVFSRYATLTGLDNWQAPDKKSLEYRDFIAL  179
GxpS_C3-A3                              IQLADEEHIFLLTQHHIISDAWSLKILESELVALYAAGLNQQPDPLLPL-TIQYPDYAAW  174
XtpS_C3-A3                              VRLAEEEHLFLLTQHHIISDGWSLGILESELTALYVACLNQQPDPLPPL-TIQYPDYAAW  174
XtpS_C3-GxpS_A3                         VRLAEEEHLFLLTQHHIISDGWSLGILESELTALYVACLNQQPDPLPPL-TIQYPDYAAW  174

                                        190       200       210       220       230       240
Consensus                               EQAAIADEKXKMXWAYXLD--TJXXLPRKLXXPVDRDRPSX----------------------  214
2VSQ.A                                  LEKQDKQASLRYWREY--------LEGFEGQTTFAEQRKKQK----------------------  209
5T3D.A                                  YQQYRESEAWQRDAAFWAEQRRQLPPPASLSPAPLPGR-------------------------  212
4ZXH.A                                  QQSSVIDDTDESLIFWKN---ALKHAPHVHSIPLDYPRPAV----------------------  206
BacA_C3-A3                              NHTFNQSAAMKKQEAYWLD--VYRDIPSKLDFPYDYKRHHI----------------------  209
BacA_C3-GxpS_A3                         NHTFNQSAAMKKQEAYWLD--VYRDIPSKLDFPYDYKRHHI----------------------  209
Xbud FclJ_C6-A6                         EQAAIADEKQKDYWAQKLDDCTILELPRKQHTDGINDVVASA---------------------  221
Xhom FclJ_C6-A6                         EQAAIADEQQKAYWTQKLDDCTILELPRKQHTDVIDDTGVPA---------------------  221
Xsze FclJ_C6-A6                         EKAAIADEKNKMYWAQKLDDCTIMELPRKLNAGVSGDAASHAGEGTVDGNNGDDNDSGTL  239
Xsze FclJ_C6-Xbud FclJ_A6               EKAAIADEKNKMYWAQKLDDCTIMELPRKLNAGVSGDAASHAGEGTVDGNNGDDNDSGTL  239
Xsze FclJ_C6-Xhom FclJ_A6               EKAAIADEKNKMYWAQKLDDCTIMELPRKLNAGVSGDAASHAGEGTVDGNNGDDNDSGTL  239
GxpS_C3-A3                              QRQVFSGEGVQAQSDYWRR--TLADAPVLLELPIDRPRPSQ----------------------  213
XtpS_C3-A3                              QRQVFSDEGVQVQSDYWRR--TLADAPVLLELPLDRPRPSQ----------------------  213
XtpS_C3-GxpS_A3                         QRQVFSDEGVQVQSDYWRR--TLADAPVLLELPLDRPRPSQ----------------------  213
                                                                  Chain I                 Chain II

                                        250       260       270       280       290       300
Consensus                               -XSXKGKSXXLXLEXTXYQGLRHLMQQLGVTLKSVLLTGHIKLLSXXSGEHDIVTGIPAA  273
2VSQ.A                                  -DGYEPKELLFSPSEAETKAFTELAKSQHTTLSTALQAVWSVLISRYQQSGDLAFGTVVS  268
5T3D.A                                  --SASADILRLKLEFTDGEFRQLATQLSGVQRTDLALALAALWLGRLCNRMDYAAGFIFM  270
4ZXH.A                                  -QQHKGSSLVFRVSESVSSGLVNLAKDYEITLFGLVLSGFYVLLHKLSNENNLVIATPVA  265
BacA_C3-A3                              -DTFEGSSVFLEMERELSDHIRKLAKHNGTTLYTVMLSAYYVLLNKYTNQTDIVVGTAAA  268
BacA_C3-GxpS_A3                         -DTFEGSSVFLEMERELSDHIRKLAKHNGTTLYTVMLSAYYVLLNKYTNQTDIVVGTAAA  268
Xbud FclJ_C6-A6                         -TSPKLKSFSLILDSTVYQGLRHLMQQLGVPLKSILLTGHIKVMSIFSGEHDILTGITAN  280
Xhom FclJ_C6-A6                         ANRPKLKSFSLILDTHIYQGLRHLMQQLSVPMKSLLLTGHIKVMSIFSGENDILTGITAN  281
Xsze FclJ_C6-A6                         ATSPKIKSFTLTLEGTVYQGLRHLMQQLGVPLKSVLLTGHIKVMSIFSGEHDILTGISTN  299
Xsze FclJ_C6-Xbud FclJ_A6               ATSPKIKSFTLTLEGTVYQGLRHLMQQLGVPLKSVLLTGHIKVMSIFSGEHDILTGISTN  299
```


Xsze FclJ_C6-Xhom FclJ_A6 ATSPKIKSFTLTLEGTVYQGLRHLMQQLGVPLKSVLLTGHIKVMSIFSGEHDILTGISTN 299
GxpS_C3-A3 -QSFTGGRVAVHIDAPLVQALKHLGQQHGATLFMTLLTAWAALLSRLSGQDDVVIGIPSA 272
XtpS_C3-A3 -QSFAGGQVAVHFDATLVQDLKHLGQQHGTTLFMTLLTAWATLLSRLSGQDDVVIGIPSA 272
XtpS_C3-GxpS_A3 -QSFAGGQVAVHFDATLVQDLKHLGQQHGTTLFMTLLTAWATLLSRLSGQDDVVIGIPSA 272
Chain III
310 320 330 340 350 360
Consensus GRP-EAQG-DNLYGLFXNTLPLRQILSPV-SWHELLRQVFANXIEAIPHQRYPXAEIQRQ 330
2VSQ.A GRPAEIKGVEHMVGLFINVVPRRVKLSEGITFNGLLKRLQEQSLQSEPHQYVPLYDIQSQ 328
5T3D.A RRLGSAAL--TATGPVLNVLPLGIHIAAQETLPELATRLAAQLKKMRRHQRYDAEQIVRD 328
4ZXH.A GRL--ERSLRNALGQFVNTIAIHMDIDADQTLRQFTQQVQEQLRQSLKHQKIAFSRVVEA 323
BacA_C3-A3 GRL-HPDL-QDVFGVFVNTLALRNEVDTSYSFKEFLQQTKERTIAAFDNSEYPFDDLIRK 326
BacA_C3-GxpS_A3 GRL-HPDL-QDVFGVFVNTLALRNEVDTSYSFKEFLQQTKERTIAAFDNSEYPFDDLIRK 326
Xbud FclJ_C6-A6 GRP-EAQGGDHLYGLFLNILPFRQILTPV-SWHELIRQVFANEIEAIPYRRYPLAEIQRQ 338
Xhom FclJ_C6-A6 GRP-EAQGGDKLYGLFLNILPFRQILTPV-SWNELIRQVFANEIEAIPYRRYPLAEIQRQ 339
Xsze FclJ_C6-A6 GRP-EAQGGDNLYGLFLNILPFRQILSPV-SWHELIRQVFANEIEAIPHRRYPLAEIQRQ 357
Xsze FclJ_C6-Xbud FclJ_A6 GRP-EAQGGDNLYGLFLNILPFRQILSPV-SWHELIRQVFANEIEAIPHRRYPLAEIQRQ 357
Xsze FclJ_C6-Xhom FclJ_A6 GRP-EAQGGDNLYGLFLNILPFRQILSPV-SWHELIRQVFANEIEAIPHRRYPLAEIQRQ 357
GxpS_C3-A3 NRN-LREI-EPLLGFFVNTLALRIDLSDMPDVATLLRRVRQTTLGAQEHQDLPFEQVVEI 330
XtpS_C3-A3 NRN-RREI-ESLLGFFVNTLALRIDLSGMPDVVTLLQRVRQTTLGAQEHQDLPFEQVVEI 330
XtpS_C3-GxpS_A3 NRN-RREI-ESLLGFFVNTLALRIDLSGMPDVVTLLQRVRQTTLGAQEHQDLPFEQVVEI 330
Chain IV
370 380 390 400 410 420
Consensus FGQQRLLDEVPLFYIXFHIYDQMAAEXXLNVXXXLXTQDVYEGXXFDL--------XVHF 382
2VSQ.A ADQPKLIDHIIVFE-NYPLQDAKNEESSENGFDMVDVH-VFEKSNYDL---------NL 376
5T3D.A SGRAAGDEPLFGPVLNIKVFDYQLDIPDVQAQTHTLATGPVNDLELAL----------F 377
4ZXH.A VSPKRDGSINPLAQIGMFWERLGGMDEFKELLLPIQTPATLVGQDLTLGSFPVRQQEGQL 383
BacA_C3-A3 LNGVRESNRNPLFDTMFVLEDARMFTKQKGDVKLSPIIFELDNAKFDM--------IFNV 378
BacA_C3-GxpS_A3 LNGVRESNRNPLFDTMFVLEDARMFTKQKGDVKLSPIIFELDNAKFDM--------IFNV 378
Xbud FclJ_C6-A6 FGQQPLLDEVLFNYIDFHIYDQMVPEVGLEVVDRLHTQDVYEGTHFTL--------DVHF 390
Xhom FclJ_C6-A6 FGQQPLLDEVLFNYIDFHIYDQMAPEIGLEVVDRLHTQDVYEGTHFTL--------DIHF 391
Xsze FclJ_C6-A6 FGQQPLLDEVLFNYIDFHIYDQMAAGLGLNVIGKLHTQDVYEGTHFAL--------TVHF 409
Xsze FclJ_C6-Xbud FclJ_A6 FGQQPLLDEVLFNYIDFHIYDQMAAGLGLNVIGKLHTQDVYEGTHFAL--------TVHF 409
Xsze FclJ_C6-Xhom FclJ_A6 FGQQPLLDEVLFNYIDFHIYDQMAAGLGLNVIGKLHTQDVYEGTHFAL--------TVHF 409
GxpS_C3-A3 VQPPRRPEHTPLFQVMFAWQESETEEWQLPELAVTPFELGYDIAKFDL--------QLEL 382
XtpS_C3-A3 VQPPRRPEHTPLFQVMFAWQESETKEWQLPELAVTPFELGYDIAKFDL--------QLEL 382
XtpS_C3-GxpS_A3 VQPPRRPEHTPLFQVMFAWQESETKEWQLPELAVTPFELGYDIAKFDL--------QLEL 382
Chain V
430 440 450 460 470 480
Consensus QHKT-------DZIXIQIDYXENLFSRETIARMAECYXAVLXAMVAEPQ-XLHCIDXFSP 434
2VSQ.A MASP------GDEMLIKLAYNENVFDEAFILRLKSQLLTAIQQLIQNPDQPVSTINLVDD 430
5T3D.A PDVH-------GDLSIEILANKQRYDEPTLIQHAERLKMLIAQFAADPALLCGDVDIMLP 430
4ZXH.A DITLEMGGEYQGELVGVLKYNTDLFSAQSAENMVQLLQAVLSEMVAHPERKIVELDIAPD 443
BacA_C3-A3 LDFE-------QKIVLNIEYSTSLFKDETIQKIAEDYFRILEEVSENLDVALHQIDMISR 431
BacA C3_GxpS A3 LDFE-------QKIVLNIEYSTSLFKDETIQKIAEDYFRILEEVSENLDVALHQIDMISR 431
Xbud FclJ_C6-A6 QHLTLSSALASDQISVQLDYDENKISRGLAANMAECYAAVFAAMVADPQ-ALHCASHFLP 449
Xhom FclJ_C6-A6 QHLTLSSALASDQISVQIDYDETKLSRGQVANMAECYSAVFASMVAEPQ-ALHCASHFLP 450
Xsze FclJ_C6-A6 QHKTLTSSLINDQVSIQIDYDENRLSRDLAADMAECYSAVFTAMVTEAQ-SLHCANHFVP 468
Xsze FclJ_C6-Xbud FclJ_A6 QHKTLTSSLINDQVSIQIDYDENRLSRDLAADMAECYSAVFTAMVTEAQ-SLHCANHFVP 468
Xsze FclJ_C6-Xhom FclJ_A6 QHKTLTSSLINDQVSIQIDYDENRLSRDLAADMAECYSAVFTAMVTEAQ-SLHCANHFVP 468
GxpS_C3-A3 TEKA-------GEIVGELNYSSALFDHETIERQVGYLQAILWAIVNQPQQPVTAIDILSS 435
XtpS_C3-A3 TEKA-------GEIVGVLNYSSALFDHETIERQMGYLQAILRAMVNQPQQPVAVIDILSS 435
XtpS_C3-GxpS_A3 TEKA-------GEIVGVLNYSSALFDHETIERQMGYLQAILRAMVNQPQQPVAVIDILSS 435
Chain VI
Chain VII
490 500 510 520 530 540
Consensus MEQQRLLEXXNATEEGYPXQVTLXQLFAEQXAKTPDACAVXY----GNQTLSYLELNQQA 490
2VSQ.A REREFLLTGLNPPAQAHETKP-LTYWFKEAVNANPDAPALTY----SGQTLSYRELDEEA 485
5T3D.A GEYAQLAQ-LNATQVEIP-ETTLSALVAEQAAKTPDAPALAD----ARYLFSYREMREQV 484
4ZXH.A YKDGIQFEALRGKATDYA-QHDLFAMILKQIDERGDNHALTS----NDHTVSYRELGQHI 498
BacA_C3-A3 QEKRTLLESWNATEEPYPTQVCVHQLFEQQIEKTPDAIAVIY----ENQTLSYAELNARA 487
BacA_C3-GxpS_A3 QEKRTLLESWNATEEPYPTQVCVHQLFEQQIEKTPDAIAVIY----ENQTLSYAELNARA 487
Xbud FclJ_C6-A6 IIQQRKIESFNQGAPGYQGQETLAALFAEQAARTPSACAVEY----GDRQLSYLELHQQS 505
Xhom FclJ_C6-A6 VTQQRLIESFSHGEAGYQGQATLADLFAEQVARSPSACVVEL----GERQLSYLALHQQS 506
Xsze FclJ_C6-A6 MAQQRKIEAFSQGASGYRGQQTLAELFAEQVARSPSACAVEFSKPSGKQQLTYLALDQQS 528
Xsze FclJ_C6-Xbud FclJ_A6 MAQQRKIEAFNQGAPGYQGQETLAALFAEQAARTPSACAVEY----GDRQLSYLELHQQS 524
Xsze FclJ_C6-Xhom FclJ_A6 MAQQRKIEAFSHGEAGYQGQATLADLFAEQVARSPSACVVEL----GERQLSYLALHQQS 524
GxpS_C3-A3 SERELLLENWNATEEPYPTQVCVHQLFEQQIEKTPDAIAVIY----ENQTLSYAELNARA 491
XtpS_C3-A3 TERTLLLKTWNATETVYPESLCIHQLFEQQVEKTPQATALIA----GEKHLSYSELNVWA 491
XtpS_C3-GxpS_A3 TERTLLLKTWNATEEPYPTQVCVHQLFEQQIEKTPDAIAVIY----ENQTLSYAELNARA 491
Chain VII
550 560 570 580 590 600
Consensus NRLAHLLIXKGVKPGQRVALLLGRSIELIIXJLAVLKAGAAYLPLDPEXPDERLXXMJED 550
2VSQ.A NRIARRLQKHGAGKGSVVALYTKRSLELVIGILGVLKAGAAYLPVDPKLPEDRISYMLAD 545
5T3D.A VALANLLRERGVKPGDSVAVALPRSVFLTLALHAIVEAGAAWLPLDTGYPDDRLKMMLED 544
4ZXH.A AGIAEYLRAHGITQGDRVGLMLDRTALLPAAILGIWAAGAAYVPLDPNFPTERLQNIIED 558
BacA_C3-A3 NRLAHQLIALGVAPDQRVAICVTRSLARIIGLLAVLKAGGAYVPLDPAYPGERLAYMLTD 547
BacA_C3-GxpS_A3 NRLAHQLIALGVAPDQRVAICVTRSLARIIGLLAVLKAGGAYVPLDPAYPGERLAYMLTD 547
Xbud FclJ_C6-A6 NQLAHLLAQKGVKQGQYVALLLGRSIELVISMLALVKLGAVYLPLNTEDPDIRLIEQIED 565
Xhom FclJ_C6-A6 NQLAHLLAQKGVKPGQRVALLLGRSIELIVSIMALAKLEAVYLPLNTEDPAARLIEQIED 566
Xsze FclJ_C6-A6 NQLAHLLVQKGVQQGQRVALLLGRSIELIVSILALVKLKAVYLPLNPEDPDARLLEQVED 588
Xsze FclJ_C6-Xbud FclJ_A6 NQLAHLLAQKGVKQGQYVALLLGRSIELVISMLALVKLGAVYLPLNTEDPDIRLIEQIED 584
Xsze FclJ_C6-Xhom FclJ_A6 NQLAHLLAQKGVKPGQRVALLLGRSIELIVSIMALAKLEAVYLPLNTEDPAARLIEQIED 584
GxpS_C3-A3 NRLAHQLIALGVAPDQRVAICVTRSLARIIGLLAVLKAGGAYVPLDPAYPGERLAYMLTD 551
XtpS_C3-A3 NRLARQLIGQGVGSGDHIALLFERSIKLVVAQLAVLKAGAVYVPLDPDMPDGRKNWLIND 551
XtpS_C3-GxpS_A3 NRLAHQLIALGVAPDQRVAICVTRSLARIIGLLAVLKAGGAYVPLDPAYPGERLAYMLTD 551

```
                                610       620       630       640       650       660
Consensus                       AQPVLLXTDXRXTAALXEXILATLTXL-----DQSTLXEXPVXDLQXSG------TPNXP 599
2VSQ.A                          SAAACLLTHQEMKEQAAELPYTGTTLFI----DDQTRFEEQASDPATAI------DPNDP 595
5T3D.A                          ARPSLLITTDDQLPRFSDVPNLTSLCY------NAPLTPQGSAPLQLSQ-------PHHT 591
4ZXH.A                          AEPKVILTQTELMDGLNVSVPRLDI-------NQAGVVALEQVRETLAF--------GDI 603
BacA_C3-A3                      ATPVILMADNVGRAALSEDILATLTVL-----DPNTLLEQPDHNPQVSG-----LTPQHL 597
BacA_C3-GxpS_A3                 ATPVILMADNVGRAALSEDILATLTVL-----DPNTLLEQPDHNPQVSG-----LTPQHL 597
Xbud FclJ_C6-A6                 TQCDLLIIDRRITDRLLPSISPHLPPILWLDEEQSALAQMPVTDLPDNLSHHAENQANLP 625
Xhom FclJ_C6-A6                 TQCHLLVSDRRITNSLLSLTEMSLPPVLWLDAEQSALAQMPVTDLTLSTNNQT---TNLP 623
Xsze FclJ_C6-A6                 AQCEFFITDQRVIDSSLPMKAANLSAVIWLDAEQSSLSRMPVSPLPDKGSHPDENQANLP 648
Xsze FclJ_C6-Xbud FclJ_A6       TQCDLLIIDRRITDRLLPSISPHLPPILWLDEEQSALAQMPVTDLPDNLSHHAENQANLP 644
Xsze FclJ_C6-Xhom FclJ_A6       TQCHLLVSDRRITNSLLSLTEMSLPPVLWLDAEQSALAQMPVTDLTLSTNNQT---TNLP 641
GxpS_C3-A3                      ATPVILMADNVGRAALSEDILATLTVL-----DPNTLLEQPDHNPQVSG-----LTPQHL 601
XtpS_C3-A3                      CAAKLLLTD-IQTAIPTDLIVPLLRLS-----DESEAVSEQASGQKNSTDLDLPRTSTEL 605
XtpS_C3-GxpS_A3                 ATPVILMADNVGRAALSEDILATLTVL-----DPNTLLEQPDHNPQVSG-----LTPQHL 601

                                670       680       690       700       710       720
Consensus                       AYVMYTSGSTGKPKGVLIGHXXIIXL-XKDVSYXDFCTFGRXLQLASVSFDASTWEIWGP 658
2VSQ.A                          AYIMYTSGTTGKPKGNITTHANIQGL-VKHVDYMAFSDQDTFLSVSNYAFDAFTFDFYAS 654
5T3D.A                          AYIIFTSGSTGRPKGVMVGQTAIVNRLLWMQNHYPLTGEDVVAQKTPCSFDVSVWEFFWP 651
4ZXH.A                          AYVMYTSGSTGKPKGVRIGHPSIINFLLSMNDRLQVTTETQLLAITTYAFDISILELLIP 663
BacA_C3-A3                      AYVIYTSGSTGRPKGVMIEHRSVVNLTLTQITQFDVCATSRMLQFASFGFDASVWEIMMA 657
BacA_C3-GxpS_A3                 AYVIYTSGSTGRPKGVMIEHRSVVNLTLTQITQFDVCATSRMLQFASFGFDASVWEIMMA 657
Xbud FclJ_C6-A6                 AYVMYTSGSTGKPKGALIGQKGIIRL-VKDVSYIDFQTFGRCLQLASVSFDASTWEIWGP 684
Xhom FclJ_C6-A6                 AYIMYTSGSTGKPKGALIGQKGIIRL-VKDVSYIDFHTFGRFLQLAAVSFDASTLEIWGP 682
Xsze FclJ_C6-A6                 AYVMYTSGSTGKPKGALIGQKGIIRL-VKDVSYADFRVFGRFLQLASVNFDASTLEIWGP 707
Xsze FclJ_C6-Xbud FclJ_A6       AYVMYTSGSTGKPKGALIGQKGIIRL-VKDVSYIDFQTFGRCLQLASVSFDASTWEIWGP 703
Xsze FclJ_C6-Xhom FclJ_A6       AYIMYTSGSTGKPKGALIGQKGIIRL-VKDVSYIDFHTFGRFLQLAAVSFDASTLEIWGP 700
GxpS_C3-A3                      AYVIYTSGSTGRPKGVMIEHRSVVNLTLTQITQFDVCATSRMLQFASFGFDASVWEIMMA 661
XtpS_C3-A3                      AYIMYTSGSTGTPKGVLVPHRAVARLVINN-GYAAIEPDDRVAFTANPAFDASTFDVWAP 664
XtpS_C3-GxpS_A3                 AYVIYTSGSTGRPKGVMIEHRSVVNLTLTQITQFDVCATSRMLQFASFGFDASVWEIMMA 661

                                730       740       750       760       770       780
Consensus                       LLNGGSLVIYPZXAIXDPQXLEXYIEEXGVTSLFLTPXLFNLIVDELPP-----ALXTVK 713
2VSQ.A                          MLNAARLIIADEHTLLDTERLTDLILQENVNVMFATTALFNLLTDAGED-----WMKGLR 709
5T3D.A                          FIAGAKLVMAEPEAHRDPLAMQQFFAEYGVTTTHFVPSMLAAFVASLTPQTARQSCATLK 711
4ZXH.A                          LMYGGVVHVCPREVSQDGIQLVDYLNAKSINVLQATPATWKMLLDSEW------SGNAGL 717
BacA_C3-A3                      LSCGAMLVIPTETVRQDPQRLWRYLEEQAITHACLTPAMFHD-GTDL-P-----AIAIKP 710
BacA_C3-GxpS_A3                 LSCGAMLVIPTETVRQDPQRLWRYLEEQAITHACLTPAMFHD-GTDL-P-----AIAIKP 710
Xbud FclJ_C6-A6                 LLNGGSIVIYPQGAISVLL-LEKIIKESGVESLFLTSSLFNLIVDERPQ-----TLQTVK 738
Xhom FclJ_C6-A6                 LLNGGSIVIYPQGAISIPQ-LEKMINDSGVESLFLTASLFNLIVDERPQ-----VLQTVK 736
Xsze FclJ_C6-A6                 LLNGGSIVIYPQSAISVLL-LEKIIKDGRVDSLFLTTSLFNLIVDERPQ-----TLQTVK 761
Xsze FclJ_C6-Xbud FclJ_A6       LLNGGSIVIYPQGAISVLL-LEKIIKESGVESLFLTSSLFNLIVDERPQ-----TLQTVK 757
Xsze FclJ_C6-Xhom FclJ_A6       LLNGGSIVIYPQGAISIPQ-LEKMINDSGVESLFLTASLFNLIVDERPQ-----VLQTVK 754
GxpS_C3-A3                      LSCGAMLVIPTETVRQDPQRLWRYLEEQAITHACLTPAMFHD-GTDL-P-----AIAIKP 714
XtpS_C3-A3                      LLNGGALVVIDRAMLLTPVELVRAVQNHGITVMWLTVGLFNRLSTELSP-----ALPQIK 719
XtpS_C3-GxpS_A3                 LSCGAMLVIPTETVRQDPQRLWRYLEEQAITHACLTPAMFHD-GTDL-P-----AIAIKP 714

                                790       800       810       820       830       840
Consensus                       QLIXGGEAMSPAHAXRJCSRYP-XDLINGYGPTENTVFTTXYCYRA-----TEGVSXPIG 767
2VSQ.A                          CILFGGERASVPHVRKALRIMGPGKLINCYGPTEGTVFATAHVVHDL---PDSISSLPIG 766
5T3D.A                          QVFCSGEAL-PADLCREWQQLTGAPLHNLYGPTEAAVDVSWYPAFGEELAQVRGSSVPIG 770
4ZXH.A                          TALCGGEALDTILAEKLLGKV--GCLWNVYGPTETTVWSSAARI-------TDAKYIDLG 768
BacA_C3-A3                      TLIFAGEAPSPALFQALCSR---ADLFNAYGPTEITVCATTWDCPAD---YTGGV-IPIG 763
BacA_C3-GxpS_A3                 TLIFAGEAPSPALFQALCSR---ADLFNAYGPTEITVCATTWDCPAD---YTGGV-IPIG 763
Xbud FclJ_C6-A6                 QLISGGEAMSSWHAARIKERYPELLLINGYGPTENTTFTTSHCYRA-----TEGNSVPIG 793
Xhom FclJ_C6-A6                 QLISGGEAMSAAHAARISKHYPQLRLINGYGPTENTTFTTSYCYRG-----DEGTSVPIG 791
Xsze FclJ_C6-A6                 QLISGGEAMSSAHANRICKYYPELHLINGYGPTENTTFTTSYRYRG-----TEGNSVPIG 816
Xsze FclJ_C6-Xbud FclJ_A6       QLISGGEAMSSWHAARIKERYPELLLINGYGPTENTTFTTSHCYRA-----TEGNSVPIG 812
Xsze FclJ_C6-Xhom FclJ_A6       QLISGGEAMSAAHAARISKHYPQLRLINGYGPTENTTFTTSYCYRG-----DEGTSVPIG 809
GxpS_C3-A3                      TLIFAGEAPSPALFQALCSR---ADLFNAYGPTEITVCATTWDCPAD---YTGGV-IPIG 767
XtpS_C3-A3                      ILIVGGDVLDPQVISQVLNTNPPQQLLNGYGPSEGTTFTTTYCIRAL---AQSATNIPIG 776
XtpS_C3-GxpS_A3                 TLIFAGEAPSPALFQALCSR---ADLFNAYGPTEITVCATTWDCPAD---YTGGV-IPIG 767
                                                                                Chain VIII

                                850       860       870       880       890       900
Consensus                       KPXANNXLYILDEXRQPVPJGTVGELYIGGDGLARGYLNRPELTAERFIEBPFSDSLSAR 827
2VSQ.A                          KPISNASVYILNEQSQLQPFGAVGELCISGMGVSKGYVNRADLTKEKFIENPF--KPGET 824
5T3D.A                          YPVWNTGLRILDAMMHPVPPGVAGDLYLTGIQLAQGYLGRPDLTASRFIADPF--APGER 828
4ZXH.A                          EPLANTQLYVLDEQQRLVPPGVMGELWIGGDGLAVDYWQRPELTDAQFRTLP-SLPNAGR 827
BacA_C3-A3                      SPVANKRLYLLDEHRQPVPLGTVGELYIGGVGVARGYLNRPELTAERFLNDPFSDETNAR 823
BacA_C3-GxpS_A3                 SPVANKRLYLLDEHRQPVPLGTVGELYIGGVGVARGYLNRPELTAERFLNDPFSDETNAR 823
Xbud FclJ_C6-A6                 KPNEGNLVYILDTFHQPVPIGTAGELYIGGDGLAMTYLNQESLYQDLFIENPFDKSLSAR 853
Xhom FclJ_C6-A6                 KPNAGNLAYILDAFRQPVPIGTVGELYIGGDGLAITYLNQEQLYQKLFIENPFDSSLSAR 851
Xsze FclJ_C6-A6                 KPNEGNLVYILNAFRQPVPIGTTGELYIGGKGLALAYLNQDSLYQNLFIENPFDKSLSTR 876
Xsze FclJ_C6-Xbud FclJ_A6       KPNEGNLVYILDTFHQPVPIGTAGELYIGGDGLAMTYLNQESLYQDLFIENPFDKSLSAR 872
Xsze FclJ_C6-Xhom FclJ_A6       KPNAGNLAYILDAFRQPVPIGTVGELYIGGDGLAITYLNQEQLYQKLFIENPFDSSLSAR 869
GxpS_C3-A3                      SPVANKRLYLLDEHRQPVPLGTVGELYIGGVGVARGYLNRPELTAERFLNDPFSDETNAR 827
XtpS_C3-A3                      RPIANTRVYLLDNHGQPVPLGAIGEIYIGGDGVACGYLNRPELTADRFLIDPFSDVPDAR 836
XtpS_C3-GxpS_A3                 SPVANKRLYLLDEHRQPVPLGTVGELYIGGVGVARGYLNRPELTAERFLNDPFSDETNAR 827
                                          Chain IX                              Chain X
```

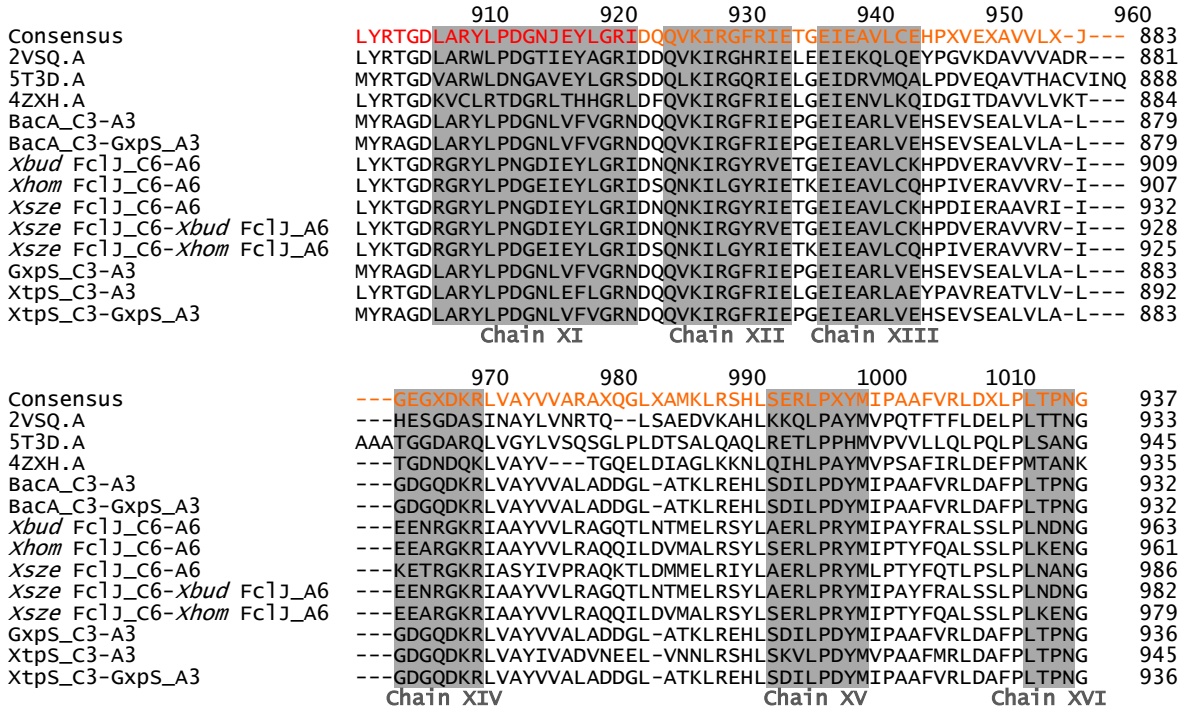


Table S8. Raw data refer to Fig. 35 of Visual TreeCmp Comparison of Phylogenetic Trees input data of weighted and rooted *Cophenetic Metric with L^2 norm* (*CML*). NRPS modules 1: XldS_C3A3T3, 2: GarS_C2A2T2, 3: XabABC$_{Xnema}$_C2A2T2, 4: XabABC$_{Xdouc}$_C2A2T2, 5: AmbS$_{Xmira}$_C3A3T3, 6: AmbS$_{Xindi}$_C3A3T3, 7: KolS_C3A3T3, 8: GxpS_C3A3T3, 9: XtpS_C3A3T3, 10: SzeS_C3A3T3, 11: HCTA_C6A6T6, 12: XeyS_C3A3T3, 13: BicA_C3A3T3, 14: PAX_C3A3T3, and 15: OldS_C5A5T5 were used for the calculation.

No	Tree1	Tree2	Tree1_taxa	Tree2_taxa	Common_taxa	Triples	*CML*
1	1	2	15	15	15	42	8.6534
2	1	3	15	15	15	175	3.5625
3	1	4	15	15	15	304	3.5033
4	1	5	15	15	15	277	3.4340
5	1	6	15	15	15	159	1.8117
6	1	7	15	15	15	146	1.9113
7	1	8	15	15	15	124	8.2375
8	1	9	15	15	15	65	6.4127
9	1	10	15	15	15	145	1.9056
10	1	11	15	15	15	153	2.3714
11	1	12	15	15	15	140	8.0717
12	1	13	15	15	15	250	4.3402
13	1	14	15	15	15	231	6.3457
14	1	15	15	15	15	251	9.8152
15	1	16	15	15	15	206	2.4278
16	1	17	15	15	15	146	2.3371
17	1	18	15	15	14	118	2.2586
18	1	19	15	15	15	290	3.4843
19	1	20	15	15	15	134	1.4654
20	1	21	15	15	14	194	3.5493
21	2	3	15	15	15	196	6.7238
22	2	4	15	15	15	329	10.8793
23	2	5	15	15	15	314	10.3203
24	2	6	15	15	15	132	8.8351
25	2	7	15	15	15	176	9.6109
26	2	8	15	15	15	119	3.7703
27	2	9	15	15	15	95	3.4280

28	2	10	15	15	15	166	9.6829
29	2	11	15	15	15	154	9.8348
30	2	12	15	15	15	157	5.5622
31	2	13	15	15	15	262	8.5274
32	2	14	15	15	15	243	9.3031
33	2	15	15	15	15	267	9.6263
34	2	16	15	15	15	227	10.4773
35	2	17	15	15	15	171	10.5588
36	2	18	15	15	14	142	6.9780
37	2	19	15	15	15	316	11.2142
38	2	20	15	15	15	159	8.4940
39	2	21	15	15	14	214	10.1675
40	3	4	15	15	15	317	6.1342
41	3	5	15	15	15	332	5.7048
42	3	6	15	15	15	126	3.1660
43	3	7	15	15	15	114	4.0442
44	3	8	15	15	15	154	5.8088
45	3	9	15	15	15	135	4.2189
46	3	10	15	15	15	136	4.3973
47	3	11	15	15	15	145	4.4953
48	3	12	15	15	15	100	6.2821
49	3	13	15	15	15	252	4.8586
50	3	14	15	15	15	291	7.0843
51	3	15	15	15	15	307	9.7605
52	3	16	15	15	15	232	5.3035
53	3	17	15	15	15	154	5.1759
54	3	18	15	15	14	118	2.5482
55	3	19	15	15	15	307	6.2587
56	3	20	15	15	15	160	3.4132
57	3	21	15	15	14	227	6.0616
58	4	5	15	15	15	141	2.1821
59	4	6	15	15	15	325	3.8888
60	4	7	15	15	15	283	3.8543
61	4	8	15	15	15	299	10.8617
62	4	9	15	15	15	264	8.6667
63	4	10	15	15	15	273	3.5929
64	4	11	15	15	15	290	3.9199
65	4	12	15	15	15	317	10.0415
66	4	13	15	15	15	239	5.1055
67	4	14	15	15	15	257	6.1863
68	4	15	15	15	15	212	9.5542
69	4	16	15	15	15	196	2.5940
70	4	17	15	15	15	284	3.1861
71	4	18	15	15	14	216	4.1771
72	4	19	15	15	15	160	1.9157
73	4	20	15	15	15	271	4.0712
74	4	21	15	15	14	203	1.7564
75	5	6	15	15	15	323	3.6796
76	5	7	15	15	15	283	3.3430
77	5	8	15	15	15	295	10.4543
78	5	9	15	15	15	278	8.1514
79	5	10	15	15	15	288	3.1174
80	5	11	15	15	15	302	3.4218
81	5	12	15	15	15	325	9.6888
82	5	13	15	15	15	290	4.6935
83	5	14	15	15	15	274	5.9490
84	5	15	15	15	15	246	9.3393
85	5	16	15	15	15	251	2.7290
86	5	17	15	15	15	288	2.9737
87	5	18	15	15	14	217	3.9149
88	5	19	15	15	15	166	1.9505
89	5	20	15	15	15	283	3.7779

90	5	21	15	15	14	162	2.3783
91	6	7	15	15	15	135	1.9673
92	6	8	15	15	15	155	8.2460
93	6	9	15	15	15	158	6.3711
94	6	10	15	15	15	153	2.3190
95	6	11	15	15	15	156	2.4580
96	6	12	15	15	15	120	8.1208
97	6	13	15	15	15	273	4.4294
98	6	14	15	15	15	286	6.4658
99	6	15	15	15	15	304	9.9947
100	6	16	15	15	15	248	2.9164
101	6	17	15	15	15	135	2,6795
102	6	18	15	15	14	96	2.4563
103	6	19	15	15	15	309	3.8655
104	6	20	15	15	15	160	2.0426
105	6	21	15	15	14	220	4.0270
106	7	8	15	15	15	147	8.9839
107	7	9	15	15	15	118	7.2097
108	7	10	15	15	15	66	1.1806
109	7	11	15	15	15	148	1.4821
110	7	12	15	15	15	126	8.8882
111	7	13	15	15	15	265	4.6073
112	7	14	15	15	15	277	6.7564
113	7	15	15	15	15	283	10.4417
114	7	16	15	15	15	187	2.2317
115	7	17	15	15	15	71	1.5364
116	7	18	15	15	14	42	3.0155
117	7	19	15	15	15	253	3.1958
118	7	20	15	15	15	68	1.7811
119	7	21	15	15	14	167	3.8773
120	8	9	15	15	15	159	4.1044
121	8	10	15	15	15	134	9.1858
122	8	11	15	15	15	140	9.2967
123	8	12	15	15	15	134	4.7454
124	8	13	15	15	15	230	8.4292
125	8	14	15	15	15	287	10.2046
126	8	15	15	15	15	282	11.4157
127	8	16	15	15	15	189	10.0902
128	8	17	15	15	15	132	10.0286
129	8	18	15	15	14	113	6.5443
130	8	19	15	15	15	286	11.0209
131	8	20	15	15	15	142	8.0067
132	8	21	15	15	14	201	9.8213
133	9	10	15	15	15	128	7.3468
134	9	11	15	15	15	151	7.5337
135	9	12	15	15	15	149	4.9919
136	9	13	15	15	15	243	6.6095
137	9	14	15	15	15	209	7.6093
138	9	15	15	15	15	224	8.7608
139	9	16	15	15	15	197	8.2429
140	9	17	15	15	15	155	8.2696
141	9	18	15	15	14	120	4.4487
142	9	19	15	15	15	258	8.9678
143	9	20	15	15	15	122	6.1648
144	9	21	15	15	14	203	8.1833
145	10	11	15	15	15	107	1.2767
146	10	12	15	15	15	75	8.6398
147	10	13	15	15	15	211	4.0769
148	10	14	15	15	15	271	6.2378
149	10	15	15	15	15	263	9.9991
150	10	16	15	15	15	143	1.7887
151	10	17	15	15	15	63	1.4350

152	10	18	15	15	14	43	2.9126
153	10	19	15	15	15	262	2.8965
154	10	20	15	15	15	64	1.7616
155	10	21	15	15	14	174	3.4608
156	11	12	15	15	15	107	8.8265
157	11	13	15	15	15	202	3.9518
158	11	14	15	15	15	272	6.4542
159	11	15	15	15	15	285	10.2453
160	11	16	15	15	15	170	2.0526
161	11	17	15	15	15	133	1.7188
162	11	18	15	15	14	104	3.2702
163	11	19	15	15	15	277	3.2541
164	11	20	15	15	15	147	2.0846
165	11	21	15	15	14	215	4.0501
166	12	13	15	15	15	195	6.9382
167	12	14	15	15	15	279	8.6058
168	12	15	15	15	15	287	9.6185
169	12	16	15	15	15	186	9.4583
170	12	17	15	15	15	106	9.6993
171	12	18	15	15	14	87	6.3152
172	12	19	15	15	15	298	10.2206
173	12	20	15	15	15	115	7.8912
174	12	21	15	15	14	208	8.8710
175	13	14	15	15	15	256	4.9327
176	13	15	15	15	15	235	7.8543
177	13	16	15	15	15	98	4.2895
178	13	17	15	15	15	249	4.8487
179	13	18	15	15	14	191	3.4091
180	13	19	15	15	15	267	5.0238
181	13	20	15	15	15	253	4.2023
182	13	21	15	15	14	224	5.0414
183	14	15	15	15	15	148	5.4583
184	14	16	15	15	15	275	6.5201
185	14	17	15	15	15	288	6.9666
186	14	18	15	15	14	216	5.4209
187	14	19	15	15	15	283	6.5970
188	14	20	15	15	15	276	6.2629
189	14	21	15	15	14	236	6.1782
190	15	16	15	15	15	241	10.1783
191	15	17	15	15	15	292	10.7456
192	15	18	15	15	14	226	8.7241
193	15	19	15	15	15	246	10.1801
194	15	20	15	15	15	262	9.7438
195	15	21	15	15	14	200	9.6367
196	16	17	15	15	15	174	1.3151
197	16	18	15	15	14	118	3.4549
198	16	19	15	15	15	220	2.0447
199	16	20	15	15	15	183	2.7768
200	16	21	15	15	14	183	2.7963
201	17	18	15	15	14	16	3.5328
202	17	19	15	15	15	250	2.3222
203	17	20	15	15	15	77	2.5191
204	17	21	15	15	14	176	3.2681
205	18	19	15	15	14	204	4.1005
206	18	20	15	15	14	58	2.3243
207	18	21	15	15	13	137	3.8290
208	19	20	15	15	15	254	3.7996
209	19	21	15	15	14	181	1.6845
210	20	21	15	15	14	167	4.1446

Table S9. Overview of all 20 proteinogenic amino acids in comparison to each other by their calculated *E*-descriptor values in *E*-descriptor vectors distances (Δ) with CAopt.py. The color gradient of the distances (Δ) ranges from the lowest (green) to the highest (red) calculated distance.

AA1	AA2	*E*-descriptor vector of AA1	*E*-descriptor vector of AA2	Δ
A	A	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,000
R	R	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,000
N	N	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,000
D	D	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,000
C	C	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,000
Q	Q	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,000
E	E	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,000
G	G	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,000
H	H	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,000
I	I	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,000
L	L	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,000
K	K	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,000
M	M	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,000
F	F	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,000
P	P	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,000
S	S	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,000
T	T	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,000
W	W	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,000
Y	Y	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,000
V	V	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,000
S	T	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,189
T	S	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,189
I	V	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,239
V	I	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,239
N	D	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,291
D	N	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,291
Q	H	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,299
H	Q	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,299
N	Q	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,304
Q	N	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,304
N	S	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,339
S	N	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,339
H	M	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,346
M	H	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,346
D	H	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,361
H	D	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,361
N	H	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,365
H	N	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,365
F	W	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,379
W	F	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,379
I	L	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,382
L	I	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,382
R	K	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,383
K	R	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,383
D	Q	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,385
Q	D	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,385
T	V	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,389
V	T	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,389
Q	K	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,412
K	Q	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,412
Q	T	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,413
T	Q	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,413
I	F	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,414
F	I	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,414
D	E	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,416

E	D	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,416
N	T	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,418
T	N	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,418
A	L	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,435
L	A	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,435
L	F	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,439
F	L	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,439
M	F	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,442
F	M	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,442
Q	S	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,446
S	Q	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,446
R	Q	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,447
Q	R	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,447
G	S	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,452
S	G	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,452
I	T	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,469
T	I	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,469
W	Y	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,474
Y	W	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,474
F	Y	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,509
Y	F	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,509
H	F	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,530
F	H	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,530
L	V	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,530
V	L	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,530
Q	E	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,532
E	Q	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,532
Q	M	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,534
M	Q	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,534
S	V	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,539
V	S	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,539
A	E	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,544
E	A	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,544
E	H	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,545
H	E	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,545
R	T	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,555
T	R	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,555
D	S	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,555
S	D	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,555
N	Y	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,555
Y	N	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,555
T	Y	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,560
Y	T	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,560
Q	Y	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,570
Y	Q	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,570
N	G	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,576
G	N	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,576
E	M	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,578
M	E	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,578
I	Y	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,578
Y	I	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,578
H	Y	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,579
Y	H	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,579
D	M	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,580
M	D	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,580
H	T	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,586
T	H	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,586
C	H	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,595
H	C	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,595
H	S	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,598
S	H	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,598
G	T	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,601

T	G	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,601
I	S	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,602
S	I	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,602
I	M	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,602
M	I	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,602
H	W	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,602
W	H	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,602
N	K	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,604
K	N	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,604
K	T	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,605
T	K	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,605
Q	F	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,607
F	Q	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,607
Q	I	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,611
I	Q	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,611
A	M	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,616
M	A	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,616
F	V	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,625
V	F	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,625
A	I	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,626
I	A	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,626
R	Y	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,626
Y	R	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,626
Q	V	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,628
V	Q	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,628
G	P	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,629
P	G	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,629
A	S	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,634
S	A	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,634
R	N	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,635
N	R	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,635
M	W	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,637
W	M	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,637
A	D	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,637
D	A	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,637
S	Y	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,639
Y	S	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,639
N	M	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,639
M	N	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,639
L	M	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,639
M	L	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,639
D	T	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,640
T	D	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,640
A	Q	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,641
Q	A	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,641
A	V	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,646
V	A	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,646
N	E	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,647
E	N	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,647
E	K	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,647
K	E	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,647
D	K	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,648
K	D	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,648
C	M	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,651
M	C	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,651
K	S	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,656
S	K	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,656
N	F	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,656
F	N	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,656
A	T	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,657
T	A	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,657
D	G	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,660

G	D	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,660
R	S	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,662
S	R	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,662
A	F	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,663
F	A	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,663
A	G	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,663
G	A	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,663
L	T	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,665
T	L	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,665
F	T	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,666
T	F	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,666
H	K	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,669
K	H	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,669
H	I	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,673
I	H	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,673
N	P	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,675
P	N	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,675
P	Y	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,676
Y	P	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,676
D	F	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,680
F	D	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,680
Q	L	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,682
L	Q	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,682
R	H	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,684
H	R	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,684
N	C	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,687
C	N	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,687
Y	V	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,695
V	Y	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,695
M	V	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,698
V	M	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,698
L	K	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,702
K	L	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,702
A	N	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,708
N	A	[[0.255 0.038 0.117 0.118 -0.055]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,708
K	Y	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,710
Y	K	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,710
A	H	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,711
H	A	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,711
N	I	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,717
I	N	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,717
F	S	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,723
S	F	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,723
E	L	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,724
L	E	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,724
M	T	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,725
T	M	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,725
L	S	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,729
S	L	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,729
F	P	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,732
P	F	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,732
I	W	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,733
W	I	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,733
P	S	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,734
S	P	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,734
P	W	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,734
W	P	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,734
R	V	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,735
V	R	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,735
N	V	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,735
V	N	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,735
H	V	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,736

V	H	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,736
Q	W	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,737
W	Q	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,737
C	T	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,737
T	C	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,737
I	K	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,741
K	I	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,741
E	F	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,743
F	E	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,743
N	W	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,744
W	N	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,744
C	S	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,744
S	C	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,744
D	Y	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,744
Y	D	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,744
M	Y	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,745
Y	M	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,745
D	P	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,747
P	D	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,747
C	Q	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,753
Q	C	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,753
R	I	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,754
I	R	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,754
H	L	[[0.023 -0.177 0.041 0.28 -0.021]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,758
L	H	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,758
D	W	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,760
W	D	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,760
M	S	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,762
S	M	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,762
A	K	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,768
K	A	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,768
K	F	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,775
F	K	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,775
K	V	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,777
V	K	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,777
L	W	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,778
W	L	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,778
Q	G	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,780
G	Q	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,780
L	Y	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,786
Y	L	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,786
R	D	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,793
D	R	[[0.303 -0.057 -0.014 0.225 0.156]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,793
N	L	[[0.255 0.038 0.117 0.118 -0.055]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,800
L	N	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.255 0.038 0.117 0.118 -0.055]]	0,800
D	L	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,803
L	D	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,803
D	C	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,804
C	D	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,804
P	T	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,805
T	P	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,805
E	S	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,811
S	E	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,811
G	V	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,812
V	G	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,812
G	F	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,813
F	G	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,813
G	I	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,814
I	G	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,814
C	V	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,820
V	C	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,820
G	H	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,820

H	G	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,820
G	L	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,827
L	G	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,827
C	Y	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,835
Y	C	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,835
D	I	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,835
I	D	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,835
K	M	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,844
M	K	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,844
E	T	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,844
T	E	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,844
Q	P	[[0.149 -0.184 -0.03 0.035 -0.112]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,849
P	Q	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.149 -0.184 -0.03 0.035 -0.112]]	0,849
C	I	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,856
I	C	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,856
C	F	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,859
F	C	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,859
L	P	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,867
P	L	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,867
R	F	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.329 -0.023 0.072 -0.002 0.208]]	0,871
F	R	[[-0.329 -0.023 0.072 -0.002 0.208]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,871
G	Y	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,875
Y	G	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,875
K	W	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,879
W	K	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,879
E	W	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,882
W	E	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,882
T	W	[[0.068 0.147 -0.015 -0.132 -0.274]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,883
W	T	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.068 0.147 -0.015 -0.132 -0.274]]	0,883
H	P	[[0.023 -0.177 0.041 0.28 -0.021]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,883
P	H	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.023 -0.177 0.041 0.28 -0.021]]	0,883
K	P	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,883
P	K	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,883
D	V	[[0.303 -0.057 -0.014 0.225 0.156]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,886
V	D	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.303 -0.057 -0.014 0.225 0.156]]	0,886
E	I	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,892
I	E	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,892
R	L	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	0,894
L	R	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,894
R	E	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,897
E	R	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,897
I	P	[[-0.353 0.071 -0.088 -0.195 -0.107]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,900
P	I	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.353 0.071 -0.088 -0.195 -0.107]]	0,900
C	G	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,901
G	C	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,901
G	M	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,902
M	G	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,902
R	M	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	0,903
M	R	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,903
E	G	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,923
G	E	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,923
S	W	[[0.199 0.238 -0.015 -0.068 -0.196]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,924
W	S	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.199 0.238 -0.015 -0.068 -0.196]]	0,924
W	V	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,932
V	W	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,932
R	W	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,939
W	R	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,939
A	P	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,949
P	A	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,949
E	V	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	0,960
V	E	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,960
C	W	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,970

W	C	[[-0.296 -0.186 0.389 0.083 0.297]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,970
A	Y	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,973
Y	A	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,973
E	Y	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.141 -0.057 0.425 -0.096 -0.091]]	0,973
Y	E	[[-0.141 -0.057 0.425 -0.096 -0.091]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,973
G	K	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	0,974
K	G	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[0.218 0.562 -0.024 0.018 0.106]]	0,974
A	R	[[0.008 0.134 -0.475 -0.039 0.181]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	0,977
R	A	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,977
A	W	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.296 -0.186 0.389 0.083 0.297]]	0,985
W	A	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,985
E	P	[[0.221 -0.28 -0.315 0.157 0.303]]	[[0.173 0.286 0.407 -0.215 0.384]]	0,994
P	E	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.221 -0.28 -0.315 0.157 0.303]]	0,994
A	C	[[0.008 0.134 -0.475 -0.039 0.181]]	[[-0.132 0.174 0.07 0.565 -0.374]]	0,996
C	A	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.008 0.134 -0.475 -0.039 0.181]]	0,996
G	W	[[0.218 0.562 -0.024 0.018 0.106]]	[[-0.296 -0.186 0.389 0.083 0.297]]	1,017
W	G	[[-0.296 -0.186 0.389 0.083 0.297]]	[[0.218 0.562 -0.024 0.018 0.106]]	1,017
P	V	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.274 0.136 -0.187 -0.196 -0.299]]	1,021
V	P	[[-0.274 0.136 -0.187 -0.196 -0.299]]	[[0.173 0.286 0.407 -0.215 0.384]]	1,021
M	P	[[-0.239 -0.141 -0.155 0.321 0.077]]	[[0.173 0.286 0.407 -0.215 0.384]]	1,024
P	M	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.239 -0.141 -0.155 0.321 0.077]]	1,024
R	C	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[-0.132 0.174 0.07 0.565 -0.374]]	1,028
C	R	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	1,028
R	P	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.173 0.286 0.407 -0.215 0.384]]	1,034
P	R	[[0.173 0.286 0.407 -0.215 0.384]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	1,034
C	E	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.221 -0.28 -0.315 0.157 0.303]]	1,051
E	C	[[0.221 -0.28 -0.315 0.157 0.303]]	[[-0.132 0.174 0.07 0.565 -0.374]]	1,051
R	G	[[0.171 -0.361 0.107 -0.258 -0.364]]	[[0.218 0.562 -0.024 0.018 0.106]]	1,081
G	R	[[0.218 0.562 -0.024 0.018 0.106]]	[[0.171 -0.361 0.107 -0.258 -0.364]]	1,081
C	L	[[-0.132 0.174 0.07 0.565 -0.374]]	[[-0.267 0.018 -0.265 -0.274 0.206]]	1,093
L	C	[[-0.267 0.018 -0.265 -0.274 0.206]]	[[-0.132 0.174 0.07 0.565 -0.374]]	1,093
C	K	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.243 -0.339 -0.044 -0.325 -0.027]]	1,153
K	C	[[0.243 -0.339 -0.044 -0.325 -0.027]]	[[-0.132 0.174 0.07 0.565 -0.374]]	1,153
C	P	[[-0.132 0.174 0.07 0.565 -0.374]]	[[0.173 0.286 0.407 -0.215 0.384]]	1,184
P	C	[[0.173 0.286 0.407 -0.215 0.384]]	[[-0.132 0.174 0.07 0.565 -0.374]]	1,184

Table S10. Clustal Omega alignment of the A domain interfaces of the used C-A didomains for the CAopt.py randomisation. Shown is the initial sequence of the GxpS A3 (bold, **Seq1**) and the corresponding reference interfaces of the desired NRPSs (italic, *Seq1*) for the engineered sequences (standard, Seq2-x). The A interface areas A1-7 are highlighted in purple.

```
                                        10        20        30        40        50        60
GxpS_A3-T3                  Seq1   P-ADYTGGVIPIGSPVANKRLYLLDEHRQPVPLGTVGELYIGGVGVARGYLNRPELTAER  59
XtpS_A3                     Seq1   RALAQSATNIPIGRPIANTRVYLLDNHGQPVPLGAIGEIYIGGDGVACGYLNRPELTADR  60
XtpS_C3-GxpS A3 HIGH1       Seq2   P-ADYTGGVIPIGSPVANKRLYLLDNHGQPQPLGAVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 HIGH2       Seq3   P-ADYTGGVIPIGSPVANKRLYLLDDHGQPVPFGAVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 HIGH3       Seq4   P-ADYTGGVIPIGSPVANKRLYLLENHGHPVPLGEVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 MEDIUM1     Seq5   P-ADYTGGVIPIGSPVANKRLYLLDEHGGPDPTCTVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 MEDIUM2     Seq6   P-ADYTGGVIPIGSPVANKRLYLLDAANTPVVLGTVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 MEDIUM3     Seq7   P-ADYTGGVIPIGSPVANKRLYLLIEGKIPVPYFTVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 LOW99       Seq8   P-ADYTGGVIPIGSPVANKRLYLLVEHRWVVRCRTVGELYIGGVGVARGYLNRPELTAER  59
XtpS_C3-GxpS A3 LOW100      Seq9   P-ADYTGGVIPIGSPVANKRLYLLDEHRCCPCRETVGELYIGGVGVARGYLNRPELTAER  59
AmbS_A3                     Seq1   M-NETGGQANTIGRPLPNMRVYLLDAYGQPVPLGAVGELYIGGAGVARGYLNRPELTAER  59
AmbS_C3-GxpS A3 HIGH1       Seq2   P-ADYTGGVIPIGSPVANKRLYLLDAFGQPVPLGTVGELYIGGVGVARGYLNRPELTAER  59
AmbS_C3-GxpS A3 HIGH3       Seq3   P-ADYTGGVIPIGSPVANKRLYLLDARGQPVPLGKVGELYIGGVGVARGYLNRPELTAER  59
BicA_A3                     Seq1   QTVVSPEDDRSIGRPLSNTRIYLLDSHGQPVPLGSVGEIYIGGVGVARGYLNQPILSEER  60
BicA_C3-GxpS A3 HIGH1       Seq2   P-ADYTGGVIPIGSPVANKRLYLLDTHGQPKPLGSVGELYIGGVGVARGYLNRPELTAER  59
BicA_C3-GxpS A3 HIGH2       Seq3   P-ADYTGGVIPIGSPVANKRLYLLDSCGRPVPLGSVGELYIGGVGVARGYLNRPELTAER  59
BicA_C3-GxpS A3 HIGH3       Seq3   P-ADYTGGVIPIGSPVANKRLYLLDKHGQGVPLGSVGELYIGGVGVARGYLNRPELTAER  59
BacA_A3                     Seq1   --IDKSHQNIPIGKPIDNVKVYILNKDLQLCPLGASGELCIAGEGLARGYVNRPELTREK  58
BacA_C3-GxpS A3 HIGH1       Seq2   P-ADYTGGVIPIGSPVANKRLYLLQVQLQLCPLGTVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 HIGH2       Seq3   P-ADYTGGVIPIGSPVANKRLYLLDKDIQLYPLGTVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 HIGH3       Seq4   P-ADYTGGVIPIGSPVANKRLYLLSEHAQICPLGAVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 MEDIUM1     Seq5   P-ADYTGGVIPIGSPVANKRLYLLYEYAQGEPLVTVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 MEDIUM2     Seq6   P-ADYTGGVIPIGSPVANKRLYLLDTGRQVVPSYTVGELYIGGVGVARGYLNRPELTAER  59
```

```
BacA_C3-GxpS A3 MEDIUM3 Seq7   P-ADYTGGVIPIGSPVANKRLYLLDEVGSPFPLKFVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 LOW98    Seq8   P-ADYTGGVIPIGSPVANKRLYLLKEVRCPVACKTVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 LOW99    Seq9   P-ADYTGGVIPIGSPVANKRLYLLWEPRCPVVCETVGELYIGGVGVARGYLNRPELTAER  59
BacA_C3-GxpS A3 LOW100   Seq10  P-ADYTGGVIPIGSPVANKRLYLLEEHRPPKECRTVGELYIGGVGVARGYLNRPELTAER  59
                                 Interface A1                 Interface A2

                                        70        80        90       100       110       120
GxpS_A3-T3               Seq1   FLNDPFSDETNARMYRAGDLARYLPDGNLVFVGRNDQQVKIRGFRIEPGEIEARLVEHSE 119
XtpS_A3                  Seq1   FLIDPFSDVPDARLYRTGDLARYLPDGNLEFLGRNDQQVKIRGFRIELGEIEARLAEYPA 120
XtpS_C3-GxpS_A3 HIGH1    Seq2   FLNDPFSDETNARMYRAGDLARYLPDGNLVFVGRNDQQLKIRGFRIELGEIEARLVEHSE 119
XtpS_C3-GxpS_A3 HIGH2    Seq3   FLNDPFSDETNARMYRAGNLARYLPDGNLVFVGRNDQQVIIRGFRIELGEIEARLVEHSE 119
XtpS_C3-GxpS_A3 HIGH3    Seq4   FLNDPFSDETNARMYRAGHLARYLPDGNLVFVGRNDQQVKVRGFRIEVGEIEARLVEHSE 119
XtpS_C3-GxpS_A3 MEDIUM1  Seq5   FLNDPFSDETNARMYRAGLLSRYDPDGNLVFVGRNDQQVKMDGFRLNSWEIEARLVEHSE 119
XtpS_C3-GxpS_A3 MEDIUM2  Seq6   FLNDPFSDETNARMYRAGDSGEYLPDGNLVFVGRNDQQQKICGFWYTFGEIEARLVEHSE 119
XtpS_C3-GxpS_A3 MEDIUM3  Seq7   FLNDPFSDETNARMYRAGDEQDYLPDGNLVFVGRNDQQVNIHGFRQELGEIEARLVEHSE 119
XtpS_C3-GxpS_A3 LOW99    Seq8   FLNDPFSDETNARMYRAGDCAPYCPDGNLVFVGRNDQQVCIMKFCIEPRPIEARLVEHSE 119
XtpS_C3-GxpS_A3 LOW100   Seq9   FLNDPFSDETNARMYRAGDCAGYCPDGNLVFVGRNDQQVCPCRFRIPPWEIEARLVEHSE 119
AmbS_A3                  Seq1   FLTDPFSDEPGARMYRSGDLARYRPDGNLVFVGRNDQQVKIRGFRVEPGEIEARLEEHPL 119
AmbS_C3-GxpS_A3 HIGH1    Seq2   FLNDPFSDETNARMYRAGDLARYRPDGNLVFVGRNDQQVKIRGIRIEPGEIEARLVEHSE 119
AmbS_C3-GxpS_A3 HIGH3    Seq3   FLNDPFSDETNARMYRAGHLARYRPDGNLVFVGRNDQQVKIRGFRIFPGEIEARLVEHSE 119
BicA_A3                  Seq1   FISDPFSPVSDARMYRTGDMARYRPDGSLEFLGRNDQQVKIRGFRVELGEIETRLLEHPA 120
BicA_C3-GxpS_A3 HIGH1    Seq2   FLNDPFSDETNARMYRAGDMARYRPDGNLVFVGRNDQQVKIRGFRYEAGEIEARLVEHSE 119
BicA_C3-GxpS_A3 HIGH2    Seq3   FLNDPFSDETNARMYRAGDHARYRPDGNLVFVGRNDQQVKIRGFRVEKSEIEARLVEHSE 119
BicA_C3-GxpS_A3 HIGH3    Seq3   FLNDPFSDETNARMYRAGHMARYRPDGNLVFVGRNDQQVKTRGFRIELSEIEARLVEHSE 119
BacA_A3                  Seq1   FIGNPFV--PGERMYRTGDLAKMLPDGNIQFLGRVDQQVKIRGYRIEPGEIENRLLKYEK 116
BacA_C3-GxpS_A3 HIGH1    Seq2   FLNDPFSDETNARMYRAGDLAKMLPDGNLVFVGRNDQQVKIRGYRAEPGEIEARLVEHSE 119
BacA_C3-GxpS_A3 HIGH2    Seq3   FLNDPFSDETNARMYRAGNLAKMLPDGNLVFVGRNDQQVYVRGYRIEPGEIEARLVEHSE 119
BacA_C3-GxpS_A3 HIGH3    Seq4   FLNDPFSDETNARMYRAGDLAKHLPDGNLVFVGRNDQQVKIRGIKIEPGEIEARLVEHSE 119
BacA_C3-GxpS_A3 MEDIUM1  Seq5   FLNDPFSDETNARMYRAGDPACILPDGNLVFVGRNDQQVKARGFEPFIYEIEARLVEHSE 119
BacA_C3-GxpS_A3 MEDIUM2  Seq6   FLNDPFSDETNARMYRAGDLSKYGPDGNLVFVGRNDQQVCIRGFRAEIWYIEARLVEHSE 119
BacA_C3-GxpS_A3 MEDIUM3  Seq7   FLNDPFSDETNARMYRAGDMGAYLPDGNLVFVGRNDQQVRIYGFWEEWGEIEARLVEHSE 119
BacA_C3-GxpS_A3 LOW98    Seq8   FLNDPFSDETNARMYRAGICARYCPDGNLVFVGRNDQQVCIRRFRICRKGIEARLVEHSE 119
BacA_C3-GxpS_A3 LOW99    Seq9   FLNDPFSDETNARMYRAGDCRRYCPDGNLVFVGRNDQQVCICGFAICMWEIEARLVEHSE 119
BacA_C3-GxpS_A3 LOW100   Seq10  FLNDPFSDETNARMYRAGDCCRYCPDGNLVFVGRNDQQVKIPWFPIECREAREARLVEHSE 119
                                                 Interface A3          Interface A4

                                       130       140       150       160       170       180
GxpS_A3-T3               Seq1   VSEALVLALGDG-QDKRLVAYVVA-----LADDGLATKLREHLSDILPDYMIPAAFVRLD 173
XtpS_A3                  Seq1   VREATVLVLGDG-QDKRLVAYIVA-----DVNEELVNNLRSHLSKVLPDYMVPAAFMRLD 174
XtpS_C3-GxpS_A3 HIGH1    Seq2   VSEALVLALGDG-QDKRLVAYVVA-----LADDGLATKLREHLSDVLPDYMVPAAFVRLD 173
XtpS_C3-GxpS_A3 HIGH2    Seq3   VSEALVLALGNG-QDKRLVAYVVA-----LADDGLATKLREHLSDVLPDYMIPAAFVRLD 173
XtpS_C3-GxpS_A3 HIGH3    Seq4   VSEALVLALGDG-QNKRLVAYVVA-----LADDGLATKLREHLSDILPDYMVPAAFVRLD 173
XtpS_C3-GxpS_A3 MEDIUM1  Seq5   VSEALVLALGIW-QSKRLVAYVVA-----LADDGLATKLREHLSDIIPNLLIPAAFVRLD 173
XtpS_C3-GxpS_A3 MEDIUM2  Seq6   VSEALVLALGPT-SDKRLVAYVVA-----LADDGLATKLREHLSDIYKDKKIPAAFVRLD 173
XtpS_C3-GxpS_A3 MEDIUM3  Seq7   VSEALVLALGDR-QLKCLVAYVVA-----LADDGLATKLREHLSDIRPDADMPAAFVRLD 173
XtpS_C3-GxpS_A3 LOW99    Seq8   VSEALVLALGDR-QDCCLVAYVVA-----LADDGLATKLREHLSDICCDAKIPAAFVRLD 173
XtpS_C3-GxpS_A3 LOW100   Seq9   VSEALVLALGDR-QDCPLVAYVVA-----LADDGLATKLREHLSDICCRYPIPAAFVRLD 173
AmbS_A3                  Seq1   VSEALVVAFDGG-QDKRLVAYVVA-----QADEALSHRLYTYLSAILPDYMVPTAFVCLP 173
AmbS_C3-GxpS_A3 HIGH1    Seq2   VSEALVLALGDG-QDKRLVAYVVA-----LADDGLATKLREHLSDILPDYMVPAAFVRLD 173
AmbS_C3-GxpS_A3 HIGH3    Seq3   VSEALVLALGDG-QDKRLVAYVVA-----LADDGLATKLREHLSDILPDYMVPAAFVRLD 173
BicA_A3                  Seq1   VREAAVLAREDG-QDKRLVAYVVAKTRVAEADETLVNNLRTHLSVVLPDYMVPAAFVRLD 179
BicA_C3-GxpS_A3 HIGH1    Seq2   VSEALVLALGDG-QDKRLVAYVVA-----LADDGLATKLREHLSDVLPDYMVPAAFVRLD 173
BicA_C3-GxpS_A3 HIGH2    Seq3   VSEALVLALGNG-QDKRLVAYVVA-----LADDGLATKLREHLSDVLPDYMIPAAFVRLD 173
BicA_C3-GxpS_A3 HIGH3    Seq4   VSEALVLALGDG-QNKRLVAYVVA-----LADDGLATKLREHLSDILPDYMVPAAFVRLD 173
BacA_A3                  Seq1   IEEAAVIAREDGDHDPYLCAYVTVKKE------VEPEKIRAFLKKSLPDYMIPQYFVQLD 170
BacA_C3-GxpS_A3 HIGH1    Seq2   VSEALVLALGDG-DDDPLVAYVVA-----LADDGLATKLREHLSDSLPDYMIPAAFVRLD 173
BacA_C3-GxpS_A3 HIGH2    Seq3   VSEALVLALGDG-QHDPLVAYVVA-----LADDGLATKLREHLSDSLPNYMIPAAFVRLD 173
BacA_C3-GxpS_A3 HIGH3    Seq4   VSEALVLALGDG-NDDPLVAYVVA-----LADDGLATKLREHLSDNLPDYMIPAAFVRLD 173
BacA_C3-GxpS_A3 MEDIUM1  Seq5   VSEALVLALGHG-QQKPLVAYVVA-----LADDGLATKLREHLSDHSPDMMHPAAFVRLD 173
BacA_C3-GxpS_A3 MEDIUM2  Seq6   VSEALVLALGQG-QDGNLVAYVVA-----LADDGLATKLREHLSDIVFDTMVPAAFVRLD 173
BacA_C3-GxpS_A3 MEDIUM3  Seq7   VSEALVLALGVG-QMKWLVAYVVA-----LADDGLATKLREHLSDNNPWYSIPAAFVRLD 173
BacA_C3-GxpS_A3 LOW98    Seq8   VSEALVLALGIR-QPKRLVAYVVA-----LADDGLATKLREHLSDICCVEMIPAAFVRLD 173
BacA_C3-GxpS_A3 LOW99    Seq9   VSEALVLALGVR-VDKRLVAYVVA-----LADDGLATKLREHLSDICCDAGIPAAFVRLD 173
BacA_C3-GxpS_A3 LOW100   Seq10  VSEALVLALGVR-QPKRLVAYVVA-----LADDGLATKLREHLSDICCDERIPAAFVRLD 173
                                           Interface A5                    Interface A6

                                       190
GxpS_A3-T3               Seq1   AFPLTPNGKL                                                  183
XtpS_A3                  Seq1   AFPLTPNGKL                                                  184
XtpS_C3-GxpS_A3 HIGH1    Seq2   AFPLTPNGKL                                                  183
XtpS_C3-GxpS_A3 HIGH2    Seq3   AFPLSPNGKL                                                  183
XtpS_C3-GxpS_A3 HIGH3    Seq4   AFPLTPDGKL                                                  183
XtpS_C3-GxpS_A3 MEDIUM1  Seq5   AFPLTYHGKL                                                  183
XtpS_C3-GxpS_A3 MEDIUM2  Seq6   AFPLGPNTKL                                                  183
XtpS_C3-GxpS_A3 MEDIUM3  Seq7   AFPLTPKGHL                                                  183
XtpS_C3-GxpS_A3 LOW99    Seq8   AFPLTPNRCL                                                  183
XtpS_C3-GxpS_A3 LOW100   Seq9   AFPLTCNRKL                                                  183
AmbS_A3                  Seq1   AFPLTPNGKL                                                  183
AmbS_C3-GxpS_A3 HIGH1    Seq2   AFPLTPNGKL                                                  183
AmbS_C3-GxpS_A3 HIGH3    Seq3   AFPLTPDGKL                                                  183
BicA_A3                  Seq1   SFPQTPNGKL                                                  189
BicA_C3-GxpS_A3 HIGH1    Seq2   AFPLTPNGKL                                                  183
BicA_C3-GxpS_A3 HIGH2    Seq3   AFPLSPNGKL                                                  183
BicA_C3-GxpS_A3 HIGH3    Seq4   AFPLTPDGKL                                                  183
BacA_A3                  Seq1   GLPLTVNGKV                                                  180
BacA_C3-GxpS_A3 HIGH1    Seq2   AFPLTVNGKL                                                  183
BacA_C3-GxpS_A3 HIGH2    Seq3   AFPLTVQGKL                                                  183
BacA_C3-GxpS_A3 HIGH3    Seq4   AFPLTVHGKL                                                  183
BacA_C3-GxpS_A3 MEDIUM1  Seq5   AFPLTINGCL                                                  183
```

```
BacA_C3-GxpS_A3 MEDIUM2 Seq6  AFPLNPPGKL                                  183
BacA_C3-GxpS_A3 MEDIUM3 Seq7  AFPLTRNHKL                                  183
BacA_C3-GxpS_A3 LOW98   Seq8  AFPLTPNECL                                  183
BacA_C3-GxpS_A3 LOW99   Seq9  AFPLTPNKCL                                  183
BacA_C3-GxpS_A3 LOW100  Seq10 AFPLTPNRCL                                  183
                              Interface A7
```

Table S11. Amino acid composition of the interface-forming residues according to TabS10.

amino acid	[%]																			
	A	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**K**	**L**	**M**	**N**	**P**	**Q**	**R**	**S**	**T**	**V**	**W**	**Y**
Consensus	*9.3*	*1.3*	*7.0*	*6.4*	*3.4*	*9.4*	*1.6*	*4.1*	*2.9*	*12.2*	*1.2*	*4.1*	*6.7*	*2.2*	*7.8*	*3.2*	*3.2*	*9.1*	*0.3*	*4.4*
GxpS_A3	9.3	0.0	7.7	6.6	3.3	9.8	1.6	4.4	2.7	13.1	1.1	3.8	7.1	2.2	8.2	2.7	3.3	8.7	0.0	4.4
XtpS_A3	9.8	0.5	7.6	5.4	3.3	8.7	1.1	6.0	2.2	13.0	1.1	5.4	7.1	2.7	8.7	2.2	3.3	7.6	0.0	4.3
XtpS_C3-GxpS_A3	9.8	0.0	7.7	6.0	3.3	10.4	1.6	3.3	2.7	14.2	1.1	4.4	6.6	2.7	7.7	2.7	2.7	8.7	0.0	4.4
XtpS_C3-GxpS_A3	9.8	0.0	7.1	6.0	3.8	10.4	1.6	4.4	2.2	13.1	1.1	4.9	6.6	2.2	7.7	3.3	2.2	9.2	0.0	4.4
XtpS_C3-GxpS_A3	9.3	0.0	6.6	7.1	3.3	10.4	2.7	3.3	2.7	13.1	1.1	4.4	6.6	1.6	7.7	2.7	2.7	10.4	0.0	4.4
XtpS_C3-GxpS_A3	8.7	0.5	7.1	6.0	3.3	9.2	2.2	4.4	2.7	13.7	1.1	4.4	6.0	1.6	7.1	4.4	3.8	8.2	1.1	4.4
XtpS_C3-GxpS_A3	9.8	0.5	7.1	6.0	3.8	9.8	1.1	3.8	4.4	12.0	0.5	4.4	6.0	1.6	6.0	3.8	4.9	8.7	0.5	4.9
XtpS_C3-GxpS_A3	9.3	0.5	7.7	7.1	3.8	9.3	2.2	4.4	2.7	12.6	1.1	3.8	6.6	2.7	7.1	2.7	3.3	8.7	0.0	4.4
XtpS_C3-GxpS_A3	9.8	5.5	7.1	6.0	3.3	7.1	1.6	4.4	2.2	10.9	1.1	3.8	6.6	1.6	8.7	2.7	3.3	9.8	0.5	3.8
XtpS_C3-GxpS_A3	9.3	6.0	7.1	6.6	3.3	7.7	1.6	3.8	1.6	10.9	0.5	3.8	7.7	1.6	9.3	2.7	3.3	8.2	0.5	4.4
AmbS_A3	9.8	0.5	5.5	6.6	3.8	10.4	1.1	2.7	1.6	11.5	2.2	3.8	8.2	3.3	8.2	2.7	3.8	8.7	0.0	5.5
AmbS_C3-GxpS_A3	9.8	0.0	7.7	6.0	3.3	10.4	1.1	4.4	2.7	12.6	1.1	3.8	7.1	2.2	8.2	2.7	3.3	9.2	0.0	4.4
AmbS_C3-GxpS_A3	9.8	0.0	7.7	5.5	3.8	10.4	1.6	3.8	3.3	12.6	1.1	3.3	7.1	2.2	8.7	2.7	2.7	9.2	0.0	4.4
BicA_A3	6.9	0.0	6.9	6.9	3.2	7.9	1.6	4.2	2.1	10.1	1.6	3.2	6.9	3.7	10.1	6.3	4.2	10.6	0.0	3.7
BicA_C3-GxpS_A3	9.8	0.0	7.7	6.0	3.3	10.4	1.6	2.7	3.3	12.0	1.6	3.8	6.6	2.2	8.2	3.3	3.3	9.2	0.0	4.9
BicA_C3-GxpS_A3	9.3	0.5	7.1	6.0	3.3	9.8	1.6	3.3	3.3	12.0	1.1	4.4	6.6	1.6	8.7	4.9	2.2	9.8	0.0	4.4
BicA_C3-GxpS_A3	9.3	0.0	7.1	6.0	3.3	10.4	2.2	3.3	3.3	12.6	1.6	3.8	6.0	2.2	8.2	3.8	3.3	9.2	0.0	4.4
BacA_A3	5.0	1.7	6.1	8.3	2.8	8.9	1.1	8.3	8.2	10.0	1.7	4.4	7.2	3.9	6.1	1.7	2.2	7.2	0.0	5.0
BacA_C3-GxpS_A3	9.8	0.5	8.2	6.0	2.7	9.8	1.1	3.3	2.7	14.2	1.6	3.8	6.6	2.7	6.6	3.3	3.3	9.3	0.0	4.4
BacA_C3-GxpS_A3	9.3	0.0	7.1	6.0	2.7	9.8	1.6	3.8	2.7	13.7	1.6	4.4	6.6	2.7	6.6	3.3	3.3	9.2	0.0	5.5
BacA_C3-GxpS_A3	10.4	0.5	7.7	6.6	2.7	9.8	2.7	4.9	3.3	13.1	1.1	4.4	6.6	1.6	6.0	3.3	2.7	8.7	0.0	3.8
BacA_C3-GxpS_A3	10.4	1.1	6.0	7.1	3.8	9.3	2.7	3.8	2.2	12.0	1.6	3.8	7.1	2.7	6.0	3.3	3.3	8.7	0.0	4.9
BacA_C3-GxpS_A3	9.3	0.5	7.1	5.5	3.8	10.4	1.1	3.8	2.2	11.5	1.1	4.4	6.0	2.7	7.1	3.8	3.8	10.4	0.5	4.9
BacA_C3-GxpS_A3	9.3	0,0	6.0	7.1	4.4	9.8	1.6	3.3	2.7	12.0	1.6	4.9	6.0	1.6	6.6	3.8	2.7	9.2	2.2	4.9
BacA_C3-GxpS_A3	9.8	4,9	4.9	6.6	3.3	7.7	1.1	5.5	3.3	10.9	1.1	3.8	6.0	1.6	9.8	2.7	3.3	9.8	0.0	3.8
BacA_C3-GxpS_A3	9.8	5,5	6.6	6.6	3.3	8.2	1.1	4.4	2.2	10.9	1.1	3.8	6.0	1.1	8.2	2.7	3.3	10.4	1.1	3.8
BacA_C3-GxpS_A3	8.7	4,4	6.0	8.2	3.3	7.1	1.6	3.8	2.7	10.9	0.5	3.8	7.7	1.6	10.4	2.7	3.3	8.7	0.5	3.8

Supplementary Figures

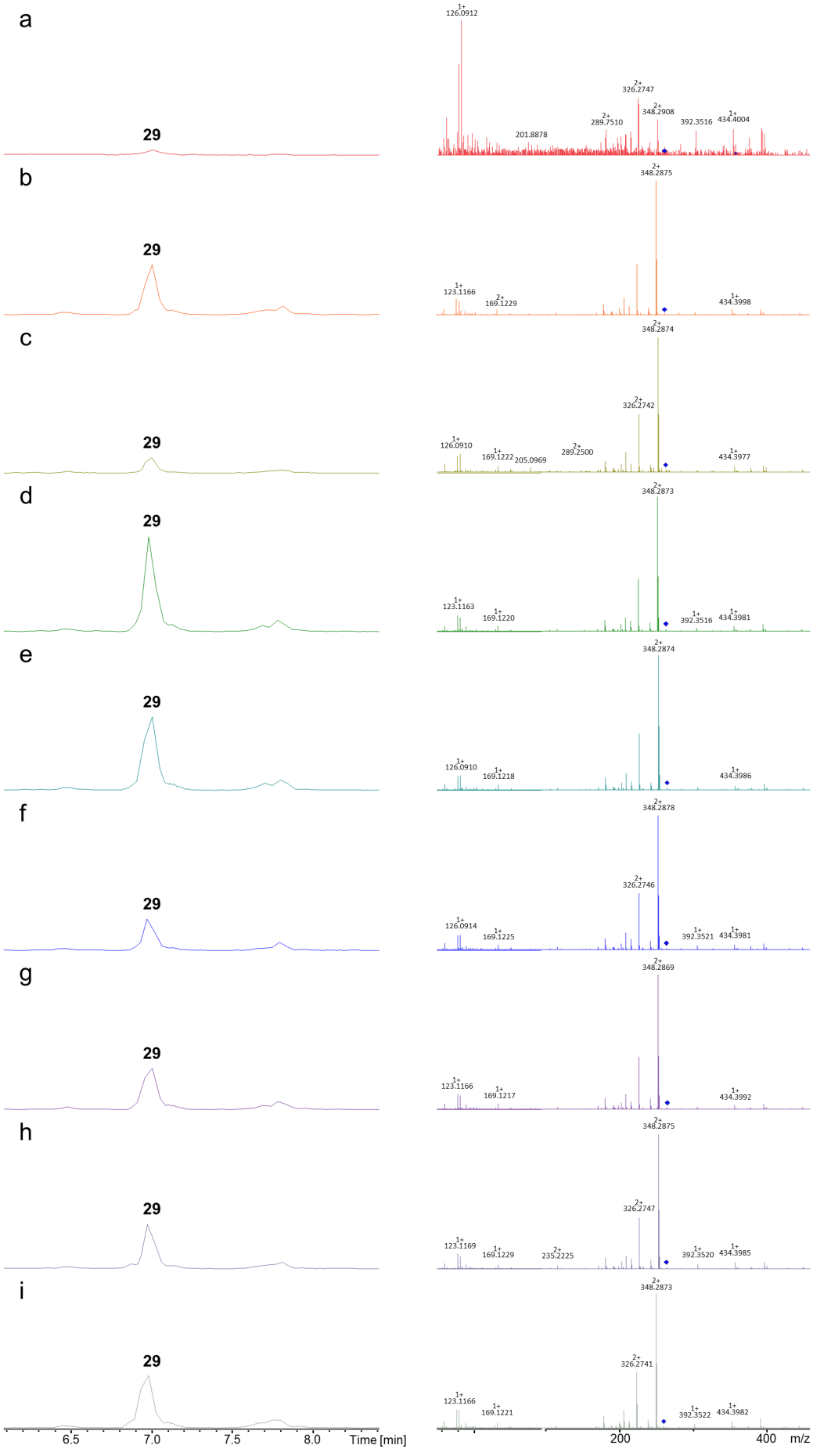

a
29
126.0912
326.2747
348.2908
289.7510
392.3516
434.4004
201.8878
b
29
348.2875
123.1166
169.1229
434.3998
c
29
348.2874
326.2742
126.0910
169.1222
205.0969
289.2500
434.3977
d
29
348.2873
123.1163
169.1220
392.3516
434.3981
e
29
348.2874
126.0910
169.1218
434.3986
f
29
348.2878
326.2746
126.0914
169.1225
392.3521
434.3981
g
29
348.2869
123.1166
169.1217
434.3992
h
29
348.2875
326.2747
123.1169
169.1229
235.2225
392.3520
434.3985
i
29
348.2873
326.2741
123.1166
169.1221
392.3522
434.3982
6.5
7.0
7.5
8.0
Time [min]
200
400
m/z

Figure S1. HPLC/MS data refers to Fig. 30 (Plasmid based FclJ complementation: NRPS-13 to -21) of compound **29** produced in *Xenorhabdus szentirmaii ΔfclJ*.

(a) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-13). (b) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-14). (c) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-15). (d) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-16). (e) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-17). (f) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-18). (g) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-19). (h) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-20). (i) Extracted ion chromatogram (EIC)/MS^2 of **29** (m/z $[M+H]^{2+}$ = 357.32; NRPS-21).

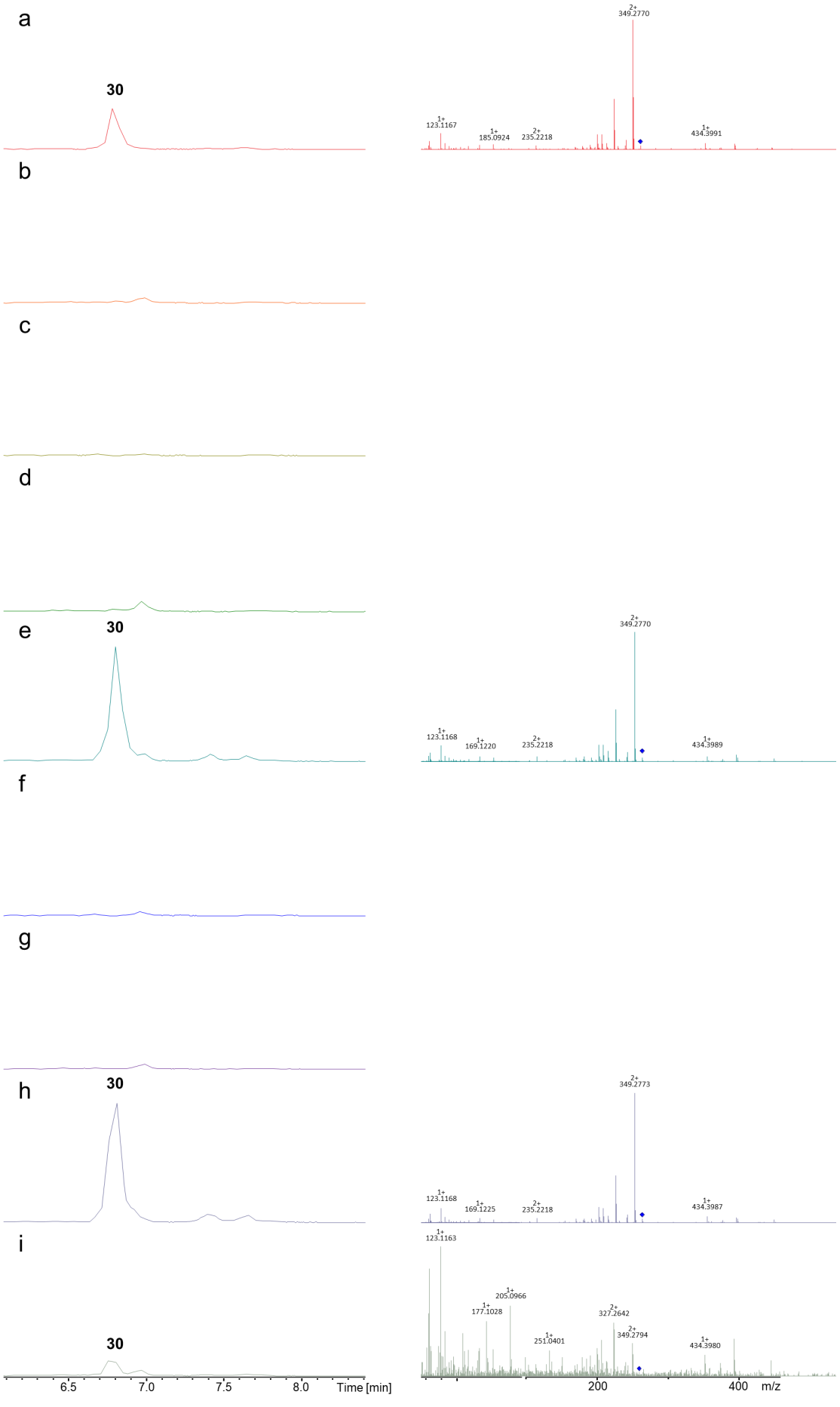

a
30
2+ 349.2770
1+ 123.1167
1+ 185.0924
2+ 235.2218
1+ 434.3991
b
c
d
e
30
2+ 349.2770
1+ 123.1168
1+ 169.1220
2+ 235.2218
1+ 434.3989
f
g
h
30
2+ 349.2773
1+ 123.1168
1+ 169.1225
2+ 235.2218
1+ 434.3987
i
1+ 123.1163
1+ 205.0966
1+ 177.1028
2+ 327.2642
2+ 349.2794
1+ 251.0401
1+ 434.3980
30
6.5
7.0
7.5
8.0
Time [min]
200
400
m/z

Figure S2. HPLC/MS data refers to Fig. 30 (Plasmid based FclJ complementation: NRPS-13 to -21) of compound **30** produced in *Xenorhabdus szentirmaii ΔfclJ*.

(a) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-13). (b) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-14). (c) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-15). (d) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-16). (e) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-17). (f) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-18). (g) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-19). (h) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-20). (i) Extracted ion chromatogram (EIC)/MS^2 of **30** (m/z $[M+H]^{2+}$ = 358.32; NRPS-21).

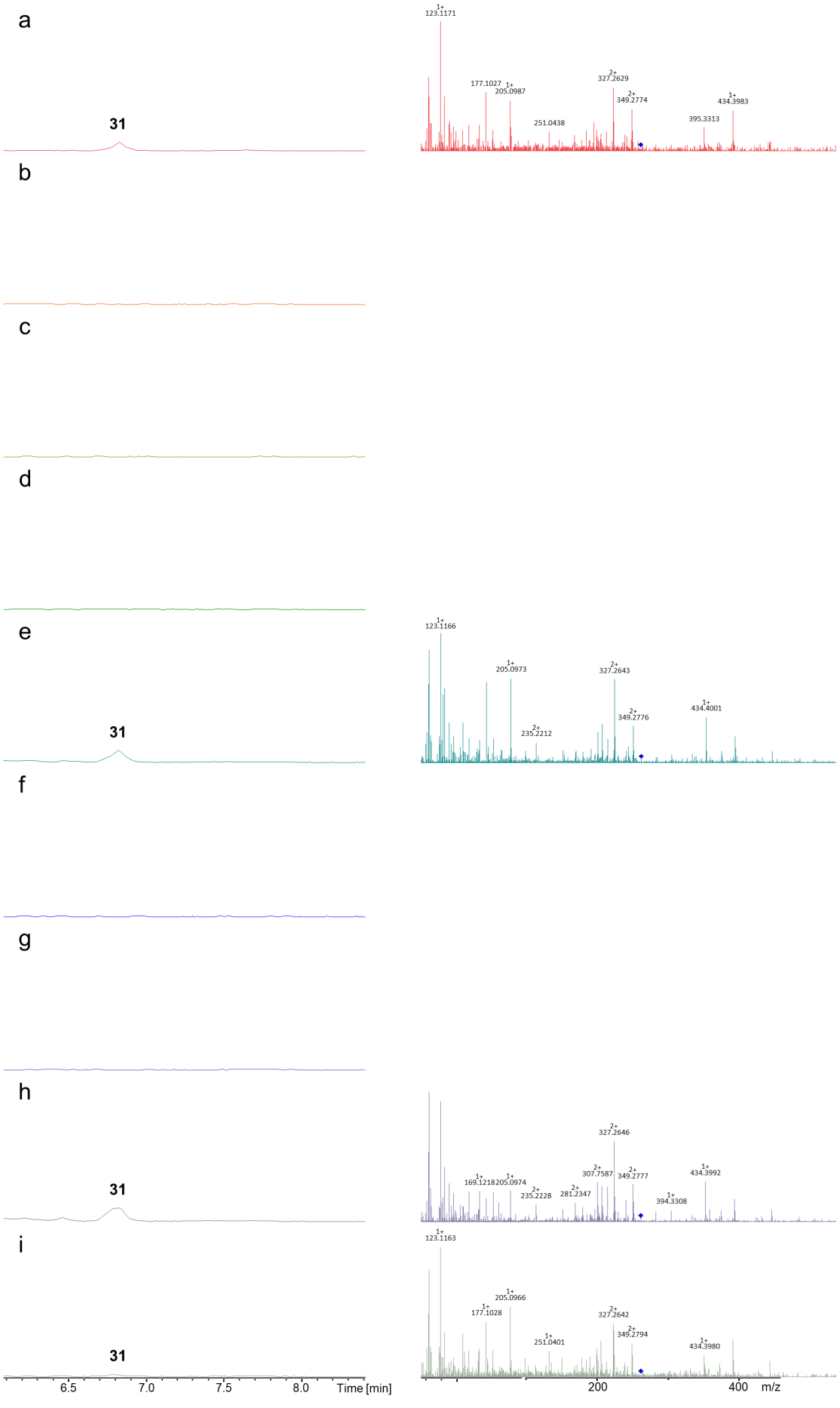

a
31
1+ 123.1171
177.1027
1+ 205.0987
251.0438
2+ 327.2629
2+ 349.2774
395.3313
1+ 434.3983
b
c
d
e
31
1+ 123.1166
1+ 205.0973
2+ 235.2212
2+ 327.2643
2+ 349.2776
1+ 434.4001
f
g
h
31
1+ 169.1218
1+ 205.0974
2+ 235.2228
2+ 281.2347
2+ 307.7587
2+ 327.2646
2+ 349.2777
1+ 394.3308
1+ 434.3992
i
31
1+ 123.1163
1+ 177.1028
1+ 205.0966
1+ 251.0401
2+ 327.2642
2+ 349.2794
1+ 434.3980
6.5
7.0
7.5
8.0
Time [min]
200
400
m/z

Figure S3. HPLC/MS data refers to Fig. 30 (Plasmid based FclJ complementation: NRPS-13 to -21) of compound **31** produced in *Xenorhabdus szentirmaii ΔfclJ*.

(a) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-13). (b) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-14). (c) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-15). (d) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-16). (e) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-17). (f) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-18). (g) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-19). (h) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-20). (i) Extracted ion chromatogram (EIC)/MS^2 of **31** (m/z $[M+H]^{2+}$ = 359.31; NRPS-21).

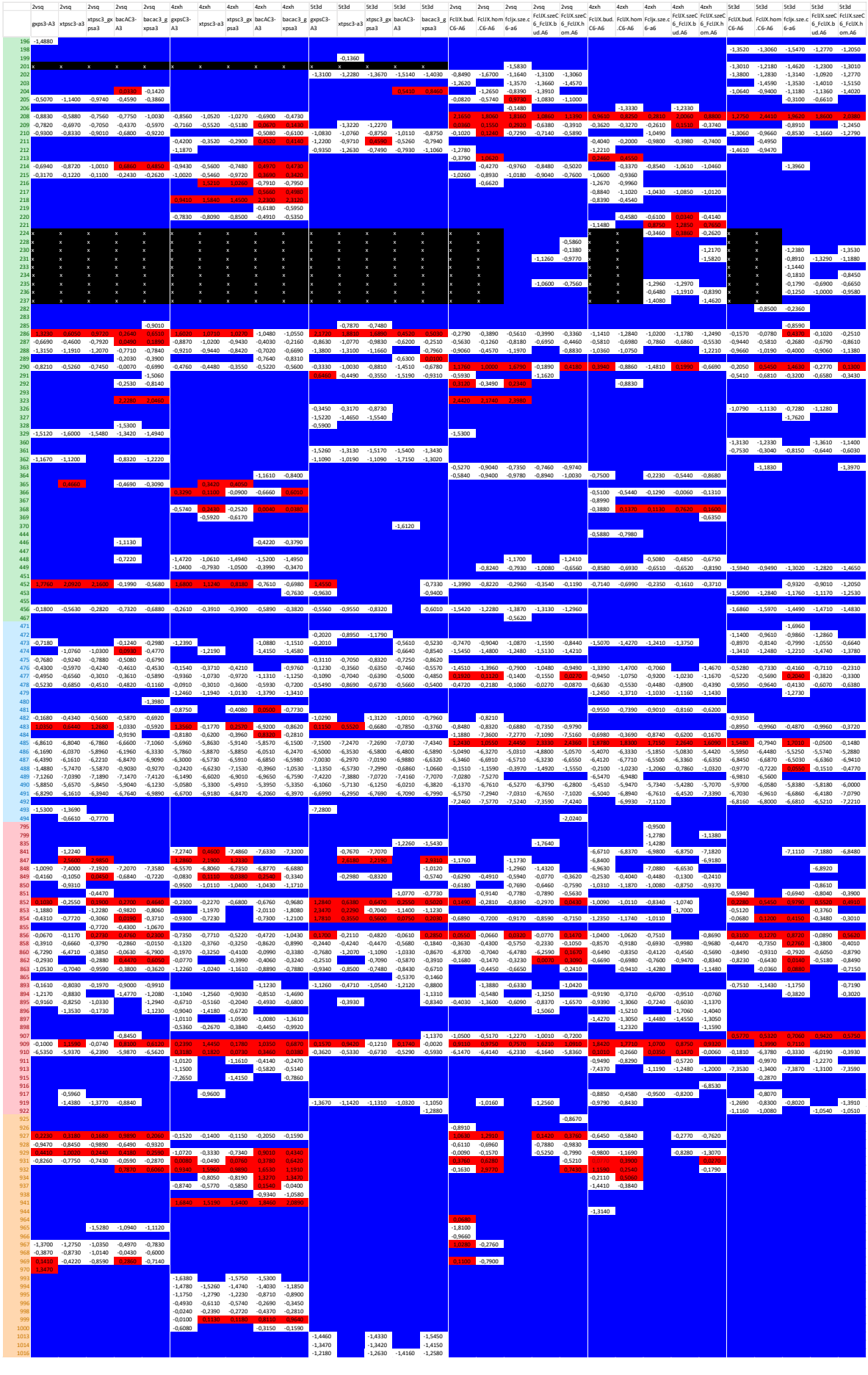


Figure S4. HSPred analysis. Related to Fig. 31 only showing the interface forming residues with highlighted hotspot residues (red), non-hotspot residues (white), non-interface residues (blue), and non-existing residues in the reference alignment (colourless/black) (Tab. S7). The colouring of the residue positions on the left is according to the C domain (green), C-A linker (blue), A_{Core} (red), and A_{Sub} (orange).

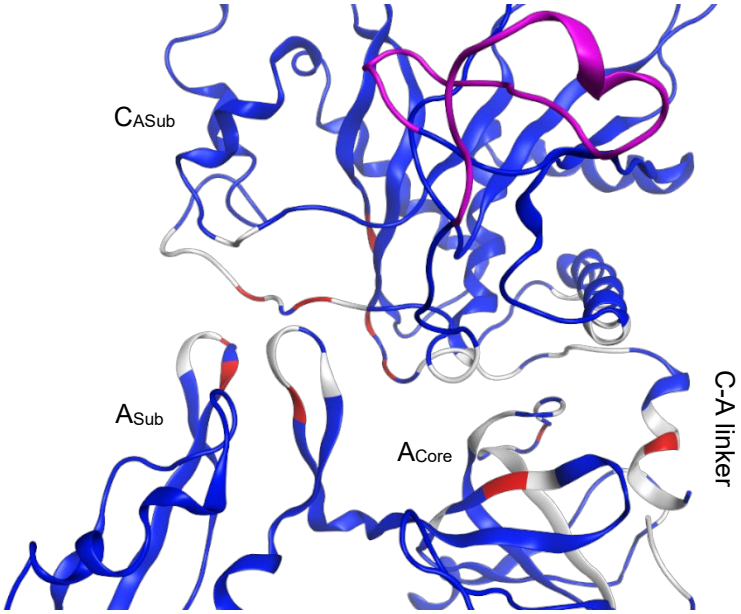


Figure S5. 2VSQ *Xsze* FclJ_C6-A6 homology model of HSPred analysis with highlighted hotspot residues (red), non-hotspot residues (white), non-interface residues (blue), and non-existing residues in the reference alignment (colourless) (Tab. S7). The novel loop 'Chain II' marked in purple.

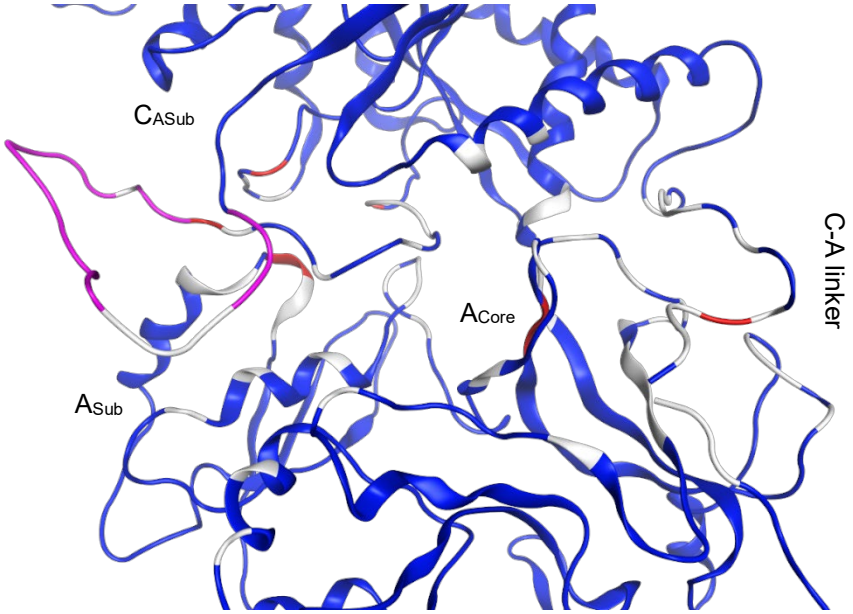


Figure S6. 4ZXH *Xsze* FclJ_C6-A6 homology model of HSPred analysis with highlighted hotspot residues (red), non-hotspot residues (white), non-interface residues (blue), and non-existing residues in the reference alignment (colourless) (Tab. S7). The novel loop 'Chain II' marked in purple.

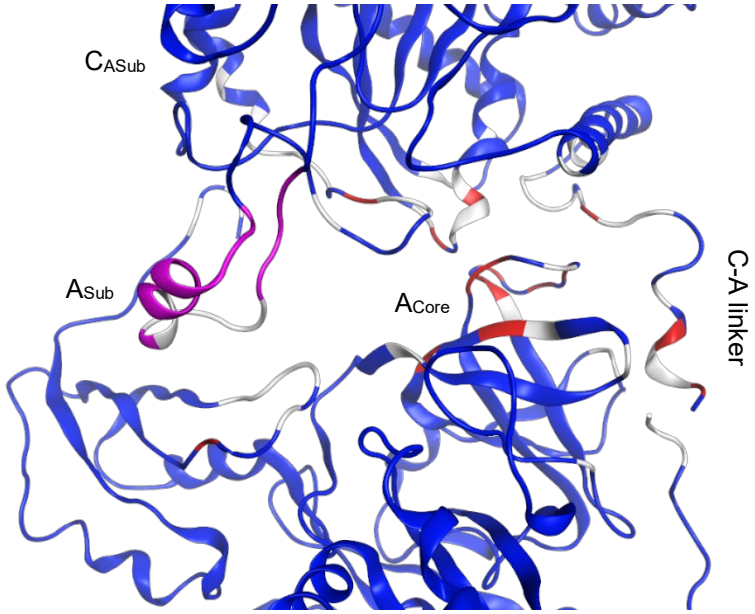


Figure S7. 5T3D *Xsze* FclJ_C6-A6 homology model of HSPred analysis with highlighted hotspot residues (red), non-hotspot residues (white), non-interface residues (blue), and non-existing residues in the reference alignment (colourless) (Tab. S7). The novel loop 'Chain II' marked in purple.

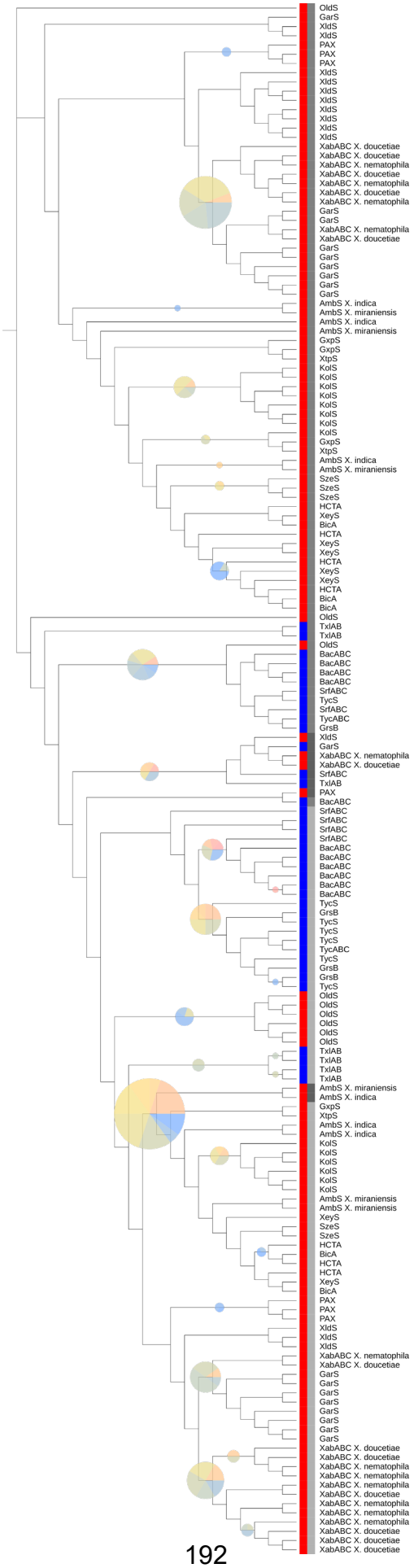

OldS
GarS
XldS
XldS
PAX
PAX
PAX
XldS
XldS
XldS
XldS
XldS
XldS
XldS
XldS
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. nematophila
GarS
GarS
XabABC X. nematophila
XabABC X. doucetiae
GarS
GarS
GarS
GarS
GarS
GarS
GarS
AmbS X. indica
AmbS X. miraniensis
AmbS X. indica
AmbS X. miraniensis
GxpS
GxpS
XtpS
KolS
KolS
KolS
KolS
KolS
KolS
KolS
KolS
KolS
GxpS
XtpS
AmbS X. indica
AmbS X. miraniensis
SzeS
SzeS
SzeS
HCTA
XeyS
BicA
HCTA
XeyS
XeyS
HCTA
XeyS
XeyS
HCTA
BicA
BicA
OldS
TxlAB
TxlAB
OldS
BacABC
BacABC
BacABC
BacABC
SrfABC
TycS
SrfABC
TycABC
GrsB
XldS
GarS
XabABC X. nematophila
XabABC X. doucetiae
SrfABC
TxlAB
PAX
BacABC
SrfABC
SrfABC
SrfABC
SrfABC
BacABC
BacABC
BacABC
BacABC
BacABC
BacABC
TycS
GrsB
TycS
TycS
TycS
TycABC
TycS
GrsB
GrsB
TycS
OldS
OldS
OldS
OldS
OldS
OldS
TxlAB
TxlAB
TxlAB
TxlAB
AmbS X. miraniensis
AmbS X. indica
GxpS
XtpS
AmbS X. indica
AmbS X. indica
KolS
KolS
KolS
KolS
KolS
KolS
AmbS X. miraniensis
AmbS X. miraniensis
XeyS
SzeS
SzeS
HCTA
BicA
HCTA
HCTA
XeyS
BicA
PAX
PAX
PAX
XldS
XldS
XldS
XabABC X. nematophila
XabABC X. doucetiae
GarS
GarS
GarS
GarS
GarS
GarS
GarS
GarS
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. nematophila
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. nematophila
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. doucetiae

Figure S8. Phylogenetic tree of selected NRPS C domains. Tree was build using RAxML of Clustal Omega alignment from selected NRPS primary sequences. The C domains' preceding A domain specificities are displayed as a pie chart (relative pie chart radius corresponds to number of included branch leaves) directly on the node branch and are color-coded according to their IP[182] as shown in the scale below for all branch leaves. Genus's affiliations and C domain types are resembled by colored strips. An interactive representation of the dendrogram can be accessed here: https://itol.embl.de/tree/14121241184571316074220 26 (accessed January 31, 2022).

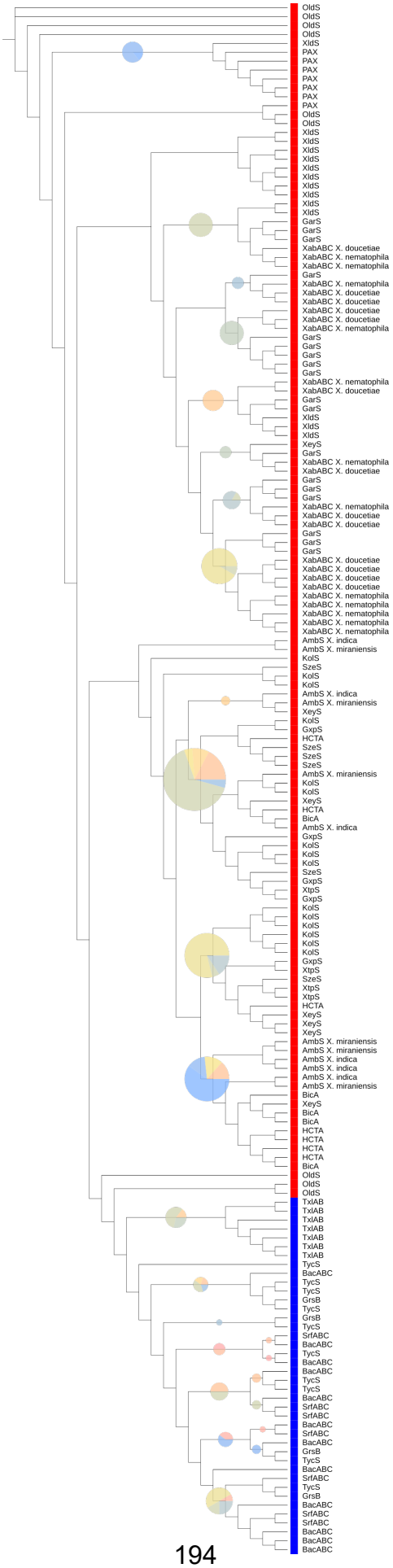

OldS
OldS
OldS
OldS
XldS
PAX
PAX
PAX
PAX
PAX
PAX
PAX
OldS
OldS
XldS
XldS
XldS
XldS
XldS
XldS
XldS
XldS
XldS
XldS
GarS
GarS
GarS
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. nematophila
GarS
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. nematophila
GarS
GarS
GarS
GarS
GarS
XabABC X. nematophila
XabABC X. doucetiae
GarS
GarS
XldS
XldS
XldS
XeyS
GarS
XabABC X. nematophila
XabABC X. doucetiae
GarS
GarS
GarS
XabABC X. nematophila
XabABC X. doucetiae
XabABC X. doucetiae
GarS
GarS
GarS
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. doucetiae
XabABC X. nematophila
XabABC X. nematophila
XabABC X. nematophila
XabABC X. nematophila
XabABC X. nematophila
AmbS X. indica
AmbS X. miraniensis
KolS
SzeS
KolS
KolS
AmbS X. indica
AmbS X. miraniensis
XeyS
KolS
GxpS
HCTA
SzeS
SzeS
SzeS
AmbS X. miraniensis
KolS
KolS
XeyS
HCTA
BicA
AmbS X. indica
GxpS
KolS
KolS
KolS
SzeS
GxpS
XtpS
GxpS
KolS
KolS
KolS
KolS
KolS
KolS
GxpS
XtpS
SzeS
XtpS
XtpS
HCTA
XeyS
XeyS
XeyS
AmbS X. miraniensis
AmbS X. miraniensis
AmbS X. indica
AmbS X. indica
AmbS X. indica
AmbS X. miraniensis
BicA
XeyS
BicA
BicA
HCTA
HCTA
HCTA
HCTA
BicA
OldS
OldS
OldS
TxlAB
TxlAB
TxlAB
TxlAB
TxlAB
TxlAB
TxlAB
TycS
BacABC
TycS
TycS
GrsB
TycS
GrsB
TycS
SrfABC
BacABC
TycS
BacABC
BacABC
TycS
TycS
BacABC
SrfABC
SrfABC
BacABC
SrfABC
BacABC
GrsB
TycS
BacABC
SrfABC
TycS
GrsB
BacABC
SrfABC
SrfABC
BacABC
BacABC
BacABC

Figure S9. Phylogenetic tree of selected NRPS A domains. Tree was build using RAxML of Clustal Omega alignment from selected NRPS primary sequences. The A domains' specificities are displayed as a pie chart (relative pie chart radius corresponds to number of included branch leaves) directly on the node branch and are color-coded according to their IP[182] as shown in the scale below for all branch leaves. Genus's affiliations are resembled by colored strips. An interactive representation of the dendrogram can be accessed here: https://itol.embl.de/tree/14121241184571116074220 25 (accessed January 31, 2022).

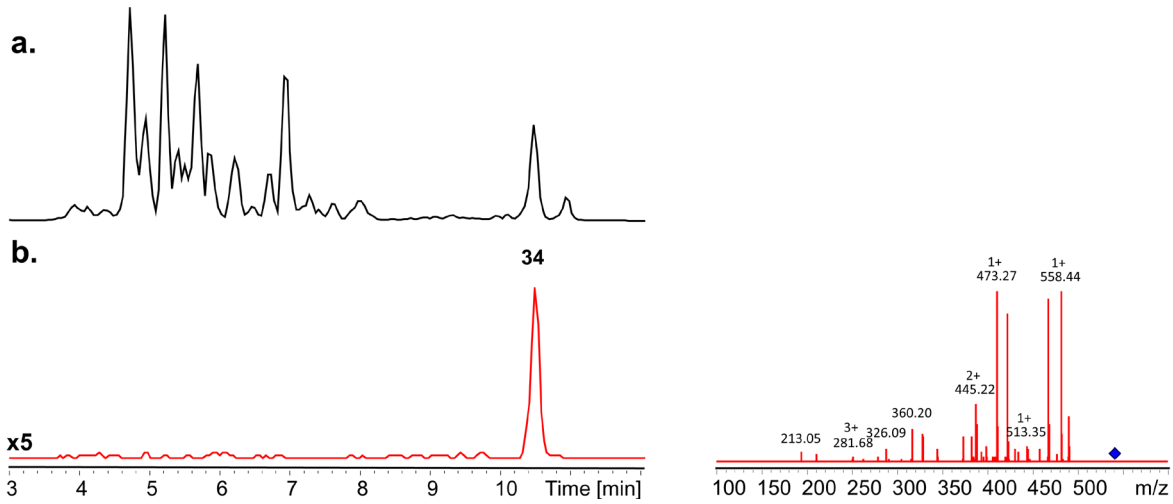


Figure S10. HPLC/MS data refers to Fig. 38a (also serves as a reference to Fig. 37) of compounds **34** produced by NRPS-23 Seq. 1 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **34** (*m/z* $[M+H]^{1+}$ = 586.40).

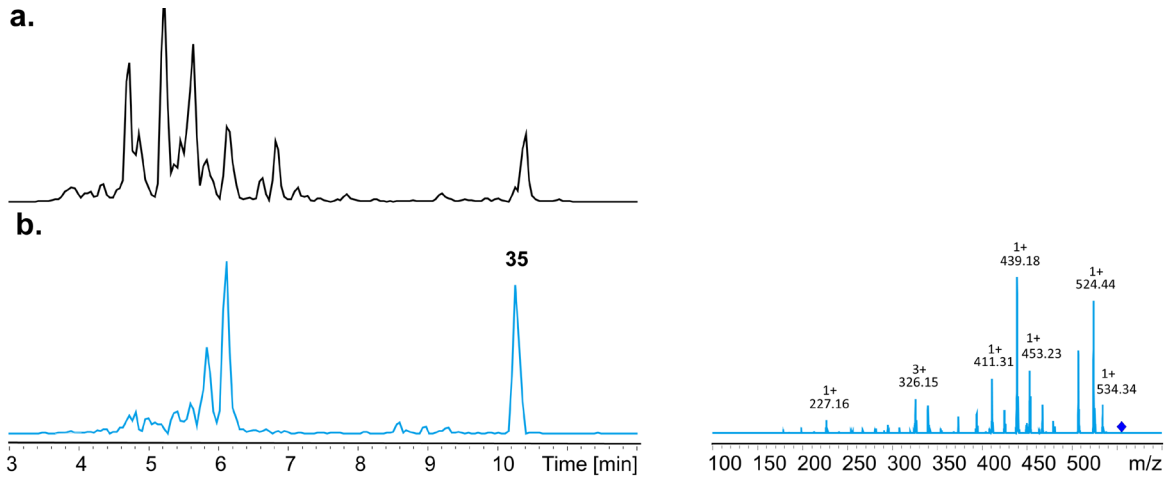


Figure S11. HPLC/MS data refers to Fig. 38a (also serves as a reference to Fig. 37) of compounds **35** produced by NRPS-24 Seq. 1 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

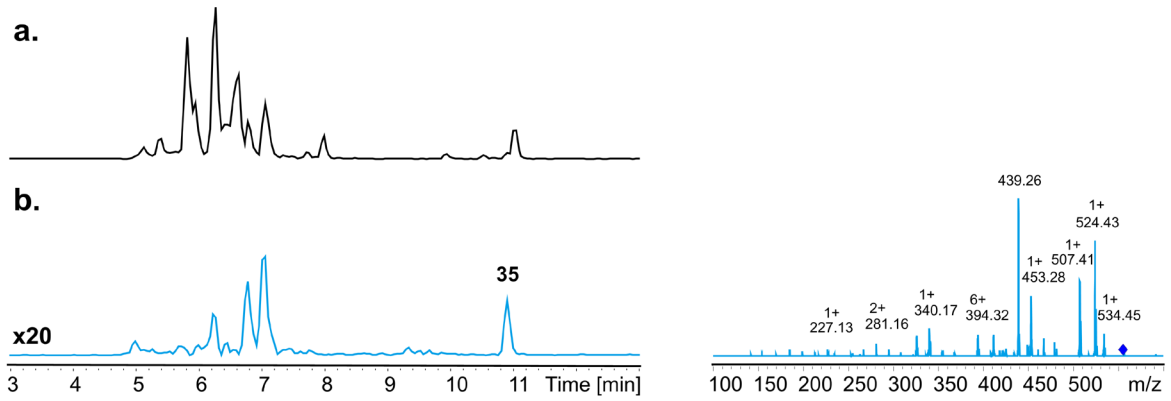


Figure S12. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 2 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

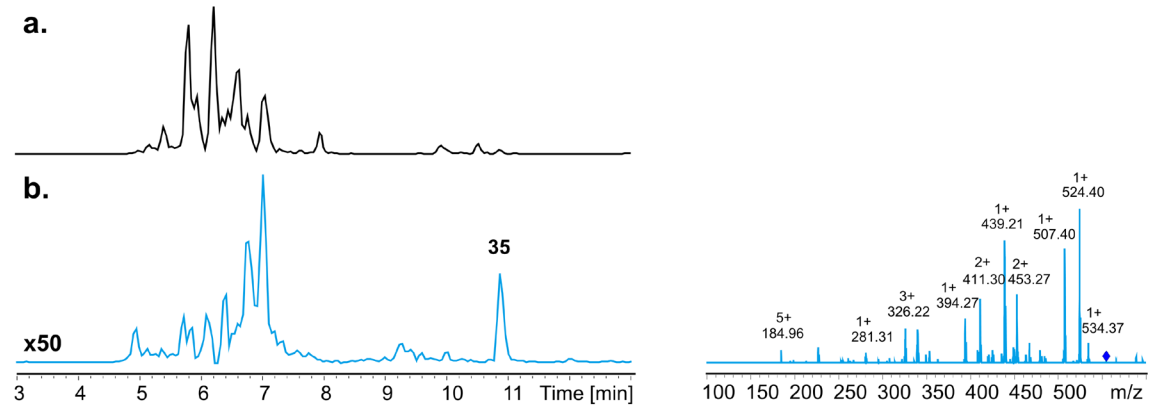


Figure S13. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 3 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

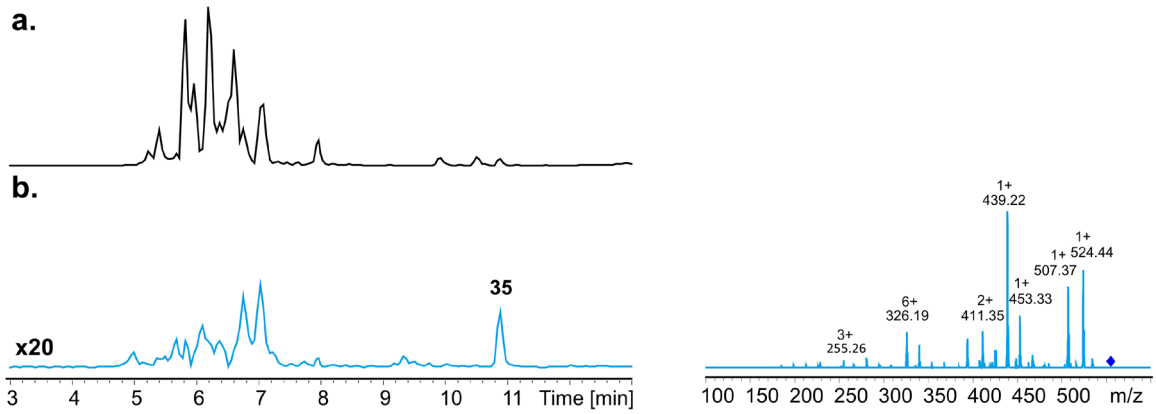


Figure S14. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 4 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

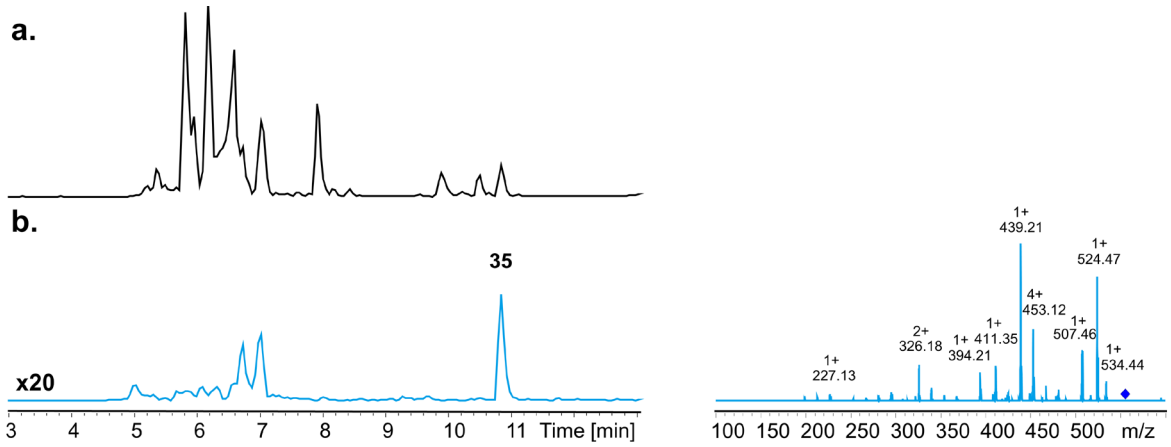


Figure S15. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 5 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

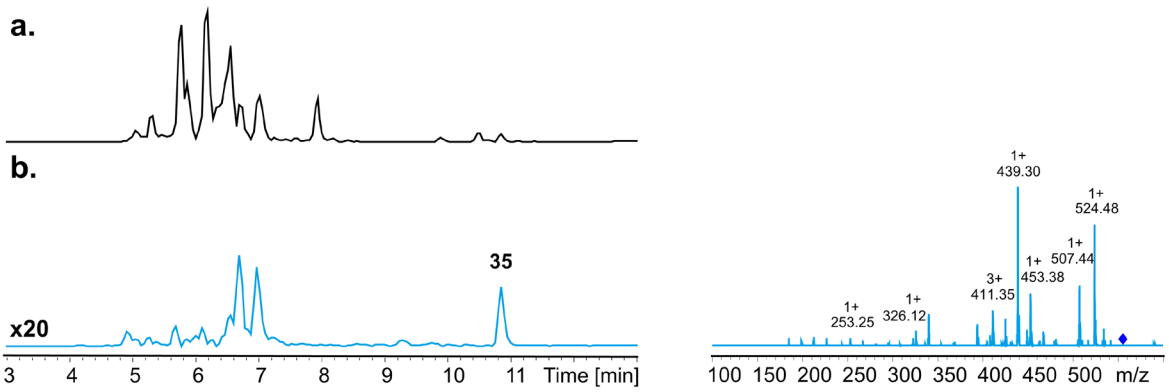


Figure S16. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 6 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

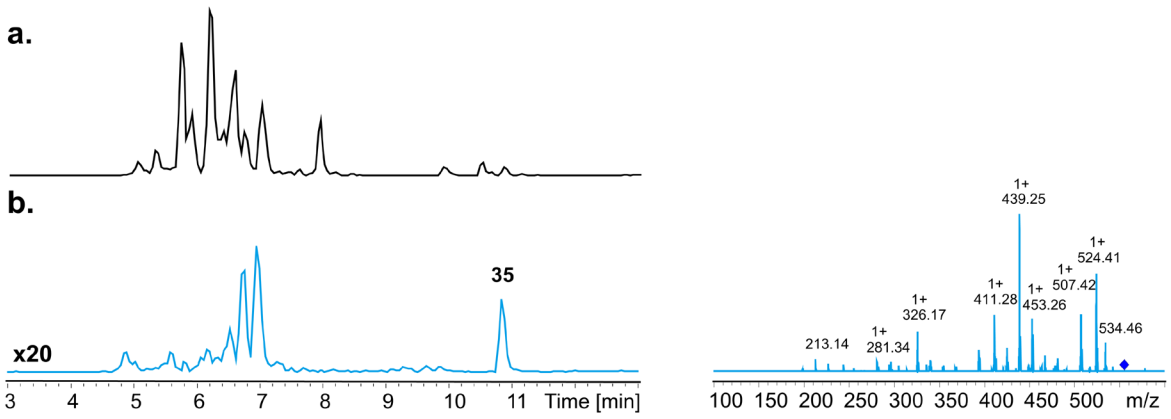


Figure S17. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 7 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

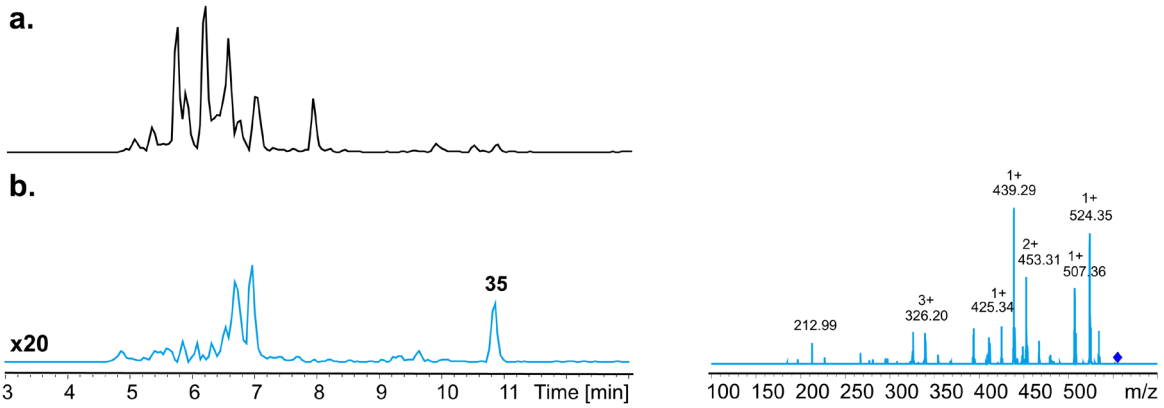


Figure S18. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 8 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

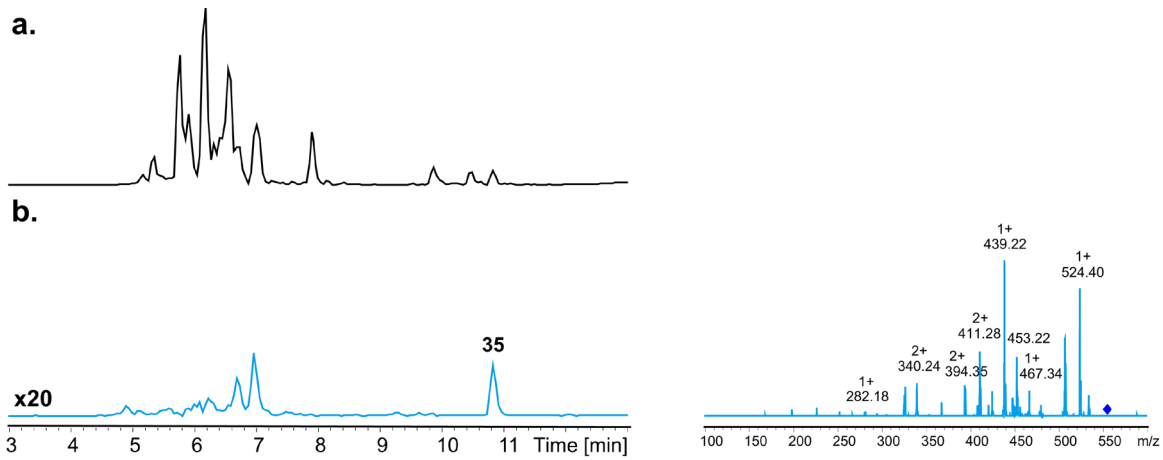


Figure S19. HPLC/MS data refers to Fig. 38a of compounds **35** produced by NRPS-24 Seq. 9 in *E. coli* DH10B::*mtaA*.(a) Base peak chromatogram (BPC) (b) EIC of **35** (*m/z* $[M+H]^{1+}$ = 552.41).

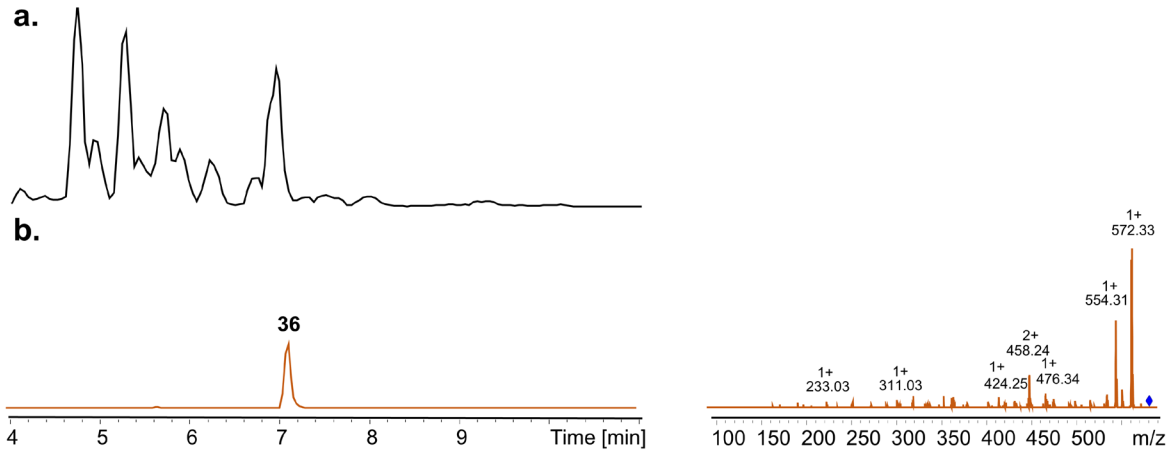


Figure S20. HPLC/MS data refers to Fig. 38a (also serves as a reference to Fig. 37) of compounds **36** produced by NRPS-25 Seq. 1 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **36** (*m/z* $[M+H]^{1+}$ = 589.33).

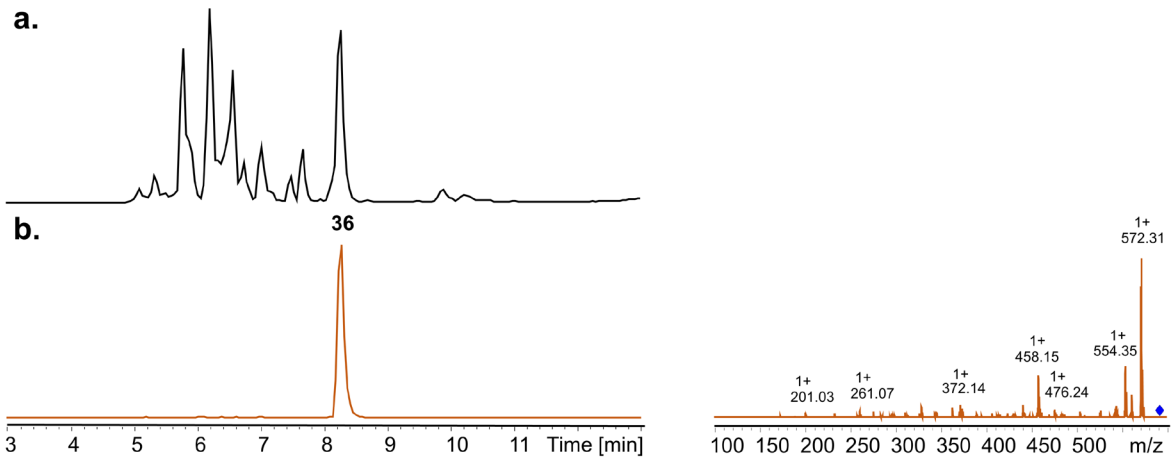


Figure S21. HPLC/MS data refers to Fig. 38a of compounds **36** produced by NRPS-25 Seq. 2 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **36** (*m/z* $[M+H]^{1+}$ = 589.33).

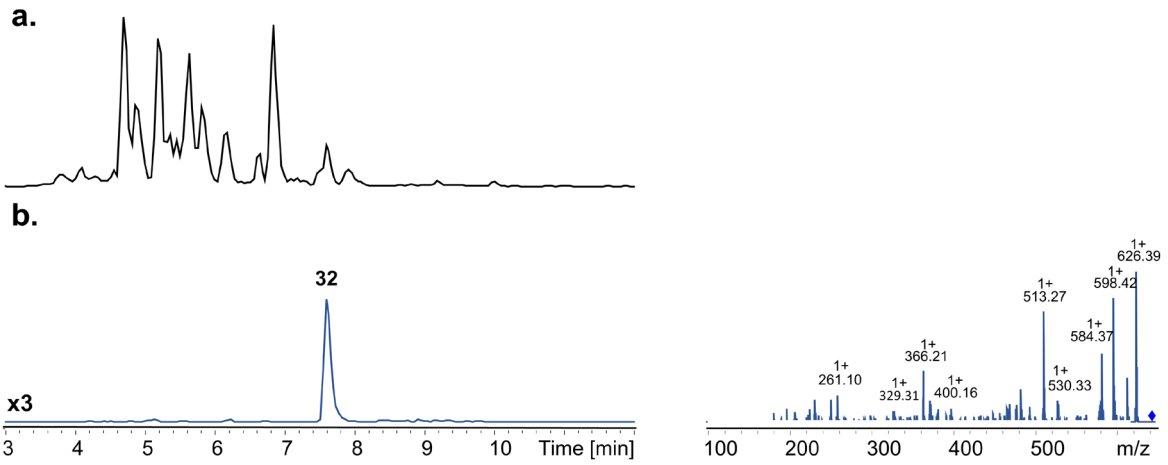


Figure S22. HPLC/MS data refers to Fig. 38a (also serves as a reference to Fig. 37) of compounds **32** produced by NRPS-26 Seq. 1 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **32** (*m/z* $[M+H]^{1+}$ = 643.43).

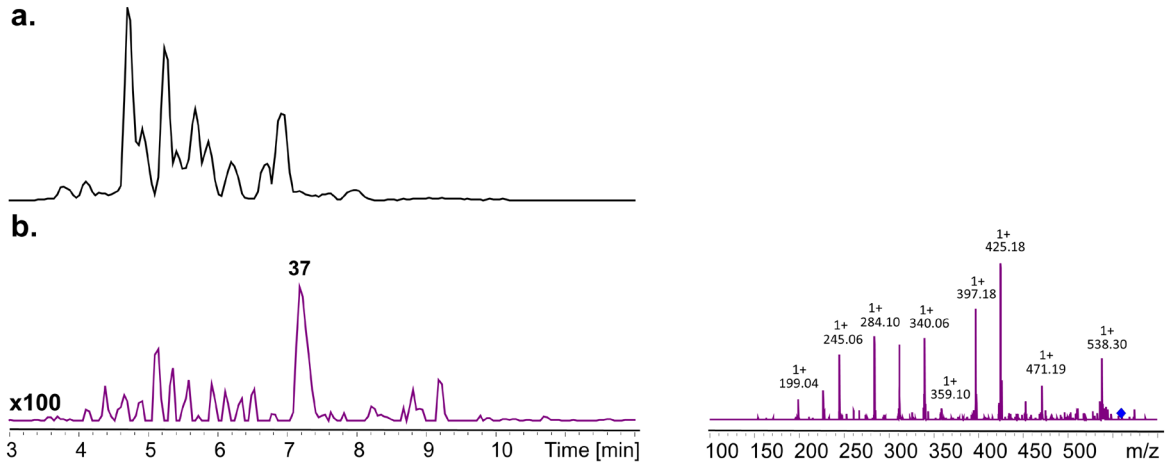


Figure S23. HPLC/MS data refers to Fig. 38a (also serves as a reference to Fig. 37) of compound **37** produced by NRPS-27 Seq. 1 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

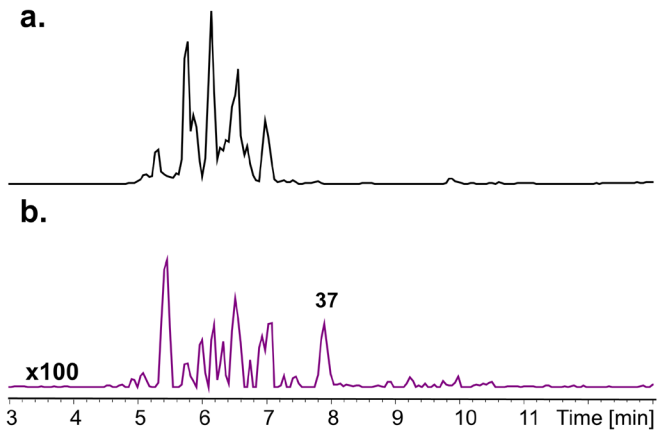


Figure S24. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 2 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

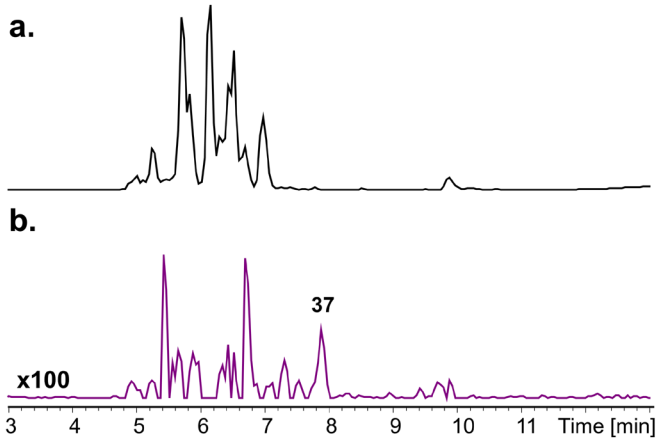


Figure S25. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 3 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

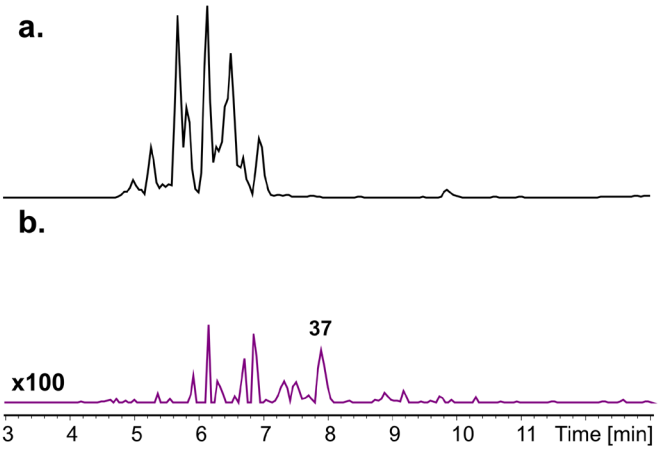


Figure S26. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 4 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

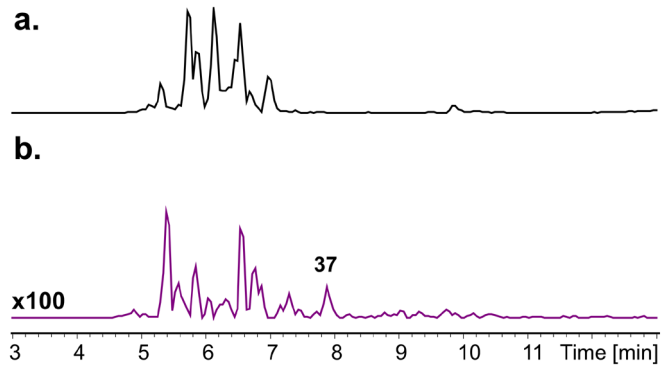


Figure S27. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 5 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

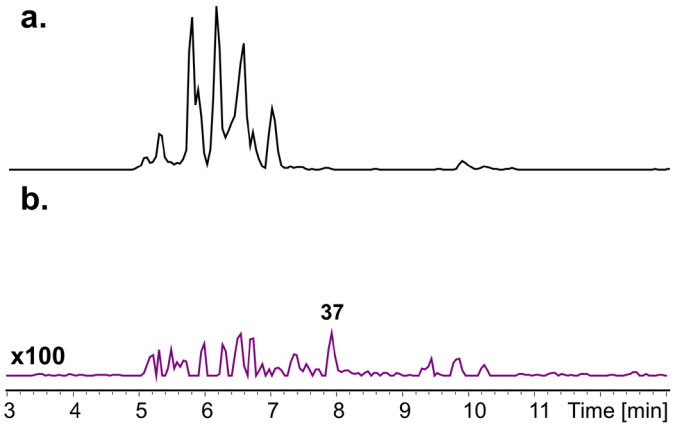


Figure S28. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 6 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

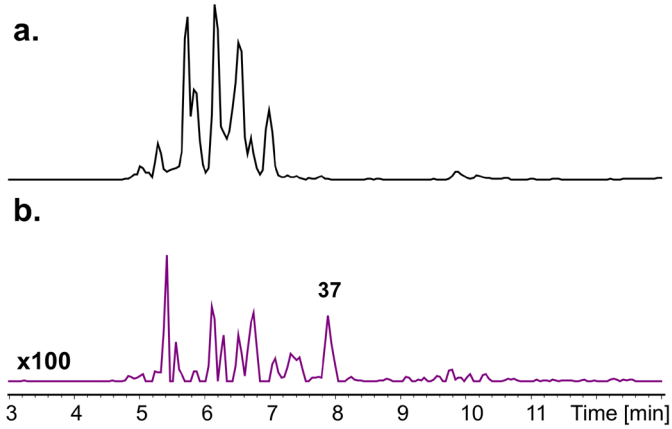


Figure S29. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 7 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

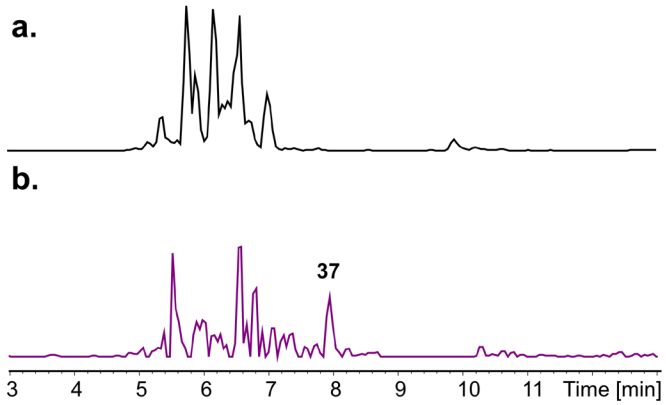


Figure S30. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 8 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

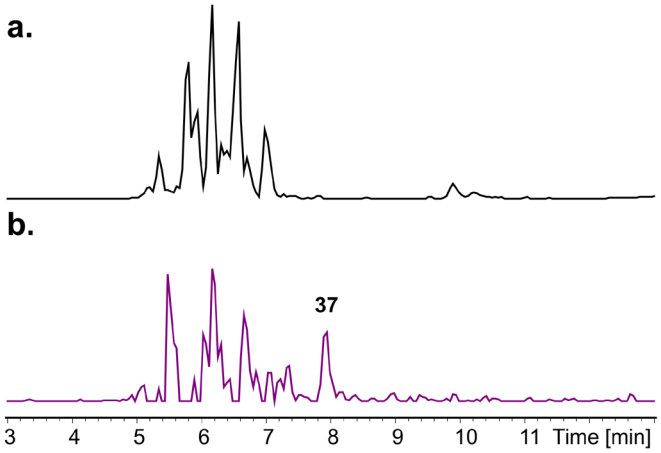


Figure S31. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 9 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

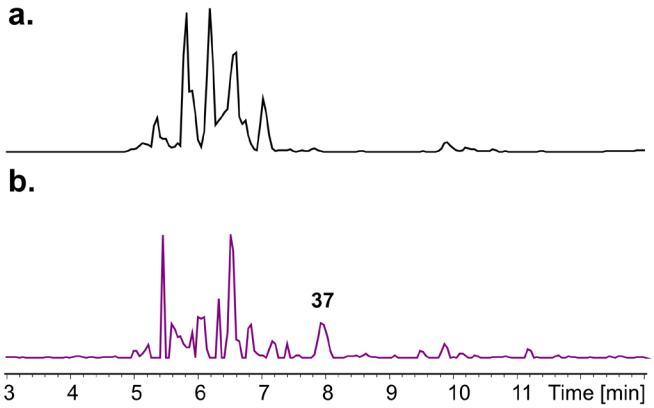


Figure S32. HPLC/MS data refers to Fig. 38a of compound **37** produced by NRPS-27 Seq. 10 in *E. coli* DH10B::*mtaA*.(a) BPC (b) EIC of **37** (*m/z* $[M+H]^{1+}$ = 556.35).

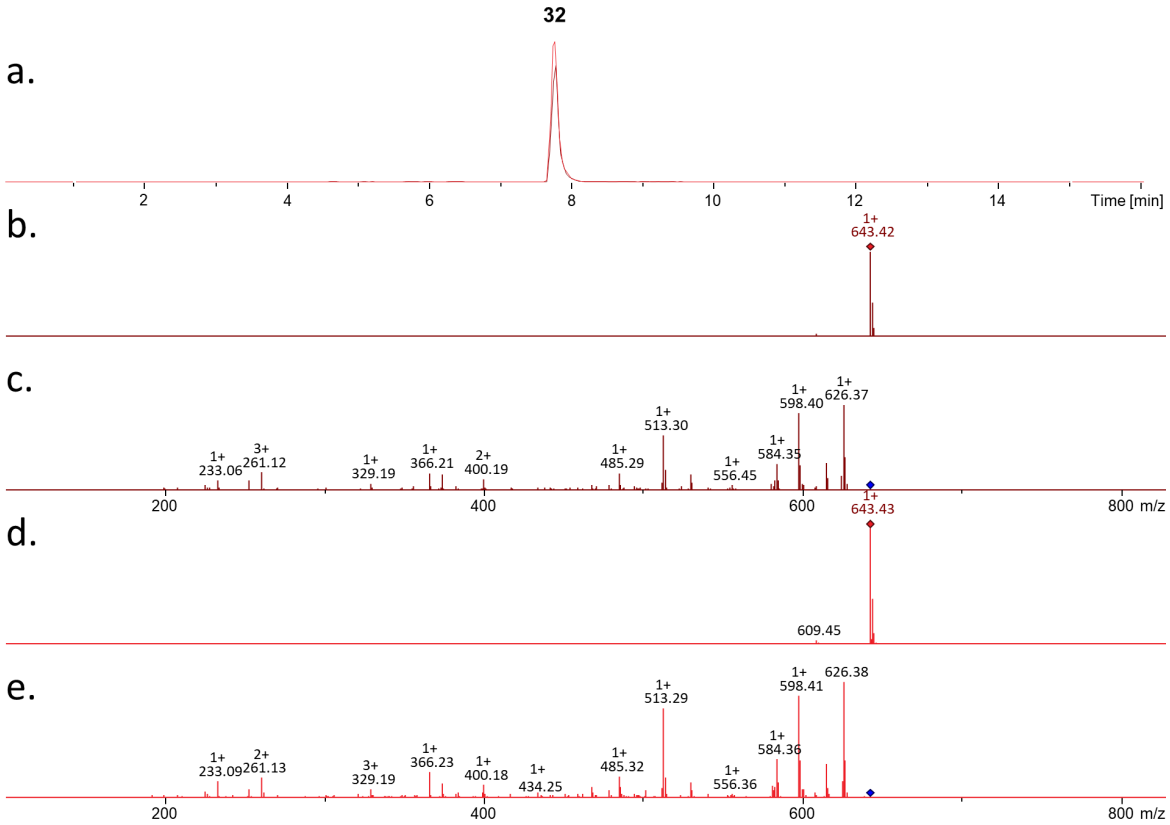


Figure S33. HPLC/MS data refers to Fig. 38c of compound **32** by NRPS-28 (dark red) & ↱NRPS-28 (red) in *E. coli* DH10B::*mtaA*.

(a) Extracted ion chromatogram (EIC) of **32** (m/z $[M+H]^{1+}$ = 643.43) of NRPS-28 & ↱NRPS-28. (b) HPLC/MS^1 data of **32** (m/z $[M+H]^{1+}$ = 643.43; NRPS-28). (c) HPLC/MS^2 data of **3326** (m/z $[M+H]^{1+}$ = 643.43; NRPS-28). (d) HPLC/MS^1 data of **32** (m/z $[M+H]^{1+}$ = 643.43; ↱NRPS-28). (e) HPLC/MS^2 data of **32** (m/z $[M+H]^{1+}$ = 643.43; ↱NRPS-28).

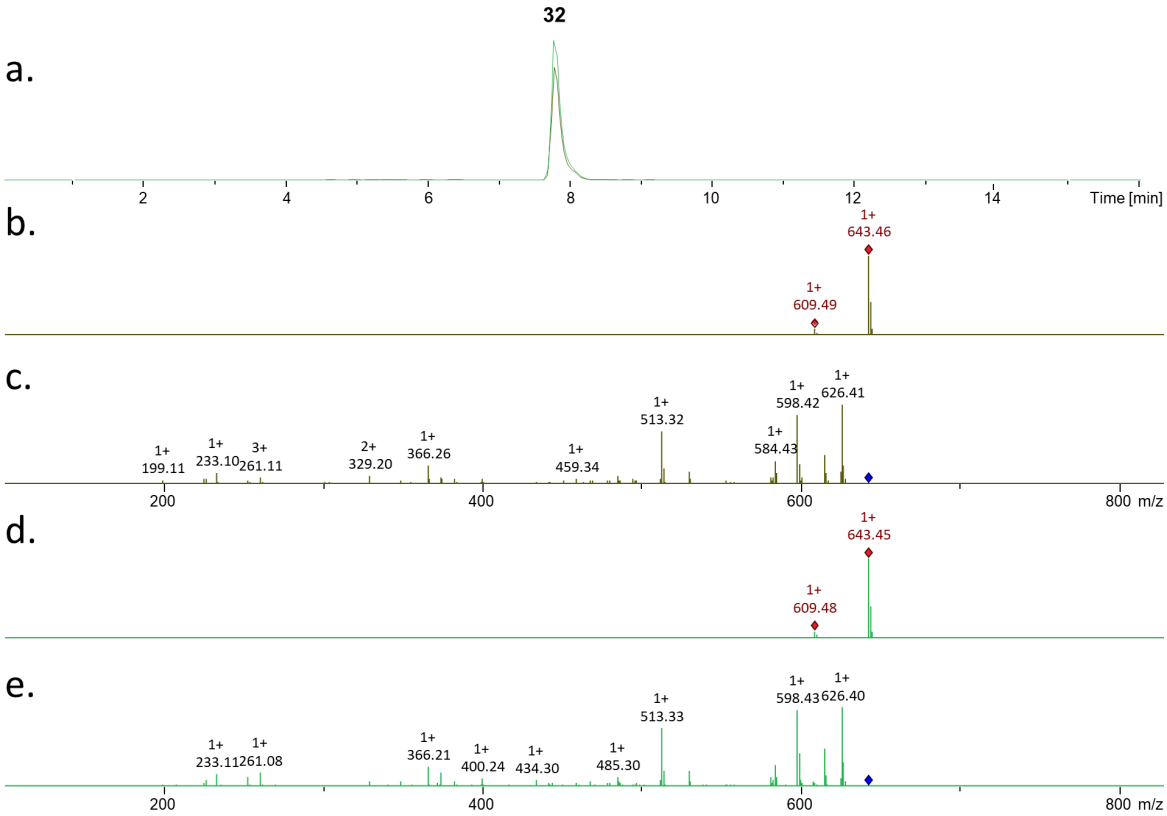


Figure S34. HPLC/MS data refers to Fig. 38c of compound **32** by NRPS-29 (dark green) & ↱NRPS-29 (green) in *E. coli* DH10B::*mtaA*.

(a) Extracted ion chromatogram (EIC) of **32** (m/z $[M+H]^{1+}$ = 643.43) of NRPS-29 & ↱NRPS-29. (b) HPLC/MS^1 data of **32** (m/z $[M+H]^{1+}$ = 643.43; NRPS-29). (c) HPLC/MS^2 data of **32** (m/z $[M+H]^{1+}$ = 643.43; NRPS-29). (d) HPLC/MS^1 data of **32** (m/z $[M+H]^{1+}$ = 643.43; ↱NRPS-29). (e) HPLC/MS^2 data of **32** (m/z $[M+H]^{1+}$ = 643.43; ↱NRPS-29).

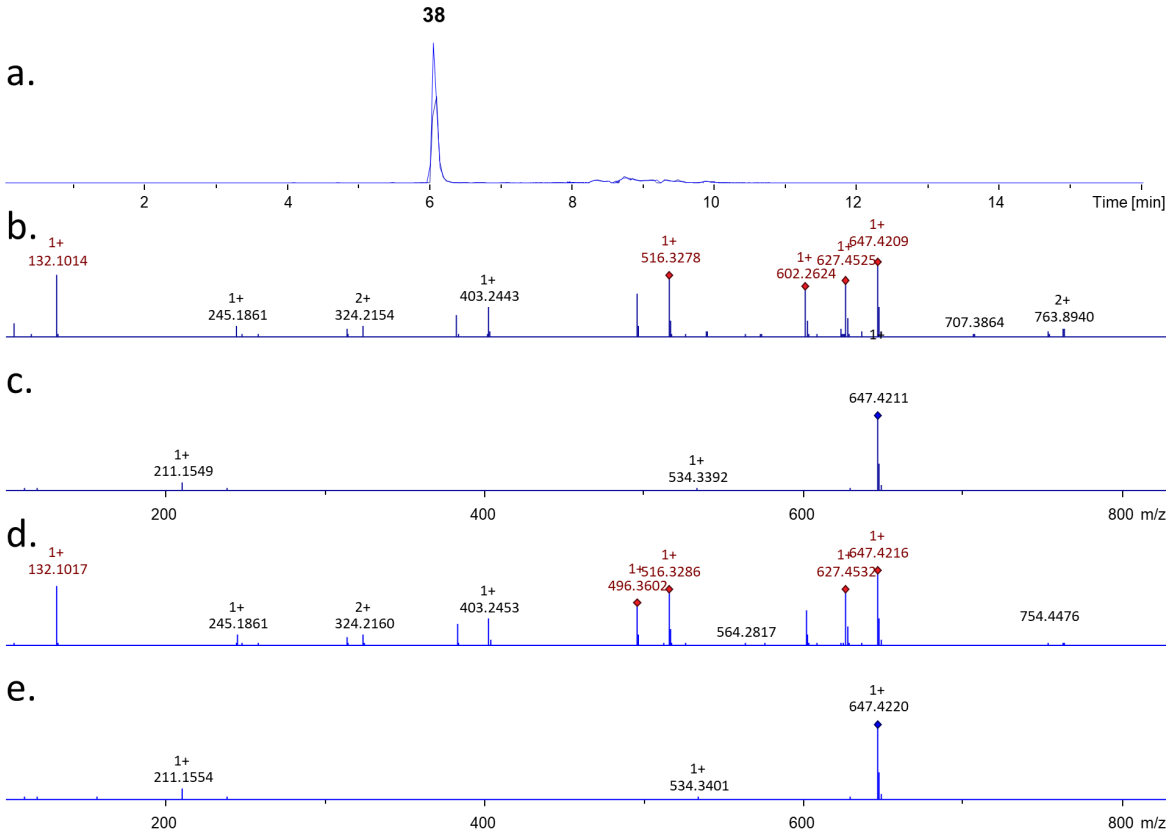


Figure S35. HPLC/MS data refers to Fig. 38c of compound **38** by NRPS-30 (dark blue) & ↱NRPS-30 (blue) in *E. coli* DH10B::*mtaA*.

(a) Extracted ion chromatogram (EIC) of **38** (m/z $[M+H]^{1+}$ = 647.42) of NRPS-30 & ↱NRPS-30. (b) HPLC/MS^1 data of **38** (m/z $[M+H]^{1+}$ = 647.42; NRPS-30). (c) HPLC/MS^2 data of **38** (m/z $[M+H]^{1+}$ = 647.42; NRPS-30). (d) HPLC/MS^1 data of **38** (m/z $[M+H]^{1+}$ = 647.42; ↱NRPS-30). (e) HPLC/MS^2 data of **38** (m/z $[M+H]^{1+}$ = 647.42; ↱NRPS-30).

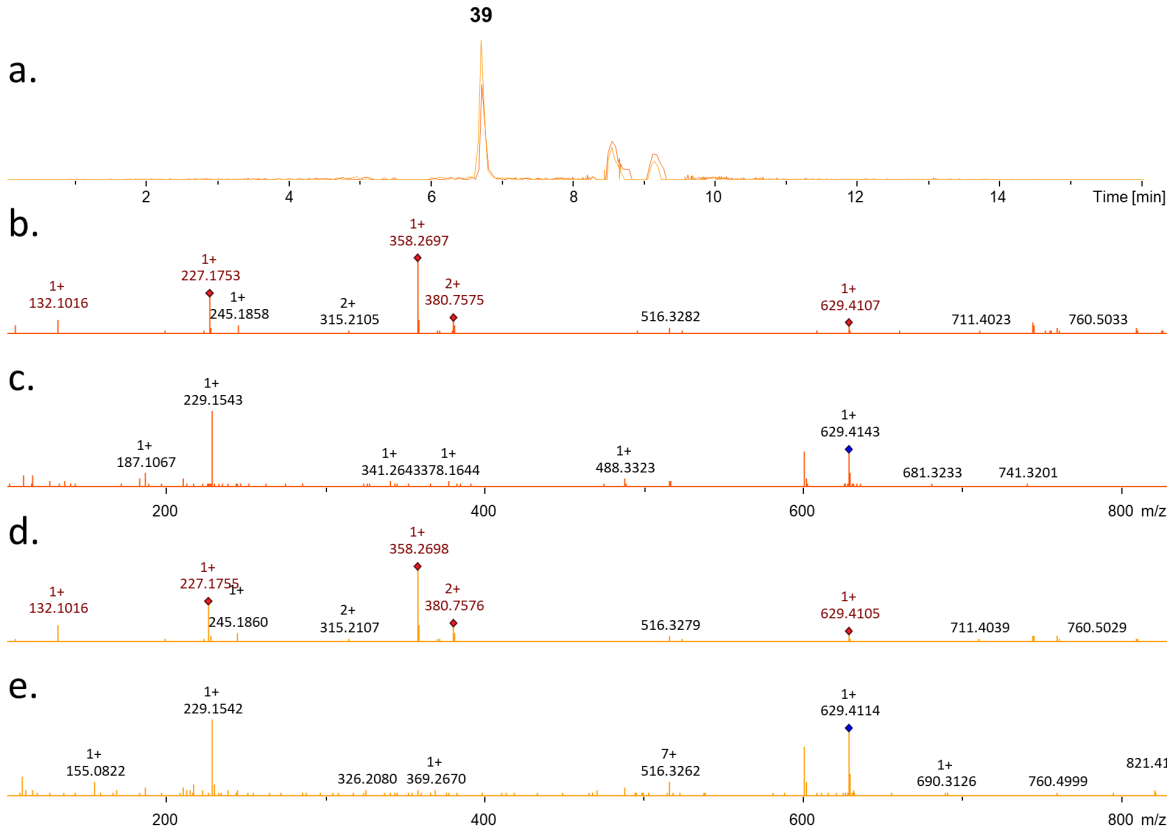


Figure S36. HPLC/MS data refers to Fig. 38c of compound **39** by NRPS-31 (dark yellow) & ↱NRPS-31 (yellow) in *E. coli* DH10B::*mtaA*.

(a) Extracted ion chromatogram (EIC) of **39** (m/z $[M+H]^{1+}$ = 670.51) of NRPS-31 & ↱NRPS-31. (b) HPLC/MS^1 data of **39** (m/z $[M+H]^{1+}$ = 670.51; NRPS-31). (c) HPLC/MS^2 data of **39** (m/z $[M+H]^{1+}$ = 670.51; NRPS-31). (d) HPLC/MS^1 data of **39** (m/z $[M+H]^{1+}$ = 670.51; ↱NRPS-31). (e) HPLC/MS^2 data of **39** (m/z $[M+H]^{1+}$ = 670.51; ↱NRPS-31).

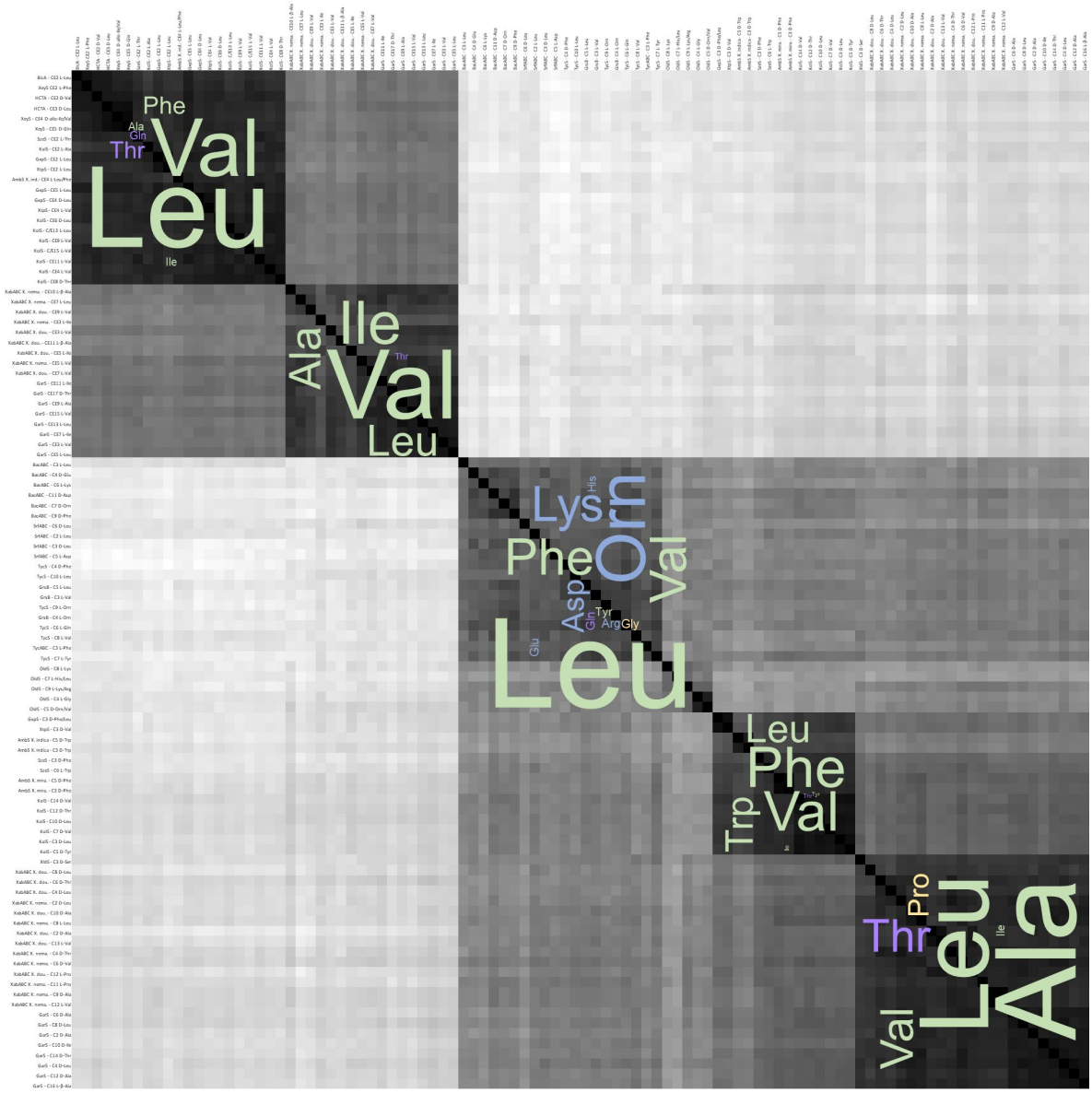


Figure S37. HeatMap of the Clustal Omega alignment of a selection of 100 C_{ASub} domains. Higher similarity is represented by darker plot colors. Blocks of highest similarity (> 75 %) were further designated with amino acid WordMaps in three-letter abbreviation of the known corresponding A domain specificities. The color-code of the amino acids are referred to their electrically charged side chains (light blue), polar uncharged side chains (light purple), hydrophobic side chains (light green), and special cases (light yellow) at physiological pH (7.4).

Supplementary Explanation

Explanation S1. Crystal structure C-A template dynamics mark major contributions of the interacting residues

The structural changes during the catalytic states are mainly manifested in the C-A interface by the conformational changes of the A domains' larger N-terminal domain (A_{Core}) in respect to the smaller C-terminal subdomain (A_{Sub}). During the 30° rotation of the A_{Sub} domain from the 2VSQ open conformation to the 4ZXH adenylate forming conformation in AOI1 Chain I and V are mainly interacting with the opposite A_{Sub} highly flexible Chain XII-XVI. This goes along with a higher abundance of hot residues mainly in Chain XII & XV and the introduction of XIII in the interface paired with a stronger, shifted connection in the C domains' Chain I & V. Meanwhile, AOI2 loosen its connections in Chain III, and entirely in Chain IV, with their opposite A_{Core} Chain IX. Likewise, in the A_{Core}/A_{Sub} transition Chain VIII, but tighten a bit up in Chain XI.

The 140° body torsion from the 4ZXH adenylate forming conformation in the 5T3D thioester forming conformation makes Chain XVI the only connection of the A_{Sub} to the C domain in AOI1 in unison with more hot residues in AOE2 and the A_{Core} and A_{Core}/A_{Sub} transition in Chain VIII-XI. Opposite the A_{Core} on the C domain side, Chain III shows more hot residues, and Chain IV with parts of Chain V (while Chain V loses its stronger connected components) is reintroduced into the interface.

During these highly complex conformational changes, the linker interactions in AOI3 with the C domain only show slight alterations in the interface, with the weakest in the 5T3D.

Emphasis should be brought to the fact that the maintenance of the C-A platform in AOI2 is mainly determined by the interaction of the C domain Chain III, IV, and VI with the A_{Core} Chain IX. Furthermore, the transition from A_{Core} to A_{Sub} occurs through Chain VIII&XI by interacting in AOI1&2.

7. Attachments

nature chemistry

ARTICLES

https://doi.org/10.1038/s41557-019-0276-z

Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains

Kenan A. J. Bozhüyük[1,3], Annabell Linck[1,3], Andreas Tietze[1,3], Janik Kranz[1,3], Frank Wesche[1], Sarah Nowak[1], Florian Fleischhacker[1], Yan-Ni Shi[1], Peter Grün[1] and Helge B. Bode[1,2]*

Non-ribosomal peptide synthetases (NRPSs) are giant enzyme machines that activate amino acids in an assembly line fashion. As NRPSs are not restricted to the incorporation of the 20 proteinogenic amino acids, their efficient manipulation would enable microbial production of a diverse range of peptides; however, the structural requirements for reprogramming NRPSs to facilitate the production of new peptides are not clear. Here we describe a new fusion point inside the condensation domains of NRPSs that results in the development of the exchange unit condensation domain (XUC) concept, which enables the efficient production of peptides, even containing non-natural amino acids, in yields up to 280 mg l^{-1}. This allows the generation of more specific NRPSs, reducing the number of unwanted peptide derivatives, but also the generation of peptide libraries. The XUC might therefore be suitable for the future optimization of peptide production and the identification of bioactive peptide derivatives for pharmaceutical and other applications.

During the past 70 years, secondary metabolite-derived drugs have become essential agents to cure infectious diseases[1,2]. Yet, infectious diseases remain the second major cause of death worldwide, and the world is facing a global public health crisis, with a growing risk of re-entering a pre-antibiotic-like era as more and more infections are caused by multi-drug-resistant bacteria[3].

Non-ribosomally made peptides (NRPs) are one source of new antibacterial agents. Their high structural diversity provides them with many properties of biological relevance. For example, peptides have been identified with antibiotic, antiviral, anticancer, anti-inflammatory, immunosuppressant and surfactant qualities[4–6]. However, natural products often need to be modified to improve their clinical properties and/or bypass resistance mechanisms[7,8]. So far, most clinically used natural product derivatives have been created by means of semi-synthesis[7,9]. A promising alternative strategy is to use engineering approaches to directly modify non-ribosomal peptide synthetases (NRPSs) to generate optimized natural products[10]. However, most attempts to achieve this have yielded impaired or non-functional biosynthetic machineries[5,11].

NRPSs are large multifunctional enzyme complexes (mega-synthases)[12,13] that form peptides not limited to the 20 proteinogenic amino acids (AA)[12]. Furthermore, these NRPSs can generate linear or cyclic peptides containing D-AAs, *N*-methylated AAs, N-terminal attached fatty acids or heterocycles[6,12,13]. NRPSs do this through having a modular architecture in which a module is defined as the catalytic unit responsible for the incorporation of one specific building block (for example, an AA) into a growing peptide chain (N → C) and associated functional group modifications[14]. Modules are composed of domains that catalyse single reaction steps such as activation, covalent attachment, optional modification of the building blocks and condensation with the amino acyl or peptidyl group on the neighbouring module[15]. At least three domains, or essential enzymatic activities, are necessary for the non-ribosomal production of peptides[6,16]. The adenylation (A) domain is needed for AA activation, the thiolation (T) domain for AA tethering and the condensation (C) domain for peptide bond formation. Finally, most NRPS termination modules harbour a thioesterase (TE) domain that releases the peptide, often in a cyclized form. These standard domains are additionally joined by tailoring domains that can catalyse epimerization (E), methylation (MT), cyclization (CY) or other modifications of the building blocks or the growing peptide chain, with dual-function C/E domains also known[16,17].

Due to the modular nature of NRPSs[16], several laboratories have tried to reprogram these systems via (1) substitution of the A or paired A-T domains, activating an alternative substrate[18,19], (2) targeted alteration of just the substrate binding pocket of the A domain[20,21] or (3) substitutions that treat C-A or C-A-T domain units as inseparable pairs[22,23]. These strategies are complemented by recombination studies, which have sought to re-engineer NRPSs by T (ref. [24]), T-C-A (ref. [25]), communication domain[26] and A-T-C swapping[27] (also see refs. [16,28,29] for further examples of NRPS engineering). However, with the exception of the A-T-C swapping strategy, denoted as the concept of exchange units (XU)[30], it has been difficult to develop clearly defined, reproducible and validated guidelines for the engineering of NRPSs.

Within the XU concept, NRPS fragments containing A-T-C or A-T-C/E domains are defined as XUs and are assembled at a specific position within the conserved C-A linker. This allowed the assembly of up to five XUs from four different natural NRPSs, resulting in fully functional de novo NRPSs that synthesize the expected peptides. In contrast to other methods, only a moderate drop in production titre is observed when one or two XUs are used. However, the great limitation of the XU concept is the specificity of the downstream C domain. The C domain has a pseudo-dimeric structure

[1]Fachbereich Biowissenschaften, Molekulare Biotechnologie, Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [2]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [3]These authors contributed equally: Kenan A. J. Bozhüyük, Annabell Linck, Andreas Tietze, Janik Kranz. *e-mail: h.bode@bio.uni-frankfurt.de

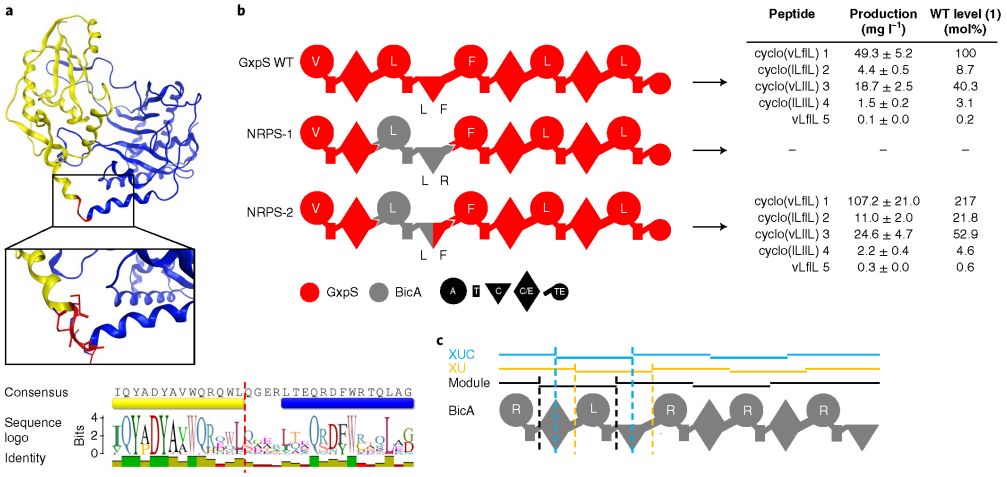


Fig. 1 | Modulation of C domain substrate specificity. a, The C domain excised from the T-C bidomain TycC 5–6 from tyrocidine syntethase (TycC) of *Brevibacillus brevis* (PDB ID: 2JGP). The C domains' N-terminal donor (yellow) and C-terminal acceptor site (blue) sub-domains are depicted in ribbon representation (top). The box shows an enlarged representation of the C_{Dsub}–C_{Asub} linker with contributing linker AAs in stick representation and the fusion site marked in red. At the bottom, a sequence logo of C_{Dsub}–C_{Asub} linker sequences from *Photorhabdus* and *Xenorhabdus* is shown. **b**, A schematic representation of WT GxpS, recombinant **NRPS-1** and **2**, as well as corresponding peptide yields obtained from triplicate experiments. For the peptide nomenclature, the standard one letter AA code (with lowercase for D-AA) is used. **c**, A schematic representation of BicA with modules, XUs and the XUCs highlighted. Specificities are assigned for all A domains. For domain assignment the following symbols are used: A, adenylation domain, large circles; T, thiolation domain, rectangle; C, condensation domain, triangle; C/E, dual condensation/epimerization domain, diamond; TE, thioesterase domain, C-terminal small circle.

with a catalytic centre between both sub-domains connecting both T-domain-tethered AAs. In NRPS biochemistry, the N-terminal sub-domain of the C domain (C_{Dsub}) is thought to bind the donor-AA (donating the peptide chain) in the catalytic centre while the C-terminal sub-domain of the C domain (C_{Asub}) is thought to bind the acceptor-AA (accepting the growing peptide chain) in the catalytic centre (Fig. 1a). The new peptide bond is then formed via a nucleophilic attack of the free amine from the acceptor-AA to the thioester of the donor-AA (Supplementary Fig. 3)[16]. Due to previous biochemical in vitro characterizations[31], it is assumed that C domains are specific for both the acceptor- and donor-AA and provide proofreading activity to ensure the correct peptide sequence.

For NRPS engineering, the C domains' proofreading/gatekeeping activity represents a severe bottleneck, as only XUs that connect the same AAs as in their original NRPS can be assembled. Therefore, at least two XUs have to be exchanged to produce a new peptide derivative that differs in one AA position from the primary sequence of the wild-type (WT) peptide[30]. Although this disadvantage can be accepted if a large number of XUs with different downstream C domains are available, a more flexible system reducing the limitations of C-domain specificities would drastically reduce the number of NRPS building blocks necessary to produce or alter particular peptides. For example, this could be used to improve NRPS-derived specialized metabolites for clinical use, to access structural diversity beyond Lipinski's rule of five, or to biosynthesize clinically relevant drugs, eliminating the need for synthetic chemistry steps and petrochemical feedstocks.

Results and discussion

C domains have acceptor site substrate specificity. To verify the influence of the C domains' acceptor site (C_{Asub}) proofreading activity, the NRPS GxpS from *Photorhabdus luminescens* TT01 (Supplementary Figs. 1 and 2) was chosen as a model system[32,33]. GxpS is responsible for the production of four cyclic peptides—GameXPeptide A–D (**1**–**4**)—that are composed of five AAs and differ in the first (Val/Leu) and third (Phe/Leu) position due to promiscuous A domains. A recombinant GxpS was constructed and expressed in *Escherichia coli*, not complying with the C domain specificity rules of the XU concept[30]. Here, XU2 of GxpS (Fig. 1b, **NRPS-1**) was exchanged against XU2 of the bicornutin producing NRPS (BicA, Fig. 1c)[34]. Although both XUs are Leu-specific, they differ in their C_{Asub} specificities—Phe for XU2 of GxpS and Arg for XU2 of BicA. This was deduced from the natural peptides in the original NRPSs. As expected, no peptide production was observed. This experiment confirmed previously published results from in vitro experiments[35–38] and illustrates the fact that C domains are indeed highly substrate-specific at their C_{Asub} domain. Although it is not yet clear how substrate specificity is conferred in C domains, the available structural data for C domains show a pseudo-dimer configuration[35–38] with their catalytic centre, including the HHxxxDG motif, having two binding sites—one for the electrophilic donor substrate and one for the nucleophilic acceptor substrate[39] (Fig. 1a and Supplementary Fig. 3). Guided by the crystal structure of the TycC5-6 T-C didomain (PDB ID: 2JGP) as well as sequence alignments of targeted *Photorhabdus* and *Xenorhabdus* C domains, we hypothesized that Gln267 and Ser268 of the four-AA-long conformationally flexible loop/linker region (Gln267–Ala270) separating the subdomains (Fig. 1a) might be an ideal fusion site to create chimaeric C domains and subsequently modulate C-domain specificities. To test this hypothesis, the Arg-specific C_{Asub} of the GxpS–BicA hybrid NRPS (Fig. 1b, **NRPS-1**) was re-exchanged to the Leu-specific C_{Asub} of GxpS, restoring the functionality of the hybrid

NRPS (**NRPS-2**) and leading to the production of GameXPeptide A–D (**1**–**5**) with 217% (107 mgl^{-1}) yield compared to the WT GxpS (Fig. 1b). The yield was confirmed by tandem mass spectrometry (MS/MS) analysis and comparison of the retention times with a synthetic standard (Supplementary Fig. 4). The high production titre was unexpected and might result from a subtle change in the overall GxpS structure, creating a more active enzyme due to the insertion of the BicA-derived fragment.

The XUC concept. From these results, along with insights from comparative structural analysis, we postulated that C_{Asub}-A-T-C_{Dsub} (XUC) units represent a self-contained catalytically active unit, without interfering in major domain–domain interfaces/interactions during the NRPS catalytic cycle[40]. To validate the proposed XUC building block (Fig. 1c) and to compare the production titres with a natural NRPS, we reconstructed GxpS (Fig. 1b) in two variants (Fig. 2a, **NRPS-3** and **-4**). Each was constructed using five XUC building blocks from four different NRPSs (XtpS, AmbS, GxpS and GarS, respectively HCTA) (Supplementary Fig. 5). **NRPS-3** was designed to contain a mixed C/E_{Dsub}-C_{Asub} domain between XUC3 and XUC4 (Fig. 2a), to reveal if C and C/E domains can be combined. In **NRPS-4**, XUC3 from HCTA, instead of GarS, was used to avoid C/E_{Dsub}-C_{Asub} domain incompatibilities between C and C/E domains (Fig. 2a).

Although **NRPS-3** (Fig. 2a) showed no detectable production of any peptide, **NRPS-4** (Fig. 2a) produced **1** and **3** in 66 and 6% yield compared to the natural GxpS, respectively (Supplementary Fig. 6). In line with expectations from domain sequences, phylogenetics and the structural differences of C/E and C domains[39], these results suggest that C/E and C domains cannot be combined with each other. Although **NRPS-4** (Fig. 2a) showed moderately reduced production titres, most probably due to the non-natural C_{Dsub}-C_{Asub} pseudo-dimer interface, the reduction was not as severe as in the XU approach, which was also based on five different NRPS building blocks[30]. The formal exchange of the promiscuous XUC1 from GxpS (for Val/Leu) against the Val-specific XUC1 from XtpS led to the exclusive production of **1** and **3** (Fig. 2a), without the production of **2** and **4** observed in the original GxpS (Fig. 1b). This indicates that the XUC can also be used to increase product specificity and to reduce the formation of side products.

Additional GameXPeptide derivatives were generated (Fig. 2a, **NRPS-5**) by combining building blocks according to the definition of XU[30] and XUC. Three fragments (1, C1-A1-T1-C/E2 of BicA; 2, A2-T2-C3-A3-T3-C/E4-A4-T4-C/E_{Dsub}5 of GxpS; 3, C/E_{Asub}5-A5-T5-C_{term} of BicA) from two NRPSs (BicA, *Xenorhabdus budapestensis* DSM 16342; GxpS, *Photorhabdus luminescens* TT01)[32,34] were used as building blocks. The expected two Arg-containing cyclic pentapeptides **6** and **7** were produced in yields of 2.2 and 0.2 mgl^{-1}, respectively, and were structurally confirmed by chemical synthesis (Supplementary Fig. 7). Both peptides only differed in Leu or Phe at position three from the promiscuous XUC3 from GxpS. Despite a drop in peptide production in comparison to the WT NRPS, we successfully demonstrated that the recently published XU[30] and the XUC strategy can be combined for successful reprogramming of NRPS and the production of tailor-made peptides.

All aforementioned recombined NRPSs are of Gram-negative origin. To show the general applicability of the novel XUC building block, we wanted to construct and express in *E. coli* artificial NRPSs also from building blocks of Gram-positive origin (using NRPSs for the production of bacitracin[41], surfactin[42], gramicidin[43] and tyrocidin[44]). The expected pentapeptide **8** containing the bacitracin NRPS-derived thiazoline ring was produced from **NRPS-6** in yields of 21 mgl^{-1} (Fig. 2b,c and Supplementary Fig. 8). For gramicidin, a 'silent' exchange of the ornithine (Orn)/Lys-specific XUC4 against the Orn/Lys-specific XUC from the tyrocidine NRPS was achieved (Fig. 2b, **NRPS-7**) that showed a different proportion of the three gramicidin derivatives **9**–**11** compared to the original GrsAB NRPS (Supplementary Fig. 9). Furthermore, new cyclic and linear gramicidin/tyrocidine hybrids **12**–**17** were produced (Fig. 2b,c, **NRPS-8** and **-9**; Supplementary Figs. 10 and 11). No peptides produced by hybrid NRPS combining XUCs from *Xenorhabdus*/*Photorhabdus* with *Bacillus* XUCs could be detected (Supplementary Fig. 12). Surprisingly, a chimaeric BacA–GxpS (**NRPS-15**) produced truncated peptides **18**–**25**, which exclusively relates to the expected activity of the GxpS portion (Supplementary Fig. 13). This suggests the presence of a correctly folded and full-length hybrid protein (Supplementary Fig. 14) that is hampered in intra-XUC communication. However, the successful assembly of chimaeric *Bacillus* NRPS (Fig. 2b) suggests the universal nature of the XUC approach when exclusively XUCs from closely related genera (only *Bacillus* or only *Photorhabdus*/*Xenorhabdus*) are used.

Increasing the number of possible starter units. A limitation for the generation of NRPSs producing any desired peptide sequence is the availability of suitable starter units because naturally their number is much smaller compared to that of extender units. This limitation could be solved if extender units could be used as starter units. However, up to now, there has been no publication describing the successful exchange of a starter unit against an internal NRPS fragment. Reasons for this might be as follows: (1) starter A domains may comprise an upstream sequence of variable length with unknown function and structure, which makes it difficult to define an appropriate artificial leader sequence or (2) necessary interactions at the C-A interface may be important for adenylation activity and A domain stability, as indicated recently[45,46]. To test whether the XUC concept can also be applied to modify starter units, three recombinant GxpS constructs (**NRPS-18** to **20**) with internal domains as starting units were created (Fig. 3). In **NRPS-18**, A1-T1-C_{Dsub}2 of GxpS was exchanged against C2-A3-linker-A3-T3-C_{Dsub}4 of XtpS because all starter A domains have a preceding C-A linker sequence. Because there are several examples of NRPSs carrying catalytically inactive starter C domains (for example, AmbS)[47], A1-T1-C_{Dsub}2 of GxpS was altered to C3-A3-T3-C_{Dsub}4 of XtpS in **NRPS-19**. In **NRPS-20**, A1-T1-C_{Dsub}2 of GxpS was altered to C_{Asub}-A3-T3-C_{Dsub}4 of XtpS as there are natural NRPSs exhibiting parts of a C domain (for example, BicA) as N-terminal parts of starter A domains.

Whereas **NRPS-18** (Fig. 3) did not show production of the desired peptides, **NRPS-19** and **NRPS-20** synthesized **1** and **3** in yields of 0.31–0.44 mgl^{-1} (Fig. 3 and Supplementary Fig. 15). This indicates that internal A domains can indeed be used as starter domains, if the upstream C_{Asub} or C domain is kept in front of the A domain, pointing to the importance of a functional C-A interface for A-domain activity. Yet, the observed low production titres might indicate that the observed difference in codon usage and/or the lower GC content at the beginning of WT NRPS encoding genes could have a major impact on transcriptional and/or translational efficiency in conjunction with protein folding, as described previously[48].

Increasing peptide diversity beyond the incorporation of natural AAs. Besides creating NRP derivatives carrying natural AAs, one useful application of NRPS reprogramming could be the incorporation of non-natural AAs. Examples include AAs containing alkyne or azide groups, allowing reactions like Cu(I)-catalysed or strain-promoted Huisgen cyclization, also known as 'click' reactions[49–52]. Although NRPS and A domains have been examined exhaustively for several years, no general method for the in vivo functionalization of NRPs is available by reprogramming NRPS templates.

Naturally, a broad range of AAs are accepted by the A3 domain of GxpS (Supplementary Fig. 1), resulting in a large diversity of natural GameXPeptides[32,33]. Moreover, by using a γ-$^{18}O_4$-ATP pyrophosphate exchange assay for A-domain activity[53,54] and adding

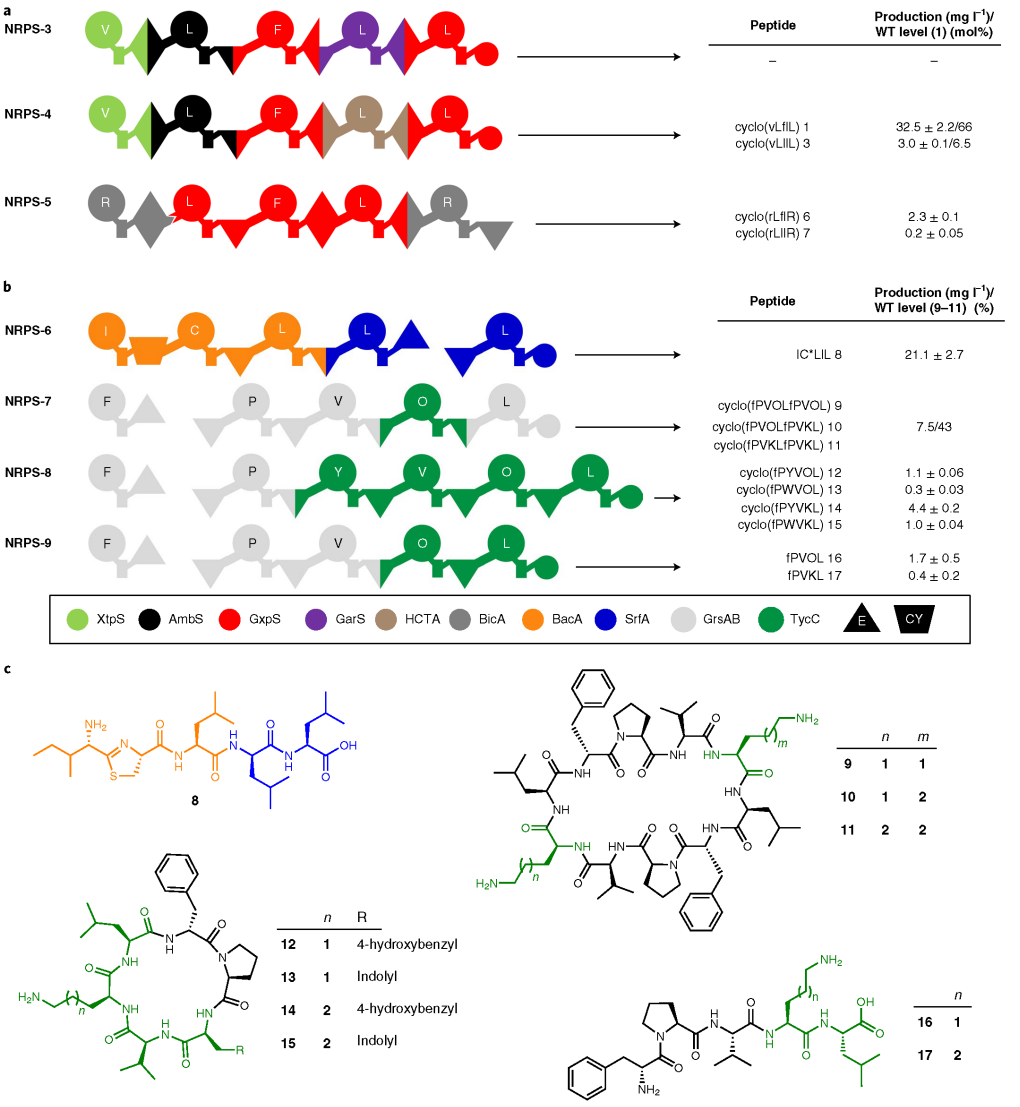


Fig. 2 | Design of recombinant NRPS for peptide production. a, The generated recombinant GxpS (**NRPS-3** to **-5**) and corresponding amounts of GameXPeptide derivatives **1**, **3**, **6** and **7** as determined in triplicates. **b**, Recombinant **NRPS-6** to **-9** using building blocks of Gram-positive origin. The gramicidin derivatives **9**–**11** were isolated as a mixture. **c**, The structures of **8**–**17** produced from **NRPS-6** to **-9** expressed in *E. coli*. For the peptide nomenclature, the standard AA one-letter code with lowercase for D-AA is used. See Fig. 1 for assignment of the domain symbols; further symbols: E, epimerization domain, inverted triangle; CY, heterocyclization domain, trapezium. The colour code identifies NRPSs used as building blocks (for details, see Supplementary Fig. 5).

substituted phenylalanine derivatives to *E. coli* cultures expressing GxpS, the respective A3 domain was shown to activate (in vitro, Supplementary Fig. 16) and incorporate (in vivo, Supplementary Fig. 17) several *ortho*- (*o*), *meta*- (*m*) and *para*- (*p*) substituted phenylalanine derivatives, including 4-azido-L-phenylalanine (pN_3-F) and *O*-propargyl-L-tyrosine (Ɣ-Y). When the Val-specific XUC3 of the xenotetrapeptide[55] (**26**) (Supplementary Fig. 18) producing NRPS (XtpS) from *X. nematophila* HGB081 was exchanged against

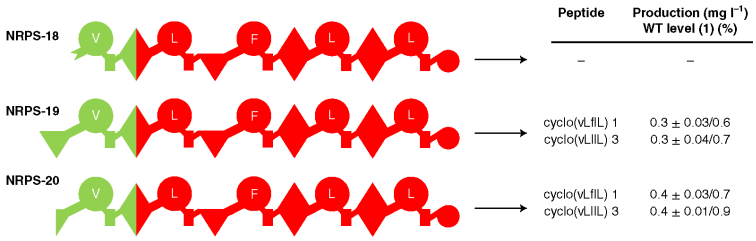


Fig. 3 | Elongation XUCs can be used as starting XUC. A schematic representation of recombinant GxpS (**NRPS-18** to **-20**) and the corresponding peptide yields obtained from triplicate experiments. For the peptide nomenclature, the standard AA one-letter code (with lowercase for D-AA) is used. See Fig. 1 for assignment of the domain symbols; GxpS NRPS is shown in red and XtpS (XUC3) in green (for details see Supplementary Fig. 5).

XUC3 of GxpS, six new xenotetrapeptide derivatives (**29**–**34**) in yields of 0.17–106 mgl^{-1} were produced, reflecting the natural promiscuity of GxpS_XUC3 (Fig. 4 and Supplementary Fig. 19). From a large-scale cultivation in shaking flasks, 52 and 47 mgl^{-1} of **29** and **30** (Fig. 4), respectively, were isolated, and their structure was confirmed by NMR analysis (Supplementary Figs. 25–34). After adding pN_3-F and ¥-Y to growing *E. coli* cultures expressing recombined XtpS (**NRPS-21**), six functionalized peptides (**35**–**37** and **38**–**40**) were produced. These peptides differed in position three and were produced in yields of 5–280 mgl^{-1} with **36**, **37** and **38** being structurally confirmed by chemical synthesis and **35** and **36** being isolated from a large-scale culture for structure confirmation by NMR in yields of 6–7 mgl^{-1} (for NMR data see Supplementary Figs. 35–39). The observed methyl ester derivatives **28**, **30**, **32**, **35** and **39** in all these experiments (Fig. 4) were derived from the use of MeOH as solvent during the work-up procedure. Similarly, linear and cyclic 4-Br-Phe derivatives were also produced (Supplementary Fig. 20) and characterized after their isolation (Supplementary Figs. 41 and 42).

Production of peptide libraries. Modern drug-discovery approaches often apply the screening of compound libraries, including natural product libraries[56], because they exhibit a wide range of pharmacophores, structural diversity and have the property of metabolite-likeness often providing a high degree of bioavailability. Yet, the natural product discovery process is as expensive as it is time-consuming[49]. Consequently, for bioactivity screenings, the random recombination of certain NRPS fragments would be a powerful tool to create focused artificial natural-product-like libraries.

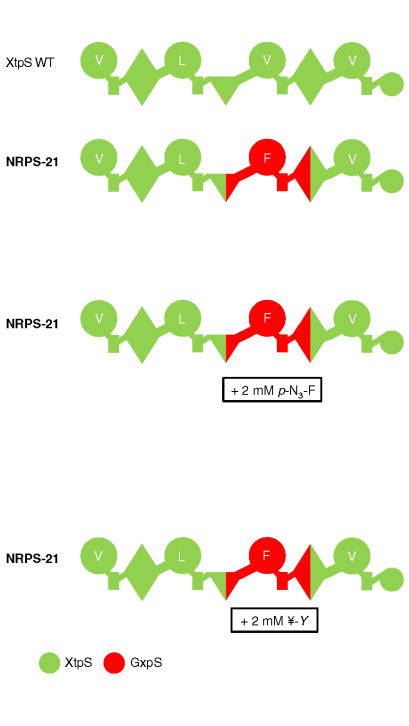
In an initial test, GxpS was chosen for the generation of a focused peptide library created via a one-shot yeast-based transformation-associated recombination (TAR) cloning approach[47,57]. Here, the third position of the peptide (D-Phe) was randomized (Fig. 5a) using six unique XUC building blocks from six NRPSs (KolS[58], BicA[34], $AmbS_{mir}$[47], Pax[59], $AmbS_{ind}$, XllS; for details see Supplementary Fig. 5), resulting in the production of **1** and four new GameXPeptide derivatives (**41**–**44**) in yields of 3–92 mgl^{-1} that were structurally confirmed by chemical synthesis (Supplementary Fig. 21) and preparative isolation of **42** from a large-scale cultivation followed by NMR analysis (Supplementary Fig. 40).

For the generation of a second and structurally more diverse peptide library, positions 1 (D-Val) and 3 (D-Phe) of GxpS were selected for parallel randomization (Fig. 5b). Theoretically, 48 different cyclic or linear peptides could be expected based on the experimental set-up. Screening of 50 *E. coli* clones resulted in the identification of 16 unique cyclic and linear peptides (**1**, **5**, **30**, **32**, **43**, **45**–**55**) from four peptide-producing clones differing in peptide length and AA composition (Supplementary Fig. 22). As only 7 out of 18 identified peptides belong to the originally expected set of peptides, it is possible that homologous recombination by TAR cloning results in the generation of unexpected NRPSs that subsequently produce unexpected peptides, resulting in an additional layer of peptide diversification, a phenomenon that has been observed previously[30].

Randomizing directly adjacent positions via a similar approach requires a standardized nucleotide sequence (39 base pairs) for homologous recombination (Supplementary Fig. 23)[47,57]. From a detailed analysis of the T-C didomain crystal structure of TycC5-6 (PDB ID: 2JGP), helix α5 (I253-F265) next to the C domain's pseudo-dimer linker was identified as an ideal target for homologous recombination. Subsequently, an artificial α5 helix was designed to randomize positions 2 (L-Leu) and 3 (D-Phe) of GxpS (Supplementary Fig. 23a), being an integral part of all resulting recombinant C domains and therefore connecting XUC2 and 3. The applied α5 helix was defined as the consensus sequence of all involved XUC building blocks (Supplementary Fig. 23b). Screening of 25 *E. coli* clones revealed the synthesis of eight cyclic and linear GameXPeptides (**1**, **42**–**43**, **50**, **53**, **56**–**58**) from three peptide-producing clones in good yields, showing the general applicability of redesigning α5 with respect to randomly reprogramming biosynthetic templates (Fig. 5c and Supplementary Fig. 24).

Conclusion. We have recently described the XU concept, enabling the efficient reprogramming of NRPSs; this is, however, limited in its applicability by downstream C-domain specificities[30]. Here we present the XUC concept, which eliminates these limitations by utilizing a direct assembly inside the C domains and allows the production of natural and artificial peptides in yields up to 280 mgl^{-1}. For the construction of any peptide based on the 20 proteinogenic AAs, only 80 XUC building blocks are necessary (only four of each: C_{Dsub}-A-T-C_{Asub}, C_{Dsub}-A-T-C/E_{Asub}, C/E_{Dsub}-A-T-C/E_{Asub} and C/E_{Dsub}-A-T-C_{Asub}), whereas 800 building blocks would be necessary to generate the same number of peptides using the XU concept. Consequently, the introduction of the XUC concept simplifies and broadens the possibilities of biotechnological applications with respect to optimizing bioactive agents via NRPS engineering (Figs. 1 and 2) or the production of functionalized peptides by incorporating XUC building blocks accepting non-natural AAs like pN_3-F and ¥-Y (Fig. 4 and Supplementary Fig. 19), allowing further derivatization. If suitable production yields can be reached as shown here, the biotechnological production of peptides could be both more sustainable and economical compared to synthetic approaches, as it avoids organic solvents and expensive modified AAs as building blocks. Moreover, once a producer strain is available, scale-up should be much easier and 'greener' compared to synthetic approaches.

Another strength of the XUC concept is its application to generate random natural-product-like peptide libraries (Fig. 5) for subsequent bioactivity screenings. The possible automation of NRPS



Peptide	Production (mg l^{-1})	WT level (26) (mol%)
cyclo(vLvV) 26	134.1 ± 26.9	100
vLvV 27	0.4 ± 0.06	0.3
vLvV-OMe 28	1.3 ± 0.1	0.9
cyclo(vLvV) 26	7.5 ± 1.2	5.6
vLfV 29	5.6 ± 0.4	3.5
vLfV-OMe 30	24.4 ± 1.9	4.1
vLIV 31	0.2 ± 0.01	0.2
vLIV-OMe 32	0.6 ± 0.08	0.1
cyclo(vLfV) 33	106.6 ± 15.6	71.2
cyclo(vLIV) 34	80.1 ± 6.3	57.8
vL-[p-N_3-f]-V-OMe 35	46.0 ± 5.4	27.2
vL-[p-N_3-f]-V 36	283.5 ± 29.5	163.4
cyclo(vL-[p-N_3-f]-V) 37	30.4 ± 2.3	18.7
cyclo(vLvV) 26	4.2 ± 1.4	3.1
vLfV 29	1.9 ± 0.2	1.2
vLfV-OMe 30	6.7 ± 0.6	4.2
vLIV 31	0.1 ± 0.01	0.1
vLIV-OMe 32	0.2 ± 0.02	0.03
cyclo(vLfV) 33	72.6 ± 0.6	48.5
cyclo(vLIV) 34	48.4 ± 2.5	35.0
vL-[¥-y]-V 38	4.7 ± 0.6	2.7
vL-[¥-y]-V-OMe 39	16.4 ± 0.7	3.9
cyclo(vL[¥-y]-V) 40	4.9 ± 0.3	2.9
cyclo(vLvV) 26	3.7 ± 1.1	2.8
vLfV 29	1.8 ± 0.2	1.2
vLfV-OMe 30	7.6 ± 1.1	4.8
vLIV 31	0.1 ± 0.01	0.05
vLIV-OMe 32	0.2 ± 0.03	0.1
cyclo(vLfV) 33	61.6 ± 6.7	41.1
cyclo(vLIV) 34	45.3 ± 3.8	32.7

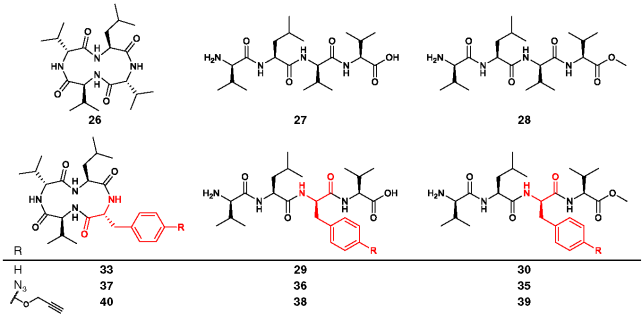
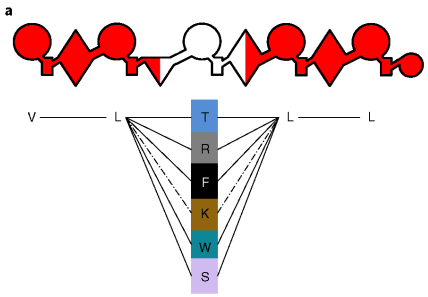


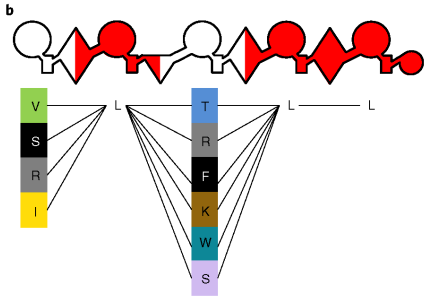
Fig. 4 | Creation of functionalized xenotetrapeptide derivatives. A schematic representation of XtpS WT, recombinant **NRPS-21**, the corresponding peptide yields obtained from triplicate experiments and selected peptide structures. For the peptide nomenclature, the standard one-letter AA code (with lowercase for D-AA) is used. See Fig. 1 for assignment of the domain symbols. The colour code of NRPS used as building blocks: Xtps, green; GxpS, red (for details, see Supplementary Fig. 5).

library design coupled to a bioactivity screening opens up entirely new opportunities of identifying novel lead compounds in the future. Particularly in the area of anti-infective research, the XUC concept might allow fast access to natural product derivatives with altered bioactivity profiles, or for the generation of producer strains with fewer side products to facilitate compound purification.

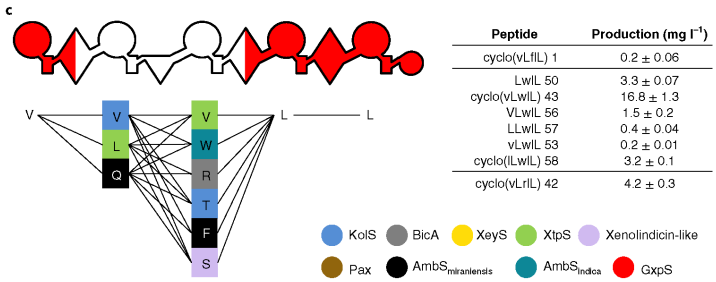
One limitation of XUC compared to XU is the missing compatibility between building blocks from different genera (here *Bacillus* and *Photorhabdus*/*Xenorhabdus*, Supplementary Figs. 12–14), while building blocks from the same genera can be fused easily (Figs. 1–5). This might be due to subtle differences in the overall structures of the C domains from these different genera that can probably be identified once more structures of C domains are available. However, this limitation is less of a concern as many building blocks (from NRPS-encoding genes) are available from well-known natural-product producers.



Peptide	Production (mg l^{-1})/ WT level (1) (mol%)
cyclo(vLtlL) 41	22.5 ± 1.0
cyclo(vLrlL) 42	33.6 ± 1.2
cyclo(vLwlL) 43	21.2 ± 0.9
cyclo(vLflL) 1	92.0 ± 17.9/187
cyclo(vLwlL) 43	9.5 ± 2.5
cyclo(vLslL) 44	2.9 ± 0.3



Peptide	Production (mg l^{-1})/ WT level (1) (mol%)
lLmL 45	0.8 ± 0.01
lLwL 46	4.4 ± 0.1
vLwL 47	1.4 ± 0.01
vLIV-OMe 32	0.7 ± 0.01
lLfL 48	9.9 ± 0.2
vLfV-OMe 30	2.2 ± 0.1
lLlL 49	1.0 ± 0.06
LwlL 50	1.1 ± 0.03
cyclo(rLwlL) 51	17.7 ± 1.3
cyclo(sLtlL) 52	0.2 ± 0.04
cyclo(vLflL) 1	122.4 ± 14.0/248
vLflL 5	0.2 ± 0.02
cyclo(vLwlL) 43	9.3 ± 1.6
vLwlL 53	0.1 ± 0.01
vLllL 54	0.1 ± 0.02
LflL 55	1.0 ± 0.04
LwlL 50	0.9 ± 0.01
vLIV-OMe 32	0.7 ± 0.01



Peptide	Production (mg l^{-1})
cyclo(vLflL) 1	0.2 ± 0.06
LwlL 50	3.3 ± 0.07
cyclo(vLwlL) 43	16.8 ± 1.3
VLwlL 56	1.5 ± 0.2
LLwlL 57	0.4 ± 0.04
vLwlL 53	0.2 ± 0.01
cyclo(lLwlL) 58	3.2 ± 0.1
cyclo(vLrlL) 42	4.2 ± 0.3

Fig. 5 | Targeted randomization of GxpS. A schematic representation of all possible recombinant NRPSs and corresponding NRPs (left). Detected peptides and corresponding yields (right) obtained from triplicate experiments are shown. For the peptide nomenclature, the standard one-letter AA code (with lowercase for D-AA) is used. See Fig. 1 for assignment of the domain symbols. **a**, Randomization of position three from GxpS. **b**, Randomization of position one and three from GxpS. **c**, Randomization of adjacent positions two and three. The colour code of NRPSs used as building blocks is shown at the bottom right (for details, see Supplementary Fig. 5).

With respect to the production of completely new-to-nature peptides, the formation of cyclic or depsi-peptides might be another limitation. This is because TE domains can be specific for AA positions or for the chain length or ring size they can act on[60]. However, we have shown previously that internal C domains can act as cyclization catalysts similar to NRPSs from fungi and some bacteria, which may avoid the substrate specificity of TE domains[30].

Taken together, XUC offers a new approach for NRPS modification and thus the generation of peptides that in the future will result in the production of novel bioactive natural products.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Received: 4 March 2018; Accepted: 26 April 2019;
Published online: 10 June 2019

References

1. Clardy, J., Fischbach, M. A. & Walsh, C. T. New antibiotics from bacterial natural products. *Nat. Biotechnol.* **24**, 1541–1550 (2006).

2. von Nussbaum, F., Brands, M., Hinzen, B., Weigand, S. & Häbich, D. Antibacterial natural products in medicinal chemistry—exodus or revival? *Angew. Chem. Int. Ed.* **45**, 5072–5129 (2006).
3. Nathan, C. Antibiotics at the crossroads. *Nature* **431**, 899–902 (2004).
4. Felnagle, E. A. et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharmaceut.* **5**, 191–211 (2008).
5. Calcott, M. J. & Ackerley, D. F. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol. Lett.* **36**, 2407–2416 (2014).
6. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
7. O'Connell, K. M. G. et al. Combating multidrug-resistant bacteria: current strategies for the discovery of novel antibacterials. *Angew. Chem. Int. Ed.* **52**, 10706–10733 (2013).
8. Bush, K. Improving known classes of antibiotics: an optimistic approach for the future. *Curr. Opin. Pharmacol.* **12**, 527–534 (2012).
9. Kirschning, A. & Hahn, F. Merging chemical synthesis and biosynthesis: a new chapter in the total synthesis of natural products and natural product libraries. *Angew. Chem. Int. Ed.* **51**, 4012–4022 (2012).
10. 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).
11. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347 (2016).
12. Caboche, S., Leclere, V., Pupin, M., Kucherov, G. & Jacques, P. Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J. Bacteriol.* **192**, 5143–5150 (2010).
13. Grünewald, J. & Marahiel, M. A. Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. *Microbiol. Mol. Biol. Rev.* **70**, 121–146 (2006).
14. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations and mutations. *Science* **282**, 63–68 (1998).
15. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490–504 (2002).
16. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis—principles and prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821 (2017).
17. Balibar, C. J., Vaillancourt, F. H. & Walsh, C. T. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **12**, 1189–1200 (2005).
18. 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).
19. Calcott, M. J., Owen, J. G., Lamont, I. L. & Ackerley, D. F. Biosynthesis of novel pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **80**, 5723–5731 (2014).
20. 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.* **51**, 7181–7184 (2012).
21. Kries, H. et al. Reprogramming nonribosomal peptide synthetases for 'clickable' amino acids. *Angew. Chem. Int. Ed.* **53**, 10105–10108 (2014).
22. Nguyen, K. T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl Acad. Sci. USA* **103**, 17462–17467 (2006).
23. Yakimov, M. M., Giuliano, L., Timmis, K. N. & Golyshin, P. N. Recombinant acylheptapeptide lichenysin: high level of production by *Bacillus subtilis* cells. *J. Mol. Microbiol. Biotechnol.* **2**, 217–224 (2000).
24. Beer, R. et al. Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol. BioSyst.* **10**, 1709–1710 (2014).
25. Calcott, M. J. & Ackerley, D. F. Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.* **15**, 1–13 (2015).
26. Chiocchini, C., Linne, U. & Stachelhaus, T. In vivo biocombinatorial synthesis of lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic complex. *Chem. Biol.* **13**, 899–908 (2006).
27. Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P. J. L. M. & Marahiel, M. A. Construction of hybrid peptide synthetases for the production of alpha-L-aspartyl-L-phenylalanine, a precursor for the high-intensity sweetener aspartame. *Eur. J. Biochem.* **270**, 4555–4563 (2003).
28. 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.* **49**, 104–119 (2018).
29. Kries, H. Biosynthetic engineering of nonribosomal peptide synthetases. *J. Pept. Sci.* **22**, 564–570 (2016).
30. Bozhüyük, K. A. J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
31. Linne, U. & Marahiel, M. A. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**, 10439–10447 (2000).
32. Bode, H. B. et al. Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry. *Chem. Eur. J.* **18**, 2342–2348 (2012).
33. Nollmann, F. I. et al. Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. *ChemBioChem* **16**, 205–208 (2015).
34. Fuchs, S. W. et al. Neutral loss fragmentation pattern based screening for arginine-rich natural products in *Xenorhabdus* and *Photorhabdus*. *Anal. Chem.* **84**, 6948–6955 (2012).
35. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
36. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
37. 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).
38. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **425**, 3137–3150 (2013).
39. 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).
40. Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140 (2016).
41. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927–937 (1997).
42. Cosmina, P. et al. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **8**, 821–831 (1993).
43. Krätzschmar, J., Krause, M. & Marahiel, M. A. 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).
44. 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).
45. Li, R., Oliver, R. A. & Townsend, C. A. Identification and characterization of the sulfazecin monobactam biosynthetic gene cluster. *Cell Chem. Biol.* **24**, 24–34 (2017).
46. Meyer, S. et al. Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem. Biol.* **23**, 462–471 (2016).
47. Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *ChemBioChem* **15**, 1290–1294 (2014).
48. Roy, A. D., Grüschow, S., Cairns, N. & Goss, R. J. M. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
49. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
50. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click-chemie: diverse chemische Funktionalität mit einer Handvoll guter Reaktionen. *Angew. Chem.* **113**, 2056–2075 (2001).
51. Pérez, A. J., Wesche, F., Adihou, H. & Bode, H. B. Solid-phase enrichment and analysis of azide-labeled natural products: fishing downstream of biochemical pathways. *Chem. Eur. J.* **22**, 639–645 (2016).
52. Pérez, A. J. & Bode, H. B. 'Click chemistry' for the simple determination of fatty-acid uptake and degradation: revising the role of fatty-acid transporters. *ChemBioChem* **16**, 1588–1591 (2015).
53. Kronenwerth, M. et al. Characterisation of taxlllaids A-G; natural products from *Xenorhabdus indica*. *Chem. Eur. J.* **20**, 17478–17487 (2014).
54. Phelan, V. V., Du, Y., McLean, J. A. & Bachmann, B. O. Adenylation enzyme characterization using gamma-(18)O(4)-ATP pyrophosphate exchange. *Chem. Biol.* **16**, 473–478 (2009).
55. Kegler, C. et al. Rapid determination of the amino acid configuration of xenotetrapeptide. *ChemBioChem* **15**, 826–828 (2014).

56. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–129 (2015).
57. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 1–4 (2007).
58. Bode, H. B. et al. Structure elucidation and activity of kolossin A, the D-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase. *Angew. Chem. Int. Ed.* **54**, 10352–10355 (2015).
59. Fuchs, S. W., Proschak, A., Jaskolla, T. W., Karas, M. & Bode, H. B. Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. *Org. Biomol. Chem.* **9**, 3130–3132 (2011).
60. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).

Acknowledgements

The authors thank M. Lindner and C. Zizka for help with the construction of selected plasmids and C. Kegler for helpful discussions. This work was funded in part by the LOEWE programme of the state of Hesse as part of the MegaSyn and TBG research clusters. H.B.B. acknowledges the Deutsche Forschungsgemeinschaft for funding of the Impact II qTof mass spectrometer (INST 161/810-1).

Author contributions

K.A.J.B. and H.B.B. designed the experiments. K.A.J.B., A.L., A.T., J.K., S.N. and F.F. performed all molecular biology and biochemical experiments. F.W. synthesized all peptide standards that were used for the high-performance liquid chromatography–mass spectrometry-based quantification performed by A.L. and A.T. J.K., Y.-N.S. and P.G. isolated selected peptides and Y.-N.S. performed their NMR analysis. All authors analysed the results and K.A.J.B., A.L., A.T., J.K. and H.B.B. wrote the manuscript. All authors saw and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-019-0276-z.

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

Correspondence and requests for materials should be addressed to H.B.B.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

nature chemistry

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41557-019-0276-z

In the format provided by the authors and unedited.

Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains

Kenan A. J. Bozhüyük[1,3], Annabell Linck[1,3], Andreas Tietze[1,3], Janik Kranz[1,3], Frank Wesche[1], Sarah Nowak[1], Florian Fleischhacker[1], Yan-Ni Shi[1], Peter Grün[1] and Helge B. Bode[1,2]*

[1]Fachbereich Biowissenschaften, Molekulare Biotechnologie, Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [2]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [3]These authors contributed equally: Kenan A. J. Bozhüyük, Annabell Linck, Andreas Tietze, Janik Kranz. *e-mail: h.bode@bio.uni-frankfurt.de

Modification and *de novo* design of non-ribosomal peptide synthetases (NRPS) using specific assembly points within condensation domains

Supplementary data

Kenan A. J. Bozhüyük, Annabell Linck, Andreas Tietze, Janik Kranz, Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-Ni Shi, Peter Grün, Helge B. Bode[†]

[†] Kenan A. J. Bozhüyük, Annabell Linck, Andreas Tietze, Janik Kranz, Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-Ni Shi, Peter Grün, Helge B. Bode

Fachbereich Biowissenschaften, Goethe-Universität Frankfurt,
Max-von-Laue-Str. 9, Frankfurt am Main 60438, Germany
Fax: (+)49 69 798 29557
E-mail: h.bode@bio.uni-frankfurt.de
Homepage: http://www.uni-frankfurt.de/fb/fb15/institute/inst-3-mol-biowiss/AK-Bode

Prof. Dr. H. B. Bode, Buchmann Institute for Molecular Life Sciences (BMLS),
Goethe Universität Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt am Main,
Germany

Materials and methods

1. Cultivation of strains

All *E. coli, Photorhabdus* and *Xenorhabdus* strains were grown in liquid or solid LB-medium (pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl). Solid media contained 1.5% (w/v) agar. *S. cerevisiae* strain CEN.PK 2-1C and derivatives were grown in liquid and solid YPD-medium (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose). Agar plates contained 1.5% (w/v) agar. Kanamycin (50 µg/mL) and G418 (200 µg/mL) were used as selection markers. *E. coli* was cultivated at 37°C all other strains were cultivated at 30°C.

2. Expression and cultivation of His-tagged proteins

For overproduction and purification of the ~72 kDa His-tagged A domain GxpS_A3,5 mL of an overnight culture in LB medium of *E. coli* BL21 (DE3)cells harboring the corresponding expression plasmid and the TaKaRa chaperone-plasmid pTf16 (TAKARA BIO INC.) were used to inoculate 500 mL of autoinduction medium (464 mL LB medium, 500 µL 1 M $MgSO_4$, 10 mL 50x5052, 25 mL 20xNPS) containing 20 µg/mL chloramphenicol, 50 µg/mL kanamycin and 0.5 mg/mL L-arabinose[1]. The cells were grown at 37°C up to an OD_{600} of 0.6. Following the cultures were cultivated for additional 48 h at 18°C. The cells were pelleted (10 min, 4,000 rpm, 4°C) and stored overnight at -20°C. For protein purification the cells were resupended in binding buffer (500 mM NaCl, 20 mM imidazol, 50 mM HEPES, 10% (w/v) glycerol, pH 8.0). For cell lysis benzonase (Fermentas, 500 U), protease inhibitor (Complete EDTA-free, Roche), 0.1% Triton-X and lysozym (0.5 mg/mL, ~20,000 U/mg, Roth) were added and the cells were incubated rotating for 30 min. After this the cells were placed on ice and lysed by sonication. Subsequently, the lysed cells were centrifuged (25,000 rpm, 45 min, 4°C). The yielded supernatant was passed through a 0.2 µm filter and loaded with a flow rate of 0.5 mL/min on a 1 mL HisTrap™ HP column (GE Healthcare) equilibrated with binding buffer. Unbound protein was washed off with 10 mL binding buffer. Impurities were washed off with 5 mL 8% elution buffer (500 mM NaCl, 500 mM imidazol,

50 mM HEPES, 10% (w/v) glycerol, pH 8.0). The purified protein of interest was eluted with 39% elution buffer. Following, the purified protein containing fraction was concentrated (Centriprep® Centrifugal Filters Ultacel® YM – 50, Merck Millipore) and the buffer was exchanged to 20 mM Tris-HCl (pH 7.5) using a PD-10 column (Sephadex™ G-25 M, GE Healthcare).

3. Cloning of GxpS_A3

The adenylation domain GxpS_A3 was cloned from *Photorhabdus luminescens* TTO1 genomic DNA by PCR using the pCOLA_Gib_A3 Insert forward and reverse oligonucleotides shown in Tab. 4. The plasmid backbone of pCOLADUET™-1 (Merck/Millipore) was amplified using the DUET_Gib forward and reverse oligonucleotides shown in Tab 4. The ~1,900 bp PCR product was cloned via Gibson Assembly® Cloning Kit (NEB) according to the manufacturers' instructions into pCOLADUET™-1.

4. γ -$^{18}O_4$–ATP Pyrophosphat Exchange Assay

The γ -$^{18}O_4$ –ATP Pyrophosphat Exchange Assay was performed as published previously [2,3]. After an incubation period of 90 min at 24°C the reactions were stopped by the addition of 6 µL 9-aminoacridine in acetone (10 mg/mL) for MALDI-Orbitrap-MS analysis.

5. MALDI-Orbitrap-MS

Samples were prepared for MALDI-analysis as a 1:1 dilution in 9-aminoacridine in acetone (10 mg/mL) and spotted onto a polished stainless steel target and air-dried. MALDI-Orbitrap-MS analyses were performed with a MALDI LTQ Orbitrap XL (Thermo Fisher Scientific, Inc., Waltham, MA) equipped with a nitrogen laser at 337 nm. The following instrument parameters were used: laser energy, 27 µJ; automatic gain control, on; auto spectrum filter, off; resolution, 30,000; plate motion, survey CPS. Mass spectra were obtained in negative ion mode over a range of 500 to 540 *m/z*. The mass spectra for ATP-PP_i exchange analysis were

acquired by averaging 50 consecutive laser shots. Spectral analysis was conducted using Qual Browser (version 2.0.7; Thermo Fisher Scientific, Inc., Waltham, MA).

6. Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus*,*Photorhabdus*and *Bacillus*strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. Polymerase chain reaction (PCR) was performed with oligonucleotides obtained from Eurofins Genomics (Tab. 4). Fragments with homology arms (40 -80 bp) were amplified in a two-step PCR program For PCR Phusion High-Fidelity DNA polymerase (Thermo Scientific), Q5 High-Fidelity DNA polymerase (New England BioLabs) and Velocity DNA polymerase (Bioline) were used. Polymeraseswere used according to the manufacturers' instructions. DNA purification was performed from 1% TAE agarose gel using Invisorb® Spin DNA Extraction Kit (STRATEC Biomedical AG). Plasmid isolation from *E. coli* was done by alkaline lysis.

7. Transformation-associated recombination (TAR) cloning

Transformation of yeast cells was done according to the protocols from Gietz and Schiestl[4,5]. 100 - 2,000 ng of each fragment was used for transformation. Constructed plasmids were isolated from yeast transformants and transformed in *E. coli* DH10B::mtaA by electroporation. Successfully transformed plasmids were isolated from *E. coli* transformants and verified by restriction digest.

8. Heterologous expression of NRPS templates and LC-MS analysis

Constructed plasmids were transformed into *E. coli* DH10B::mtaA. Strains were grown overnight in LB medium containing50 µg/mLkanamycin. 100 µL of an overnight culture were used for inoculation of 10 mL cultures, containing 0.02 mg/mL L-arabinose and 2% (v/v) XAD-16.50 µg/mL kanamycin were used as selection markers. After incubation for 72 h at 22°C, respectively, the XAD-16 was harvested. One culture volume methanol was added and incubated for 30 min. The organic phase was filtrated and evaporated to dryness under

reduced pressure. The extract was diluted in 1 mL methanol and a 1:10 dilution was used for LC-MS analysis as described previously[6,7]. All measurements were carried out by using an Ultimate 3000 LC system (Dionex) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. High-resolution mass spectra were obtained on an Ultimate 3000 RSLC (Dionex) coupled to an Impact II qTof (Bruker) equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

9. Homology-Modelling

The homology-modelling was performed as described previously[7]. For homology modelling, the 1.85 Å crystal structure of PCP-C bidomain TycC 5-6 from tyrocidine syntethase (TycC) of *Brevibacillus brevis* (PDB-ID: 2JGP) were used[8]. The sequence identity of GxpS_C3 in comparison to TycC 5-6 is 34.8%, respectively. The final models have a root-mean-square deviation (RMSD) of 1.4 Å respectively, in comparison to the template structures.

10. Peptide quantification

All peptides were quantified using a calibration curve and HPLC/MS measurements. Triplicates of all experiments were measured. As standards, either synthetic **1** (for quantification of **1**-**4**), **5** (for quantification of **5**, **53**,**54**,**56**,**57** and **58**), **7** (for quantification of **6** and **7**), I-thiazoline-L [9](for quantification of **8**), **12**(for quantification of **12** and**13**),**14** (for quantification of **14** and**15**), **16**, **17**, **26** (for quantification of **26**, **33** and **34**), **29** (for quantification of **29**, **31**, **45**, **46**, **47**, **48**, **49**, **50**and **55**), **36**, **37**, **38** (for quantification of **38** and **40**), **41**, **42**, **43**(for quantification of **43** and **58**), **44**(for quantification of **44** and **52**), and cyclo[RLflL][9] (for quantification of **51**and **59**) or purified **30**(for quantification of **28**, **30** and **32**) and **35**from 6 L LB culture of *E. coli* DH10B::mtaA pFF1_*NRPS*_21 respectively supplemented with 2 mM ¥-Y were used.

9, **10** and **11** were purified in one fraction from 1 L LB culture of *E. coli* DH10B::mtaA pFF1_*grsTAB* respectively pFF1_*NRPS*_7 and used for determination of the production titer. For quantification of **39**, the proportion of all eleven values from the calibration curve of **29** to the respective values from the calibration curve of **38**was used to calculate the calibration curve of **37**from all values of the calibration curve of **40**.

13. Chemical synthesis

Chemical synthesis of all peptides was performed as described previously[9].

14. Peptide isolation and structure elucidation

Seven peptides (**29**, **30**, **35**, **36**, **42**, **71**, **72**) were isolated from *E. coli* DH10B::mtaA (Supplementary Table 2). The strains were cultivated and the extracts were generated as described above from 1 L cultures. Compared to the small scale cultivations, different ratios of linear to cyclic peptides were observed for some peptides resulting in the linear forms as the main derivatives.

Compounds were isolated in a first chromatography using either Sephadex LH20 (MeOH, 25–100 µm, Pharmacia Fine Chemical Co. Ltd.) or a 1260 Infinity II LC system coupled to a G6125B LC/MSD ESI-MS (Agilent). A 25-55% water/acetonitrile gradient was applied over 25 min on a Agilent Eclipse XDB-C18, 7 µm, 21.2 x 250 mm column using a flow rate of 20 mL/min. Subsequently, **35**, **36**, **71**, **72** were purified in an additional chromatographic step using a 1260 Semiprep LC system coupled to a G6125B LC/MSD ESI-MS (Agilent). A 35-65% water/acetonitrile gradient was applied over 25 min on an Eclipse Plus Phenyl-Hexyl, 5 µm, 9.4 x 250 mm column using a flow rate of 3 mL/min. If required an additional semi-preparative HPLC was performed on an Agilent 1260 Infinity II LCMS Systems with a Cholester column (1OID × 250 mm, COSMOSIL).

The structures of all isolated compoundswere elucidated by detailed 1D and 2D NMR experiments (Supplementary Figures25-42).^{1}H, ^{13}C, HSQC, HMBC, ^{1} H-^{1}H COSY, and

ROESY spectra were measured on Bruker AV500 and AV600 spectrometers, using DMSO as solvent. Coupling constants are expressed in Hz and chemical shifts are given on a ppm scale. High-resolution MS analysis was performed as described above.

Supplementary Tables

Supplementary Table 1. HR-ESI-MS data of all produced peptides.

Compound	MS detected $[M+H]^+$ (* = $[M+2H]^{2+}$)	MS calculated $[M+H]^+$ (* = $[M+2H]^{2+}$)	Molecular formular	Δppm	Reference
1	586.3952	586.3962	$C_{32}H_{51}N_5O_5$	1.9	[10]
2	600.4103	600.4119	$C_{33}H_{53}N_5O_5$	2.7	[10]
3	552.4106	552.4119	$C_{29}H_{53}N_5O_5$	2.4	[10]
4	566.4259	566.4275	$C_{30}H_{55}N_5O_5$	3.0	[10]
5	604.4054	604.4069	$C_{32}H_{53}N_5O_6$	2.5	[10]
6	343.7255*	343.7267*	$C_{33}H_{55}N_{11}O_5$	3.3	
7	326.7336*	326.7345*	$C_{30}H_{57}N_{11}O_5$	2.9	
8	556.3521	556.3527	$C_{27}H_{50}N_5O_5S$	1.0	
9	571.3604*	571.3602*	$C_{60}H_{92}N_{12}O_{10}$	0.5	
10	578.3677*	578.3680*	$C_{61}H_{94}N_{12}O_{10}$	1.1	
11	585.3752*	585.3758*	$C_{62}H_{96}N_{12}O_{10}$	1.1	
12	734.4203	734.4236	$C_{39}H_{55}N_7O_7$	4.5	
13	757.4360	757.4396	$C_{41}H_{56}N_8O_6$	4.7	
14	748.4358	748.4392	$C_{40}H_{57}N_7O_7$	4.6	
15	771.4510	771.4552	$C_{42}H_{58}N_8O_6$	5.5	
16	295.1895*	295.1890*	$C_{30}H_{48}N_6O_6$	1.6	
17	302.1974*	302.1969*	$C_{31}H_{50}N_6O_6$	2.9	
18	358.2701	358.2700	$C_{18}H_{36}N_3O_4$	0.1	
19	358.2699	358.2700	$C_{18}H_{36}N_3O_4$	0.4	
20	372.2855	372.2857	$C_{19}H_{38}N_3O_4$	0.5	
21	372.2854	372.2857	$C_{19}H_{38}N_3O_4$	0.7	
22	392.2539	392.2544	$C_{21}H_{34}N_3O_4$	1.3	
23	392.2540	392.2544	$C_{21}H_{34}N_3O_4$	1.1	
24	406.2702	406.2700	$C_{22}H_{36}N_3O_4$	0.5	
25	406.2697	406.2700	$C_{22}H_{36}N_3O_4$	0.9	
26	411.2966	411.2965	$C_{21}H_{38}N_4O_4$	0.2	
27	429.3066	429.3071	$C_{21}H_{40}N_4O_5$	1.4	
28	443.3224	443.3228	$C_{22}H_{42}N_4O_5$	0.8	
29	477.3065	477.3071	$C_{25}H_{40}N_4O_5$	1.3	
30	491.3223	491.3228	$C_{26}H_{42}N_4O_5$	1.0	
31	443.3220	443.3228	$C_{22}H_{42}N_4O_5$	1.8	
32	457.3377	457.3384	$C_{23}H_{44}N_4O_5$	1.5	

33	459.2964	459.2966	$C_{25}H_{38}N_4O_4$	0.4
34	425.3123	425.3122	$C_{22}H_{40}N_4O_4$	0.2
35	532.3236	532.3242	$C_{26}H_{41}N_7O_5$	1.1
36	518.3082	518.3085	$C_{25}H_{39}N_7O_5$	0.6
37	500.2978	500.2979	$C_{25}H_{37}N_7O_4$	0.2
38	531.3164	531.3177	$C_{28}H_{42}N_4O_6$	2.4
39	545.3321	545.3334	$C_{29}H_{44}N_4O_6$	2.3
40	513.3068	513.3071	$C_{28}H_{40}N_4O_5$	0.7
41	540.3748	540.3756	$C_{27}H_{49}N_5O_6$	1.5
42	595.4276	595.4290	$C_{29}H_{54}N_8O_5$	2.3
43	625.4054	625.4072	$C_{34}H_{52}N_6O_5$	2.9
44	526.3595	526.3599	$C_{26}H_{47}N_5O_6$	0.8
45	489.3100	489.3105	$C_{23}H_{44}N_4O_5S$	1.0
46	544.3484	544.3493	$C_{29}H_{45}N_5O_5$	1.8
47	530.3327	530.3337	$C_{28}H_{43}N_5O_5$	1.9
48	505.3379	505.3384	$C_{27}H_{44}N_4O_5$	1.0
49	471.3533	471.3541	$C_{24}H_{46}N_4O_5$	1.8
50	544.3484	544.3493	$C_{29}H_{45}N_5O_5$	1.7
51	682.4381	682.4399	$C_{35}H_{55}N_9O_5$	2.7
52	528.3366	528.3392	$C_{25}H_{45}N_5O_7$	4.9
53	643.4160	643.4178	$C_{34}H_{54}N_6O_6$	2.8
54	570.4246	570.4225	$C_{29}H_{55}N_5O_6$	3.7
55	505.3363	505.3384	$C_{27}H_{44}N_4O_5$	4.2
56	643.4170	643.4178	$C_{34}H_{54}N_6O_6$	1.2
57	657.4320	657.4334	$C_{35}H_{56}N_6O_6$	2.2
58	639.4205	639.4228	$C_{35}H_{54}N_6O_5$	3.7
59	601.4065	601.4072	$C_{32}H_{52}N_6O_5$	1.1
60	604.3834	604.3869	$C_{32}H_{50}N_5O_5F$	5.7
61	627.3971	627.3977	$C_{32}H_{50}N_8O_5$	0.9
62	640.4079	640.4069	$C_{35}H_{53}N_5O_6$	1.6
63	620.3560	620.3573	$C_{32}H_{50}N_5O_5Cl$	2.1
64	604.3856	604.3869	$C_{32}H_{50}N_5O_5F$	2.2
65	620.3563	620.3573	$C_{32}H_{50}N_5O_5Cl$	1.6
66	664.3053	664.3068	$C_{32}H_{50}N_5O_5Br$	2.3
67	620.3542	620.3573	$C_{32}H_{50}N_5O_5Cl$	5.0
68	604.3862	604.3869	$C_{32}H_{50}N_5O_5F$	1.1
69	644.3042	664.3068	$C_{32}H_{50}N_5O_5Br$	3.9

70	392.2540	392.2540	$C_{21}H_{34}N_3O_4$	1.1
71	555.2166	555.2177	$C_{25}H_{40}BrN_4O_5$	2.0
72	569.2321	569.2333	$C_{26}H_{42}BrN_4O_5$	2.1
73	537.2062	537.2062	$C_{25}H_{38}BrN_4O_4$	1.7

Supplementary Table 2. Strains used and generated in this work.

Strain	Genotype	Reference
E. coli DH10B	F_ mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lacZ*Δ, M15, Δ*lacX*74 *rec*A1 *end*A1 *ara*D 139 Δ(*ara, leu*)7697 *gal*U *gal*K λ*rpsL* (*Strr*) *nup*G	[11]
E. coli DH10B::mtaA	DH10B with mtaA from pCK_mtaA ΔentD	[12]
E. coli BL21 (DE3) Star	F- *omp*T *hsd*SB(rB-, mB-) *galdcmrne*131 (DE3)	Invitrogen
E. coli BL21 (DE3) Star pCOLA_g*xpS*_A3	BL21 (DE3) Star:pCOLA_*gxpS*_A3, pTf16, Km^R, Cm^R	This work
S. cerevisiae CEN.PK 2-1C	MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2	Euroscarf
P. luminescens TT01		DSMZ
X. nematophila ATCC 19061		ATCC
X. miraniensis DSM 17902		DSMZ
X. budapestensis DSM 16342		DSMZ
X. indica DSM 17382		DSMZ
X. szentirmaii DSM 16338		DSMZ
X. bovienii SS2004		DSMZ
X. doucetiae DSM 17909		DSMZ
B. licheniformis ATCC 10716		M. A. Marahiel / ATCC
B. subitlis MR 168		ATCC
A. migulanus ATCC9999		ATCC
B. brevis ATCC 8185		ATCC
E. coli DH10B::mtaA pFF1_*grsTAB*_WT	*E. coli* DH10B::mtaA pFF1_*grsAB*_WT, Kan^R	This work
E. coli DH10B::mtaA pFF1_*gxpS*_WT	*E. coli* DH10B::mtaA pFF1_*gxpS*_WT, Kan^R	[9]
E. coli DH10B::mtaA pFF1_*xtpS*_WT	*E. coli* DH10B::mtaA pFF1_*xtpS*_WT, Kan^R	[9]
E. coli DH10B::mtaA pFF1_*NRPS*_0	*E. coli* DH10B::mtaA pFF1_*NRPS*_0, Kan^R	[9]
E. coli DH10B::mtaA pFF1_*NRPS*_1	*E. coli* DH10B::mtaA pFF1_*NRPS*_1, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_2	*E. coli* DH10B::mtaA pFF1_*NRPS*_2, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_3	*E. coli* DH10B::mtaA pFF1_*NRPS*_3, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_4	*E. coli* DH10B::mtaA pFF1_*NRPS*_4, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_5	*E. coli* DH10B::mtaA pFF1_*NRPS*_5, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_6	*E. coli* DH10B::mtaA pFF1_*NRPS*_6, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_7	*E. coli* DH10B::mtaA pFF1_*NRPS*_7, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_8	*E. coli* DH10B::mtaA pFF1_*NRPS*_8, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_9	*E. coli* DH10B::mtaA pFF1_*NRPS*_9, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_18	*E. coli* DH10B::mtaA pFF1_*NRPS*_18, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_19	*E. coli* DH10B::mtaA pFF1_*NRPS*_19, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_20	*E. coli* DH10B::mtaA pFF1_*NRPS*_20, Kan^R	This work
E. coli DH10B::mtaA pFF1_*NRPS*_21	*E. coli* DH10B::mtaA pFF1_*NRPS*_21, Kan^R	This work
E. coli DH10B::mtaA pFF1_*library*_1	*E. coli* DH10B::mtaA pFF1_*library*_1, Kan^R	This work
E. coli DH10B::mtaA pFF1_*library*_2	*E. coli* DH10B::mtaA pFF1_*library*_2, Kan^R	This work
E. coli DH10B::mtaA pFF1_*library*_3	*E. coli* DH10B::mtaA pFF1_*library*_3, Kan^R	This work

Supplementary Table 3. Plasmids used and generated in this work.

Plasmid	Genotype	Reference
pAT41	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	This work
pTF16	Chaperone tig, L-arabinose inducible Promotor *araB*, Cm^R	TaKaRa Bio Inc., Singapore
pCOLADuet-1	3719 bp vector, T7 promotor-1, T7 promotor-2, His•Tag® coding sequence, Multiple cloning sites-1 (*Nco* I-*Afl* II), Multiple cloning sites-2 (*Nde* I-*Avr* II), T7 transcription start-1, T7 transcription start-2, S•Tag™ coding sequence, T7 terminator, Kan^R, ColA ori, *lacI* coding sequence	Merck/Millipore KGaA, Darmstadt
pCOLA_*gxpS*_A3	ColA ori, Kan^R, T7 promotor, *gxpS_A3*	This work
pFF1_Ypet	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	g
pFF1_*grsTAB*_WT	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*	This work
pFF1_*gxpS*_WT	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*	g
pFF1_*xtpS*_WT	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*	g
pFF1_*NRPS*_0	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bicA*-A1T1C2_*gxpS*-A2T2C3A3T3C4A4T4C5A5T5TE	g
pFF1_*NRPS*_1	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*-A1T1C2_*bicA*-A2T2C3_*gxpS*-A3T3C4A4T4C5A5T5TE	This work
pFF1_*NRPS*_2	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*-A1T1C2_*bicA*-A2T2C_{Dsub}3_*gxpS*-C_{Asub}3A3T3C4A4T4C5A5T5TE	This work
pFF1_*NRPS*_3	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*-A1T1C_{Dsub}1_*ambS*-C_{Asub}4A4T4C_{Dsub}5_*gxpS*-C_{Asub}3A3T3C_{Dsub}4_*garS*-C_{Asub}4A4T4C_{Dsub}5_*gxpS*-C_{Asub}5A5T5TE	This work

pFF1_*NRPS*_4	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R,*xtpS*-A1T1C_{Dsub}1_*ambS*-C_{Asub}4A4T4C_{Dsub}5_*gxpS*-C_{Asub}3A3T3C_{Dsub}4_*hctaA*-C_{Asub}34A4T4C_{Dsub}5_*gxpS*-C_{Asub}5A5T5TE	This work
pFF1_*NRPS*_5	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bicA*-A1T1C2_*gxpS*-A2T2C3A3T3C4A4T4C_{Dsub}5_*bicA*-C_{Asub}5A5T5C_{term}	This work
pFF1_*NRPS*_6	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bacA*-A1T1CyA2T2C3A3T3C_{Dsub}4_*sfrA-BC*-C_{Asub}6A6T6E6C7A7T7TE	This work
pFF1_*NRPS*_7	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*-A1T1E2C2A2T2C3A3T3 C_{Dsub}4_ *tycC*-C_{Asub}9A9T9C_{Dsub}10_ *grsAB*-C_{Asub}5A5T5TE	This work
pFF1_*NRPS*_8	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*-A1T1E2C2A2T2C_{Dsub}3_ *tycC*-C_{Asub}7A7T7C8A8TC9A9T9C10A10T10TE	This work
pFF1_*NRPS*_9	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*-A1T1E2C2A2T2C3A3T3C_{Dsub}4_*tycC*-C_{Asub}9A9T9C10A10T10TE	This work
pFF1_*NRPS*_18	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*-A3T3C_{Dsub}4_*gxpS*-C_{Asub}2A2T2C3A3T3C4A4T4C5A5T5TE	This work
pFF1_*NRPS*_19	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*-C_{Asub}2A3T3C_{Dsub}4_*gxpS*-C_{Asub}2A2T2C3A3T3C4A4T4C5A5T5TE	This work
pFF1_*NRPS*_20	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*-C2A3T3C_{Dsub}4_*gxpS*-C_{Asub}2A2T2C3A3T3C4A4T4C5A5T5TE	This work
pFF1_*NRPS*_21	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*-A1T1C2A2T2C_{Daub}3_*gxpS*-C_{Asub}3A3T3C_{Dsub}4_*xtpS*-C_{Asub}4A4T4TE	This work

pFF1_*library*_1	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, random sequences	This work
pFF1_*library*_2	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, random sequences	This work
pFF1_*library*_3	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, random sequences	This work

Supplementary Table 4. Oligonucleotides used in this work.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pCOLA_*gxpS*_A3	pCOLA_Gib_A3 Insert FW	CATCACCATCATCACCACCCTCAACAACCTGTCACGGC	*P. luminescens* TT01
	pCOLA_Gib_A3 Insert RV	CAGCCTAGGTTAATTAAGCTGTTAAGTCAGATCAATCAGCGGCAAC	
	DUET_Gib_FW	CAGCTTAATTAACCTAGGCTG	pCOLADUET-1
	DUET_Gib_RV	GTGGTGATGATGGTGATG	
pFF1_*grsTAB*_WT	grs fw1	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	grs rv1	GTCTTTCCATCCAACTGCAAC	
	grs fw2	CAGAAATCGAGATATTGTCTGAAG	*A. migulanus* ATCC9999
	grs rv2	ACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGA TCCGGCGCGCCGAGCTCTTATTTTACTACAAATGTCCCTTGTAG	
pFF1_*NRPS*_1	KB-RT-6	TCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAATTC	*P. luminescens* TT01
	KB-RT-7	TGGAATGCGACCGAAGAACC	
	KB-RT-8	CACATACCTGAGTAGGATACGGTTCTTCGGTCGCATTCCAAGTTTTCAGCAACAACTGGC	*X. budapestensis* DSM 16342
	KB-RT-9	TGTTTTGCCTGCATCGGAACGCACGTTGTTGCTGGAAACGTGGAATACAACGGAAACTGC	
	KB-RT-10	CGTTTCCAGCAACAACG	*P. luminescens* TT01
	KB-RT-11	TTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG G	
pFF1_*NRPS*_2	KB-RT-6	TCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAATTC	*P. luminescens* TT01
	KB-RT-7	TGGAATGCGACCGAAGAACC	
	KB-RT-15	AAGCCATTGACGCTGCCAG	*X. budapestensis* DSM 16342
	KB-RT-9	TGTTTTGCCTGCATCGGAACGCACGTTGTTGCTGGAAACGTGGAATACAACGGAAACTGC	
	KB-RT-10	CGTTTCCAGCAACAACG	*P. luminescens* TT01
	KB-RT-11	TTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG G	
	KB-RT-13	CACATACCTGAGTAGGATACGGTTCTTCGGTCGCATTCCAATTTTCCAGTAATAACTCCCGCTC	*P. luminescens* TT01
	KB-RT-14	TCAATATCCTGATTATGCGGTCTGGCAGCGTCAATGGCTTTCAGGTGAAGGAGTACAGGC	
pFF1_*NRPS*_3	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTA AAAAGG	*X. nematophila* ATCC 19061
	AL-GxpS-2-2	CCCAATCAACATATCGGTAAAAAAGCGAGTATGTTCCATCTGGCTCACCCCCTGGTGGGCC	
	AL-GxpS-2-3	CCCGTACCTTTCCGTAATCTGGTCGCTCAGGCCCACCAGGGGGTGAGCCAGATGGAACATACT CG	*X. miraniensis* DSM 17902
	AL-GxpS-2-4	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAAAACCACTGGCGTTGCC	
	AL-GxpS-2-5	TCAGGTGAAGGAGTACAGGCAC	*P. luminescens* TT01
	AL-GxpS-2-6	GACACCCTGCCGAGCC	
	AL-GxpS-2-7	CCGGTCCCGTTCCGCCATTTAGTGGCACAGGCTCGGCAGGGTGTCCAAGGCGCTGTTCTCACT	*X. bovienii* SS2004

		G	
	AL-GxpS-2-8	CTCAGCCAACATTTCAGTAAAGAAACGGGTATGTTCAGCCTGACTCACGCTCAGGGTCTGGG	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	*P. luminescens* TT01
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_*NRPS*_4	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTA AAAAGG	*X. nematophila* ATCC 19061
	AL-GxpS-2-2	CCCAATCAACATATCGGTAAAAAAGCGAGTATGTTCCATCTGGCTCACCCCCTGGTGGGCC	
	AL-GxpS-2-3	CCCGTACCTTTCCGTAATCTGGTCGCTCAGGCCCACCAGGGGGTGAGCCAGATGGAACATACT CG	*X. miraniensis* DSM 17902
	AL-GxpS-2-4	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAAAACCACTGGCGTTGCC	
	AL-GxpS-2-5	TCAGGTGAAGGAGTACAGGCAC	*P. luminescens* TT01
	AL-GxpS-2-6	GACACCCTGCCGAGCC	
	AL-GxpS-2-11	GCCGGTCCCGTTCCGCCATTTAGTGGCACAGGCTCGGCAGGGTGTCAGTCAGGAAGCCCACA CC	*X. miraniensis* DSM 17902
	AL-GxpS-2-12	CTCAGCCAACATTTCAGTAAAGAAACGGGTATGTTCAGCCTGACTCACCCCCAACCGAACC	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	*P. luminescens* TT01
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_*NRPS*_5	ML020	AGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTT TTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTG	pFF1_*NRPS*_0
	FF_305	CCAATAATCACTCTGTGCCTG	
	FF_306	GAACTGGTTGCACTTTACGC	
	FF_307	CATCCCTAACCGGGACTG	
	FF_308	CATACCTTGACGGACAGGGCGGCAACCTGCCAGCGCCGGCACCCTTCCGCAATCTGGTAGCG CAGTCCCGGTTAGGGATGAGTCAGGCAGCCCATACC	*X. budapestensis* DSM 16342
	FF_309	AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCC AGCGGATCCGGCGCGCCCTATTGCTCTGCTGATATCAGAA	
pFF1_*NRPS*_6	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	*B. licheniformis* ATCC 10716
	AT_107	GGCATGGCTATTTTCCCATTC	
	AT_157	GGAGCGCTGTATGAAGGGAAATCACTGAAGCCGTTCCATATTCAATATAAAGATTACGCAGAAT GGGAAAATAGCCATGCCCAAAAAGAACGGATGAAGGAGC	*B. subtilis* MR 168
	AT_118	AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCC AGCGGATCCGGCGCGCCTTATGAAACCGTTACGGTTTGTGTATTAAG	
pFF1_*NRPS*_7	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	Grs Bib P4 rv	AAACCATTCGTTTTGCCATAC	
	tycC eXU9_grs P4 fw	GAGCTTGCTGAACTGCATATTCAGTATAAAGATTTTGCTGTATGGCAAAACGAATGGTTTCAGTC GGATCGCTTCCA	*B. brevis* ATCC 8185
	tycC eXU9_grs P4 rv	GCGTTCATCTGCGAAGACATTTACCCAGTATTCTTCTTGCTTTTTAAAGGCTTCTGACTGGGAAA ACTCTTGTTGCCACAC	
	grs Bib P4 fw2	CAGTCAGAAGCCTTTAAAAAGC	*A. migulanus* ATCC9999
	FF_319	ACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGA	

		TCCGGCGCGCCGAGCTCTTATTTTACTACAAATGTCCCTTGTAG	
pFF1_*NRPS*_8	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACC GAGGAATTTGCC	*B. brevis* ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCGGCGCGCCTTATT TCAGGATGAACAGTTCTTGC	
pFF1_*NRPS*_9	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_376	AAACCATTCGTTTTGCCATAC	
	FF_377	GAGCTTGCTGAACTGCATATTCAGTATAAAGATTTTGCTGTATGGCAAAACGAATGGTTTCAGTC GGATCGCTTCCAAAAAC	*B. brevis* ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCGGCGCGCCTTATT TCAGGATGAACAGTTCTTGC	
pFF1_*NRPS*_18	ML_P1	GACCAGACAGAACATCACCG	pFF1_*gxpS*_WT
	ML_P2	GGCCCAATCCTATACGCC	
	ML_P3	CTTACCAAGCGCCACAAGG	
	ML_P4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCA	
	ML_P5	TCGGTCAGCCCAAACGGTAATGTCGGTTCATCCACTTCTGCCAACATGTCGGTAAAGAATCGGT GATGTTCTGTCTGGTCGACACCCTGCCGAGCC	*X. nematophila* HGB081
	ML_P6.1	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCATGCGGGCAATGGTGAACC	
pFF1_*NRPS*_19	ML_P1	GACCAGACAGAACATCACCG	pFF1_*gxpS*_WT
	ML_P2	GGCCCAATCCTATACGCC	
	ML_P3	CTTACCAAGCGCCACAAGG	
	ML_P4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCA	
	ML_P5	TCGGTCAGCCCAAACGGTAATGTCGGTTCATCCACTTCTGCCAACATGTCGGTAAAGAATCGGT GATGTTCTGTCTGGTCGACACCCTGCCGAGCC	*X. nematophila* HGB081
	ML_P6.3	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCATGTCTGATGAAGGCGTGC	
pFF1_*NRPS*_20	ML_P1	GACCAGACAGAACATCACCG	pFF1_*gxpS*_WT
	ML_P2	GGCCCAATCCTATACGCC	
	ML_P3	CTTACCAAGCGCCACAAGG	
	ML_P4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCA	
	ML_P5	TCGGTCAGCCCAAACGGTAATGTCGGTTCATCCACTTCTGCCAACATGTCGGTAAAGAATCGGT GATGTTCTGTCTGGTCGACACCCTGCCGAGCC	*X. nematophila* HGB081

	ML_P6.2	ATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCATGGCATTTACCGAAAAGATCTGCG	
pFF1_*NRPS*_21	ML_P7	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* HGB081
	ML_P8	CTATCGGCAATTCAAGTAACACCGGTGCATCTGCCAACGTCCGACGCCAATAATCACTCTGTGC CTGTACTCCTTCACCTGAAAATACCTGCCGCTGCC	
	ML_P9	GCTTGTCTGAATCAACAACCTGATCCGCTGCCGCCATTGACCATTCAATATCCTGATTATGCTG CCTGGCAGCGGCAGGTATTTTCAGGTGAAGGAGTACAGGC	
	ML_P10	GCACTTCCGACAATCCAAATGACAACGTTGGCTCATCTACCTCAGCCAACATATCGGTAAAGAA ACGGGTATGTTCTGCCTGACTGACACCCTGCCGAGCC	
	ML_P11	GGCTTGCCTCTTGGGGCAAATGGATAGCCTGCCTGCGCCGGTCCCGTTCCGCCATTTAGTGGC ACAGGCTCGGCAGGGTGTCAGTCAGGCAGAACATACCCG	*X. nematophila* HGB081
	ML_P12	CCAGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGC AGCGGAGCCAGCGGATCCCAGCGCCTCCACTTCG	
pFF1_*library*_1	KB-AmbF-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGTTGAACGCTTACAAT CC	*X. miraniensis* DSM 17902
	KB-AmbF-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGATACCCAGCCGGGCTTG	
	KB-AmbW-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCATCGGAACGGGTACAAA TTC	*X. indica* DSM 17382
	KB-AmbW-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACCCCCATCCGTGCCTG	
	KB-Thr1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGCACAGATACAGT CTC	*P. luminescens* TT01
	KB-Thr2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Arg1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGG TGCAG	*X. budapestensis* DSM 16342
	KB-Arg2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Ser1	GTCTAAATCAACAGCCAGATCCGTTGTTGCCATTGACCATTCAATATCCTGATTATGCGGCTTG GCAGCGGCAGGTATTTCAGGGTGACCGCCTGAC	*X. szentirmaii* DSM 16338
	KB-Ser2	TCCGCCAACCCAAATAGCAGCGTCGGTTCATCCACCTCAGCCAACATGTCAGTAAAGAAGCGA GTATGTTCTGCCTGACTCACACTCAGGATTTGAGCGATAAAG	
	KB-Lys1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGAGGTACTGGAA AAGC	*X. nematophila* HGB081
	KB-Lys2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACACTGCGGGTTTGGGC	
	AL-GxpS-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTA AAAAGG	pFF1_*gxpS*_WT
	AL-GxpS-2	AAATACCTGCCGCTGCC	
	AL-GxpS-3	AGTCAGGCAGAACATACTCGCTTCTTTAC	
	AL-GxpS-4	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_*library*_2	KB-xeyS-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCTTCCCCCAACCAGGACTG	*X. indica* DSM 17382
	KB-xeyS-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATGGCTAC AACG	
	KB-BicA-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCCATCCCCAACCAGGACTG	*X. budapestensis* DSM 16342

	KB-BicA-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTGG	
	KB-17902-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCAACGCCCAGCCGGGCTTGAGC	*X. miraniensis* DSM 17902
	KB-17902-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAAATGATAAGGTGATGACTCTGC	
	KB-2022-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCCACCCCCTGGTGGGCC	*X. nematophila* HGB081
	KB-2022-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	
	KB-XLSer1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGACCGCCTGAC	*X. szentirmaii* DSM 16338
	KB-XLSer2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACACTCAGGATTTGAGCGATAAAG	
	KB-AmbF-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGTTGAACGCTTACAATCC	*X. miraniensis* DSM 17902
	KB-AmbF-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGATACCCAGCCGGGCTTG	
	KB-AmbW-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCATCGGAACGGGTACAAATTC	*X. indica* DSM 17382
	KB-AmbW-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACCCCCATCCGTGCCTG	
	KB-Thr1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGCACAGATACAGTCTC	*P. luminescens* TT01
	KB-Thr2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Arg1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGGTGCAG	*X. budapestensis* DSM 16342
	KB-Arg2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Lys1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGAGGTACTGGAAAAGC	*X. nematophila* HGB081
	KB-Lys2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACACTGCGGGTTTGGGC	
	AL-GxpS P3	AGTCAGGCAGAACATACTCGCTTCTTTAC	pFF1_*gxpS*_WT
	AL-GxpS P4	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
	KB-Lib3-1	GACCAGACAGAACATCACCG	
	KB-Lib3-2	AAATACCTGCCGCTGCC	
pFF1_*library*_3	KB-XL-X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACACTCAGGATTTGAGCG	*X. szentirmaii* DSM 16338
	KB-XL-X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGACCGCCTGACC	
	KB-Xl-amb X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCATACCTAGACGTGCCTGTGC	*X. indica* DSM 17382
	KB-Xl-amb X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGAACGGATACAAATTC	
	KB-Kol X3 - RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	*P. luminescens* TT01
	KB-Kol X3 - FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGCACAGATACAGTC	
	KB-Kol X2 RV	AAATACCTGCCGCTGCCAAGCCGCATAATCAGGATATTGAATCGTCAACGGTAGCAACGG	*P. luminescens* TT01
	KB-Kol X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGTCAGGCTGAGCATACCCG	
	KB-BicA X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	*X. budapestensis*

			DSM 16342
	KB-BicA X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGGTGC	
	KB-Bb 2 RV	TCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAATTC	pFF1_*gxpS*_WT
	KB-Bb 2 FW	AGTCAGGCAGAACATACTCGC	
	KB-Bb 1 RV	AACCCCCTGCCGAGCC	pFF1_*gxpS*_WT
	KB-Bb 1 FW	TTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	
	KB-amb X2 RV	AAATACCTGCCGCTGCCAAGCCGCATAATCAGGATATTGAATTGCCAATGGTGGCAAGGG	*X. indica* DSM 17382
	KB-amb X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGCCAGACAGAGCACACCCG	
	KB-17902 X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGATACCCAGCCGGGCTTGTGC	*X. miraniensis* DSM 17902
	KB-17902 X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGTTGAACGCTTACAATCC	
	KB-2022 X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGACACCCTGCCGAGCC	*X. nematophila* HGB081
	KB-2022 X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGATGAAGGCGTGCAGG	
	KB-2022 X2 RV	AAATACCTGCCGCTGCCAAGCCGCATAATCAGGATATTGAATGGTCAATGGCGGCAGCGG	*X. nematophila* HGB081
	KB-2022 X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGTCAGGAAGCGTACACGCG	

Supplementary Table 5. Strains used and generated for the fusion of NRPS from Gram-positive and –negative origin.

Strain	Genotype	Reference
E. coli DH10B::mtaA pAT41_*NRPS*_10	*E. coli* DH10B::mtaA pAT41_*NRPS*_10, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_11	*E. coli* DH10B::mtaA pAT41_*NRPS*_11, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_12	*E. coli* DH10B::mtaA pAT41_*NRPS*_12, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_13	*E. coli* DH10B::mtaA pAT41_*NRPS*_13, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_14	*E. coli* DH10B::mtaA pAT41_*NRPS*_14, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_15	*E. coli* DH10B::mtaA pAT41_*NRPS*_15, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_16	*E. coli* DH10B::mtaA pAT41_*NRPS*_16, Kan^R	This work
E. coli DH10B::mtaA pAT41_*NRPS*_17	*E. coli* DH10B::mtaA pAT41_*NRPS*_17, Kan^R	This work

Supplementary Table 6. Plasmids used and generated for the fusion of NRPS from Gram-positive and –negative origin.

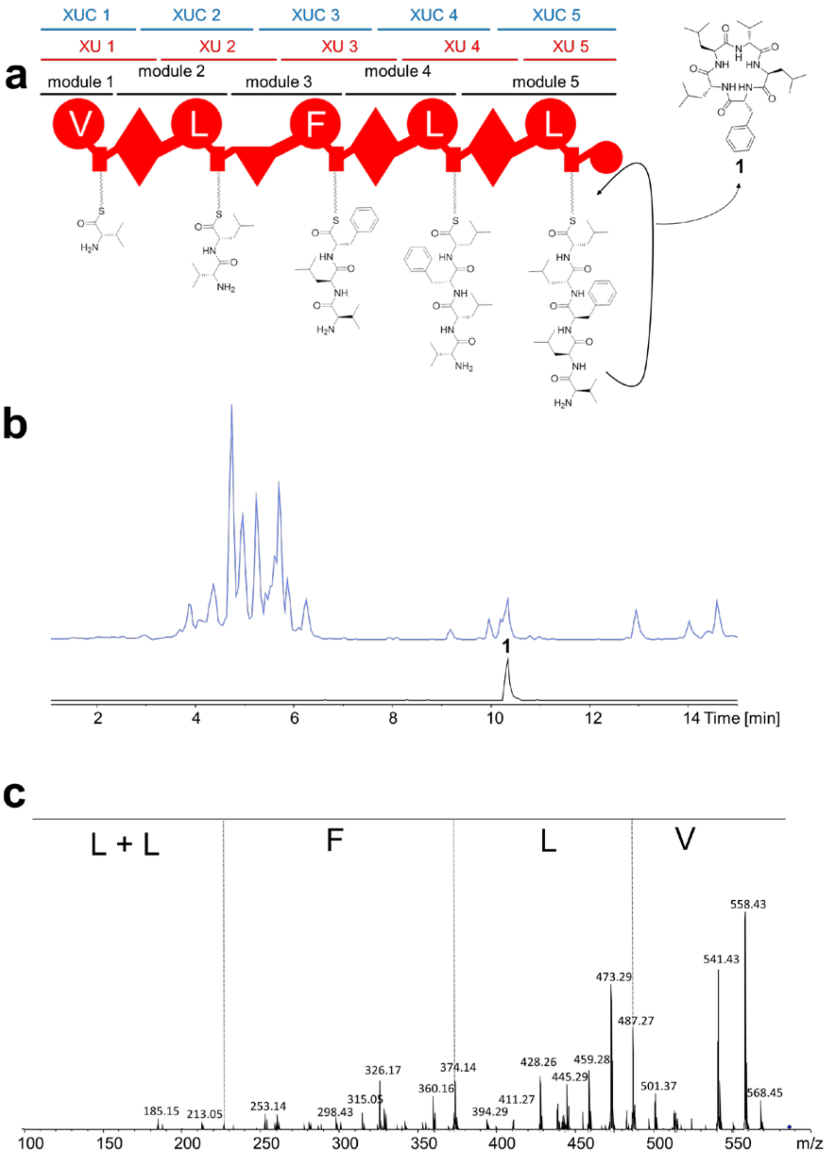
Plasmid	Genotype	Reference
pFF1_*NRPS*_10	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*_A1T1E2C2A2T2C_{Dsub}3-*tycC*_C_{Asub}7A7T7C8A8T8C_{Dsub}9-*odl4*_C_{Asub}9A9T9C_{Dsub}10-*tycC*_C_{Asub}10A10T10TE	This work
pFF1_*NRPS*_11	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*_A1T1E2C2A2T2C_{Dsub}3-$xabA_{dou}$_C_{Asub}4A4T4CE5A5T5C_{Dsub}6-*tycC*_C_{Asub}9A9T9C10A10T10TE	This work
pFF1_*NRPS*_12	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*_A1T1E2C2A2T2C_{Dsub}3-*tycC*_C_{Asub}7A7T7C_{Dsub}8-*txlA*_C_{Asub}2A2T2C_{Dsub}3-*tycC*_C_{Asub}9A9T9C10A10T10TE	This work
pFF1_*NRPS*_13	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*_A1T1E2C2A2T2C_{Dsub}3-*xmaS*_C_{Asub}2A2T2C_{Dsub}3-*tycC*_C_{Asub}8A8T8C9A9T9C10A10T10TE	This work
pFF1_*NRPS*_14	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *grsAB*_A1T1E2C2A2T2C_{Dsub}3-*tycC*_C_{Asub}7A7T7C_{Dsub}8-*xabC*_C_{Asub}8A8T8C_{Dsub}9-*tycC*_C_{Asub}9A9T9C10A10T10TE	This work
pFF1_*NRPS*_15	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bacA*_ A1T1Cy2A2T2C_{Dsub}3-*gxpS*_C_{Asub}3A3T3CE4A4T4CE5A5T5TE	This work
pFF1_*NRPS*_16	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bacA*_ A1T1Cy2A2T2C3A3T3C_{Dsub}4-*gxpS*_C_{Asub}3A3T3CE4A4T4CE5A5T5TE	This work
pFF1_*NRPS*_1	2µ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1CE2A2T2C_{Dsub}3-*tycC*_C_{Asub}8A8T8C9A9T9C10A10T10TE	This work

Supplementary Table 7. Oligonucleotides used for the fusion of NRPS from Gram-positive and –negative origin.

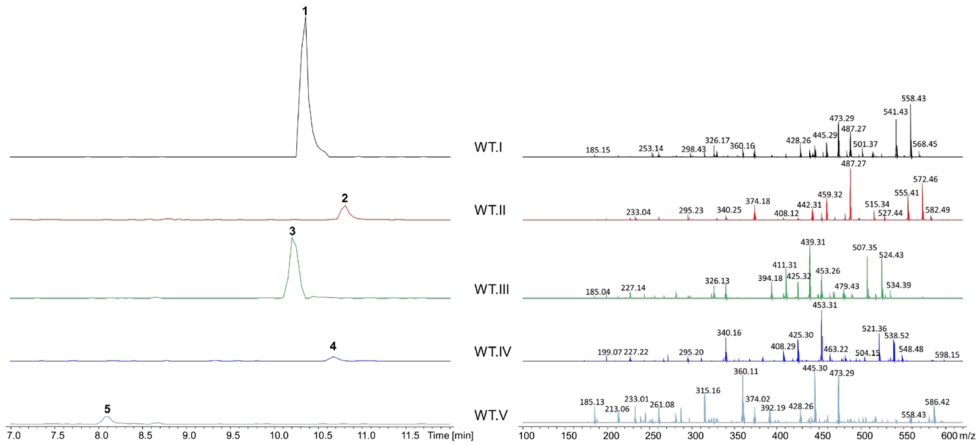
Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pFF1_*NRPS*_10	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACCGA GGAATTTGCC	*B. brevis* ATCC 8185
	AL_C9_rv	AGCCAGCTTGGTCTGCC	
	AL_odl_fw	CCAGTACAAAGACTTTGCTGTGTGGCAGACCAAGCTGGCTGCTGATGGCATATTTGATGC	*X. nematophila* ATCC 19061
	AL_odl_rv	TGCAACCAGTACGCTTCCTGCTTGTGGAAGGCAGCCGACTGTAATTGTTGTCGCTGCCAG	
	AL_C10_fw	CAGTCGGCTGCCTTCC	*B. brevis* ATCC 8185
	AL_tycC_rv	CTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_*NRPS*_11	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	AL_xabdou_fw	CAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCCGGAAGCTAAACTGACGG	*X. doucetiae* DSM 17909
	AL_xabdou_rv	CCGGGTCCAAAAATCCTCCTGTTTTTGGAAGCGATCCGACTGCAGCCTGTCACGCTGC	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	*B. brevis* ATCC 8185
	AL_tycC_rv	CTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_*NRPS*_12	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACCGA GGAATTTGCC	*B. brevis* ATCC 8185
	AL_C8_rv	GAACAGTTCAGACTGCCAGAC	
	AL_txl_fw	CCATTACAAAGATTTCGCCGTCTGGCAGTCTGAACTGTTCCAGGGGGAGGTGGGGG	*X. bovienii* SS2004
	AL_txl_rv	GGGTCCAAAAATCCTCCTGTTTTTGGAAGCGATCCGACTGCAAATAGTCACGCTGCCATTT	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	*B. brevis* ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACA GTTCTTGC	
pFF1_*NRPS*_13	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	

	AL_xma_fw	TCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGGGTGACATACGTGAACG	*X. nematophila* ATCC 19061
	AL_xma_rv	TCAGCCAGTAGTTTTCATGCTCGGTATAGACGTCGCTCTGTAAATAGTCGCGTTGCCATG	
	AL_C8_fw	CAGAGCGACGTCTATACCGAGCATGAAAACTAC	*B. brevis* ATCC 8185
	AL_tycC_rv	CTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_*NRPS*_14	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	*A. migulanus* ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACCGA GGAATTTGCC	*B. brevis* ATCC 8185
	AL_C8_rv	GAACAGTTCAGACTGCCAGAC	
	AL_xabnem_fw	CGTCCATTACAAAGATTTCGCCGTCTGGCAGTCTGAACTGTTCCAGGGCAATGCCCTGAC	*X. nematophila* ATCC 19061
	AL_xabnem_rv	GGGTCCAAAAATCCTCCTGTTTTTGGAAGCGATCCGACTGCAACATATCATGTTGCCAGACAG	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	*B. brevis* ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACA GTTCTTGC	
pFF1_*NRPS*_15	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	*B. licheniformis* ATCC 10716
	ALAT_1	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAATTAAATGTATGATTCCATTCCA CATAATC	
	AT_99	TCAGGTGAAGGAGTACAGGCAC	pFF1_*gxpS*_WT
	ABCR14_neu	GAAACGGGTATGTTCAGCCTGAC	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	pFF1_*gxpS*_WT
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_*NRPS*_16	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAC AGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	*B. licheniformis* ATCC 10716
	ALAT_2	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAGGCATGGCTATTTTCCCATTC	
	AT_99	TCAGGTGAAGGAGTACAGGCAC	pFF1_*gxpS*_WT
	ABCR14_neu	GAAACGGGTATGTTCAGCCTGAC	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	pFF1_*gxpS*_WT
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_*NRPS*_17	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAA GG	pFF1_*gxpS*_WT
	AT_217	AAATACCTGCCGCTGCC	
	ALAT_3	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGTCGGATCGCTTCCAAAAAC	*B. brevis* ATCC 8185
	ALAT_4	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTG CAGG	

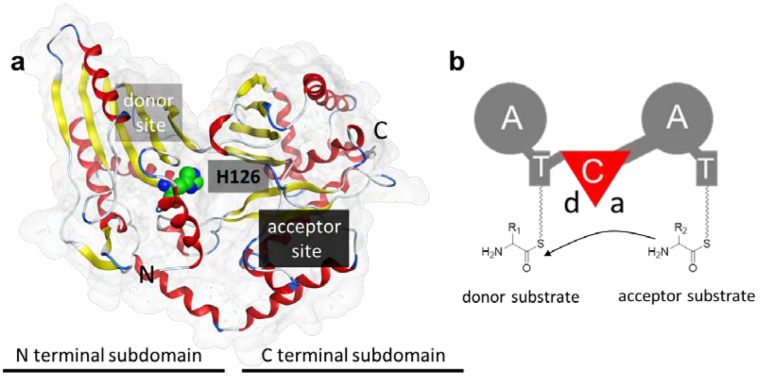
Supplementary Figures



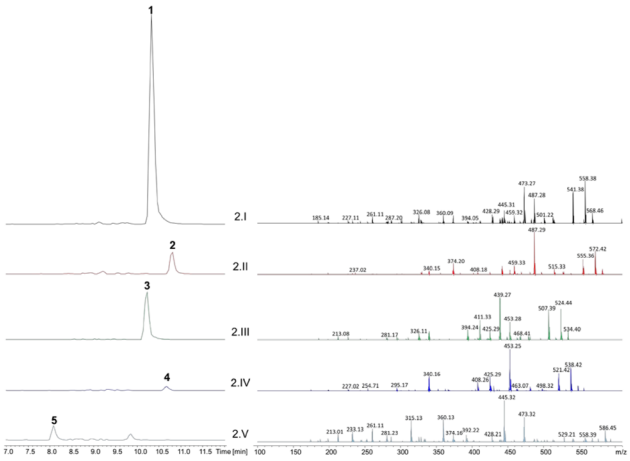
Supplementary Figure 1. Heterologous production of GameXPeptide in *E. coli* DH10B::mtaA. Schematic representation of the GxpS assembly line (**a**). Base peak chromatogram (blue) and extracted ion chromatogram (black) of **1** (*m/z* $[M+H]^+$ = 586.4) (**b**). MS-MS spectra of **1** (**c**).



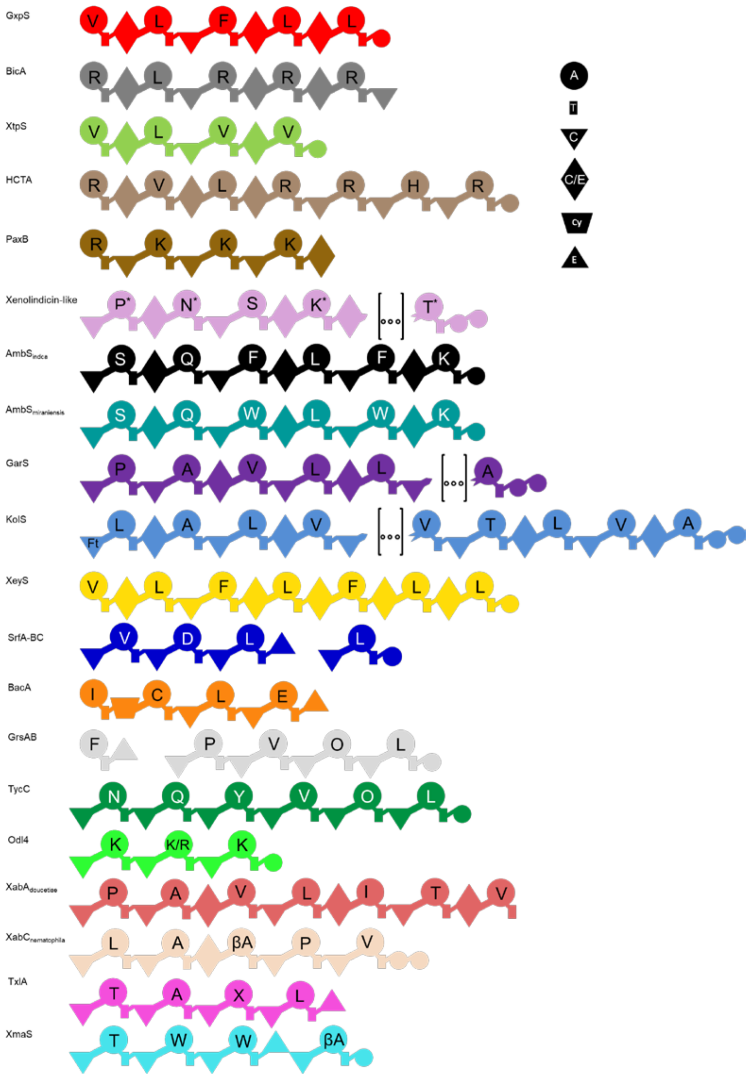
Supplementary Figure 2. HPLC/MS data of GameXPeptides.Extracted ion chromatograms (left) and HPLC/MS^2 data (right)of GameXPeptide (**1**, *m/z* $[M+H]^+$ = 586.4) and its derivatives**2** (*m/z* $[M+H]^+$ = 600.4), **3** (*m/z* $[M+H]^+$ = 552.4), **4** (*m/z* $[M+H]^+$ = 566.4) and the linear GameXPeptide**5** (*m/z* $[M+H]^+$ = 604.4) all from *E. coli* DH10B::mtaA extracts.



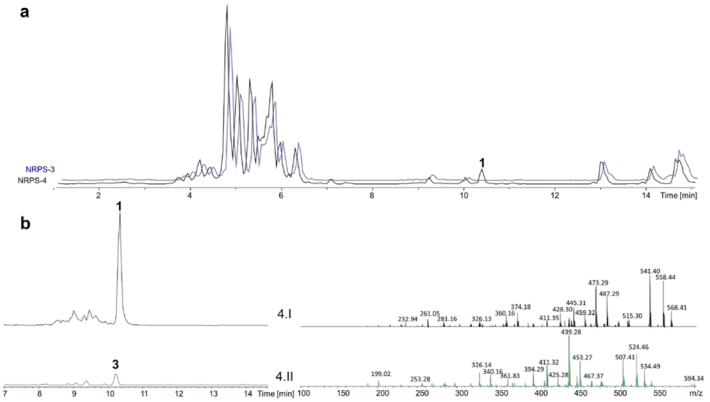
Supplementary Figure 3. Structure and mechanism of a C domain.(**a**) Crystal structureof VibH, a C domain of *V. cholera* vibriobactin synthethase (PDB-ID: 1L5A)[13],subdivided into N terminal subdomain (donor site) and C terminal subdomain (acceptor site). The catalytic center (H126) is highlighted in green. (**b**) C domain catalyzes the nucleophilic attack of the T domain bound acceptor substrate to the T domain bound donor substrate. During peptide bond formation the donor substrate ("donating" the peptide chain) is attackedby the amino-group of the acceptor substrate (that thereby "accepts" the peptide chain).



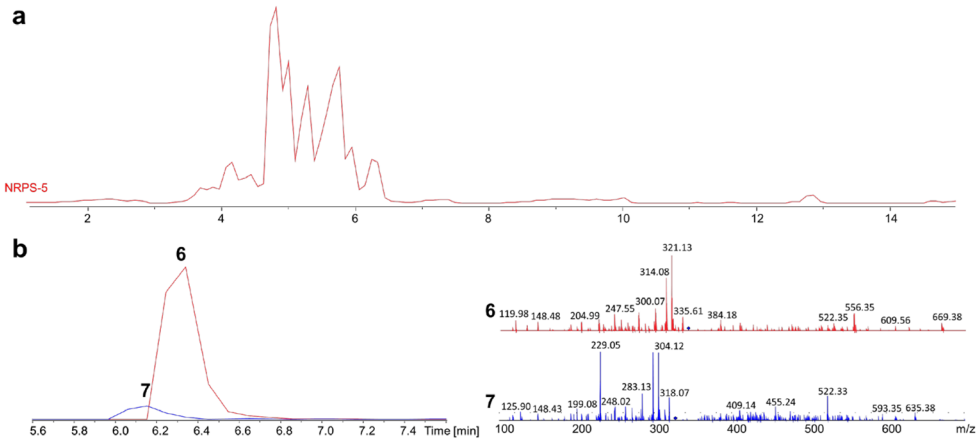
Supplementary Figure 4. HPLC/MS data of compounds **1-5** produced by NRPS-2 in *E. coli* DH10B::mtaA.Extracted ion chromatograms (left) and HPLC/MS^2 data (right) of GameXPeptide (**1**, *m/z* $[M+H]^+$ = 586.4) and its derivatives **2** (*m/z* $[M+H]^+$ = 600.4), **3** (*m/z* $[M+H]^+$ = 552.4), **4** (*m/z* $[M+H]^+$ = 566.4) and **5**(*m/z* $[M+H]^+$ = 604.4) produced by NRPS-2 (Figure 1b) in *E. coli* DH10B::mtaA.



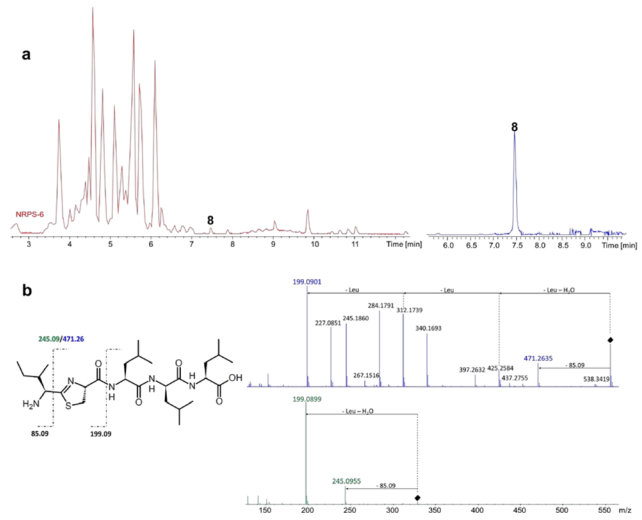
Supplementary Figure 5. Schematic overview of all NRPS used in this work. GxpS[14], BicA[15], XtpS[16], HCTA[15], PaxB[17], KolS[18], $AmbS_{mir}$ from *X. miraniensis*[12], $AmbS_{ind}$ from *X. indica*[19], SrfA-BC[20], BacA[21], GrsAB[22], TycC[23], Odl4[24], $XabB_{dou}$ from *X. doucetiae*[25], $XabB_{nem}$ from *X. nematophila*[19], TxlA[2] and XmaS[26]have been described previously. For GarS producing gargantuanin see Genbank accession number PRJNA224116. For Xenolindicin-like syntethase see Genbank accession number PRJNA328553. For XeyS producing xindeyrin see Genbank accession number PRJNA328572.



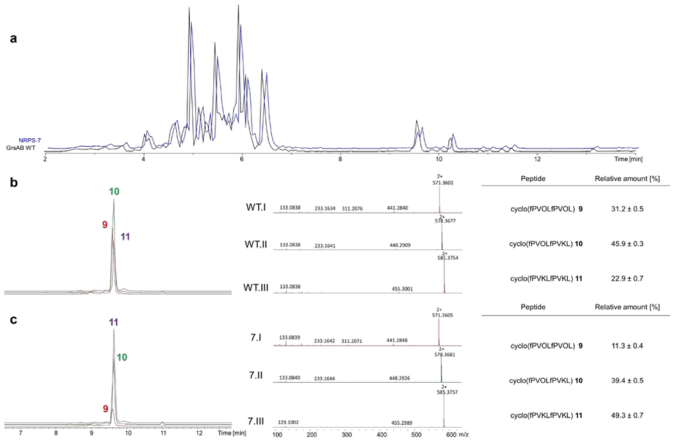
Supplementary Figure 6.HPLC/MS data of compounds **1** and **3** produced by NRPS-4 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-3 and NRPS-4 (Figure 2) in *E. coli* DH10B::mtaA. (**b**)EIC (left) and HPLC/MS^2 data (right) of GameXPeptide (**1**, *m/z* $[M+H]^+$= 586.4) and its derivative**3** (*m/z* $[M+H]^+$ = 552.4) produced by NRPS-4 and (Figure 1b) in *E. coli* DH10B::mtaA.



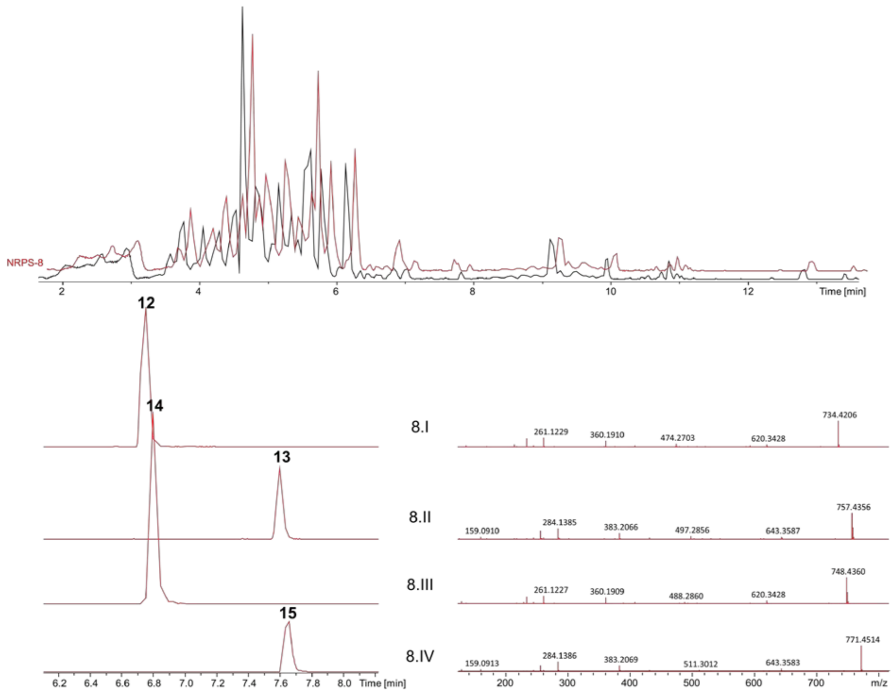
Supplementary Figure 7.HPLC/MS data of compounds **6** and **7** produced by NRPS-5 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-5 (Figure 2) in *E. coli* DH10B::mtaA. (**b**) EIC (left) and HPLC/MS^2 data (right) of arginine containing GameXPeptide **6**, *m/z* $[M+2H]^{2+}$ = 352.7 and its derivative **7** (*m/z* $[M+2H]^{2+}$ = 335.7).



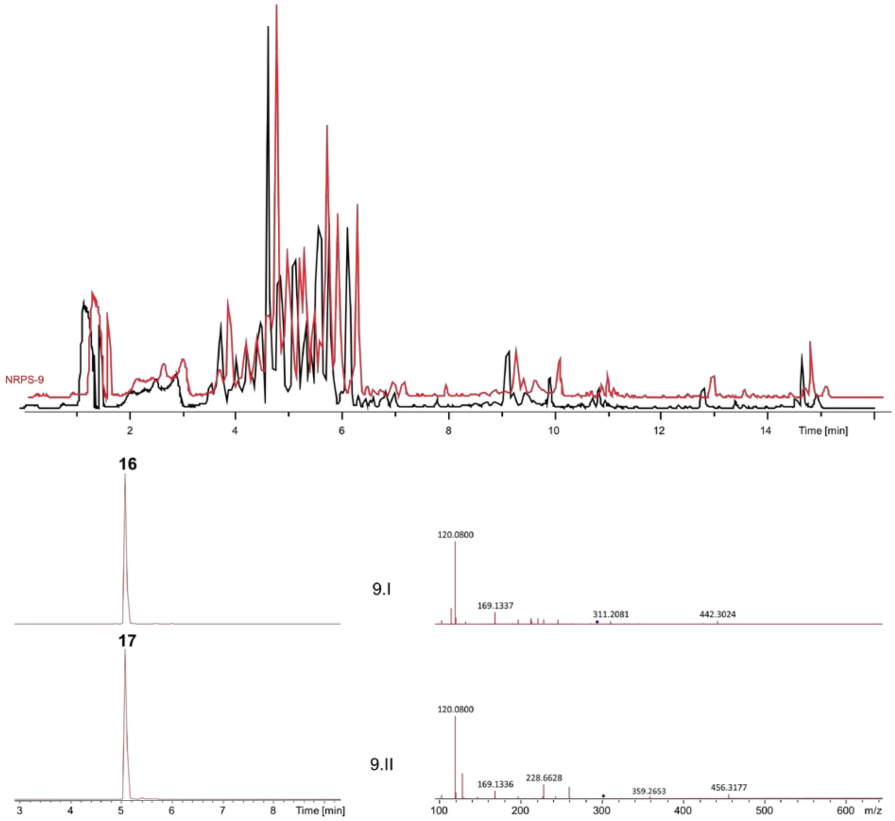
Supplementary Figure 8.HPLC/MS data of compound **8** produced by NRPS-6 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram (left) of production from NRPS-6 (Figure 2) in *E. coli* DH10B::mtaA and EIC (right) of **8** (*m/z* $[M+H]^+$ = 556.35). (**b**) Structure, fragmentation and HPLC/HR-MS^2 data of **8**(blue) compared to Ile-Thiazoline-Leu (green; *m/z* $[M+H]^+$ = 330.18)[9].



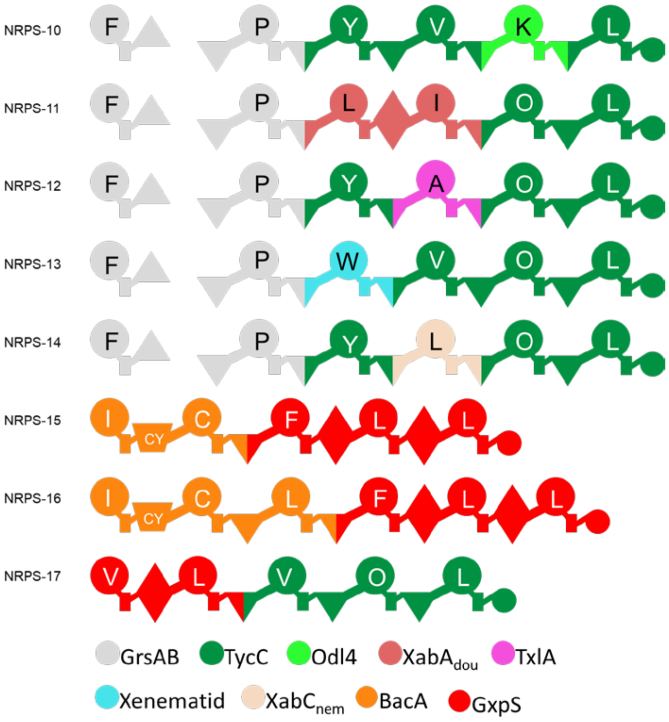
Supplementary Figure 9.HPLC/MS data of compounds **9-11** produced by GrsAB and NRPS-7 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-7 (Figure 2) and Gramicidin S-producing synthetase GrsAB in *E. coli* DH10B::mtaA. (**b**) overlaid EIC (left) and HR-HPLC/MS^2 data (middle) of Gramicidin S **9** (*m/z* $[M+2H]^{2+}$ = 571.360, red) and its derivatives **10** (*m/z* $[M+2H]^{2+}$ = 578.368, green) and **11** (*m/z* $[M+2H]^{2+}$ = 585.375, purple) from Gramicidin S-producing synthetase GrsAB. Relative amount of derivatives (right) calculated from triplicates. (**c**) overlaid EIC (left) and HR-HPLC/MS^2 data (middle) of Gramicidin S **9** (*m/z* $[M+2H]^{2+}$ = 571.360, red) and its derivatives **10** (*m/z* $[M+2H]^{2+}$ = 578.368, green) and **11** (*m/z* $[M+2H]^{2+}$ = 585.375, purple) from NRPS-7. Relative amount of derivatives (right)calculated from triplicates.



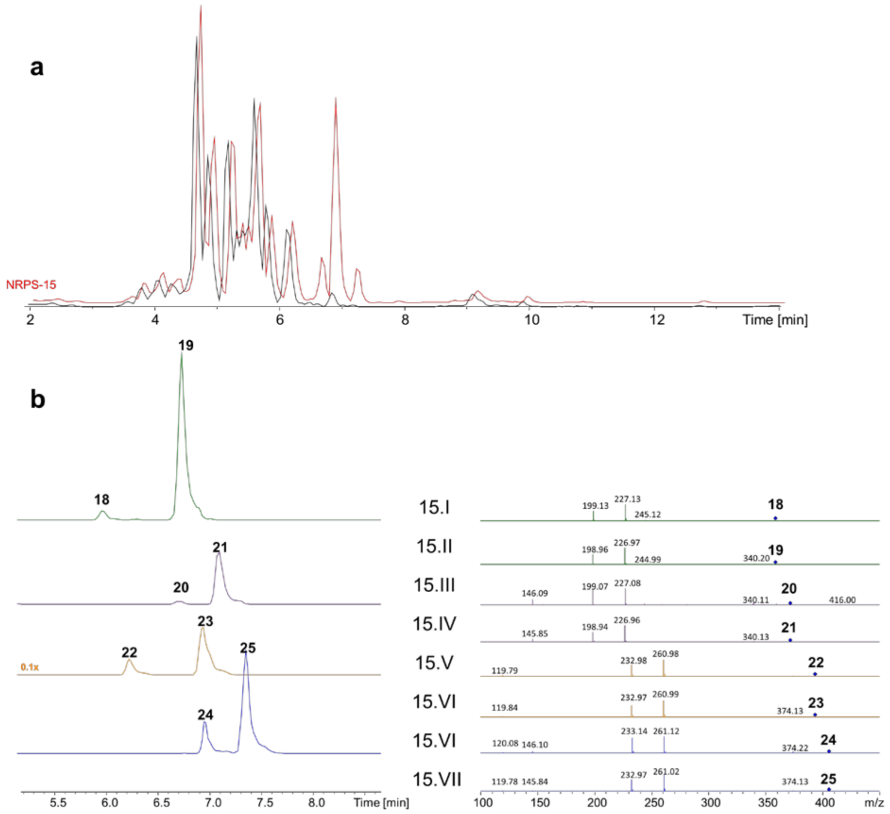
Supplementary Figure 10. HPLC/MS data of compounds **12-14** produced by NRPS-8 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-8 (Figure 2) in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/HR-MS^2 data (right) of **12** (*m/z* $[M+H]^+$= 734.420), **13** (*m/z* $[M+H]^+$ = 757.4396), **14** (*m/z* $[M+H]^+$ = 748.435) and **15** (*m/z* $[M+H]^+$ = 771.451).



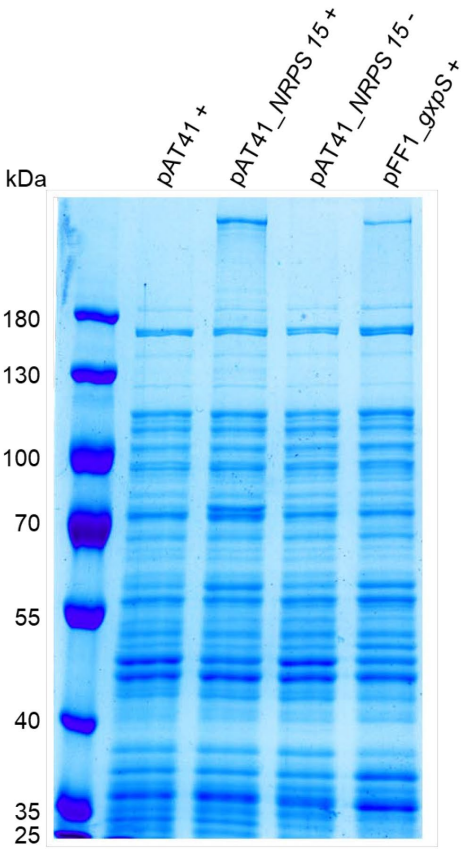
Supplementary Figure 11. HPLC/MS data of compounds **16** and **17** produced by NRPS-9 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-9 (Figure 2) in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/HR-MS^2 data (right) of **16** (*m/z* $[M+2H]^{2+}$ = 295.189) and **17** (*m/z* $[M+2H]^{2+}$ = 302.197).



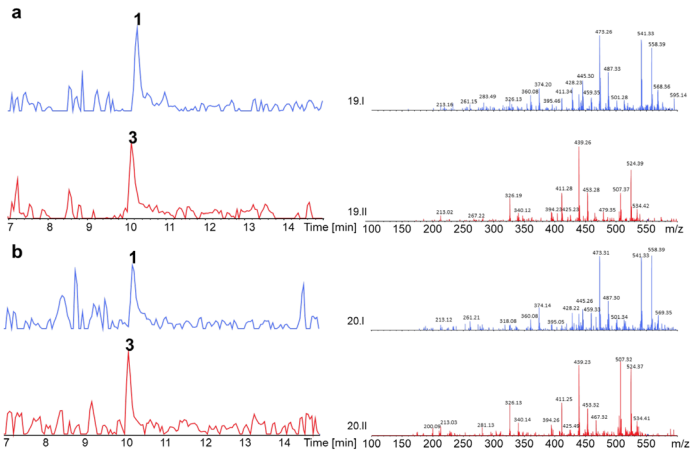
Supplementary Figure 12. Generated recombinant NRPS from Gram-positive and -negative origin. For assignment of domain symbols see Fig. 1; further symbols are E (epimerization; inverted triangle), CY (heterocyclization; trapezium). Bottom: Color code of NRPS used as building blocks (for details see Supplementary Figure 5).



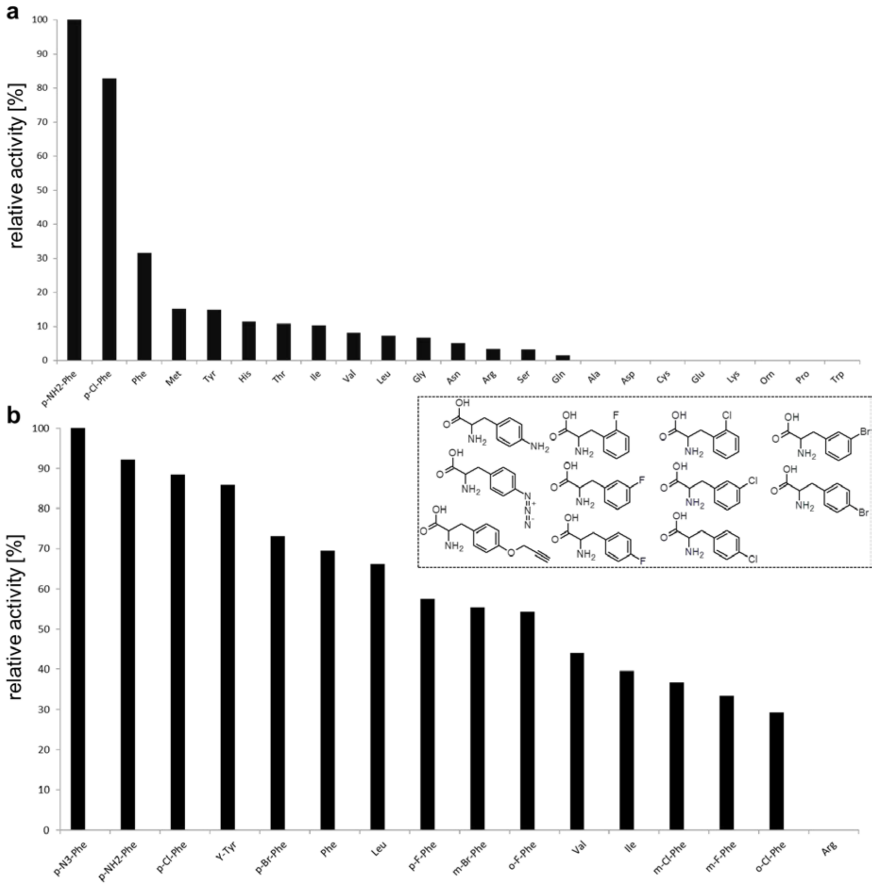
Supplementary Figure 13.HPLC/MS data of compounds **18-25** produced by NRPS-15 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-15 in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/MS^2 data (right) of **18** (*m/z* $[M+H]^+$ = 358.22), **19** (*m/z* $[M+H]^+$ = 358.22), **20** (*m/z* $[M+H]^+$ = 372.22), **21** (*m/z* $[M+H]^+$ = 372.22), **22** (*m/z* $[M+H]^+$ = 392.22), **23** (*m/z* $[M+H]^+$ = 392.22), **24** (*m/z* $[M+H]^+$ = 406.22) and **25** (*m/z* $[M+H]^+$ = 406.22). Peptides **20**, **21**, **24** and **25** are methoxy derivatives derived from MeOH use during the work-up procedure. The shift of the retention time of **18**, **20**, **22** and **24** compared to **19**, **21**, **23** and **25** respectively is supposed to be due to partial epimerization by NRPS-15.



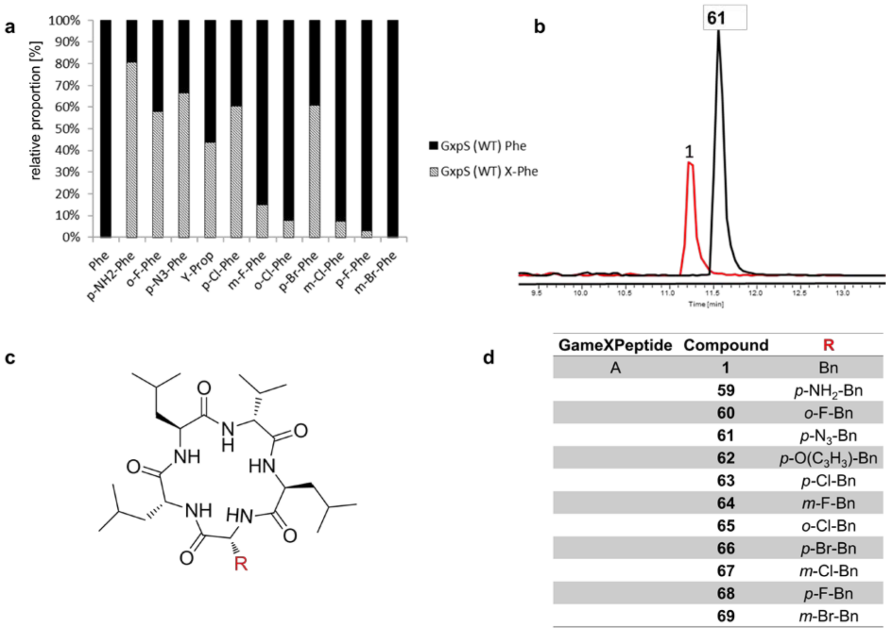
Supplementary Figure 14. SDS-PAGE assay of protein extracts of *E. coli* DH10B::mtaA harboring pAT41 with no insert (control), pAT41_*NRPS 15* and pFF1_*gxpS*. All samples marked with "+" were inoculated with 0.02% arabinose and "-" without arabinose. The expected molecular weight for NRPS 15 and GxpS is 572 kDa and 514 kDa, respectively.



Supplementary Figure 15.HPLC/MS data of compounds **1** and **3** produced by NRPS-19 and -20 in *E. coli* DH10B::mtaA.(**a**) EIC (left) and HPLC/MS^2 data (right) of GameXPeptide **1**, *m/z* $[M+H]^+$= 586.4 and its derivative **3** (*m/z* $[M+H]^+$ = 552.4) produced by NRPS-19 (Figure 3) in *E. coli* DH10B::mtaA. (**b**) EIC (left) and HPLC/MS^2 data (right) of GameXPeptide **1**, (*m/z* $[M+H]^+$= 586.4) and **3** (*m/z* $[M+H]^+$ = 552.4) produced from NRPS-20 (Figure 3) in *E. coli* DH10B::mtaA.

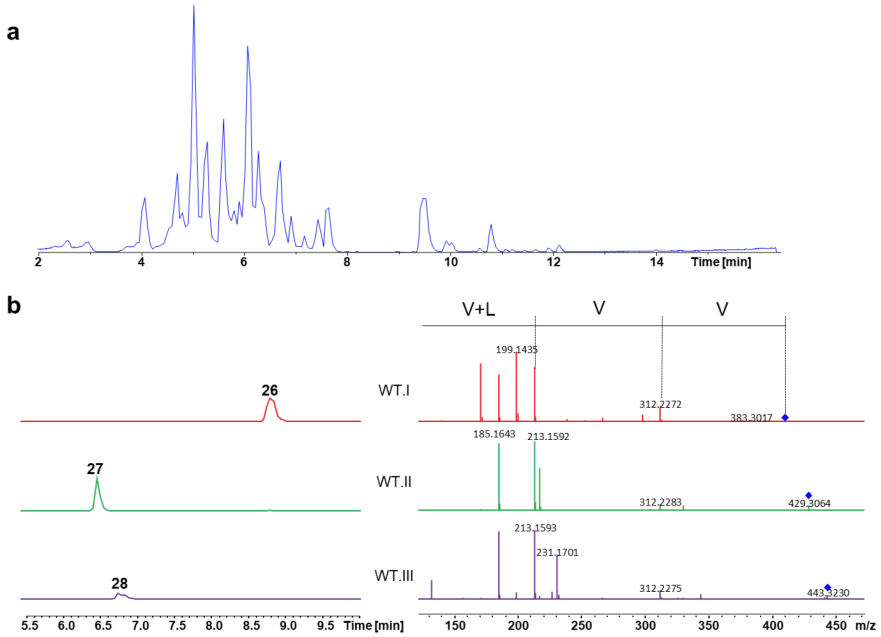


Supplementary Figure 16.*In vitro* adenylation activity of GxpS_A3.(**a**)Adenylation activity[3] of GxpS_A3tested with proteinogenic AAs and *para* substituted phenylalanine.The activities are calculated relative to the substrate with the highest activity (*p*-NH_2-Phe). (**b**) GxpS_A3 adenylation activity tested with non-proteinogenic AAs *ortho- meta-* and *para* substituted phenylalanine.The activities are calculated relative to the substrate with the highest activity (*p*-N_3-Phe).

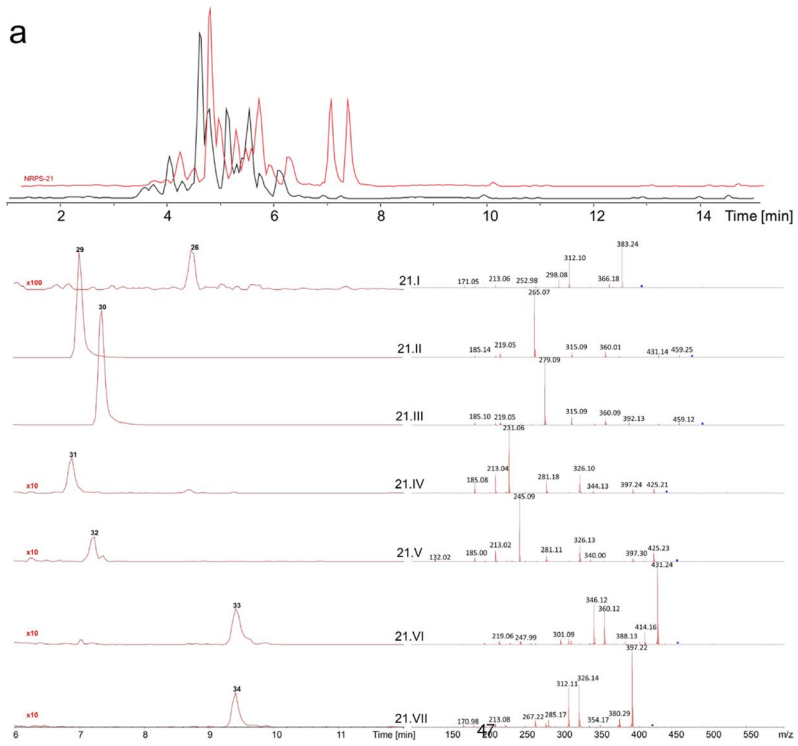


GameXPeptide	Compound	R
A	**1**	Bn
	59	*p*-NH_2-Bn
	60	*o*-F-Bn
	61	*p*-N_3-Bn
	62	*p*-$O(C_3H_3)$-Bn
	63	*p*-Cl-Bn
	64	*m*-F-Bn
	65	*o*-Cl-Bn
	66	*p*-Br-Bn
	67	*m*-Cl-Bn
	68	*p*-F-Bn
	69	*m*-Br-Bn

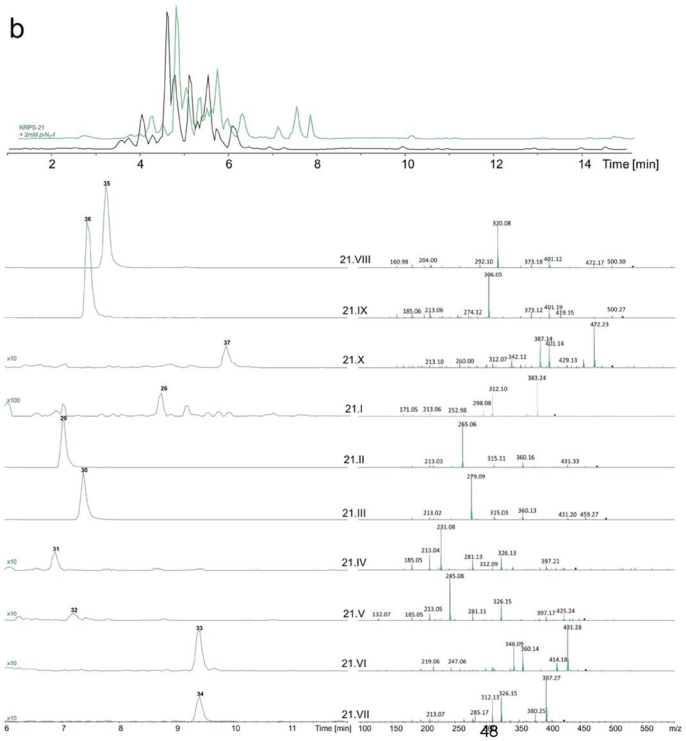
Supplementary Figure 17. *In vivo* characterization of GxpS in *E. coli* DH10B::mtaA. (**a**) *In vivo* feeding experiments with the GameXPeptide producing WT and mutant (W239S) GxpS. The relative proportions of the HPLC/MS detected signals of peptides (**1, 59-69**) are shown, according to the supplemented substituents (X-Phe). In case of the control (Phe) no AA are fed. (**b**) EIC of an extract of *E. coli* DH10B::mtaA with pFF1_*gxpS*_WTshowing the production of **1** (*m/z* $[M+H]^+$ = 586.4) and **61** (*m/z* $[M+H]^+$ = 627.4), when fed with *p*-N_3-Phe. (**c**) GameXPeptide structure. (**d**) Expected GameXPeptide derivatives.



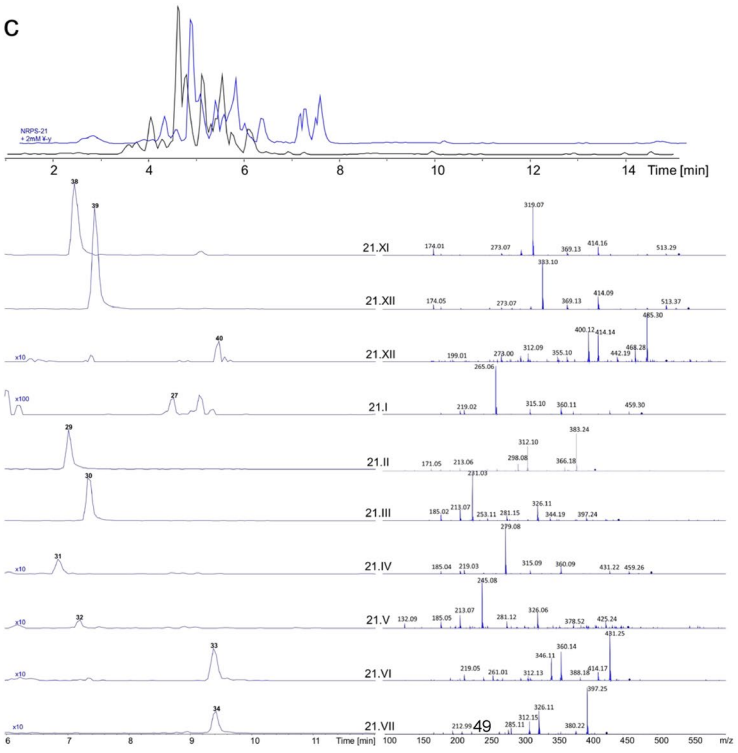
Supplementary Figure 18. Heterologous production of xenotetrapeptide in *E. coli* DH10B::mtaAand HPLC/HR-MS analysis. (**a**) Base peak chromatogram (blue) and (**b**) extracted ion chromatograms (left) of **26**(*m/z* $[M+H]^+$ = 411.29), **27** (*m/z* $[M+H]^+$ = 429.30), **28** (*m/z* $[M+H]^+$ = 443.32) and MS-MS spectra.


a
NRPS-21
Time [min]
21.I
21.II
21.III
21.IV
21.V
21.VI
21.VII
m/z

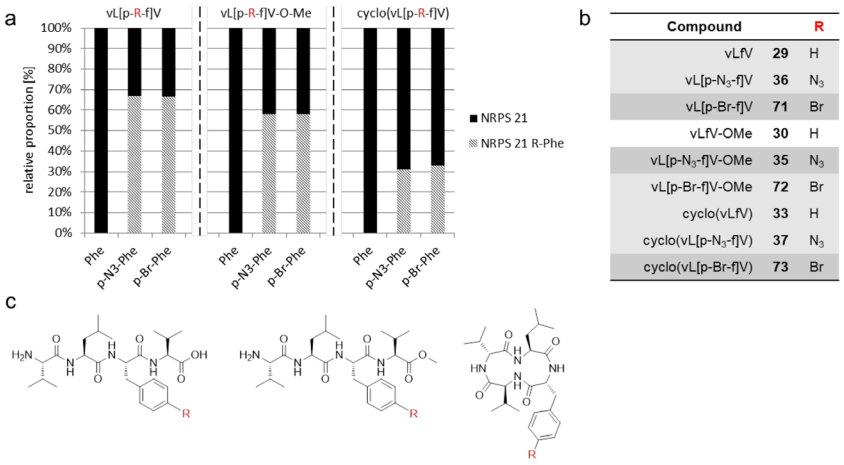
47


b
NRPS-21
Time [min]
21.VIII
21.IX
21.X
21.I
21.II
21.III
21.IV
21.V
21.VI
21.VII
m/z

48

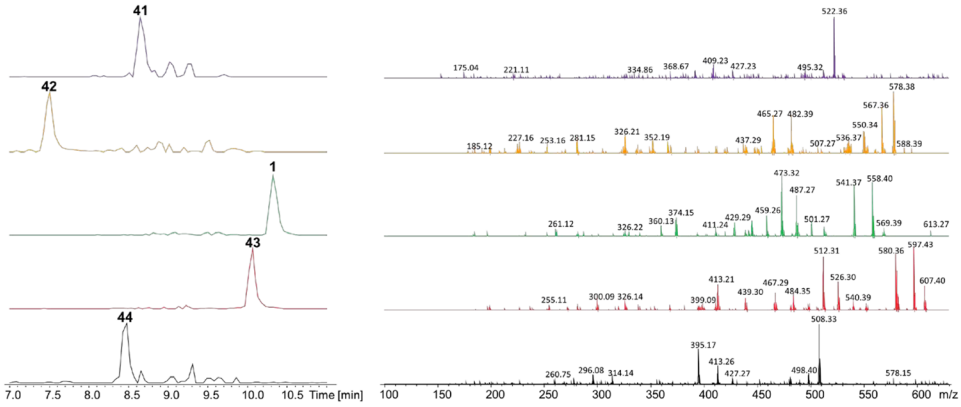


Supplementary Figure 19.HPLC/MS data of xenotetrapeptide derivatives produced by NRPS-21 in *E. coli* DH10B::mtaA.BPCs (top), EICs (left) and HPLC/MS^2 data (right). (**a**)**26** (*m/z* $[M+H]^+$ = 411.2)**29** (*m/z* $[M+H]^+$ = 477.2), **30** (*m/z* $[M+H]^+$ = 491.2),**31** (*m/z* $[M+H]^+$ = 443.2),**32** (*m/z* $[M+H]^+$ = 457.2),**33** (*m/z* $[M+H]^+$ = 459.2) and **34** (*m/z* $[M+H]^+$ = 425.3); (**b**)**35** (*m/z* $[M+H]^+$ = 532.2),**36** (*m/z* $[M+H]^+$ = 518.2),**37** (*m/z* $[M+H]^+$ = 500.2)**26** (*m/z* $[M+H]^+$ = 411.2) **29** (*m/z* $[M+H]^+$ = 477.2), **30** (*m/z* $[M+H]^+$ = 491.2),**31** (*m/z* $[M+H]^+$ = 443.2),**32** (*m/z* $[M+H]^+$ = 457.2), **33** (*m/z* $[M+H]^+$ = 459.2) and **34** (*m/z* $[M+H]^+$ = 425.3); (**c**)**38** (*m/z* $[M+H]^+$ = 531.2),**39** (*m/z* $[M+H]^+$ = 545.2), **40** (*m/z* $[M+H]^+$ = 513.2)**26** (*m/z* $[M+H]^+$ = 411.2) **29** (*m/z* $[M+H]^+$ = 477.2), **30** (*m/z* $[M+H]^+$ = 491.2),**31** (*m/z* $[M+H]^+$ = 443.2),**32** (*m/z* $[M+H]^+$ = 457.2), **33** (*m/z* $[M+H]^+$ = 459.2) and **34** (*m/z* $[M+H]^+$ = 425.3).

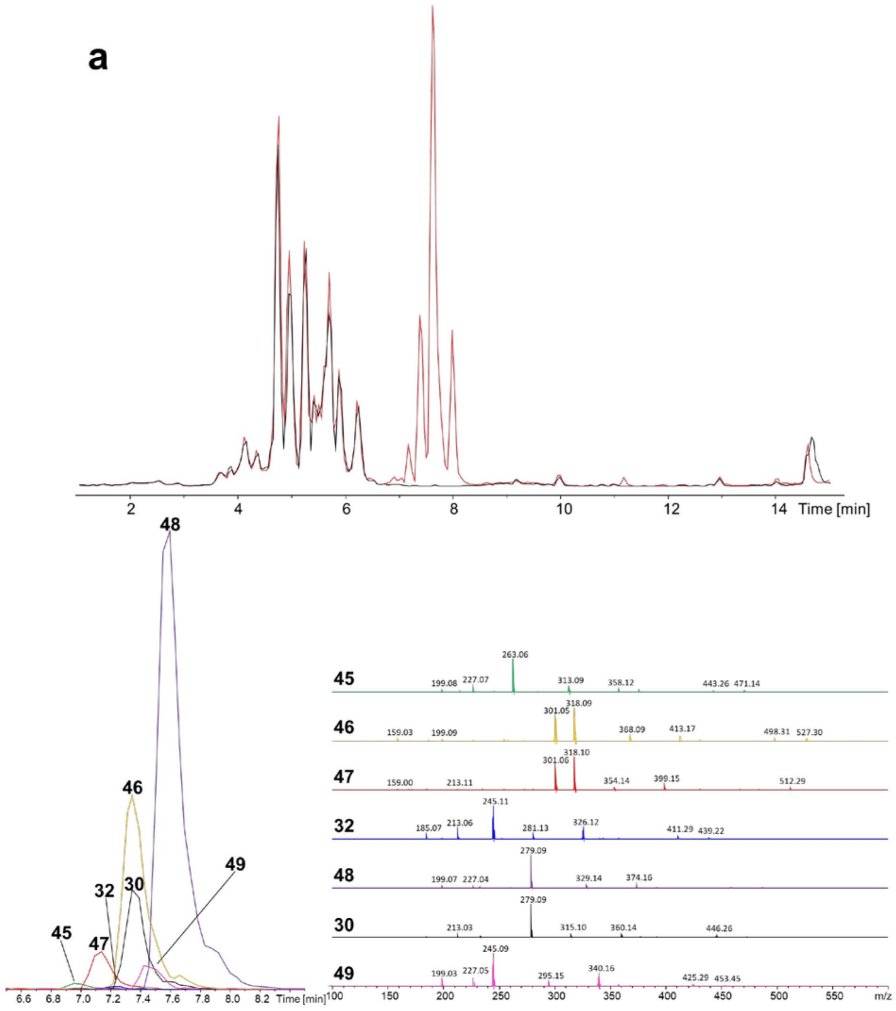


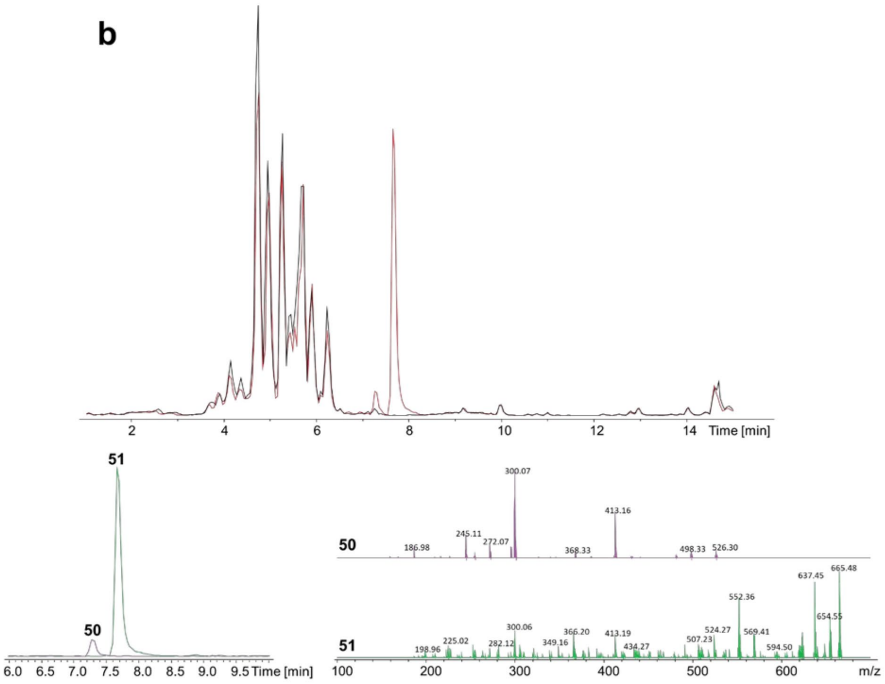
Compound		R
vLfV	29	H
vL[p-N_3-f]V	36	N_3
vL[p-Br-f]V	71	Br
vLfV-OMe	30	H
vL[p-N_3-f]V-OMe	35	N_3
vL[p-Br-f]V-OMe	72	Br
cyclo(vLfV)	33	H
cyclo(vL[p-N_3-f]V)	37	N_3
cyclo(vL[p-Br-f]V)	73	Br

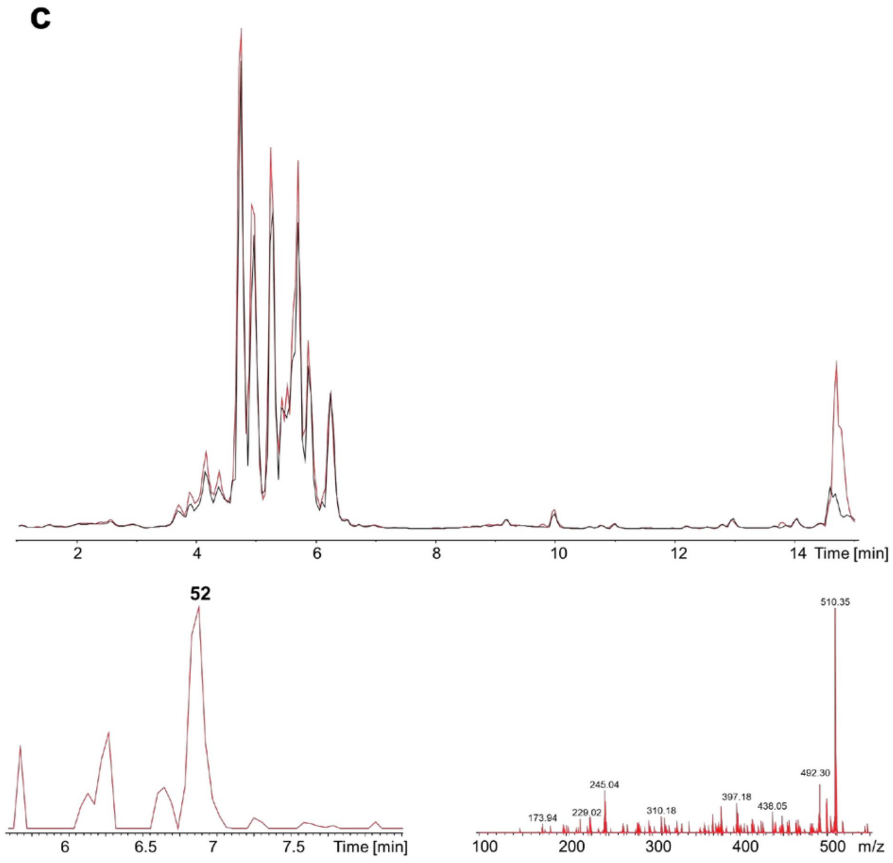
Supplementary Figure 20. *In vivo* characterization of NRPS-21 in *E. coli* DH10B::mtaA. (**a**) *In vivo* feeding experiments with NRPS-21incorporating substituted Phe derivatives (R). The relative proportions of the HPLC/MS detected signals of peptides (**29, 30, 33, 35-37,71-73**) are shown, according to the supplemented Phe derivatives showing different para substituents(R). In case of the control (R=H) no AA are fed. (**b**) Structure of the phenylalanine containing NRPS-21 derivatives. (**c**) Expected NRPS-21 derivatives.

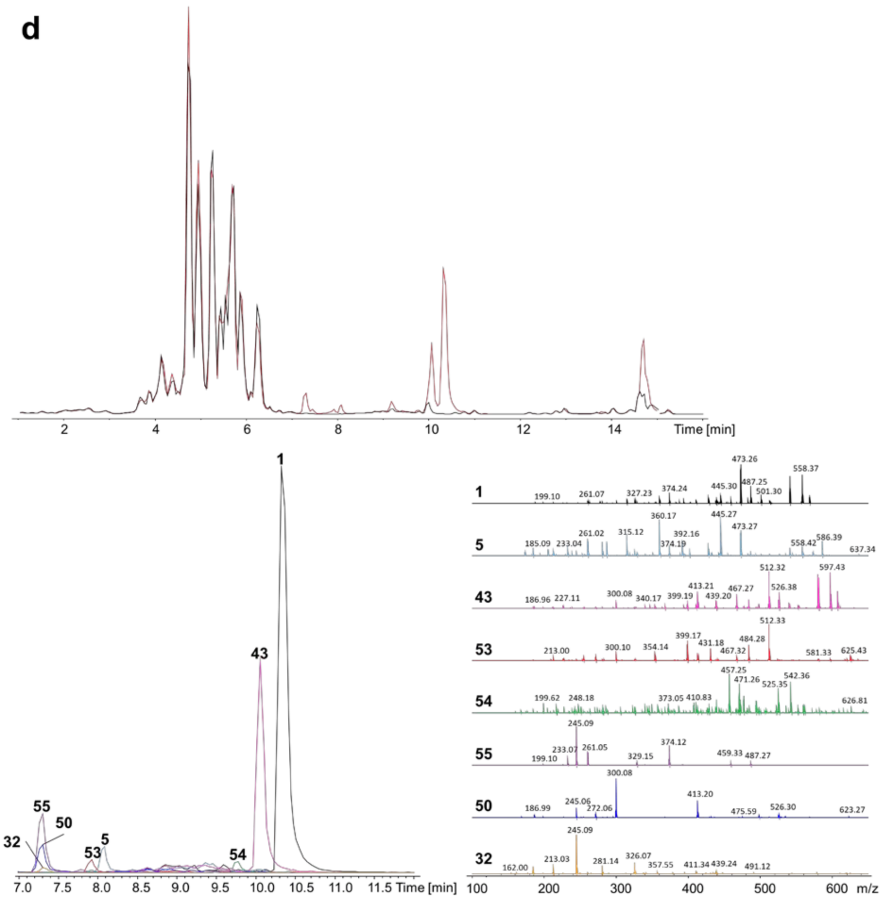


Supplementary Figure 21. Targeted randomization of GxpS at position three. EIC (left) and HPLC/MS^2 data (right) of GameXPeptide derivatives produced by library 1 (Figure 4)**41** (*m/z* $[M+H]^+$ = 540.3),**42** (*m/z* $[M+H]^+$ = 595.4),**1** (*m/z* $[M+H]^+$ = 586.4)**43** (*m/z* $[M+H]^+$ = 625.4)and **44** (*m/z* $[M+H]^+$ = 526.3) in *E. coli* DH10B::mtaA.


a
2
4
6
8
10
12
14 Time [min]
48
46
49
32
30
45
47
6.6 6.8 7.0 7.2 7.4 7.6 7.8 8.0 8.2 Time [min]
45
263.06
199.08 227.07 313.09 358.12 443.26 471.14
46
301.05 318.09
159.03 199.09 368.09 413.17 498.31 527.30
47
301.06 318.10
159.00 213.11 354.14 399.15 512.29
32
245.11
185.07 213.06 281.13 326.12 411.29 439.22
48
279.09
199.07 227.04 329.14 374.16
30
279.09
213.03 315.10 360.14 446.26
49
245.09
199.03 227.05 295.15 340.16 425.29 453.45
100 150 200 250 300 350 400 450 500 550 m/z

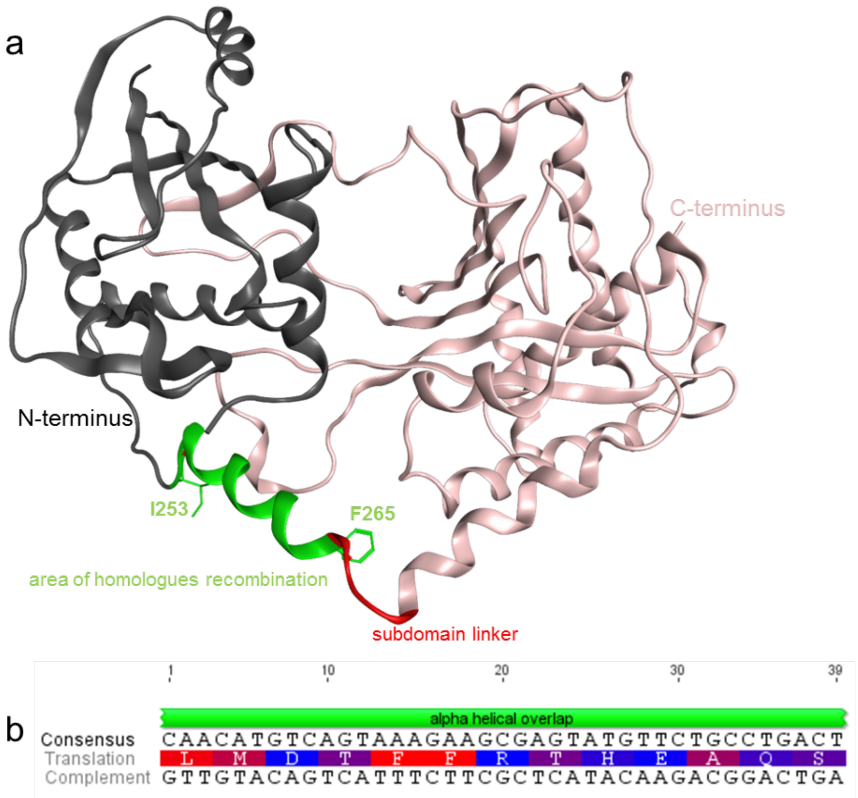

b
2
4
6
8
10
12
14
Time [min]
51
50
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
Time [min]
300.07
413.16
245.11
272.07
186.98
368.33
498.33
526.30
50
637.45
665.48
552.36
654.55
524.27
569.41
300.06
366.20
413.19
225.02
282.12
349.16
434.27
507.23
198.96
594.50
51
100
200
300
400
500
600
m/z


c
2
4
6
8
10
12
14 Time [min]
52
6
6.5
7
7.5 Time [min]
510.35
492.30
245.04
397.18
438.05
310.18
229.02
173.94
100
200
300
400
500 m/z

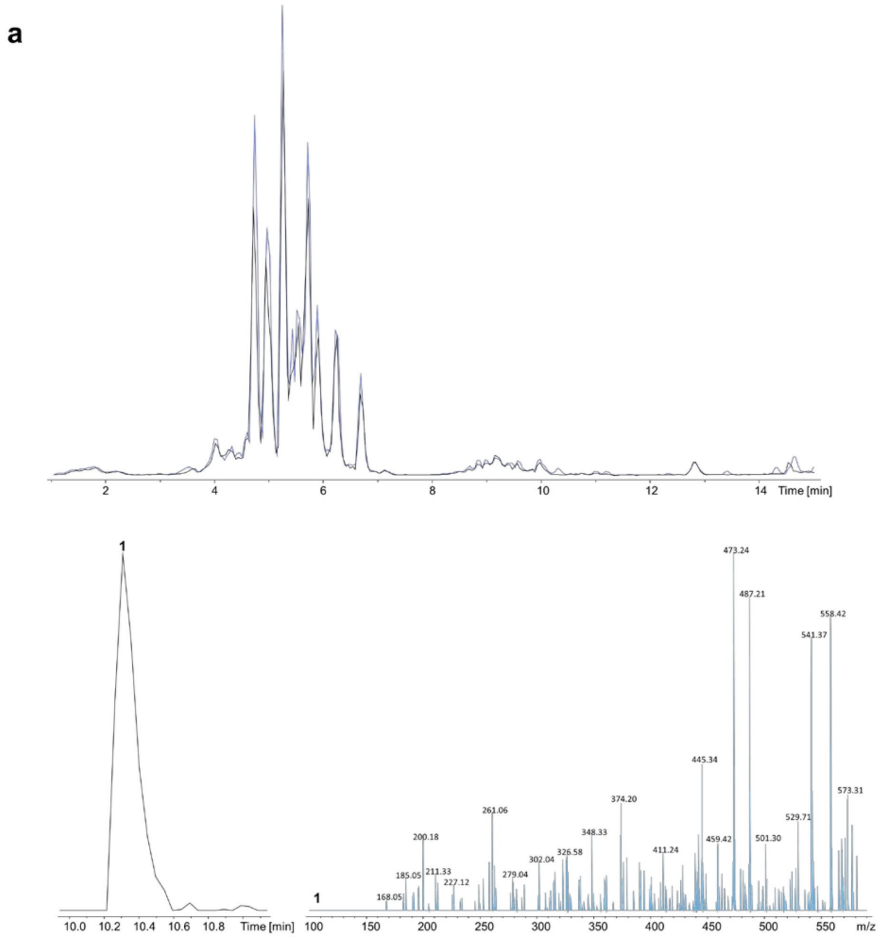


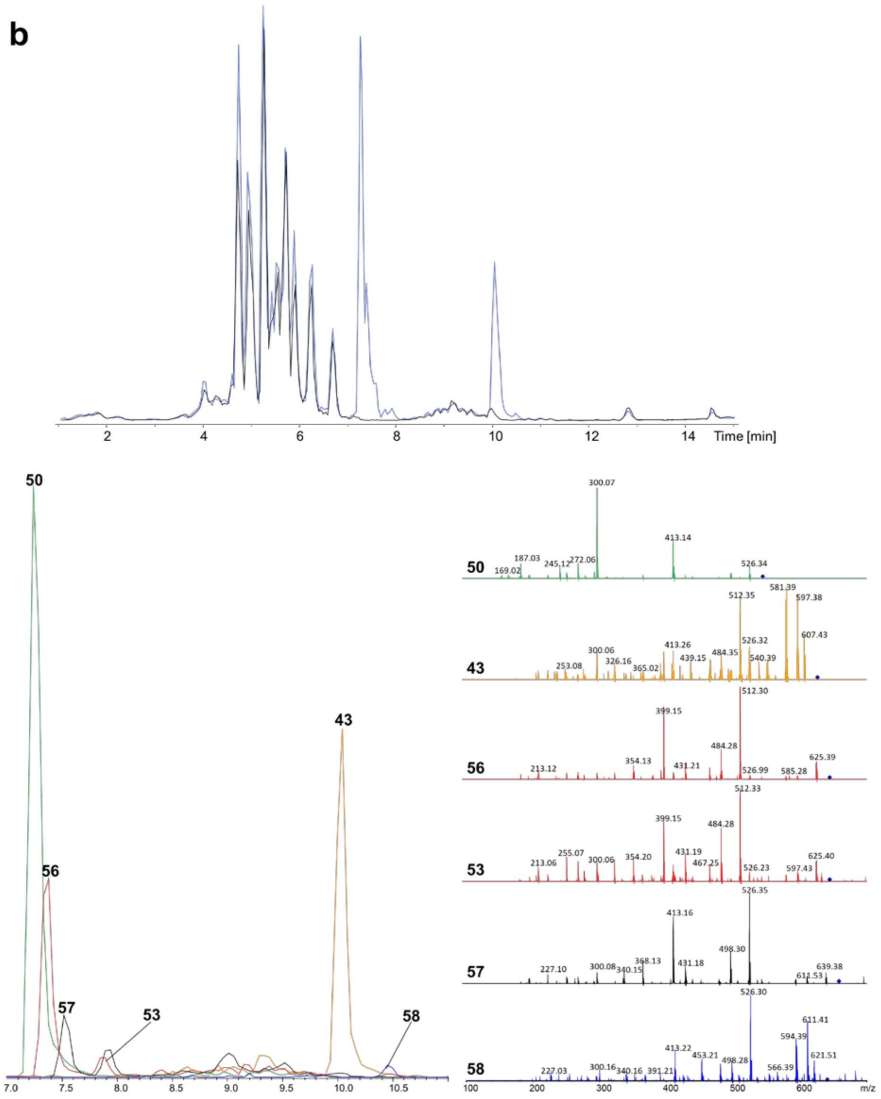
Supplementary Figure 22. The creation of a library of GxpS where position one and three were randomized. (**a**) Top: BPCs of production from Lib2_NRPS-1in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom:EICs (left) and HPLC/MS^2 data (right)of peptides produced by Lib2_NRPS-1 (**45**, *m/z* $[M+H]^+$ = 489.3), (**46**, *m/z* $[M+H]^+$ = 544.3), (**47**, *m/z* $[M+H]^+$ = 530.3), (**32**, *m/z* $[M+H]^+$ = 457.3), (**48**, *m/z* $[M+H]^+$ = 505.3), (**30**, *m/z* $[M+H]^+$ = 491.3), (**49**, *m/z* $[M+H]^+$ = 471.3). (**b**) Top: BPCs of production from Lib2_NRPS-2 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib2_NRPS-2 (**50**, *m/z* $[M+H]^+$= 544.4), (**51**, *m/z* $[M+H]^+$ = 683.4). (**c**) Top: BPCs of production from Lib2_NRPS-3 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by

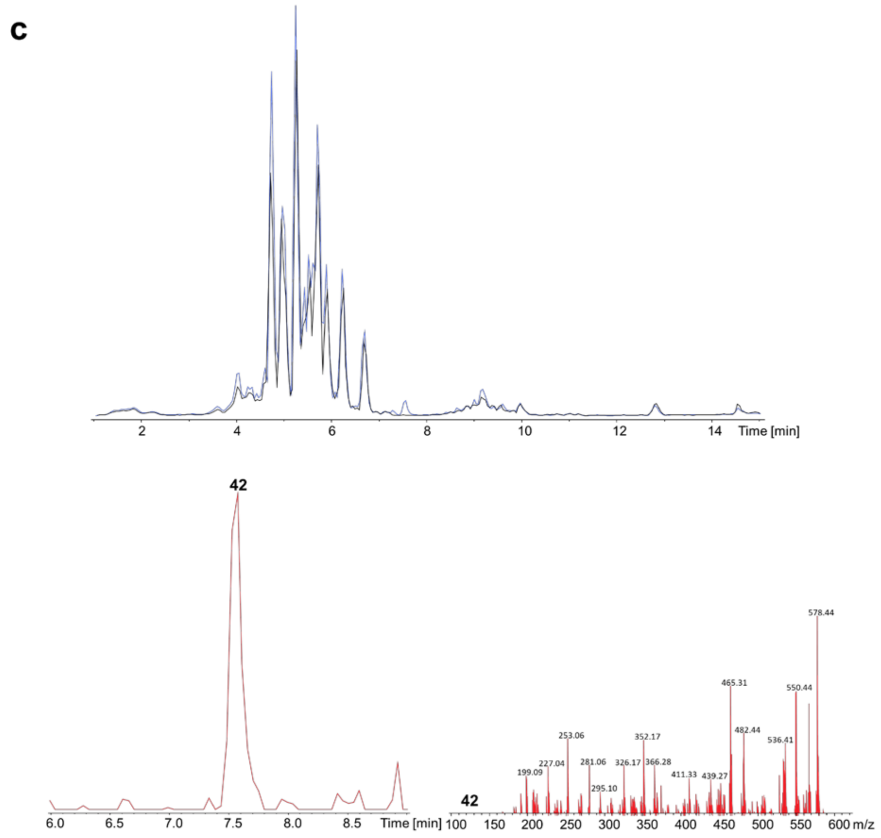
Lib2_NRPS-3 (**52**, *m/z* $[M+H]^+$= 528.4). (**d**) Top: BPCs of production from Lib3_NRPS-4 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib2_NRPS-4 (**1**, *m/z* $[M+H]^+$ = 586.4), (**5**, *m/z* $[M+H]^+$ = 604.4), (**43**, *m/z* $[M+H]^+$ = 625.4), (**53**, *m/z* $[M+H]^+$ = 643.4), (**54**, *m/z* $[M+H]^+$ = 570.4), (**55**, *m/z* $[M+H]^+$ = 505.3), (**50**, *m/z* $[M+H]^+$ = 544.3), (**32**, *m/z* $[M+H]^+$ = 457.3).



Supplementary Figure 23. Design of an artificial α5 helix.(**a**) Crystal structure of TycC6 (PDB-ID: 2JGP)[8], subdivided into N terminal subdomain (grey) and C terminal subdomain (light red). The subdomain linker is highlighted in red and the targeted area (I253 – F265) for homologous recombination in yeast is highlighted in green (39 nucleotides). (**b**) Consensus sequence used to generate library three (Figure 5c).


a
2
4
6
8
10
12
14
Time [min]
1
10.0
10.2
10.4
10.6
10.8
Time [min]
1
100
150
200
250
300
350
400
450
500
550
m/z
168.05
185.05
200.18
211.33
227.12
261.06
279.04
302.04
326.58
348.33
374.20
411.24
445.34
459.42
473.24
487.21
501.30
529.71
541.37
558.42
573.31

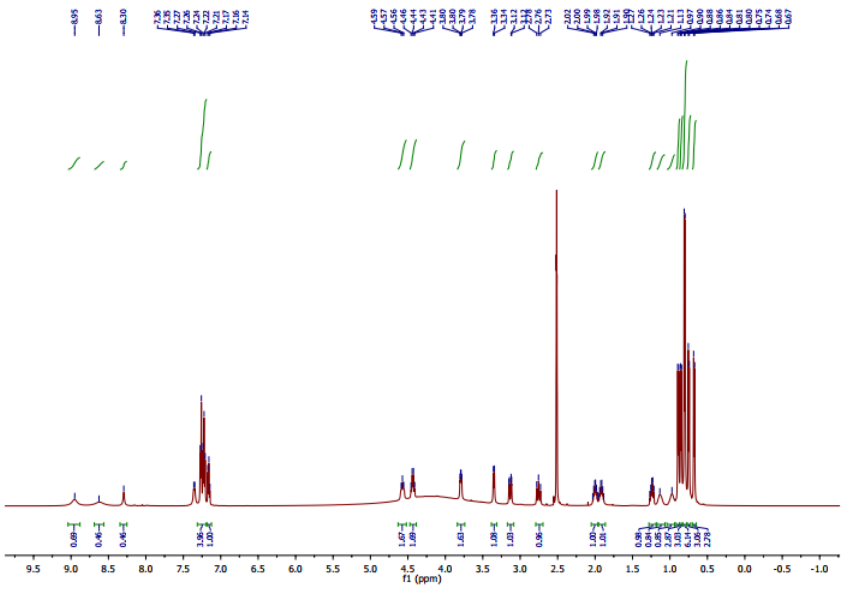

b
2
4
6
8
10
12
14
Time [min]
50
56
57
53
43
58
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
m/z



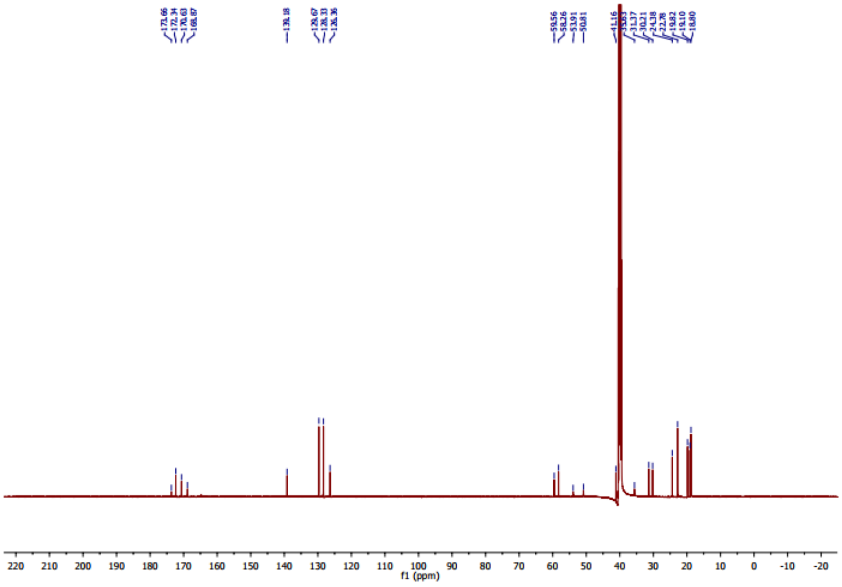
Supplementary Figure 24. The creation of a random library via an artificial α5 helix. (**a**) Top: BPCs of production from Lib3_NRPS-1 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib3_NRPS-1 (**1**, *m/z* $[M+H]^+$= 586.4). (**b**) Top:BPCs of production from Lib3_NRPS-2 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib3_NRPS-2 (**50**, *m/z* $[M+H]^+$ = 544.3), (**43**, *m/z* $[M+H]^+$ = 625.4), (**56**, *m/z* $[M+H]^+$ = 643.4), (**53**, *m/z* $[M+H]^+$ = 643.4), (**57**, *m/z* $[M+H]^+$ = 657.4) and (**58**, *m/z* $[M+H]^+$ = 639.4). (**c**) Top:BPCs of production from Lib3_NRPS-3 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib3_NRPS-3 (**42**, *m/z* $[M+H]^+$= 595.4).

NMR data

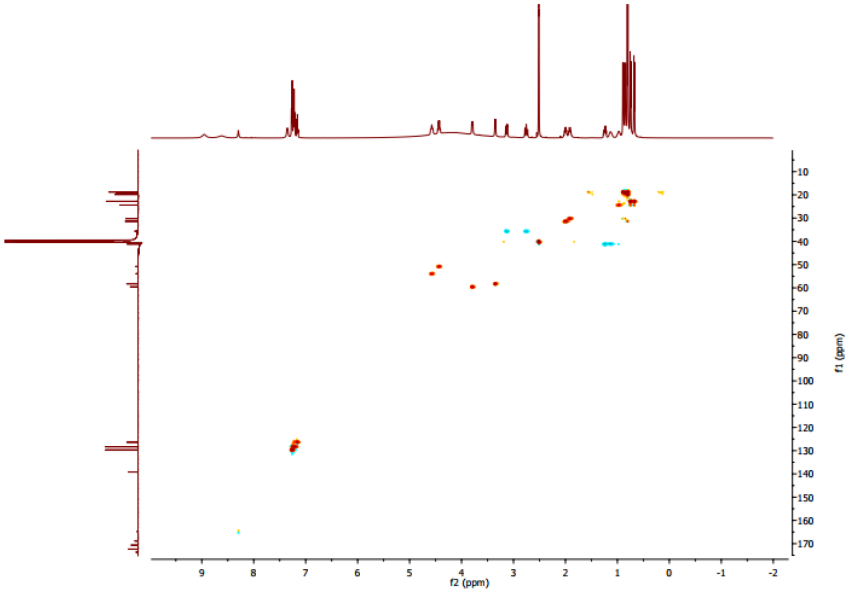
Compound **29** (52.6 mg/L; white powder): HR-ESI-MS (found *m/z*477.3068 $[M + H]^+$, calcd. for $C_{25}H_{41}N_4O_5$, 477.3071 $[M + H]^+$, Δppm 0.8). ^{1}H NMR (500 MHz, DMSO) δ in ppm: 0.68 (d, 3H, J = 6.5 Hz), 0.75 (d, 3H, J = 6.4 Hz), 0.80 (d, 6H, J = 6.8 Hz), 0.85 (d, 3H, J = 6.8 Hz), 0.89 (d, 3H, J = 6.4 Hz), 0.97 (brs, 1H), 1.13 (brs, 1H), 1.24 (m, 1H), 1.92 (m, 1H), 2.00 (m, 1H), 2.75 (m, 1H), 3.13 (dd, 1H, J = 13.9, 3.6 Hz), 3.35 (d, 1H, J = 7.3 Hz), 3.79 (dd, 1H, J = 7.3, 4.5 Hz), 4.43 (dd, 1H, J = 16.5, 8.0 Hz), 4.57 (m, 1H), 7.16 (t, 1H, J = 7.2 Hz), 7.22 (t, 2H, J =7.4 Hz), 7.26 (t, 2H, J = 7.1 Hz), 8.63 (s, 2H), 8.95 (s, 1H). ^{13}C NMR (125 MHz, DMSO) δ in ppm: 18.8, 18.8, 19.1, 19.8, 22.8, 22.8, 24.4, 30.2, 31.4, 35.6, 41.6, 50.8, 53.9, 58.3, 59.6, 126.4, 128.3, 128.3, 129.7, 129.7, 139.2, 168.9, 170.6, 172.3, 173.7.



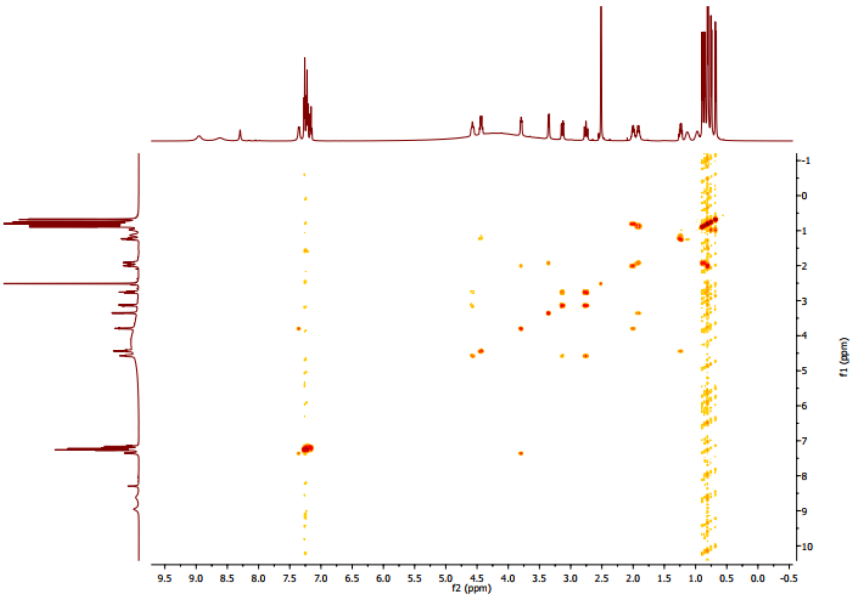
Supplementary Figure 25. ^{1}H NMR spectrum of compound **29**.



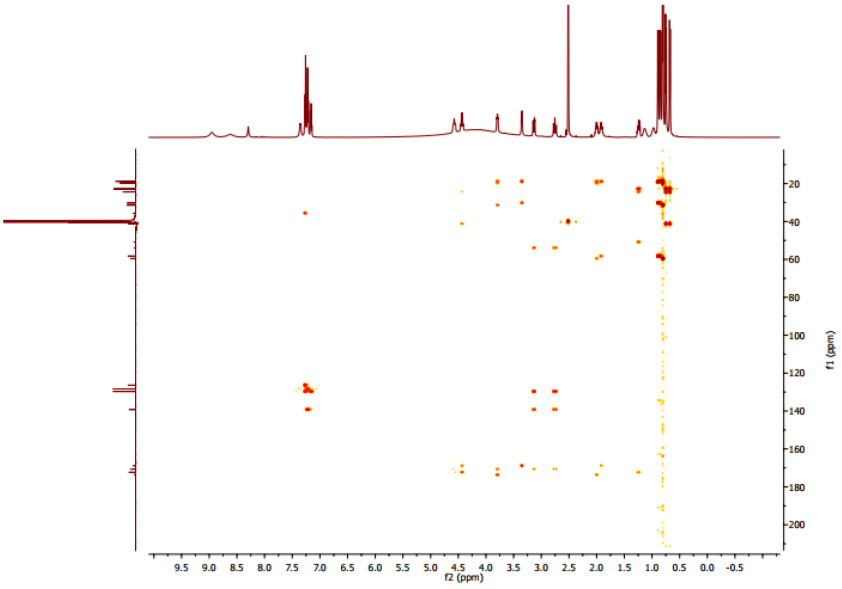
Supplementary Figure 26. ^{13}C NMR spectrum of compound **29**.



Supplementary Figure 27. HSQC spectrum of compound **29**.

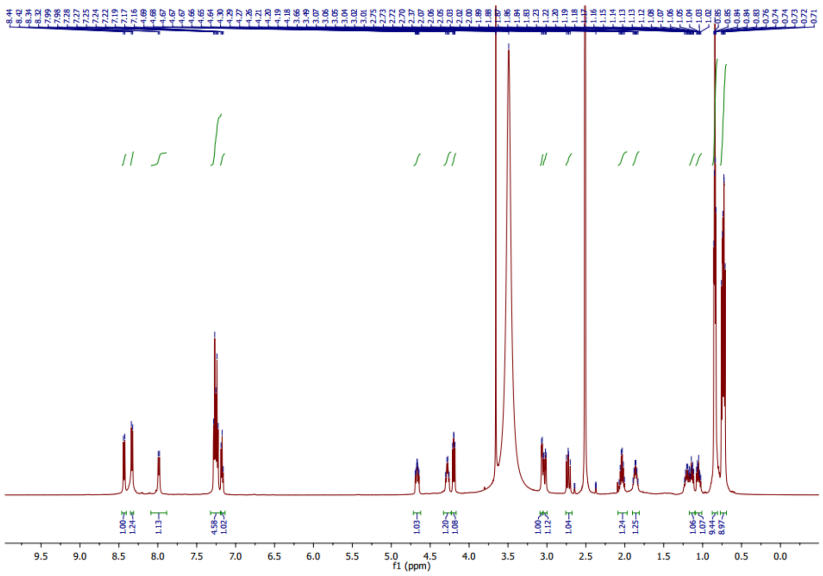


Supplementary Figure 28. COSY spectrum of compound **29.**

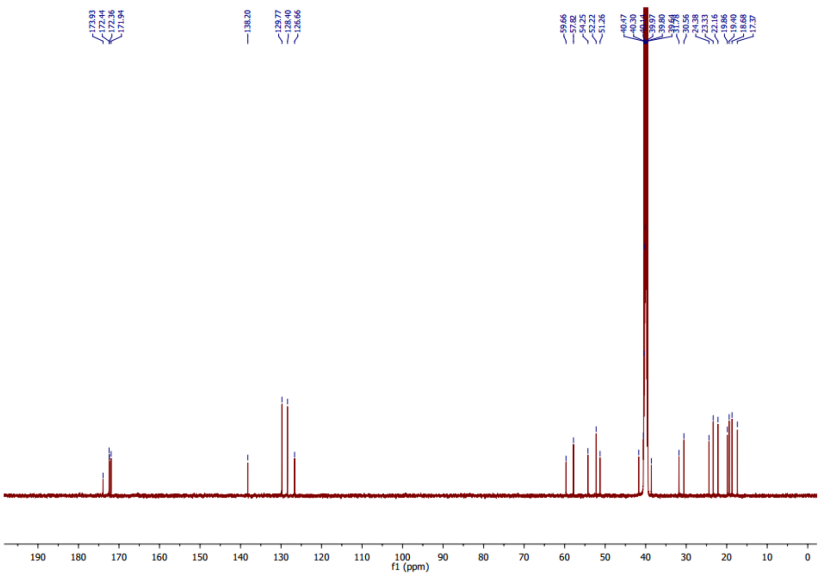


Supplementary Figure 29. HMBC spectrum of compound **29**.

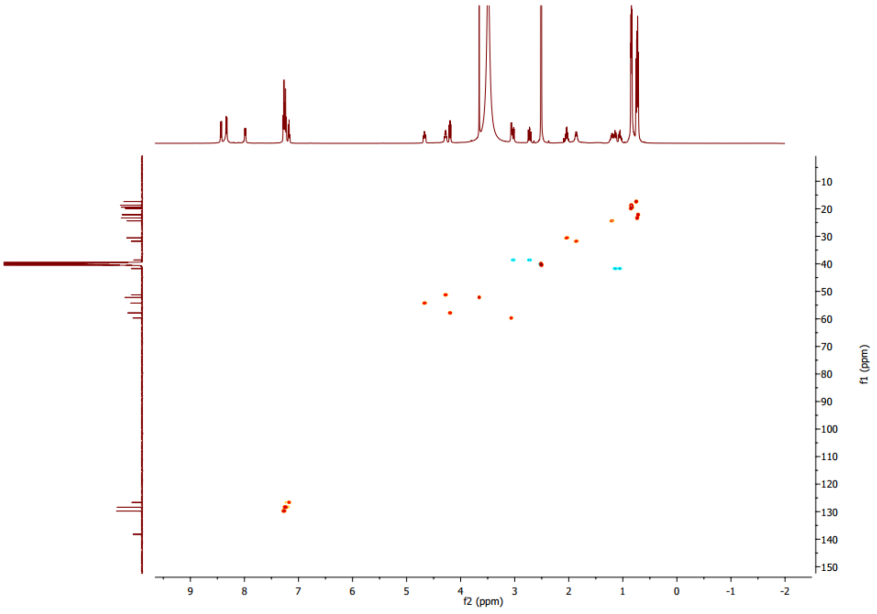
Compound **30** (47.3 mg/L; white powder): HR-ESI-MS (found *m/z* 491.3225 $[M + H]^+$, calcd. for $C_{26}H_{43}N_4O_5$, 491.3228 $[M + H]^+$, Δppm 0.5).^{1}H NMR (500 MHz, DMSO) δ in ppm: 0.72 (d, 3H, J = 6.8 Hz), 0.73 (d, 3H, J = 6.8 Hz), 0.75 (d, 3H, J = 6.9 Hz), 0.84 (d, 3H, J = 6.7 Hz), 0.84 (d, 3H, J = 6.6 Hz), 0.85 (d, 3H, J = 6.5 Hz), 1.05 (ddd, 1H, J = 13.6, 8.1, 5.7 Hz), 1.14 (m, 1H), 1.21 (m, 1H), 1.86 (m, 1H), 2.04 (m, 1H), 2.72 (dd, 1H, J = 13.3, 11.0 Hz), 3.03 (dd, 1H, J = 13.3, 4.2 Hz), 3.06 (d, 1H, J = 4.7 Hz), 4.20 (dd, 1H, J = 8.3, 6.7 Hz), 4.28 (m, 1H), 4.67 (ddd, 1H, J = 11.0, 4.2, 9.0 Hz), 7.17 (t, 1H, J = 7.1 Hz), 7.24 (t, 2H, J =7.5 Hz), 7.27 (t, 2H, J = 7.1 Hz), 7.99 (d, 1H, J = 7.9 Hz), 8.33 (d, 2H, J = 8.4 Hz), 8.43 (d, 1H, J = 8.8 Hz). ^{13}C NMR (125 MHz, DMSO) δ in ppm: 17.4, 18.7, 19.4, 19.9, 22.2, 23.3, 24.4, 30.6, 31.8, 38.6, 41.7, 51.3, 52.2, 54.3, 57.8, 59.7, 126.7, 129.8, 129.8, 128.4, 128.4, 138.2, 171.9, 172.3, 172.4, 173.9.



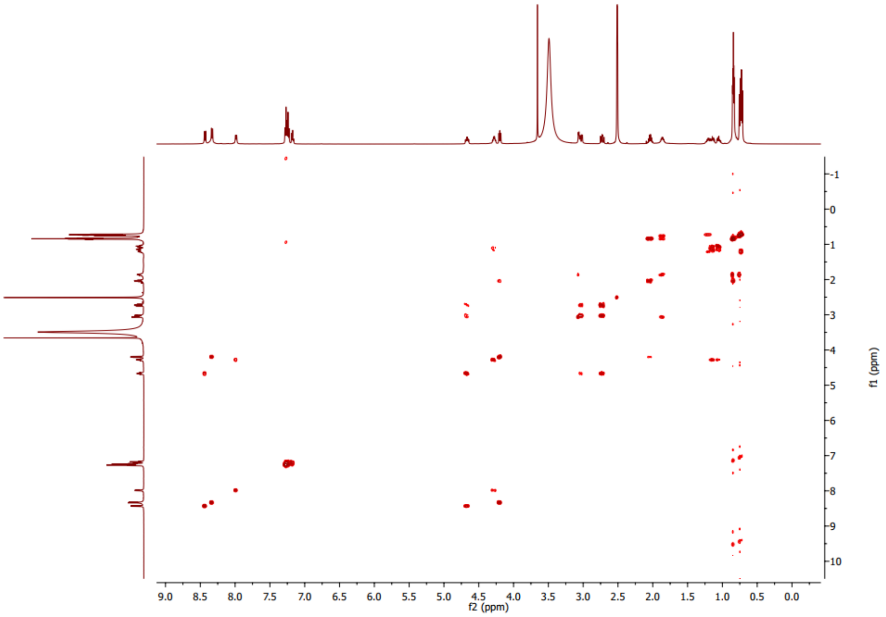
Supplementary Figure 30. ^{1}H NMR spectrum of compound **30**.



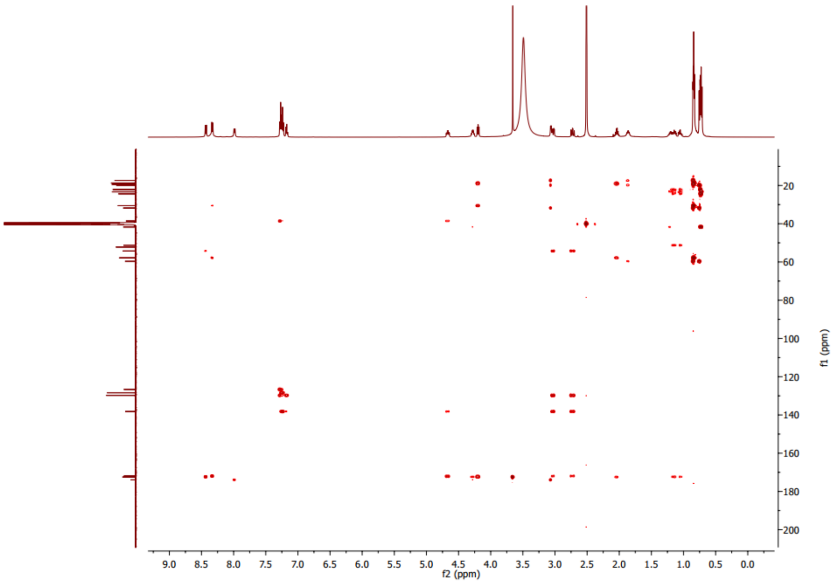
Supplementary Figure 31. ^{13}C NMR spectrum of compound **30**.



Supplementary Figure 32. HSQC spectrum of compound **30**.

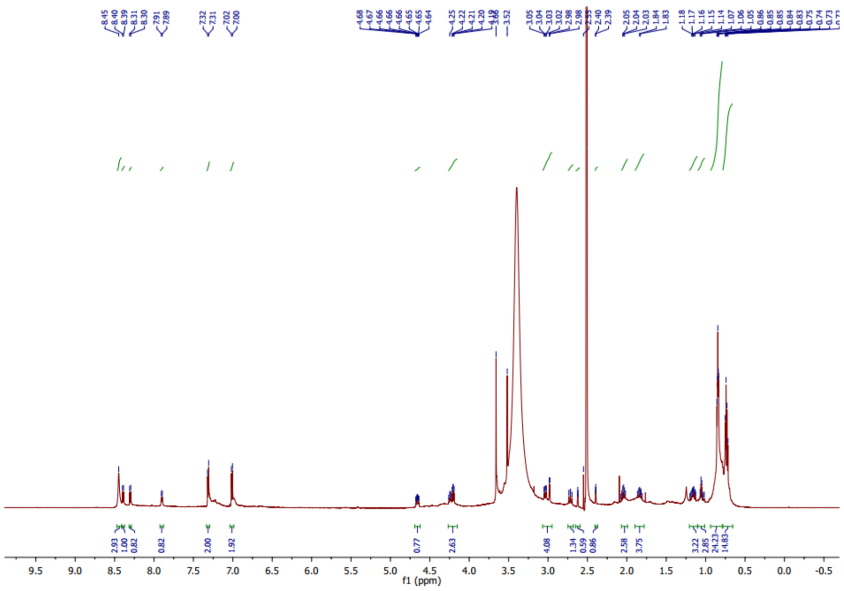


Supplementary Figure 33. COSY spectrum of compound **30**.



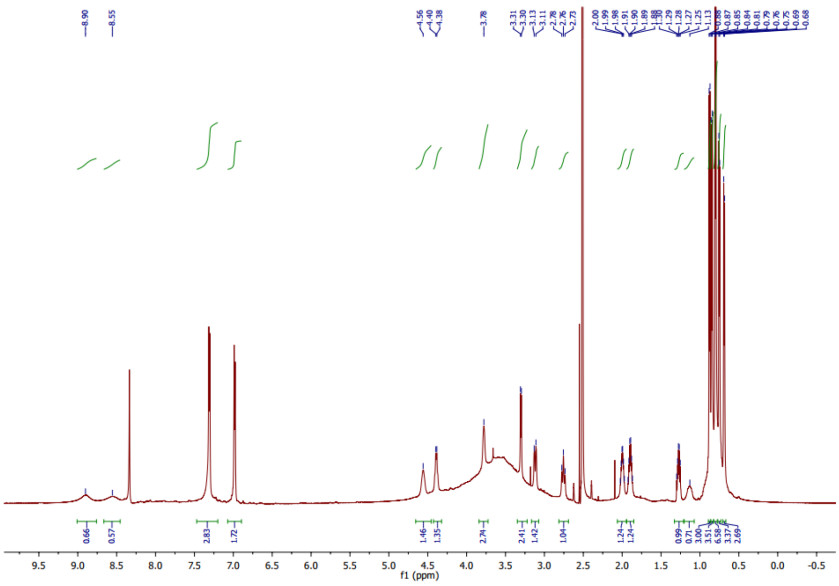
Supplementary Figure 34. HMBC spectrum of compound **30**.

Compound **35** (4.8 mg/L; white powder): HR-ESI-MS (found *m/z* 532.3236 $[M + H]^+$, calcd. for $C_{26}H_{42}N_7O_5$, 532.3242 $[M + H]^+$, Δppm 0.6).^{1}H NMR (600 MHz, DMSO) δ in ppm: 0.72 (d, 3H, J = 7.3 Hz), 0.73 (d, 3H, J = 7.3 Hz), 0.75 (d, 3H, J = 6.4 Hz), 0.84 (d, 3H, J = 6.6 Hz), 0.84 (d, 3H, J = 6.6 Hz), 0.85 (d, 3H, J = 6.7 Hz), 1.04 (m, 1H), 1.15 (m, 1H), 1.25 (br s, 1H), 1.83 (m, 1H), 2.05 (m, 1H), 2.72 (dd, 1H, J = 13.4, 11.1 Hz), 2.98 (d, 1H, J = 5.0 Hz), 3.04 (d, 1H, J = 13.4, 4.2 Hz), 3.66 (s, 3H), 4.20 (m, 1H), 4.25 (m, 1H), 4.66 (m, 1H), 7.01 (d, 2H, J = 8.3 Hz), 7.31 (d, 2H, J = 8.3 Hz), 7.90 (d, 1H, J = 7.9 Hz), 8.30 (d, 2H, J = 8.4 Hz), 8.39 (d, 1H, J = 8.8 Hz).

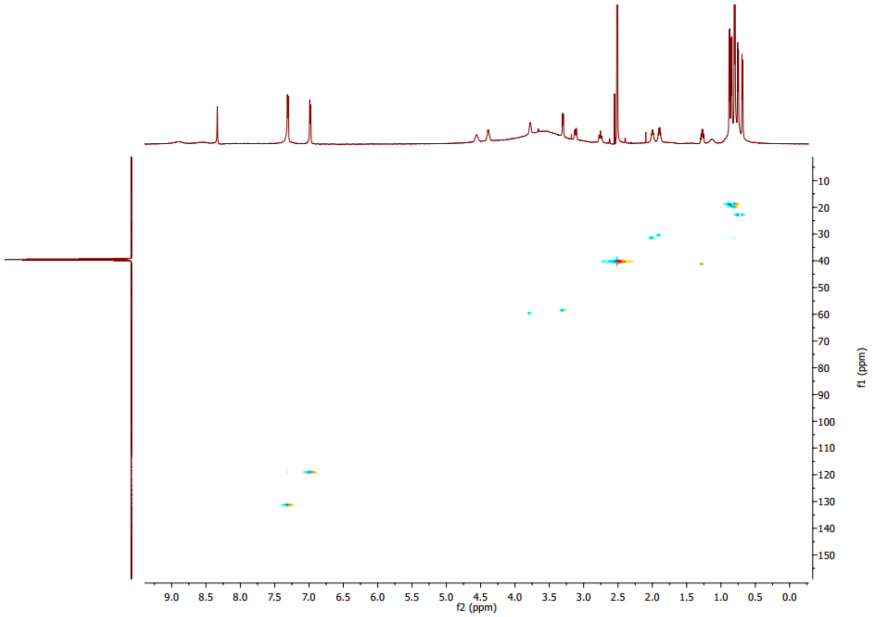


Supplementary Figure 35. ^{1}H NMR spectrum of compound **35**.

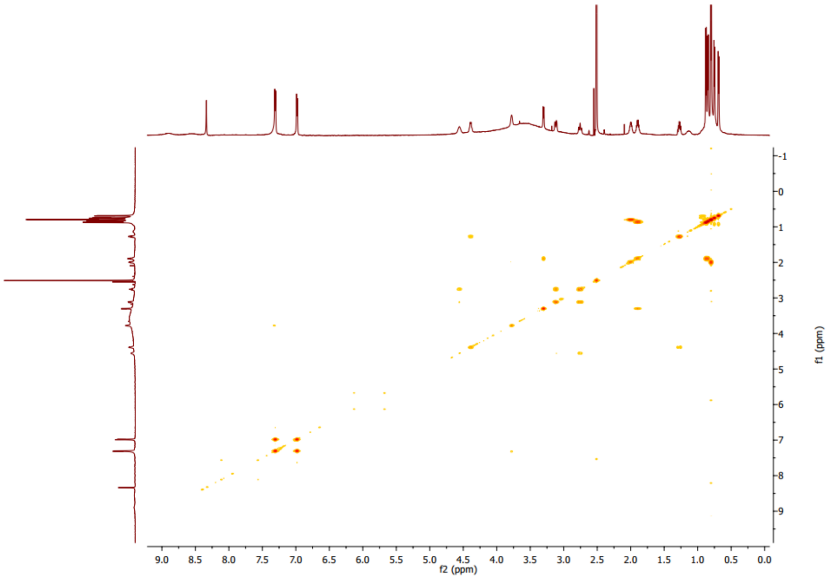
Compound **36** (6.7 mg/L; white powder): HR-ESI-MS (found *m/z* 518.3082 $[M + H]^+$, calcd. for $C_{25}H_{40}N_7O_5$, 518.3085 $[M + H]^+$, Δppm 0.7). ^{1}H NMR (600 MHz, DMSO) δ in ppm: 0.69 (d, 3H, J = 6.5 Hz), 0.75 (d, 3H, J = 6.4 Hz), 0.80 (d, 6H, J = 6.8 Hz), 0.85 (d, 3H, J = 6.7 Hz), 0.88 (d, 3H, J = 6.7 Hz), 0.95 (br s, 1H), 1.13 (br s, 1H), 1.28 (m, 1H), 1.90 (m, 1H), 2.00 (m, 1H), 2.76 (dd, 1H, J = 14.4, 12.6 Hz), 3.12 (dd, 1H, J = 14.4, 3.0 Hz), 3.30 (d, 1H, J = 7.3 Hz), 3.78 (br s, 1H), 4.39 (m, 1H), 4.56 (br s, 1H), 6.98 (d, 2H, J = 8.0 Hz), 7.31 (d, 2H, J = 8.0 Hz), 8.55 (s, 2H), 8.88 (s, 1H).



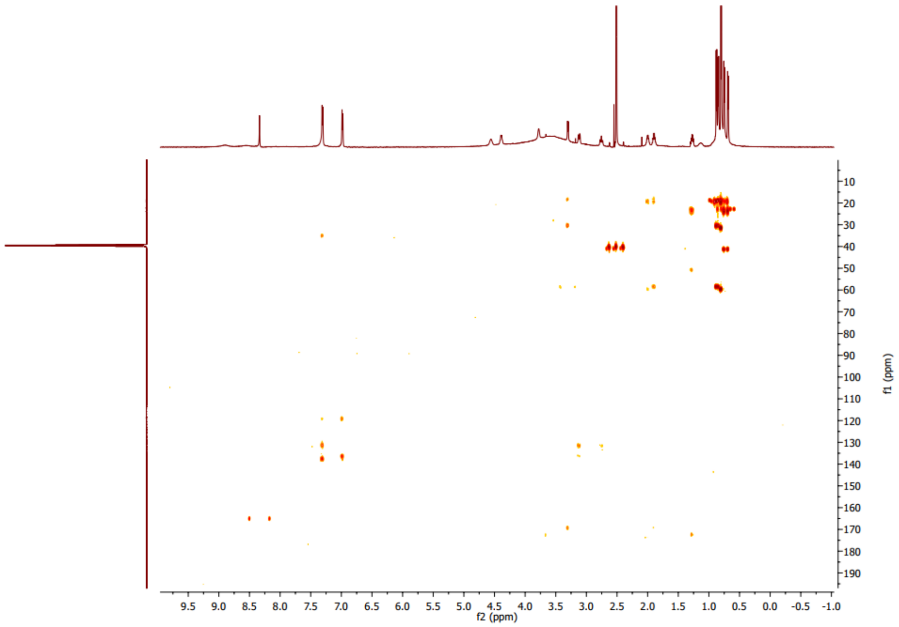
Supplementary Figure 36. ^{1}H NMR spectrum of compound **36**.



Supplementary Figure 37. HSQC spectrum of compound**36**.

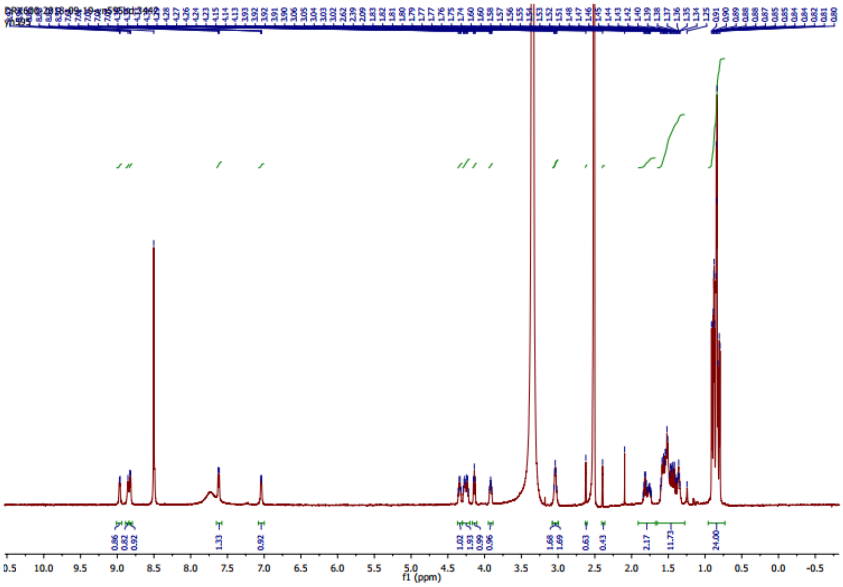


Supplementary Figure 38. COSY spectrum of compound **36**.



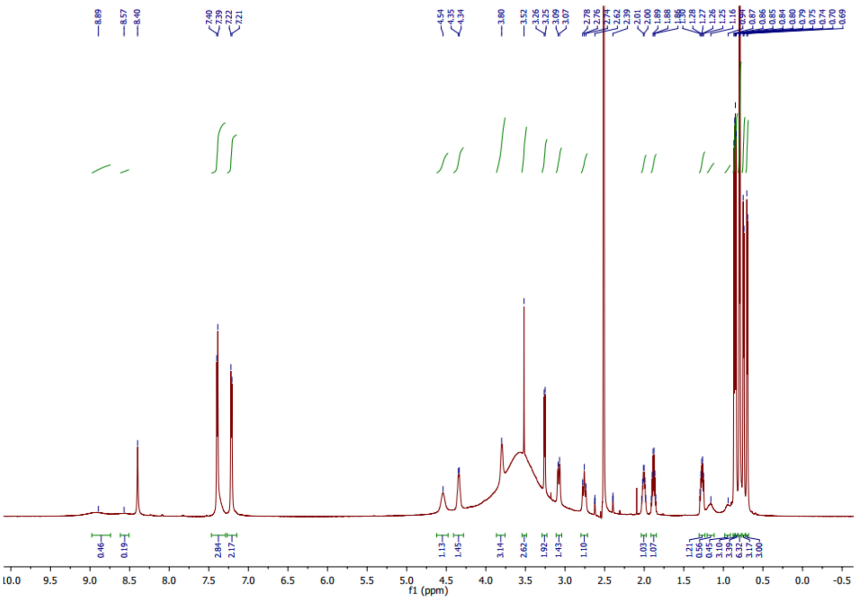
Supplementary Figure 39. HMBC spectrum of compound**36**.

Compound **42** (10.7 mg/L; white powder): HR-ESI-MS (found *m/z* 595.4289 $[M + H]^+$, calcd. for $C_{29}H_{55}N_8O_5$, 595.4290 $[M + H]^+$, Δppm 0.1). ^{1}H NMR (600 MHDMSO)δin ppm: 0.80 (d, 3H, *J* = 6.5 Hz), 0.83 (d, 3H, *J* = 7.9 Hz), 0.84 (d, 6H, *J* = 6.5 Hz), 0.85 (d, 3H, *J* = 6.6 Hz), 0.87 (d, 3H, *J* = 5.8 Hz), 0.88 (d, 3H, *J* = 6.3 Hz), 0.91 (d, 3H, *J* = 6.5 Hz), 1.32-1.61 (m, 12H), 1.75 (br s, 1H), 1.82 (m, 1H), 3.04 (br s, 2H), 3.92 (br s, 1H), 4.14 (t, 1H, *J* = 8.1 Hz), 4.24 (dd, 1H, *J* = 15.6, 7.3 Hz), 4.27 (br s, 1H), 4.34 (dd, 1H, *J* = 14.9, 7.6 Hz), 7.04 (d, 1H, *J* = 4.9 Hz), 7.62 (d, 1H, *J* = 9.0 Hz), 7.73 (br s, 1H), 8.82 (d, 1H, *J* = 7.6 Hz), 8.85 (br s, 1H), 8.97 (d, 1H, *J* = 7.12 Hz).



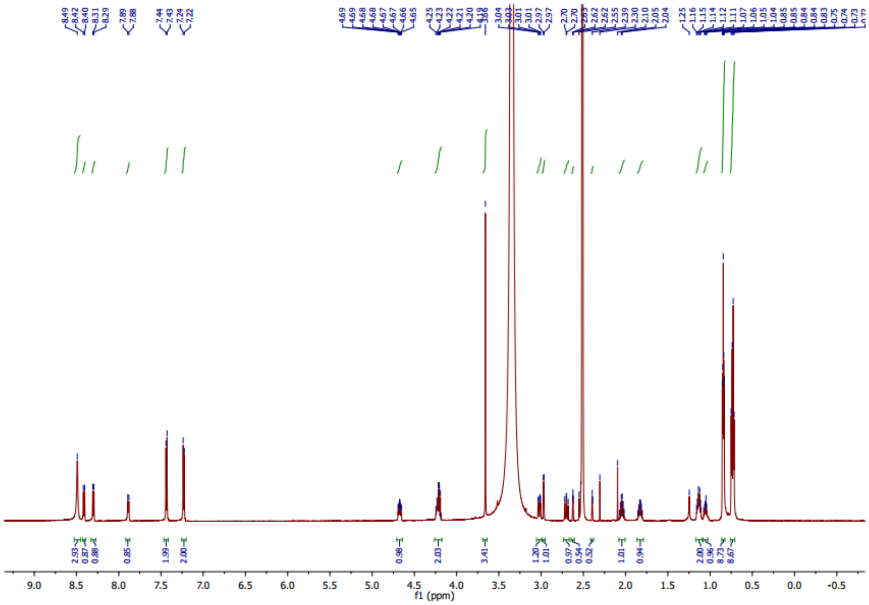
Supplementary Figure 40. ^{1}H NMR spectrum of compound **42**.

Compound **71** (6.5 mg/L; white powder): HR-ESI-MS (found *m/z* 555.2177 $[M + H]^+$, calcd. for $C_{25}H_{40}BrN_4O_5$, 555.2177 $[M + H]^+$, Δppm 0.0). ^{1}H NMR (600 MHz, DMSO) δ in ppm: 0.70 (d, 3H, J = 6.4 Hz), 0.74 (d, 3H, J = 6.3 Hz), 0.79 (d, 6H, J = 6.8 Hz), 0.84 (d, 3H, J = 6.8 Hz), 0.86 (d, 3H, J = 6.8 Hz), 0.95 (br s, 1H), 1.16 (br s, 1H), 1.27 (m, 1H), 1.88 (dq, 1H, J = 13.6, 7.0 Hz), 2.01 (dq, 1H, J = 13.1, 6.6 Hz), 2.76 (dd, 1H, J = 14.4, 12.2 Hz), 3.08 (dd, 1H, J = 14.4, 3.1 Hz), 3.80 (br s, 1H), 4.34 (m, 1H), 4.54 (br s, 1H), 7.21 (d, 2H, J = 7.9 Hz), 7.39 (d, 2H, J = 7.9 Hz).



Supplementary Figure 41. ^{1}H NMR spectrum of compound **71**.

Compound **72** (2.0 mg/L; white powder): HR-ESI-MS (found *m/z* 569.2329 [M + H]+, calcd. for $C_{26}H_{42}BrN_4O_5$, 569.2333 $[M + H]^+$, Δppm 0.7). ^{1}H NMR (600 MHz, DMSO) δ in ppm: 0.72 (d, 3H, J = 6.5 Hz), 0.73 (d, 3H, J = 7.2 Hz), 0.74 (d, 3H, J = 7.3 Hz), 0.84 (d, 3H, J = 6.9 Hz), 0.84 (d, 3H, J = 6.7 Hz), 0.85 (d, 3H, J = 6.8 Hz), 1.06 (m, 1H), 1.12 (m, 1H), 1.13 (m, 1H), 1.83 (dq, 1H, J = 13.2, 6.5 Hz), 2.04 (dq, 1H, J = 13.2, 6.6 Hz), 2.70 (dd, 1H, J = 13.3, 11.3 Hz), 2.97 (d, 1H, J = 4.9 Hz), 3.02 (dd, 1H, J = 13.3, 4.2 Hz), 3.66 (s, 3H), 4.20 (m, 1H), 4.22 (m, 1H), 4.67 (m, 1H), 7.23 (d, 2H, J = 8.3 Hz), 7.43 (d, 2H, J = 8.3 Hz), 7.89 (d, 1H, J = 7.9 Hz), 8.30 (d, 1H, J = 8.4 Hz), 8.41 (d, 1H, J = 8.8 Hz).



Supplementary Figure 42. ^{1}H NMR spectrum of compound **72**.

References

1. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl. Environ.Microbiol.***66,** 884–889 (2000).
2. Kronenwerth, M. *et al.* Characterisation of taxlllaids A-G; natural products from *Xenorhabdus indica*. *Chem. Eur. J.***20,** 17478–17487 (2014).
3. Phelan, V. V., Du, Y., McLean, J. A. & Bachmann, B. O. Adenylation enzyme characterization using gamma -(18)O(4)-ATP pyrophosphate exchange. *Chem. Biol.***16,** 473–478 (2009).
4. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.***2,** 1–4 (2007).
5. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.***2,** 31–34 (2007).
6. Fuchs, S. W., Grundmann, F., Kurz, M., Kaiser, M. & Bode, H. B. Fabclavines: bioactive peptide-polyketide-polyamino hybrids from *Xenorhabdus*. *ChemBioChem***15,** 512–516 (2014).
7. Fuchs, S. W. *et al.* Formation of 1,3-Cyclohexanediones and Resorcinols Catalyzed by a Widely Occuring Ketosynthase. *Angew. Chem. Int. Ed.***52,** 4108–4112 (2013).
8. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure***15,** 781–792 (2007).
9. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.***10,** 275–281 (2018).
10. Nollmann, F. I. *et al.* Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. *ChemBioChem***16,** 205–208 (2015).
11. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.***166,** 557–580 (1983).
12. Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *ChemBioChem***15,** 1290–1294 (2014).
13. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* 1–5 (2002). doi:10.1038/nsb810
14. Bode, H. B. *et al.* Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry. *Chem. Eur. J.***18,** 2342–2348 (2012).
15. Fuchs, S. W. *et al.* Neutral Loss Fragmentation Pattern Based Screening for Arginine-Rich Natural Products in Xenorhabdusand Photorhabdus. *Anal. Chem.***84,** 6948–6955 (2012).
16. Kegler, C. *et al.* Rapid Determination of the Amino Acid Configuration of Xenotetrapeptide. *ChemBioChem***15,** 826–828 (2014).
17. Fuchs, S. W., Proschak, A., Jaskolla, T. W., Karas, M. & Bode, H. B. Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. *Org. Biomol. Chem.***9,** 3130–3132 (2011).
18. Bode, H. B. *et al.* Structure Elucidation and Activity of Kolossin A, the D-/L-Pentadecapeptide Product of a Giant Nonribosomal Peptide Synthetase. *Angew. Chem. Int. Ed. Engl.***54,** 10352–10355 (2015).
19. Tobias, N. J. *et al.* Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.***2,** 1676–1685 (2017).

20. Cosmina, P. *et al.* Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis. *Mol. Microbiol.***8,** 821–831 (1993).
21. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.***4,** 927–937 (1997).
22. Krätzschmar, J., Krause, M. & Marahiel, M. A. 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).
23. 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).
24. Pantel, L. *et al.* Odilorhabdins, Antibacterial Agents that Cause Miscoding by Binding at a New Ribosomal Site. *Mol. Cell.***70,** 83–94.e7 (2018).
25. Zhou, Q. *et al.* Structure and Biosynthesis of Xenoamicins from Entomopathogenic *Xenorhabdus*. *Chem. Eur. J.***19,** 16772–16779 (2013).
26. Crawford, J. M., Portmann, C., Kontnik, R., Walsh, C. T. & Clardy, J. NRPS Substrate Promiscuity Diversifies the Xenematides. *Org. Lett.***13,** 5144–5147 (2011).

List of Publications and Manuscripts

Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains.

Kenan A. J. Bozhüyük*, Annabell Linck*, Andreas Tietze*, Janik Kranz*, Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-Ni Shi, Peter Grün, Helge B. Bode
Nature chemistry **10**:275–281. doi: 10.1038/nchem.2890.

*equally shared first authors

Influence of condensation domains on activity and specificity of adenylation domains

Janik Kranz*, Sebastian L. Wenski, Alexander A. Dichter, Helge B. Bode, Kenan A. J. Bozhüyük
manuscript in preparation; preprinted (non-peer-reviewed) version of the manuscript is available on *bioRxiv* (doi: 10.1101/2021.08.23.457306).

Phylogenetic analysis and rational modification of the condensation-adenylation interface sheds light on its importance for NRPS engineering

Janik Kranz*, Nadya Abbood*, Mohammad Alanjary, Helge B. Bode, Kenan A. J. Bozhüyük
manuscript in preparation

*equally shared first authors

Record of Conferences

MegaSyn Symposium on megasynthases

Poster presentation: Improving the production of non-ribosomal peptides by optimizing the interface between the condensation and adenylation didomain – from a stable platform to a multi-functional working bench

Bad Nauheim, Germany, 29.09.2020 - 01.10.2020

International VAAM Workshop on Biology of Microorganisms Producing Natural Products

Poster presentation: Influence of condensation-adenylation domain interactions on the substrate specificity and production rate of NRPS

Jena, Germany, 15.09 - 17.09.2019

3rd European Conference on Natural Products

Poster presentation: Dividing condensation domains by half – A new strategy for the engineered biosynthesis of non-ribosomal peptides

Frankfurt am Main, Germany, 02.09 - 05.09.2018

International VAAM Workshop on Biology of Bacteria Producing Natural Products

Poster presentation: Dividing condensation domains by half – A new strategy for the engineered biosynthesis of non-ribosomal peptides (winner of the poster award)

Frankfurt am Main, Germany, 31.08 - 02.09.2018

Acknowledgements

Zuallererst möchte ich mich bei meinem Doktorvater **Helge B. Bode** bedanken. Ich kam als einer der wenigen Externen an die Goethe Uni in der AK Bode und bin sehr dankbar für diese Chance.

Für die Übernahme des Zweitgutachtens dieser Arbeit möchte ich **Martin Grininger** herzlich danken.

Ein besonderer Dank gilt **Kenan A. J. Bozhüyük**, der nicht nur mich, sondern die gesamte NRPS-Engineering-Crew in der Arbeitsgruppe hervorragend bereut. Ohne die vielen hilfreichen fachlichen Gespräche, Tipps, und besonders die Zeit für unzähligen Korrekturen der Manuskripte hätte diese Arbeit nicht diese Qualität erreicht. Leider hat es dann doch nicht für die schlussendliche Publikation gereicht, aber für den Versuch und den Glauben an die Coolness meiner Ergebnisse bin ich sehr dankbar.

Ein großer Dank geht an **Carsten Kegler** und **Kenan A. J. Bozhüyük** für das Korrekturlesen dieser Arbeit. Des Weiteren danke ich **Nadya Abbood** und **Mohammad Alanjary**, die mit mir sehr viel Arbeit und Zeit in die Manuskripte gesteckt haben.

Insbesondere möchte ich mich bei dem gesamten **AK Bode** bedanken. Ohne die gute Arbeitsatmosphäre hätte ich mich vermutlich nie entschieden noch einen Doktor zu machen. Dabei danke ich besonders **Lukas Kreling**, **Nick Neubacher** und **Moritz Drechsler** für unsere illustren Runden zum dienstäglichen Feierabend; **Sebastian Wenski** neben der Zusammenarbeit für die Manuskripte für die vielen konstruktiven Pausenraumgespräche; **Margaretha Westphalen** für die fachlichen wie auch nicht-fachlichen Gespräche und Telefonate; **Gina Grammbitter** für die Hilfe besonders in meiner Anfangszeit und für die gute Nachbarschaft; und **Peter Grün**, der sich immer Zeit für uns Doktoranden genommen hat und ich dem in den Großteil meiner HPLC/MS Expertise zu verdanken habe.

Zum Schluss möchte ich noch meinen Freunden und meiner Familie danken. Meiner Freundin **Laura Paechnatz** danke ich besonders dafür, dass sie mich in der Schreibphase ausgehalten hat und viel Nachsicht mit mir hatte.

Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im Mathematisch-Naturwissenschaftlichen Bereich unterzogen habe.

Ort/Datum

Janik Kranz

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation mit dem Titel

Non-ribosomal peptide synthetase engineering focusing on the condensation domain and the condensation/adenylation domain interface

selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Ort/Datum	Janik Kranz

Curriculum Vitae

Personal Information

Name	Janik Kranz
Address:	Andréstraße 22b, 63067 Offenbach am Main
Date and place of birth:	09/10/1992 in Frankfurt am Main
Nationality:	german
Contact:	janik.kranz@gmail.com

Education

since 2018	**PhD Student** Faculty Biological Science Johann Wolfgang Goethe-University, Frankfurt a.M.
2015 – 2018	**Master of Science – Pharma-Biotechnology** Ernst-Abbe-University of Applied Sciences, Jena
2012 – 2015	**Bachelor of Engineering – Biotechnology** Ernst-Abbe-University of Applied Sciences, Jena
2009 – 2012	**A-levels** Albert-Einstein-Schule, Schwalbach (Majors: Biology and Music)

Work experience

09/2015	**Internship** at Biotechnikum/Genetic Engineering laboratory of Ernst-Abbe-University of Applied Sciences, Jena
02/2013 – 03/2013 & 07/2013 – 08/2013	**Internship** at the Institute of Pathology at Klinikum Frankfurt Höchst GmbH

Hobbies

	Rugby player Guitar player