Engineering and characterisation of non-ribosomal peptide synthetases

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich für Biowissenschaften (15)

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Andreas Tietze

aus Lindenberg i. Allgäu

Frankfurt am Main 2020

D30

vom Fachbereich für Biowissenschaften (15) der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel

Gutachter: Prof. Dr. Helge B. Bode

Zweitgutachterin: Prof. Dr. Claudia Büchel

Datum der Disputation: 16.11.2020

Acknowledgements

Mein erster Dank gebührt **Helge Bode**, der mich als Doktorvater in den letzten Jahren als Wissenschaftler großgezogen und begleitet hat. Hierzu zählen neben der Bereitstellung des Themas auf diesem hochinteressanten Forschungsgebiet auch die Rahmenbedingungen, unzählige Ideen und Motivationsstöße sowie auch mal ein Wachrütteln, wenn es nicht so läuft.

Für die Übernahme des Zweitgutachtens dieser Arbeit möchte ich **Claudia Büchel** ganz herzlich danken.

Mein Dank gilt allen **ehemaligen** und **aktuellen Mitgliedern** aus dem **AK Bode** und den **Koautoren der Publikationen**, die für die fachliche Unterstützung und die gute Arbeitsatmosphäre gesorgt haben. Vom „kannst du mir mal den Stamm ausstreichen?" über die 11:30 Uhr-Mensa-Crew bis zum Volleyball-Team *onlyhereforthebeer* und (Feierabend)Bierchen - jede*r einzelne hat einen Teil dazu beigesteuert und manche Personen möchte ich ganz besonders hervorheben:

Yan-Ni und **Yi-Ming Shi** haben mir sehr bei der Auswertung von NMR Daten geholfen. Vielen Dank dafür.

Danke an **Tien Duy Vo**, der mich bei der chemischen Synthese angeleitet hat und mit mir auch zu Beginn des PhD das Thema WG-Neugründung angegangen ist.

Ebenso sollen **Sebastian Wenski**, **Nick Neubacher** und **Jonas Watzel** nicht unerwähnt bleiben, die mit mir aus einem Jahrgang entsprungen sind und es vom Anfang bis zum nun nahenden Ende mit allen Höhen und Tiefen durchgezogen haben.

Unendlichen Dank an **Peter Grün**, der mir unglaublich viel über die Bedienung und Wartung unserer LC-MS Geräte vermittelt hat, bei jedem analytischen Problem stets kompetent zur Seite stand und immer verlässlich für eine Kaffeepause zu haben war. Diese Runde wird mit **Moritz Drechsler** und **Margaretha Westphalen** vervollständigt, die nicht nur als Teil des MS-Teams, sondern auch darüber hinaus in jeder freien Minute wesentlich zu meiner guten Laune beigetragen haben.

Acknowledgements

Table of contents

Table of contents

Table of abbreviations

4'-PPant	4'phosphopantetheine
AA	amino acid
A	adenylation
ACN	acetonitrile
ACP	acyl carrier protein
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
ATRed	non-ribosomal peptide synthetase consisting of an adenylation, thiolation and reductase domain
atred	ATRed coding sequence
BGC	biosynthetic gene cluster
BicA	bicornutin-producing synthetase
BPC	base peak chromatogram
BraB	brabantamide A-producing synthetase
C	condensation
CAR	carboxylic acid reductase
C_{Asub}	*N*-terminal acceptor condensation subdomain
CAT	chloramphenicol acetyltransferase
CDA	calcium-dependent antibiotic
C_{Dsub}	*C*-terminal donor condensation subdomain
C/E	dual condensation/epimerization
CoA	coenzyme A
COSY	correlation spectroscopy
C_{start}	starter condensation
C_{term}	terminal condensation
Cy	heterocyclization
DCM	dichlormethane

VI

DEBS	6-deoxyerythronolide B-producing synthase
DHA	dehydroalanine
DIPEA	*N,N*-diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
Dpt	daptomycin-producing synthetase
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
E	epimerization
EIC	extracted ion chromatogram
ESI	electrospray ionization
Ent	enterobactin-producing synthetase
FACS	fluorescence-activated cell sorting
Fmoc	fluorenylmethoxycarbonyl
Grs	gramicidin S-producing synthetase
GxpS	GameXPeptide-producing synthetase
HATU	*O*-(7-azabenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HCTU	*O*-(6-chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HMBC	heteronuclear multiple-bond correlation spectroscopy
HOAt	1-hydroxy-7-azabenzotriazole
HPLC-MS	high performance liquid chromatography coupled mass spectrometry
HR	high resolution
HSQC	heteronuclear single-quantum correlation spectroscopy
Hrm	hormaomycin-producing synthetase
KolS	kolossin-producing synthetase
IC_{50}	half maximal inhibitory concentration
KR	ketoreductase

Table of abbreviations

LB	lysogeny broth
LC	liquid chromatography
Lgr	linear gramicidin-producing synthetase
M	methyltransferase
MLP	MbtH-like protein
mlp	MbtH-like protein coding sequence
MS	mass spectrometry
NAD(P)$^+$	oxidised nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NMP	*N*-methylpyrrolidone
NMR	nuclear magnetic resonance
Noc	nocardicin A-producing synthetase
NRP	non-ribosomal peptide
Nrp	probable peptide synthetase
NRPS	non-ribosomal peptide synthetase
MOE	Molecular Operating Environment
MxaA	myxalamid-producing synthetase
PCR	polymerase chain reaction
PDB-ID	RCSB Protein Data Bank identification number
PFBHA	*O*-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine
PPtase	phosphopantheteine transferase
PKS	polyketide synthetase
PP$_i$	pyrophosphate
Pvd	pyoverdine-producing synthetase
PxaA	pyrrolizixenamide A-producing synthetase
R	reductase
RiPP	ribosomally synthesized and post-translationally modified peptide
RMSD	root-mean-square deviation

SAM	*S*-adenosyl-methionine
SDR	short-chain dehydrogenase/reductase
Ser-AVS	serine adenosine vinylsulfonamide
SfrA	surfactin-producing synthetase
SimC7	ketoreductase 7 of simocyclinone D8-producing synthase
T	thiolation
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TAR	transformation-associated recombination
Tyc	tyrocidin-producing synthetase
TE	thioesterase
VAAM	Vereinigung für Allgemeine und Angewandte Mikrobiologie
VibH	condensation domain from vibriobactin-producing synthetase
WT	wild type
XabS	xenoamicin-producing synthetase
Xtv	tilivalline-producing synthetase
XU	exchange unit
XUC	exchange unit condensation domain
YPD	yeast extract peptone dextrose

Summary

The growing number of infections with multi-resistant bacteria or the current COVID-19 pandemic put compounds with therapeutic properties into the public focus. Non-ribosomal peptides (NRPs) are natural products that are already marketed as antibiotics, cytotoxic agents or immunosuppressants. Their biological activities rely on the structural diversity including non-proteinogenic amino acids (AAs), heterocycles or modifications like methylation or acylation.

The biosynthesis of NRPs is carried out by non-ribosomal peptide synthetases (NRPSs). These multifunctional megaenzymes show a modular architecture like in an assembly-line. Each module is thereby responsible for the incorporation and modification of one AA and therefore contains different catalytic domains. The adenylation (A) domain recognizes and activates its specific substrate in an ATP-dependent manner which is transferred to a 4'-phosphopantetheine cofactor post-translationally attached to the thiolation (T) domain. Peptide bond formation between two T domain bound substrates catalysed by the condensation (C) domain transfers the growing peptide chain to the following module. Such a C-A-T module can be extended with optional domains to integrate structural diversity and a terminal thioesterase (TE) domain usually releases the peptide via hydrolysis or intramolecular attack of nucleophiles.

Inspired by the modular architecture, NRPS engineering deals with the modification of NRPs in order to increase biological activities, circumvent bacterial resistances or create *de novo* peptides. This can be achieved by mutasynthesis or modification of the substrate binding pocket as well as single and multiple domain substitution. However, the few successful approaches led to impaired enzymes and did not establish a general applicable guideline.

In the first publication as part of this work, the development of such a guideline comprising three rules is addressed. First, the A-T-C tridomain named exchange unit (XU) is seen as a catalytic unit instead of a module. When using them as building blocks, the C domain's specificity for the AA of the following XU has to be considered as second rule. Third, a conserved WNATE motif within the C-A linker depicts the fusion point of the XUs. Upon heterologous expression of the cloned plasmids in *E. coli* and high performance liquid

chromatography coupled mass spectrometry-based analysis of the extracts, the ambactin-producing NRPS from *Xenorhabdus* was reprogrammed with one and two XUs. This only leads to a moderate loss of production titre or an even higher one when the AA configuration was changed by introducing a dual condensation/epimerization (C/E) domain. The pentamodular GameXPeptide-producing NRPS was reconstructed using up to five XUs of four different NRPSs and even completely *de novo* synthetases were created.

The second publication describes the exchange unit condensation domain (XUC) concept and relies on a fusion point between the two subdomains (*N*-terminal C_{Dsub} and *C*-terminal C_{Asub}) of the C domain's V-shaped pseudodimeric structure which generates A-T didomains with flanking C_{Asub} and C_{Dsub}. These hybrid C domain-forming building blocks depict an improvement to the XU concept by avoiding the drawback of C domain specificity. This allows a more flexible NRPS engineering that can e.g. enable peptide library design. Furthermore, beside a combination of both concepts within one NRPS and a transfer to *Bacillus* NRPSs, the use of XUC with relaxed A domain specificity allowed further peptide modifications by introducing non-natural AAs.

The third publication deals with aldehyde and alcohol-generating reductase (R) domains which depict an alternative for peptide release in NRPSs. A promoter exchange in *X. indica* identified a pyrazine-producing NRPS with a minimal architecture of an A, T and R domain and was therefore termed ATRed. R domains were additionally used in engineered NRPSs to produce pyrazinones and derivatives thereof by XU substitution although most constructs failed to show production.

Beyond that, an R domain has been shown to replace a TE domain in wild type synthetases leading to slightly modified NRPs and the postulated biosynthesis was incidentally revised. Furthermore, an NRPS with terminal R domain was engineered to produce a free peptide aldehyde, which are known to be potent proteasome inhibitors. For the above mentioned ATReds, the presence of up to three coding regions was further identified in 20 different *Xenorhabdus* strains but only six of them were verified to produce pyrazines. All ATReds share variable sequence similarities among each other and were subsequently divided into three subtypes. One subtype is supposed to perform the pyrazine biosynthesis via a non-canonical catalytic triad.

Zusammenfassung

Die wachsende Gefahr durch antibiotikaresistente Keime sowie die aktuelle COVID-19 Pandemie rücken Wirkstoffe mit therapeutisch wirksamen Eigenschaften in den Fokus der Öffentlichkeit. Die nichtribosomalen Peptide (NRPs) umfassen Naturstoffe, die in allen drei Domänen des Lebens vorzufinden sind und heute bereits beispielsweise als Antibiotika, Zytostatika oder Immunsuppressiva vermarktet werden. Ihre biologische Aktivität basiert auf der strukturellen Diversität wie zum Beispiel der Präsenz nichtproteinogener Aminosäuren und Heterozyklen oder Modifikationen wie Methylierung und Acylierungen. Da NRPs als Teil des Sekundärstoffwechsels gebildet werden, sind sie nicht essenziell für das Überleben des jeweils produzierenden Mikroorganismus, ermöglichen es diesem jedoch sich einen Vorteil in ihrem Lebensraum zu verschaffen.

Die Biosynthese dieser Peptide wird mRNA-unabhängig durch nichtribosomale Peptidsynthetasen (NRPSs) vermittelt. Diese multifunktionalen Megaenzyme zeigen einen modularen Aufbau, weshalb sie umgangssprachlich mit der Fließbandproduktion eines Autos verglichen werden können. Jedes Modul ist hierbei verantwortlich für den Einbau einer Aminosäure und besitzt dazu verschiedene Domänen mit unterschiedlichen katalytischen Aktivitäten. Eine Adenylierungs (A)-Domäne erkennt über eine Bindetasche ihr spezifisches Substrat und wandelt dieses unter ATP-Verbrauch in einem ersten Schritt in ein reaktives Aminoacyladenylat um. Dieses wird dann auf die nachfolgende Thiolierungs (T)-Domäne kovalent an einen posttranslational angefügten 4'-Phopshopantethein-Kofaktor übertragen. Die T-Domäne besitzt keine eigene katalytische Aktivität, ist aber für den Transfer der wachsenden Peptidkette zwischen den Domänen und Modulen unerlässlich. Die Weitergabe auf das folgende Modul erfolgt mit Bildung der Peptidbindung zwischen zwei T-Domänen-gebundenen Substraten zweier benachbarter Module durch die Kondensations (C)-Domäne. Ein solches C-A-T Modul kann mit optionalen Domänen erweitert werden, um dadurch strukturelle Vielfalt zu erzeugen. Das letzte Modul besitzt meist eine Thioesterase (TE)-Domäne, über die das Peptid freigesetzt wird.

Inspiriert durch den modularen Aufbau von NRPSs befasst sich die Forschung zur Reprogrammierung dieser Enzyme seit nunmehr 25 Jahren mit der Modifizierung der pharmazeutisch relevanten NRPs, um dadurch beispielsweise eine bestehende biologische

Aktivität dieser Stoffe zu verbessern, bakterielle Resistenzen zu umgehen oder vollkommen neue Peptide zu generieren. Zu diesem Zweck können die Substrate der A-Domänen modifiziert, die Bindetasche innerhalb dieser Domänen selbst verändert oder einzelne und mehrere Domänen ausgetauscht werden. Allerdings führten die wenigen Erfolge meist zu einer verringerten Aktivität des Enzymes und es konnte kein allgemein gültiger Leitfaden entwickelt werden.

Die erste Publikation im Zuge dieser Arbeit, welche unter dem Titel „*De novo* design and engineering of non-ribosomal peptide synthetases" veröffentlicht wurde, befasst sich mit der Entwicklung eines solchen Leitfadens. Hierzu wurden drei Regeln formuliert: Erstens wird eine A-T-C-Tridomäne anstelle eines Moduls als katalytische Einheit betrachtet und als Exchange Unit (XU) bezeichnet. Die XUs werden als Bausteine für die Modifizierung und Neuorganisation von NRPSs verwendet. Da innerhalb einer XU nicht nur die A-Domäne auf ihr Substrat eine Spezifität, sondern auch die C-Domäne auf die eingebaute Aminosäure der nachfolgenden XU besitzt, muss als zweite Regel diese Spezifität einer XU berücksichtigt werden. Drittens definiert ein konserviertes WNATE-Motiv innerhalb des Linkers, welcher die C- und A-Domäne verbindet, den Fusionspunkt zwischen den XUs.

Die Umsetzung des XU Konzepts erfolgte über heterologe Expression der klonierten Plasmide in *E. coli* und Hochleistungsflüssigkeitschromatographie mit Massenspektrometrie-Kopplung (HPLC-MS) basierter Analyse der Extrakte. Am Beispiel der Ambactin-produzierenden NRPS aus *Xenorhabdus* führte dies nur zu einer schwachen Minderung der Produktionsrate, wenn ein oder zwei XUs ersetzt wurden und zu einer Steigerung, wenn durch den Austausch einer C-Domäne mit einer dualen Kondensations/Epimerisierungs (C/E)-Domäne die Aminosäure-Konfiguration an der dazugehörigen Position verändert wurde. Der Austausch von A-T-Einheiten, Modulen oder die Nichtbeachtung der zweiten Regel führte in diesem System zum Verlust der Peptidproduktion. Eine fünfmodulare NRPS, welche für die Produktion von GameXPeptid verantwortlich ist (GxpS), wurde aus bis zu fünf XUs von vier verschiedenen NRPSs rekonstruiert. Auch wenn hier die Produktionsrate abhängig von der Anzahl der Substitutionen sinkt, konnte durch die Wahl von XUs mit spezifischeren A-Domänen die Biosynthese auf ein spezifisches Derivat gelenkt werden. Die bisher genannten Beispiele

behandelten die Rekonstruktion oder die leichte Modifizierung von natürlich vorkommenden NRPSs. Mithilfe des XU-Leitfadens konnten weiterhin auch neuartige Synthetasen geschaffen werden, was die generelle Anwendbarkeit dieses Konzepts unterstreicht.

In der zweiten Publikation „Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains" wird aufbauend auf den XUs die Exchange Unit Kondensations-Domäne (XUC) beschrieben, die eine Weiterentwicklung zur vorangehenden Methode darstellt. Den Nachteil des XU-Konzeptes stellt die C-Domänenspezifität dar, welche daher als zweite Regel aufrechterhalten werden muss. Dies führt dazu, dass beispielsweise beim Austausch einer XU durch eine XU mit anderer Aminosäurespezifität auch die vorangehende XU angepasst werden musste und generell eine größere Bandbreite an XUs notwendig ist. Das Konzept der XUCs beruht auf der V-artigen Pseudodimerstruktur von C-Domänen, die in eine *N*-terminale Donorseite (C_{Dsub}) und *C*-terminale Akzeptorseite (C_{Asub}) mit dem dazwischenliegenden Substrattunnel unterteilt werden. XUCs selbst definieren sich über A-T-Didomänen, die *N*- und *C*-terminal von der C_{Asub} beziehungsweise C_{Dsub} flankiert werden. Die Hybrid-C-Domäne erlaubt nun die Kombination von XUCs aus *Xenorhabdus* und *Photorhabdus* ungeachtet der Aminosäure, welche in der nativen NRPS durch die folgende A-Domäne aktiviert wird und die Bausteine aufgrund der C-Domänenspezifität daher möglicherweise inkompatibel gewesen wären.

Auch wenn die Anwendung des XUC-Konzepts bei C-Domänen unterschiedlicher Subtypen (C und C/E) oder unterschiedlichen Ursprungs (Gram-positiv und -negativ) zu keinem Erfolg führte, konnte gezeigt werden, dass sowohl XU als auch XUC innerhalb einer NRPS anwendbar sind, dass das XUC System auf *Bacillus* NRPS übertragbar ist und dass Elongations-XUCs eines NRPS auch zur Initiation der Biosynthese eingesetzt werden können. Eine erweiterte Diversität von NRPs konnte erreicht werden, indem die Xenotetrapeptid-produzierende Synthetase mit einer XUC mit flexibler Aminosäurespezifität reprogrammiert wurde. Nach Zugabe von *O*-Propargyl-L-Tyr, *p*-Bromo-L-Phe oder *p*-Azido-L-Phe in das Kultivierungsmedium konnte die Produktion der jeweiligen Derivate beobachtet werden. Da im Gegensatz zum XU-System die XUCs

nicht von den benachbarten XUCs abhängig sind, ermöglicht dies auch die Erstellung einer Peptidbibliothek mittels randomisierter Rekombination mehrerer DNA-Fragmente.

Die dritte Publikation „Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains" behandelt monomodulare und reprogrammierte NRPSs mit terminaler Reduktase (R)-Domäne. Diese Domäne kann anstelle einer TE-Domäne für die Freisetzung des NRP zu Ende der Biosynthese verantwortlich sein und zugleich die Reduktion dessen *C*-terminaler Carboxylgruppe zu einem Aldehyd oder Alkohol katalysieren. In *X. indica* wurde ein Gen identifiziert, welches eine hypothetische NRPS bestehend aus einer A-, T- und R-Domäne kodiert. Diese minimale NRPS wurde daher ATRed genannt und konnte durch Promotoraustausch im Wildtyp der Biosynthese von Pyrazinen zugeordnet werden. Dies erfolgt über die Freisetzung der Aminosäure als Aldehyd und anschließender Dimerisierung.

R-Domänen wurden daraufhin auch im Kontext der Reprogrammierung von NRPSs untersucht und es konnte mit der R-Domäne der Tilivallin-produzierenden Synthetase (XtvB) eine Pyrazinon-produzierende Synthetase aus einem Teil der GxpS reprogrammiert werden. Diese NRPS wurde zudem mit dem XU-Konzept modifiziert und führte folglich zu einem modifizierten, nicht-natürlichen Naturstoff. Der Fusionspunkt direkt *C*-terminal der T-Domäne erwies sich hier als produktionssteigernd im Vergleich zur Aufrechterhaltung der T-R-Didomäne. Allerdings konnte weder bei derselben R-Domäne im monomodularen Kontext noch bei anderen getesteten R-Domänen eine Produktion beobachtet werden.

Darüber hinaus wurde gezeigt, dass eine R-Domäne auch anstelle einer TE-Domäne verwendet werden kann. Für die Pyrrolizixenamid-produzierende Synthetase konnte in beiden Fällen eine Produktion von Peptiden beobachtet werden, welche sich anhand einer Hydroxyl- beziehungsweise Ketogruppe unterscheiden. Dies ist auf die unterschiedlichen biochemischen Mechanismen der beiden Terminationsdomänen zurückzuführen und verdeutlicht, dass auch dadurch eine Modifikation von NRPs vorgenommen werden kann. Dies bestätigt zudem, dass die Biosynthese von Pyrrolizixenamid anders verläuft als zuvor postuliert. Es wurde angenommen, dass die TE-Domäne für die Generierung eines Dihydroalanins verantwortlich sei, wohingegen die vorliegenden Daten auf eine unübliche C-Domäne mit einem zusätzlichen Histidin im Substrattunnel hindeuten.

Zusammenfassung

Alle bisher genannten Beispiele einer reprogrammierten Synthetase mit terminaler R-Domäne ziehen eine nukleophile, intramolekulare Reaktion mit der Aldehydgruppe mit sich, sodass diese nicht frei vorliegt. Da Peptide mit freien Aldehydgruppen als Proteasominhibitoren gegen beispielsweise *M. tuberculosis* bekannt sind, wurde mit Hilfe des XU-Konzeptes eine Synthetase für die Produktion eines NRP ohne nukleophile Gruppen reprogrammiert. Folglich konnte nach heterologer Expression ein freier Peptidaldehyd nachgewiesen werden. In fortlaufenden Experimenten soll dieser nun im Komplex mit seiner potenziellen, biologischen Bindestelle kristallisiert und auf mögliche Bioaktivität evaluiert werden.

Zu guter Letzt setzt sich diese Arbeit detaillierter mit den ATReds auseinander. Entsprechende Gene wurden insgesamt 36 Mal in 20 verschiedenen *Xenorhabdus*-Stämmen identifiziert. Allerdings konnte nur bei sechs dieser Wildtyp-Stämme die Produktion eines Pyrazins detektiert werden. Bei genauerer Betrachtung der Proteinsequenzen und eines Homologie-Modelles der R-Domäne stellte sich heraus, dass jeweils genau ein ATRed aus diesen sechs Stämmen abweichende Aminosäuren in dem aktiven Zentrum der R-Domäne trägt. Dies betrifft ein zusätzliches Histidin anstelle eines Leucins und mit einer Ausnahme ist das konservierte Lysin der beschriebenen katalytischen Triade gegen ein Glutamin substituiert. Hierfür wurde ein Reaktionsmechanismus postuliert, welcher für die Pyrazinproduktion in *Xenorhabdus* von Bedeutung zu sein scheint.

Basierend auf der Proteinsequenz aller ATReds werden diese hier in drei verschiedene Subtypen unterteilt. Dies sind zum einen ATReds mit oben genannter Veränderung im aktiven Zentrum und nur Stämme mit diesem Subtyp zeigten eine Pyrazinproduktion. Weiterhin gruppieren sich ATReds mit einem kanonischen aktiven Zentrum. Der dritte Subtyp kodiert ein zusätzliches, direkt stromabwärts des ATRed-Gens kodierten, MbtH-ähnliches Protein. Die ATReds weisen jeweils eine hohe Ähnlichkeit innerhalb der drei Subtypen sowie wenig Ähnlichkeit zu den anderen Subtypen auf. Zudem weißt ein *Xenorhabdus*-Stamm mit mehreren kodierten ATReds nur maximal einen Vertreter pro Subtyp auf.

1 Introduction

"It is not difficult to make microbes resistant to penicillin in the laboratory by exposing them to concentrations not sufficient to kill them, and the same thing has occasionally happened in the body. The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant."[1]

- Alexander Fleming, Nobel lecture 1945 -

When enumerating the greatest achievements of mankind in recent history, the discovery and making use of antibiotics deserves to be named.[2] 75 years later, Alexander Fleming's warning has by now become the truth as the golden age of antibiotic discovery has faded away, the number of microorganisms with antimicrobial resistance increases and the pharmaceutical industry does not prioritize anti-infective drug development anymore because of economic reasons.[3] Since antibiotic-resistant bacteria do not respect geographical borders or people's prosperity,[4] this is a global challenge for today if there should be a solution for tomorrow.[5] In 2015, over 33 000 deaths have already been attributed to antibiotic-resistant bacteria in the European Union[6] and there is an estimated number of 10 million victims worldwide in 2050.[7] Although being a viral disease, we can also witness impacts on socio-economic areas during the current COVID-19 pandemic.[8,9]

1.1 Natural products - a source for drug development

Reports on the use of nature for human health date back to 2 600 BC when plant-derived natural products were used for treatment of ailments ranging from coughs and colds to parasitic infections and inflammation.[10] In 2010, over 33 000 natural products with biological activities have been characterised[11] and over 50 % of all approved drugs in the last decades are natural products, semisynthetic modifications thereof or inspired by natural pharmacophores.[12] Nevertheless, there is a drop in drug discovery rate caused by a decrease in natural product discovery efforts.[13] Advances in screening techniques as well as metabolomics and -genomics open up the opportunity to identify new bioactive natural products since only less than 1 % of the microbial biodiversity has been investigated so far.[14] Beside this, insights into the structural biology of natural product producing enzymes like polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs)[15] and the

modification of their biosynthetic machineries[16] can contribute to a second golden age of antibiotics.[17]

With almost 30 drugs in clinical use, non-ribosomal peptides (NRPs) contribute € billions in sales to the chemical and pharmaceutical industry.[18] In contrast to ribosomally synthesized and post-translationally modified peptides (RiPPs), NRPs are synthesized from non-ribosomal origin what has already been discovered in the 1960s at the example of the polypeptide antibiotic tyrocidine from *B. brevis*.[19] When considering all three domains of life, NRPs are found in all of them[20] with Actinobacteria, Firmicutes, α-, β-, γ-Proteobacteria and Ascomycota being the most prolific contributors.[18] For multicellular organisms, examples from *C. elegans* and *D. melanogaster* are known.[21,22]

NRPs are associated with a high structural diversity (Fig. 1; the underlying biosynthesis is discussed in chapter 1.2.4). Besides their overall arrangement as either linear or (partial) cyclic peptides, including macrolactones and -lactams[23], they can contain D-amino acids (AAs) and are not restricted to the 20 proteinogenic AAs[24], like β-AAs (e.g. xenoamicin)[25], hydroxylated AAs (e.g. bacillibactin)[26] or aminobenzoic acids (e.g. tilivalline)[27]. Furthermore, attachments of fatty acids (e.g. flavopeptin)[28], sugars (e.g. vancomycin)[29] or *C*-terminal amines (e.g. rhabdopeptides)[30] as well as N-methylations (e.g. bassianolide)[31] or heterocycles like β-lactam (e.g. nocardicin)[32] or thiazolidin (e.g. lugdunin)[33] moieties are known. This structural variety leads to different biological activities such as siderophores (e.g. enterobactin)[34], toxins (e.g. microcystin)[35], immunosuppressive (e.g. cyclosporine)[36], cytotoxic (e.g. bleomycin)[37], antiviral (e.g. feglymyin)[38] or antibacterial (e.g. bacitracin)[39].
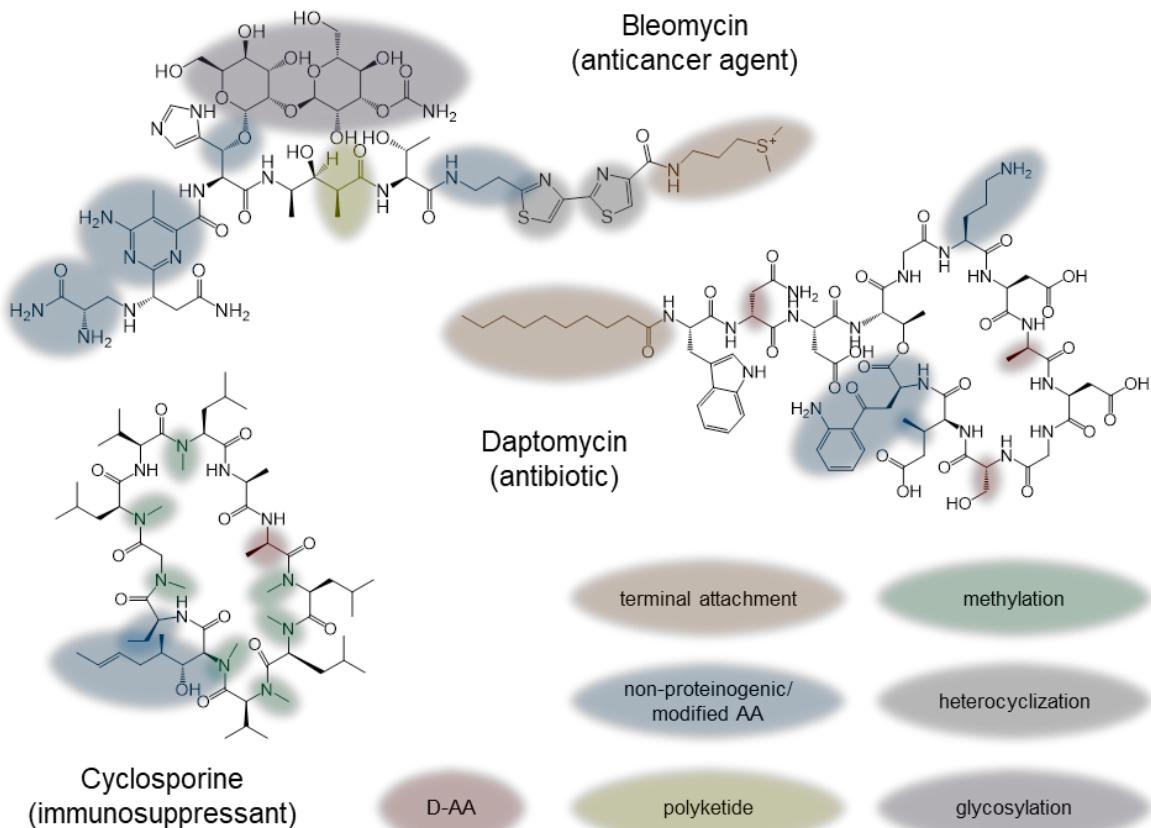
Figure 1. Structural features of NRPs. Some characteristic structural features of the anticancer agent bleomycin, antibiotic daptomycin and immunosuppressant cyclosporine are highlighted in orange (terminal attachment of fatty acids or amines), green (methylation), blue (non-proteinogenic or modified AA), grey (heterocyclization), red (D-AA), yellow (polyketide-derived elements) and purple (glycosylation).

1.2 Non-ribosomal peptide synthetases – structure and mechanism

The biosynthesis of NRPs is carried out by large multifunctional NRPSs.[24] The underlying biochemical and mechanistic principles were firstly described by the groups of Søren Laland and especially Nobel laureate Fritz Lipmann.[18] Their insights comprise the two-step activation and binding of the AA substrates under adenosine triphosphate (ATP) consumption to the NRPS as a thioester.[40–42] This was shown to be mediated by a 4'-phosphopantetheine (4'-PPant) cofactor.[43–45] Furthermore, an observed correlation between the NRPS protein size and the number of AAs within the produced NRP[43,46] led to the hypothesis that NRPSs consist of repetitive catalytic units, each being responsible for the incorporation of one AA. Although the earlier assumption that the intermediate is shuffled within the NRPS on only one 4'-PPant[46–48] has been revised to a multiple carrier thiotemplate mechanism[49], the modular assembly line-fashioned biosynthesis from the

amino to the carboxyl terminus of the peptide chain (N→C) was already proposed in the early 1970s.[50]

As mentioned above, NRPSs harbour a modular architecture and a module is defined as the catalytic unit responsible for incorporation of one AA into the growing peptide chain and optionally modification of the AA. To fulfil this function, one module is composed of domains with defined tasks (a detailed mechanism along with the structural basis will be given in the following chapters).[51] The adenylation (A) domain activates its specific substrate (Fig. 2A) which is subsequently bound onto the following thiolation (T) domain (Fig. 2B). Peptide bond formation between two T domain-bound substrates is then facilitated by the condensation (C) domain and the elongated peptidyl chain is transferred to the downstream T domain (Fig. 2C). These three core domains occur in the arrangement of C-A-T (N→C) and are typically denoted as elongation module. The first module is denoted as initiation module and can lack the C domain; the last module of an NRPS is denoted as termination module and often contains a terminal thioesterase (TE) domain for peptide release.[51]

According to their biosynthetic logic, NRPSs are classified into three groups.[51] Linear NRPSs (type A) follow the collinearity rule with one module incorporating one AA. As a consequence, the number and sequence of AAs in the final NRP is analogous to the number and order of modules. Examples are the linear gramicidin-producing synthetase (Lgr) from *B. brevis*[52] or the GameXPeptide-producing synthetase (GxpS) from *P. laumondii*[53]. NRPSs of the iterative group (type B) deviate from the single use of modules.[51] This multiple utilization of modules creates a molecular symmetry within the final product[18] like in the decapeptide gramicidin S[54] (two-time iteration of a pentapeptide) or octapeptide bassianolide[31] (four-time iteration of a dipeptide). However, repetitive use of the assembly line requires a "waiting position" of the intermediate after every iteration.[18] In gramicidin S biosynthesis, this is achieved by transfer of the pentapeptide on the TE domain which has been shown to catalyse ligation and cyclization with another T domain-bound pentapeptide[54] or at the example of bassianolide by an unusual C-A-T-T-C termination module[31]. NRPSs like the lugdunin- or capreomycin-producing synthetase[33,55] are called non-linear NRPS (type C). Within this group, single domains (not

4

whole modules) are used more than once and deviation from the standard core domain architecture occurs.[51]
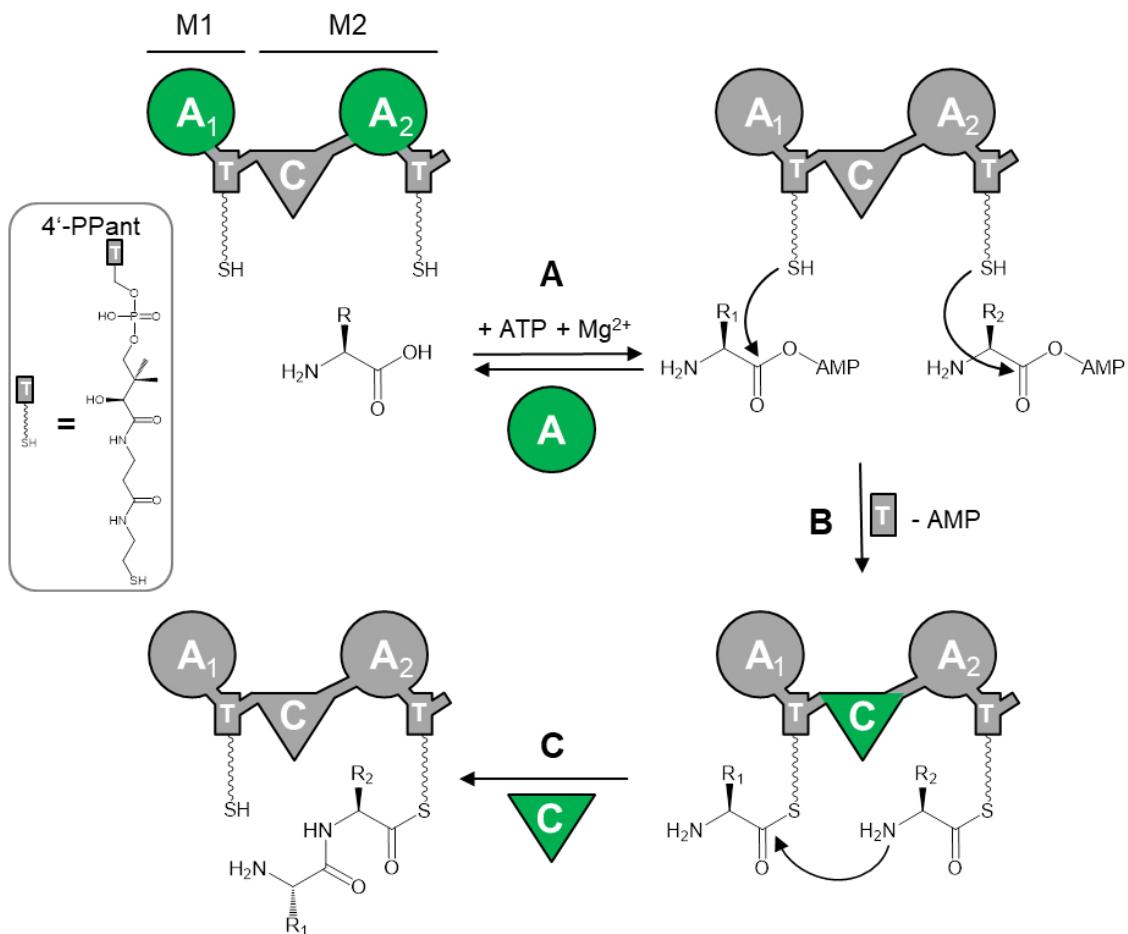
Figure 2. Basic mechanisms of the NRPS core domains. The respective active domain is highlighted in green. Domain abbreviations are A domain, circle; T domain, rectangle; C domain, triangle. The structure of the 4'-PPant moiety is shown in the box. **A.** The A domain recognizes the AA and catalyses the formation of an aminoacyl adenylate under consumption of ATP and Mg^{2+}. The definition of a module (as exemplified for module M1 and M2) is indicated by bars. **B.** The T domain was post-translationally modified with CoA for covalent attachment of the activated substrates as a thioester on the 4'-PPant moiety. **C.** The C domain catalyses peptide bond formation via nucleophilic attack of the amine of the acceptor substrate onto the electrophilic thioester of the donor substrate. Based on [24].

All aforementioned NRPSs are considered as multimodular enzymes but also monomodular NRPSs (e.g. rhabdopeptides in *Xenorhabdus*)[30], NRPS-like or minimal NRPSs lacking a C domain[56] (e.g. chloramphenicol biosynthesis in *S. venezuela*)[57] as well as even single stand-alone domain (e.g. stand-alone C domain as part of the fabclavine biosynthesis in *Xenorhabdus*)[58] are known. The latter example is also a representative of widespread hybrid NRPS/PKS enzymes.[20] Notably, Walsh and co-workers identified a combination of the non-ribosomal and ribosomal route of peptide synthesis in *S. coeruleorubidus*.[59]

1.2.1 Adenylation domain

A domains belong to the superfamily of adenylate forming enzymes along with acyl-coenzyme A (CoA) synthetases or firefly luciferases and catalyse a two-step reaction.[60] The first reaction step comprises the activation of the substrate to form a highly reactive aminoacyl adenylate (Fig. 2A). Here, the carboxy group of the AA is adenylated using ATP, requiring Mg^{2+} and releasing pyrophosphate (PP_i).[24] A domains are, with a few exceptions,[61] specific for L-AAs. Transfer of the activated substrate to the T domain is achieved by nucleophilic attack of the thiol group of the 4'-PPant moiety to form a thioester (Fig. 2B). In this second step, adenosine monophosphate (AMP) is released.[62] The function of A domains can be compared with those of aminoacyl tRNA synthetases from the ribosomal pathway although they share no similarity in structure or sequence.[24]

Conti *et al.* presented the first structure of a phenylalanine-activating A domain from the gramicidin S-producing synthetase (Grs) in *B. brevis* (PDB-ID: 1AMU) and revealed the sub-division in a larger *N*-terminal A_{core} (ca. 50 kDa) and *C*-terminal A_{sub} (ca. 10 kDa) subdomain (Fig. 3).[63] Within the binding pocket, several residues have been identified which are facing towards the substrate and are important for substrate recognition and positioning.[64] While Lys517 and Asp235 are responsible for positioning the AA by stabilizing the carboxy and amino group respectively, another eight crucial residues determine the A domains substrate specificity and are therefore referred to as specificity-conferring or Stachelhaus code (Fig. 3). This knowledge can be used to predict the specificity of A domains *in silico*.[65,66]
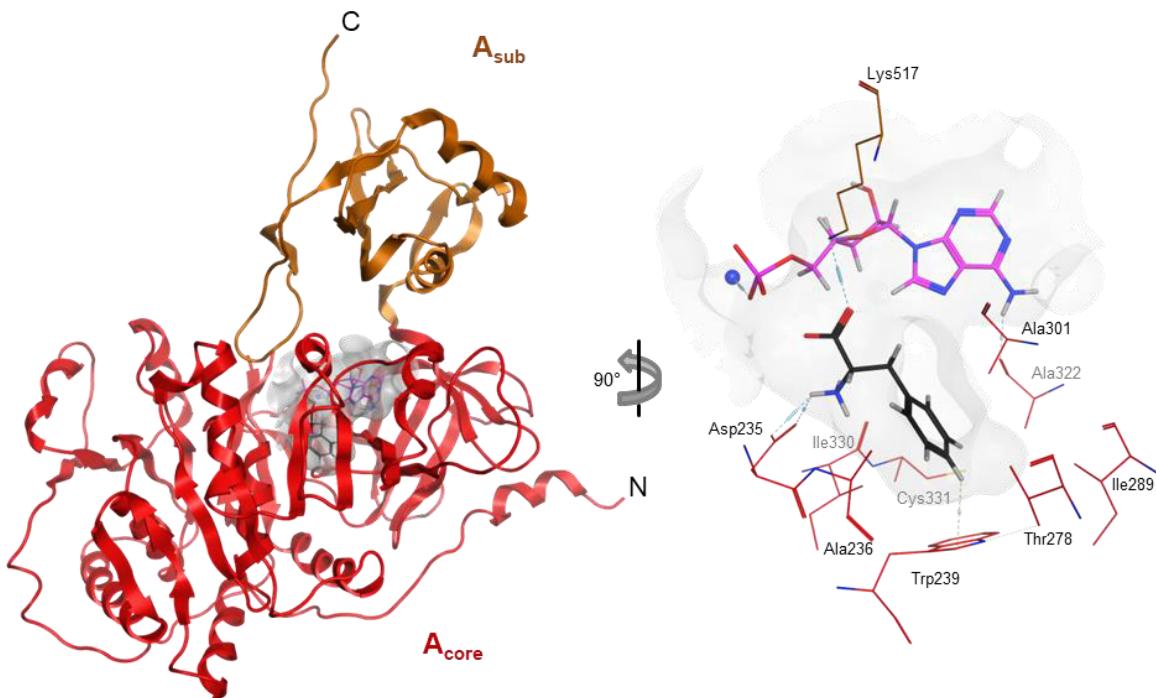
Figure 3. Structure and specificity-conferring code of an A domain.[63,64] **Left.** Ribbon diagram of the A domain from gramicidin-producing NRPS in *B. brevis* (PDB-ID: 1AMU) with the A_{core} (red), A_{sub} (orange) domain, phenylalanine (black), ATP (purple) and Mg^{2+} (blue). The binding pocket is indicated by a grey sphere. **Right.** Close-up view of the binding pocket with the specificity-conferring side chains in their respective subdomain colour. Based on [63,64] and processed with Molecular Operating Environment (MOE) 2016 (Chemical Computing Group).

Both A domain reactions are mediated by different conformations of the A domain.[67,68] The open conformation allows binding of the AA and ATP in the active site since the flexible A_{sub} subdomain is facing away from the A_{core} subdomain and the binding pocket is accessible. In the adenylation state, the formation of the aminoacyl adenylate intermediate takes place and the A_{sub} subdomain rotates 140° towards the A_{core} subdomain to concurrently displace the conserved Lys from the carboxy group. This rotation is mediated by one of ten (A1 - A10) conserved A domain core motifs and allows the 4'-PPant moiety to form the thioester.[62] This is called the thiolation or closed conformation. With conversion back to the open conformation, a new cycle can start.[69]

In addition, A domains can interact with small interaction partners called MbtH-like proteins (MLPs).[70] Their role is still not fully explored but MLPs are shown to be essential for some A domains for their activity[71] or solubility[72] but are also exchangeable against each other[73].

1.2.2 Thiolation domain

Although the T domain does not have a catalytic activity itself, it occupies a central role within the NRPS machinery and other biosynthesis pathways by carrying and shuttling the growing intermediates between the different active sites (the overall interactions and catalytic cycle will be discussed in chapter 1.2.6).[74] The T domain is the smallest (ca. 10 kDa) of the three core domains and adopts a four helix-bundle with hydrophobic interactions.[70] A conserved GGxS core motif is located at the start of the second α-helix.[75] The shift from the inactive *apo*- to the active *holo*-form is achieved by post-translational attachment of a CoA-derived 4'-PPant moiety to the serine residue of the core motif.[76]

This is catalysed by 4'-phosphopantetheine transferases (PPtases) of the Sfp-type.[77] These are named after the gene *sfp* of the firstly described PPtase of the surfactin-producing synthetase (SfrA) from *B. subtilis*. Sfp-like enzymes generally are involved in activating T domains from NRPS systems.[78] Alignments of Sfp-type PPtases and T domains revealed widely conserved hydrophobic residues responsible for interaction and support a promiscuous specificity of PPtases for 4' PPant modification of different NRPS systems.[78]

1.2.3 Condensation domain

The C domain (50 kDa) belongs to the family of chloramphenicol acetyltransferases (CATs) and catalyses nucleophilic attack of the α-amino group of the 4'-PPant-bound acceptor substrate of the *N*-terminal T_n domain onto the thioester of the 4'-PPant-bound donor substrate of the *C*-terminal T_{n+1} domain (Fig. 2C).[79] The latter one then carries the elongated peptide intermediate. The condensation reaction was assigned to C domains for the first time when alignments identified a conserved HHxxxDG motif within NRPSs analogous to the number of condensation reactions performed by the respective NRPS.[80] Mutational analysis of C domains confirmed the peptide bond-forming task and highlighted the importance of the second histidine and the aspartate within this His-motif.[81,82] The histidine has been postulated to act as a general base catalyst, however mutation of this critical residue did not abolish the activity of the stand-alone C domain in vibriobactin biosynthesis (VibH) in *V. cholerae*[83] and there is evidence that the His-motif may occupy different roles in different C domains[70,79] like positioning the substrate[84] or enhancing the solubility of the enzyme[82].

First structural insights have been gained in 2002.[83] VibH and subsequent solved structures like the C domain from LgrA[85] share the canonical overall architecture of a pseudodimer with two lobes in a "V"-shape (Fig. 4).[79] These lobes are referred to as *N*-terminal donor (C_{Dsub}) and *C*-terminal acceptor (C_{Asub}) condensation subdomains. They allow the interaction of the two substrate-loaded 4'-PPant moieties of the *C*-terminal T domain (T_D) as donor substrate and the *N*-terminal T domain (T_A) as acceptor substrate with the His-motif which is located in a tunnel ~15 Å inside the "V" and surrounded by the so-called latch and floor loop.[79]
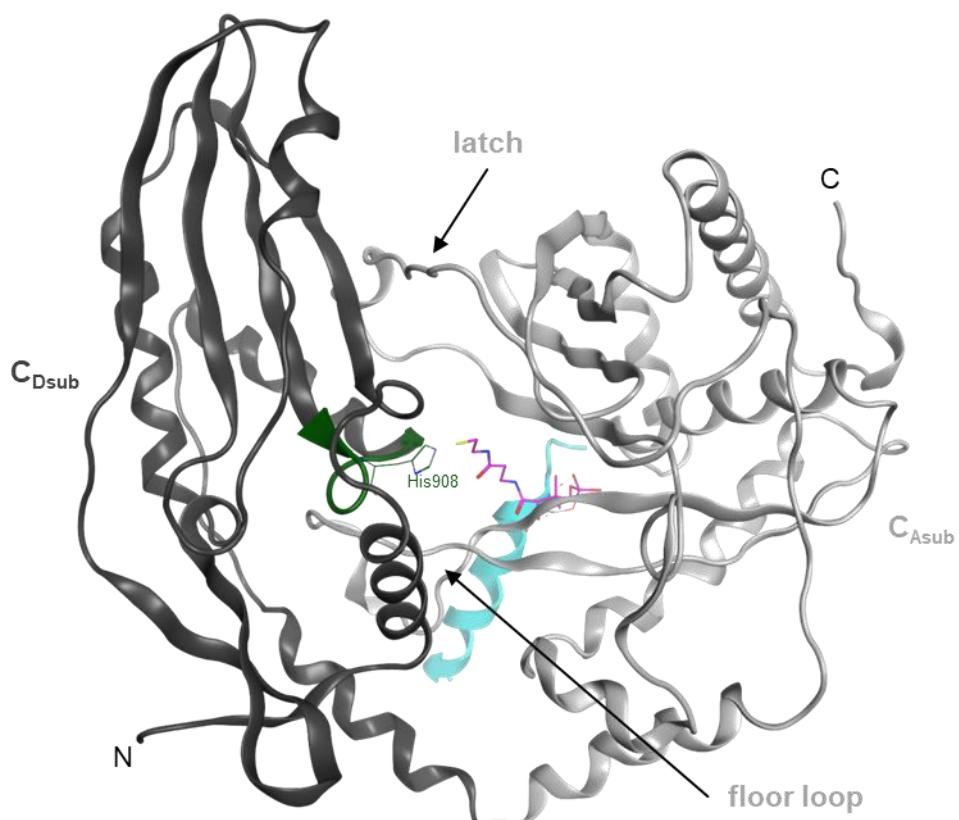
Figure 4. Structure of a C domain.[85] Ribbon diagram of the C domain of LgrA from *B. parabrevis* (PDB-ID: 6MFY) with the *N*-terminal C_{Dsub} (black) and *C*-terminal C_{Asub} (grey). The HHxxxDG motif with the catalytic active His908 is highlighted in green, the T domain (excised α-helix 2; light blue)-bound 4'-PPant moiety of the donor site in purple and the latch and floor loop are indicated by arrows. Based on [85] and processed with MOE 2016 (Chemical Computing Group).

A domains are responsible for the recognition of a specific AA and therefore the specificity of NRPSs.[64] Beyond that, it has been shown that C domains can influence their adenylation activity *in vitro*.[86,87] C domains also contribute to a proofreading mechanism preventing the enzyme of synthesising incorrect NRPs and facilitating the N→C biosynthesis direction.[24,79] It is known that the C domain exhibits side chain and configuration

selectivity for the acceptor substrate, although the underlying principle, like the presence of a Stachelhaus code in A domains, is not understood.[88,89] At the donor side, a selectivity for the AA configuration was shown and the C domains are accordingly denoted as ^DC_L or ^LC_L (condensation between a donor D- respectively L-substrate with an acceptor L-substrate).[90] Beside the classification in D/L catalysts, there are further subclasses of C domains like epimerization (E), dual condensation/epimerization (C/E), starter C (C_{start}), terminal C (C_{term}), heterocyclization (Cy) or X domains known, which are involved in the modification process of NRPs.[91]

1.2.4 Editing domains

Modification processes, associated with increasing the structural diversity of NRPs, are realised by editing domains.[18] Besides precursor modification (e.g. synthesis of 3-hydroxy-5-methyl-*O*-methyltyrosine in safracin biosynthesis)[92] or post-NRPS modification (e.g. cross-linking of vancomycin)[29], editing domains can act during NRP synthesis *in cis* or *in trans* and are optional domains beside the three core domains. In this section, a few examples will be highlighted; for further modification possibilities please refer to a detailed review.[18]

1.2.4.1   Fatty acid attachment

C domains in initiation modules are responsible for the *N*-terminal incorporation of fatty acids in lipopeptides.[91] These C_{start} domains catalyse acylation of the first 4'-PPant-bound AA with acyl carrier protein (ACP)-bound carboxylic acids[93] which can include branched[94], hydroxy and unsaturated[95] fatty acids or carboxylic acids, i.e. phenylacetic acid[96].

1.2.4.2   Epimerization

In *Bacillus*-derived linear gramicidin A, an alternating sequence of D- and L-AAs forms a helical structure and induces membrane permeabilization.[52] The presence of these D-AA is due to E domains which are located *N*-terminal of C domains. They harbour the same CAT structure as well as the catalytic HHxxxDG motif as C domains with an additional conserved Glu residue[97] and create a racemate by deprotonation of the α-carbon of the T domain-bound donor AA.[98] To ensure the incorporation of the D-AA from the D/L

equilibrium, E domains are followed by C domains of the $^{D}C_{L}$ subtype.[90] Furthermore, E domains in elongation modules act preferentially on donor peptidyl substrates (elongated intermediate of the condensation reaction of the *N*-terminal C domain) than donor aminoacyl substrates (AA tethered to the T domain before condensation reaction) which is in sum another checkpoint for keeping the directionality of NRPS biosynthesis.[99]

In 2005, the Walsh group observed D-AAs in athrofactin without an E domain in the respective NRPS.[100] They identified that some C domains are able to perform both, epimerization and condensation reaction. These C/E domains contain an extended HH[I/L]xxxxGD motif next to the N-terminus in addition to the His-motif of C domains within the tunnel.

1.2.4.3   Heterocyclization

Cy domains substitute C domains in NRPS assembly lines and introduce heterocycles like thiazoline in bacitracin[39] or oxazoline in mycobactin[101]. They catalyse both, condensation between donor and acceptor (Ser, Thr or Cys) substrate as well as nucleophilic attack of the acceptor side chain onto the newly-formed peptide bond to form a hydroxylated thiazolidine (for Cys) and oxazolidine (for Thr and Serine), respectively, which is subsequently dehydrated.[102] This can be followed by further reduction or oxidation.[103,104]

1.2.4.4   Methylation

Methyltransferase (M) domains transfer a methyl group from *S*-adenosyl-methionine (SAM) to the NRP.[24] Although most of the methylations are carried out on nitrogen atoms (e.g. bassianolide)[31] of the adjacent 4'-PPant-bound AA, *O*- (e.g. kutzneride)[105], *C*- (e.g. yersiniabactin)[106] and *S*- (e.g. thiocoraline)[107] methylations are additionally known. M domains are mainly embedded within the A_{sub} domain between core motifs A8 and A9.[18] But also other A domain interruptions[107], separated from the A domain within the assembly line[108] or stand-alone M domains[109] have been observed.

1.2.5 Peptide release

Since the growing intermediate is tethered by a 4'-PPant moiety, the last step in NRP assembly has to be the release from the enzyme. In this section, two fundamental examples for peptide release will be highlighted; for further routes please refer to a detailed review.[18]

1.2.5.1   Thioesterase domain

TE domains can be divided into two types.[110] Type I TEs of the α/β-hydrolase family are located *C*-terminal within a termination module and enable access of the peptidyl-4'-PPant substrate via a flexible lid to the catalytic triad (Ser-Asp-His) (Fig. 5A left).[111] Here, the catalytic base/acid His is stabilized by Asp and enhances the nucleophilic character of the Ser which attacks the thioester in order to bind the substrate as peptidyl-*O*-Ser-TE on the TE domain. Subsequently, hydrolysis of the ester bond releases a linear peptide but intramolecular attack of nucleophiles can also release cyclic structures as macrolactons (e.g. xenoamicin)[25] or -(thio)lactams (e.g. thiocoralin and GameXPeptide respectively)[53,112] (Fig. 5A right).[113] As a consequence, TE domains are not only responsible for peptide release; but provide additional possibilities for increasing structural diversity of NRPs.

T domains with acetylated 4'-PPant moieties or wrongly loaded AAs would bring the NRPS assembly line to a deadlock by covering active residues or the C domain's proofreading mechanism. Regeneration of such misprimed T domains is achieved via hydrolysis of the 4'-PPant thioester by type II TEs.[114] In contrast to type I TEs, they act *in trans* and exhibit low substrate specificity.[115]
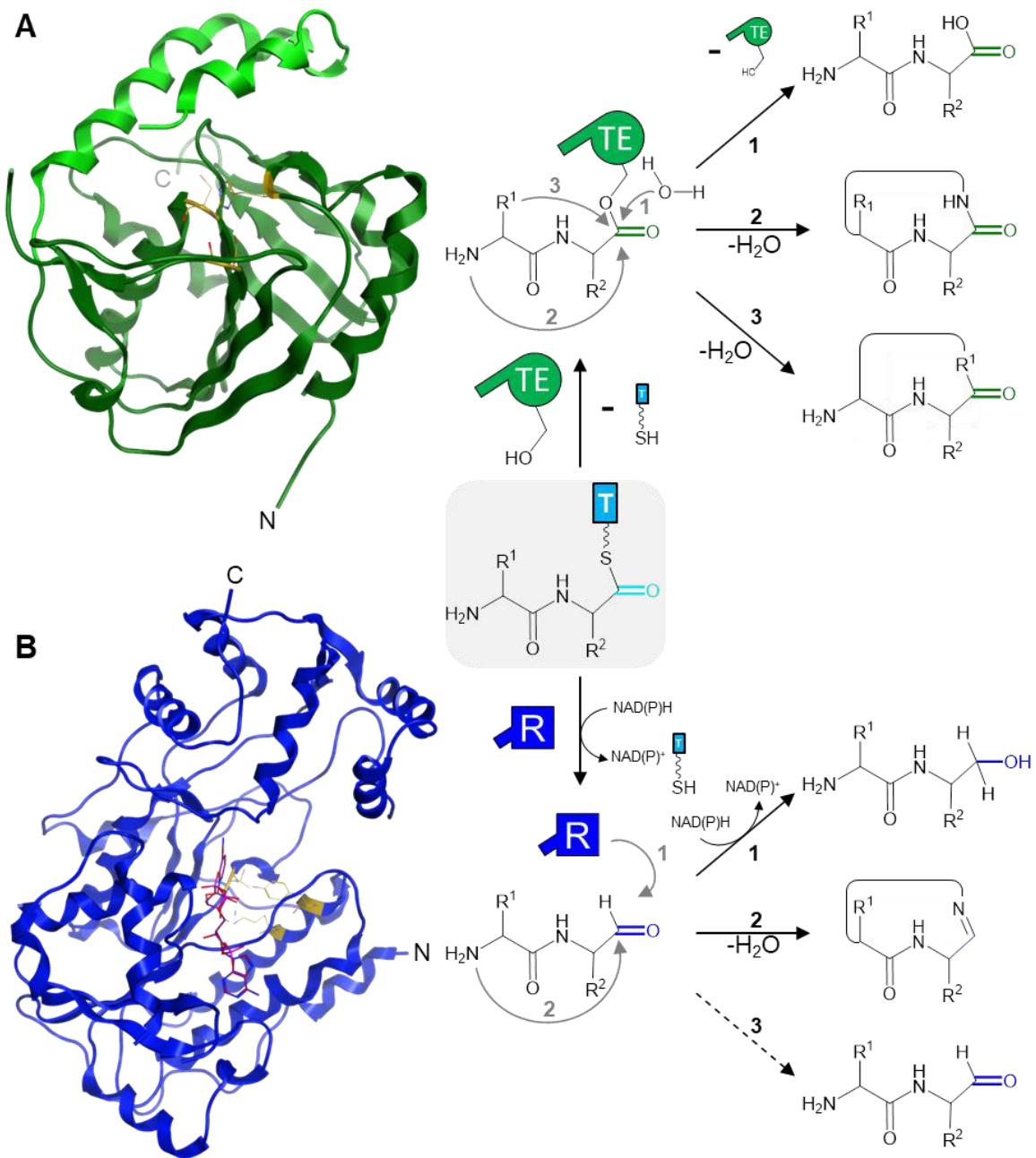
Figure 5. TE and R domain-mediated peptide release in NRP biosynthesis.[18,110] The starting point is the peptidyl-4'-PPant thioester as exemplified by a dipeptide (highlighted in grey). The *C*-terminal carbonyl group is shown in light blue and during further processing in green and blue (TE (circle) or R domain (square)-mediated release respectively) **A. Left.** Ribbon diagram of the TE domain from SfrA in *B. subtilis* (PDB-ID: 1JMK).[111] The lid is highlighted in lighter green and the active site residues in yellow. **Right.** Nucleophilic attack of the active site Ser on the thioester, formation of a TE domain-bound ester and release from the T domain. The ester can be cleaved by (1) hydrolysis on water to release a linear peptide, (2) intramolecular attack of the *N*-terminal amino group to release a cyclic peptide (head-to-tail) or (3) intramolecular attack of another nucleophile to release a branched cyclic peptide (side-chain-to-tail). **B. Left.** Ribbon diagram of the R domain from myxalamid-producing PKS-NRPS in *S. aurantiaca* (PDB-ID: 4U7W).[116] The active site residues are highlighted in yellow, NADPH in purple and the *C*-terminal hydrophobic insertion in brown. **Right.** NAD(P)H-derived hydride transfer on the thioester and release from the T domain. The aldehyde can be exposed to (1) a second reduction to an alcohol by the R domain, (2) intramolecular attack of the *N*-terminal amino group to form a cyclic imine or (3) no further modification. Based on [18,110,111,116] and processed with MOE 2016 (Chemical Computing Group).

1.2.5.2   Reductase domain

Peptide release and further structural diversity can also be realised by terminal reductase (R) domains.[117] Structural insights gained similarities to the short-chain dehydrogenase/reductase (SDR) superfamily with a Rossmann fold (TGxxGxxG) for cofactor nicotinamide adenine dinucleotide (phosphate) (NAD(P)H) binding and a catalytic triad (Thr/Ser-Tyr-Lys) (Fig. 5B left).[116] These two motifs are located within a larger *N*-terminal subdomain and the less conserved *C*-terminal subdomain contains a hydrophobic insertion.[118]

R domains catalyse reduction of the peptidyl-4'-PPant thioester to an alcohol without covalent binding of the substrate onto the domain (Fig. 5B right). A hydride is transferred from NAD(P)H to the thioester via a thiohemiacetal intermediate (stabilized by Tyr and a proton relay system) to form an aldehyde which can subsequently be further reduced to an alcohol (e.g. myxalamid A).[116,119] This is a nonprocessive $[2+2]e^-$ mechanism, i.e. the aldehyde dissociates from the R domains active site for cofactor exchange and reassociates.[118] Some R domains are capable of performing only the first round of reduction to the aldehyde (e.g. flavopeptin).[28] This has been shown for carboxylic acid reductases (CAR) to be regulated by an absent reorientation of the nicotinamide moiety to a catalytically competent position in the second round due to no 4'-PPant binding.[120] Because of the high reactivity of aldehydes, such products can undergo intramolecular cyclization with e.g. the *N*-terminal amino group to form an imine which can be maintained (e.g. nostocyclopeptide)[121] or further modified (e.g. tilivalline)[27]. Free aldehydes groups (e.g. flavopeptin or fellutamide B)[28,122] can function as warheads since they can mediate potent protease inhibitor activity due to a nucleophilic attack onto the aldehyde.[123]

1.2.6 Multidomain aspects of non-ribosomal peptide synthetases

In 2008, the crystal structure of SrfA-C from *B. subtilis* provided first insights into the overall structure of NRPSs and revealed the interaction between the domains.[124] These are connected via linker regions ranging in length from 9 (T-TE) and 15 (A-T) up to 32 AAs (C-A). Marahiel and co-workers highlighted the numerous interactions of the L-shaped C-A linker with its adjacent domains as well as among themselves and stated this domain interface as inseparable. Furthermore, they noted that the distances from the T domain's core motif to the active sites of the C, A and TE domain (up to 57 Å) exceed the length of the 4'-PPant moiety (20 Å), although the T domain is the protagonist in the shuttle of the intermediates. This implicated an overall structural rearrangement and flexibility during the catalytic cycles of NRPSs[124,125] since the T domain itself remains rigid regardless of whether being in the *apo* or *holo* state or in different catalytic steps what has been proven by following studies.[70]

Drake *et al.* described the structures of two *holo* NRPSs revealing distinct steps in the catalytic cycle of NRPSs.[126] The termination module from enterobactin-producing NRPS (EntF) in *E. coli* is covalently trapped with serine adenosine vinylsulfonamide (Ser-AVS) within the A domain for thioester formation (Fig. 6A) whereas the 4'-PPant moiety of the termination module from *holo*-AB3403-NRPS in *A. baumannii* resides in the C domain depicting the condensation state (Fig. 6B). Notably, the C-A interface and overall conformations of EntF, *holo*-AB3403 and SrfA-C[124] differ and e.g. EntF is incompatible of the aminoacyl adenylate forming state. Based on this, the 140° rotation of the A_{sub} domain (see chapter 1.2.1) is concluded to be a structural mechanism and guides the T domain with the 4'-PPant moiety between the active sites.[60,126]

In addition, Drake *et al.* established a model which connects the three active sites, where the T domain bound 4'-PPant moiety has to interact with the four catalytic structural states.[126] The latter ones are (I) substrate binding in the A domain in the open conformation, (II) thiolation reaction with the 4'-PPant moiety in the closed conformation (Fig. 6A), (III) delivery to the upstream C domain for condensation as acceptor aminoacyl (Fig. 6B) and (IV) delivery to the downstream C domain as donor peptidyl for condensation or peptide release in case of an termination module. Strikingly, states (I) and (III) are structurally identical i.e. peptide bond formation in the C domain as acceptor

substrate and AA adenylation in the A domain can occur simultaneously within one NRPS module.[126]
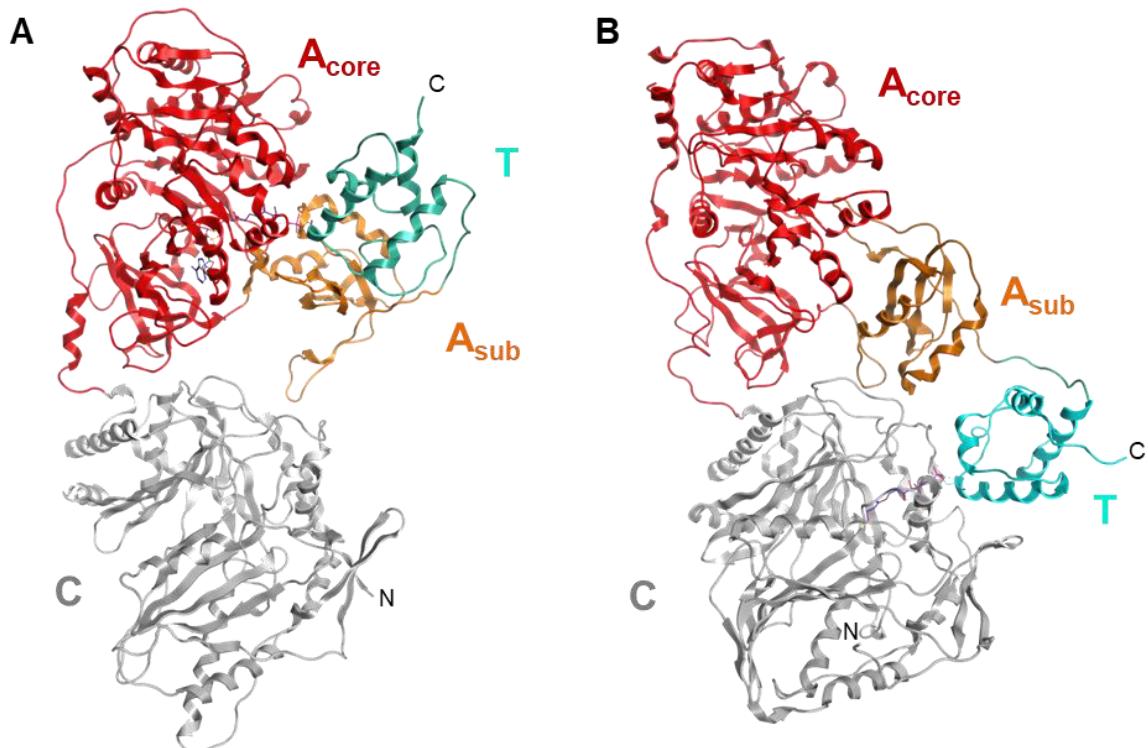
Figure 6. Structure of an NRPS module and domain movement.[126] Comparison of two superposed termination modules with the C (grey), A_{core} (red), A_{sub} (orange) and T (light blue) domain as well as the 4'-PPant moiety (purple). The TE domains are not shown. **A.** Ribbon diagram of the termination module from EntF in *E. coli* (PDB-ID: 5T3D).[126] The catalytic state of thiolation within the A domain is trapped by the inhibitor Ser-AVS (black). **B.** Ribbon diagram of the termination module from *holo*-AB3403-NRPS in *A. baumannii* (PDB-ID: 4ZXI).[126] Here, the T domain delivers the 4'-PPant moiety to the condensation state within the C domain. These two structures implicate a domain rearrangement during the catalytic cycle of NRPSs. Based on [126] and processed with MOE 2016 (Chemical Computing Group).

A great contribution to the understanding of multimodular NRPSs during the catalytic cycles has been achieved by Schmeing and co-workers.[127,128] In a current and remarkable study, they determined five independent structures of LgrA including the initiation module up to the three following domains of the elongation module.[85] In sum, they concluded that multimodular NRPSs are very flexible as no strict coupling between the catalytic states of a particular module and the overall conformation of the multimodular NRPS has been observed. Different models for the higher order architecture of multimodular NRPSs have been proposed,[129,130] however it is becoming obvious that NRPSs do not possess constant and rigid supermodular architecture.[85]

1.3   Approaches for modifying non-ribosomal peptides

Inspired by the fact that NRPSs harbour a modular architecture and a module contributes its respective substrate to the final NRP, engineering of NRPSs is in focus of several laboratories since 1995.[16,131,132] Overcoming resistance mechanisms (in case of peptides with antimicrobial properties), increasing biological activities or decreasing side effects are some incentives to modify NRPS assembly-lines[133–135] as nature is thought to do similarly during the evolutionary process.[136] To achieve this, different approaches have been published including modification of the precursors, engineering of the gatekeeping function of A domains or substitution of whole (di)domains or modules (Fig. 7).[137]
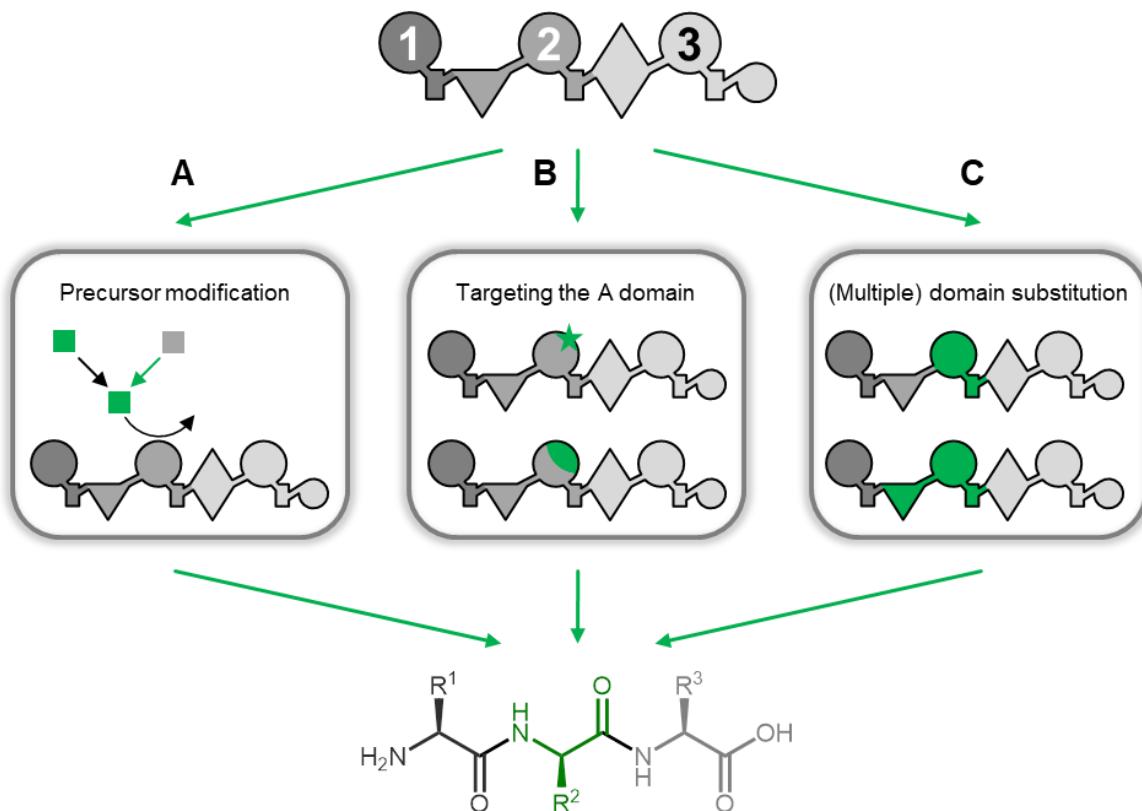
Figure 7. Approaches for modifying NRPs. A schematic trimodular NRPS (top; grey) and the production of a modified NRP (bottom) by three different engineering strategies (highlighted in green). **A.** Precursor modification covers providing modified substrates to the NRPS which can be either directly added or provided by manipulated biosynthesis pathways. **B.** Targeting the A domain include modification of the binding pocket by mutations or exchange. **C.** (Multiple) domain substitution base on exchange of whole (di)domains or modules. For domain assignment see Figs. 2 and 5; further symbols: diamond, dual C/E domain. Based on [137].

1.3.1 Precursor modification

Although the mutasynthesis is not engineering of the enzyme itself, it provides access to more diversity of NRPs. The probably simplest approach is to feed a modified precursor to the strain expressing the NRPS of interest which is subsequently activated and processed by the enzyme. However, this approach relies on the requirement of the gatekeeping domains to accept the non-natural substrate. Another drawback is that the added precursors compete with the natural substrates. When adding e.g. 3-fluoro-L-Tyr to iturin-producing *B. subtilis*, the incorporation of the non-natural halogenated substrate beside the natural substrate L-Tyr has been observed.[138] This can be solved by reducing the naturally occurring substrates in a pathway engineering approach. In the lipopeptide calcium-dependent antibiotic (CDA) biosynthetic gene cluster (BGC) in *S. coelicolor*, a deletion of *hmaS* abolished the production of CDA since *hmaS* is involved in the synthesis of the A6 domain's substrate 4-hydroxymandelic acid.[139] Exogenous supply of 4-hydroxymandelic acid or derivatives thereof, restored the CDA biosynthesis in the Δ*hmaS* mutant and led to the production of non-natural lipopeptides.

Instead of the exogenous supply of non-natural substrates to the production strain, the endogenous biosynthesis of modified precursors is also possible. E.g. the biosynthesis of the antibiotic pacidamycin in *S. coeruleorubidus* was directed to the major production of chlorinated pacidamycin by integration of a halogenase into the host.[140] Here, the chlorinated Trp residue was furthermore applicable for synthetic diversification with phenyl boronic acid derivatives. Modification of existing enzymes involved in the biosynthesis has been reported e.g. by Micklefield and co-workes.[141] They were able to shorten the fatty acid side chain of CDA by site directed mutagenesis of an active site residue in β-ketoacyl ACP synthase of the fatty acid biosynthesis operon.

1.3.2 Targeting the adenylation domain

Identification of the specificity-conferring code by Stachelhaus *et al.* gave rise to NRPS engineering approaches targeting the A domain.[64] As shown by the same group, a targeted mutation within the binding pocket led to a shift of the specificity of Glu- or Asp-activating A domains from SrfA to Gln and Asn, respectively.[142] For the alteration of Glu to Gln, only one mutation was necessary. This position was also targeted in a

(methyl)-Glu-activating module of CDA NRPS and showed a significant effect on incorporation of (methyl)-Gln although wild type (WT) CDA was still produced in minor amounts.[143]

Another single mutation enabled the activation of the non-natural AAs *O*-propargyl-L-Tyr and *p*-azido-L-Phe in GrsA.[144] Such AAs are "clickable" and can undergo bioorthogonal click reactions for further labelling or enrichment.[145,146] The AAs were incorporated *in vitro* and *in vivo* up to 10^5 fold more efficient than in the WT even upon L-Phe competition. The engineered A domain was furthermore able to interfere with the downstream GrsB1 module for diketopiperazine production.[144] Latter finding highlighted that one has to consider downstream biochemical reactions for engineered NRPSs as single mutations in A domain has also led to insufficient thioester formation.[147] Notably, the experiment was also transferred to tyrocidin-producing NRPS (Tyc) from *B. brevis* although with lower catalytic efficiency.[144]

In a directed evolution approach, the L-Phe-activating A domain of TycA was successively suspended to saturation mutagenesis to finally activate L-Ala but with low catalytic efficiency.[148] In 2018, Hilvert and co-workers presented a high-throughput assay for testing the adenylation as well as thioesterification reaction of just mentioned A domain.[149] They combined rational shortening of structural elements in the A domain to prefer β-AAs, fluorescence-activated cell sorting (FACS) via a yeast surface display for library screening and biorthogonal click chemistry to label active constructs. Finally, the engineered TycA accepts and processes (*S*)-β-Phe with a 220-fold preference over the native substrate, exhibits a very high fold switch in α/β-AA specificity and leads to modified tyrocidine with production titre of 120 mg/L in *E. coli*.

A subdomain swap within A domains takes advantage of maintaining the native environment for altered substrate binding.[132] Inspired by a bioinformatic view on enzyme evolution,[150] Crüsemann *et al.* focused on exchanging parts of A domains which emerged by evolutional recombination.[151] When exchanging this subdomain in an A domain of the hormaomycin-producing NRPS (Hrm) from *S. griseoflavus* against three other Hrm A subdomains, they observed the same AA activation by the engineered A domains as in their respective parental A domain *in vitro*. However, subdomains from CDA NRPS did not yield active engineered Hrm enzymes.[151] The same approach was performed by

Kries *et al.* in Phe-activating A domain from GrsA but they were guided by a structural definition of subdomains which resulted in slightly different segments.[152] Four out of nine tested subdomains showed significant adenylation activity with two being from different species than *B. subtilis*. Additionally, engineered Val-specific GrsA fused to the following GrsB1 module was shown to produce modified diketopiperazines *in vitro*, although 300-fold slower than WT GrsA-GrsB1.

1.3.3 (Multiple) domain substitutions

If one relates the engineered part to the whole NRPS, substitution of domains up to whole modules depicts the biggest variances of the mentioned approaches (Fig. 7). Pioneers were Marahiel and co-workers with their A-T substitutions in SrfA. By exchanging the Leu-specific didomain against five Phe-, Orn-, Leu, Cys- and Val-specific didomains from *B. brevis* and *P. chrysogenum*, they reported the production of all peptides including four non-natural surfactin derivatives *in vivo*.[131] When subjecting another Leu-specific didomain of SrfA to these substitutions, only the Orn-containing lipopeptide was detected.[153] However, product yield was lowered compared to the WT and also undesired byproducts have been observed due to domain interactions or the subsequently discovered C domain specificity.[88,89]

Different strategies of domain assembling have been performed by Duerfahrt *et al.* in a hybrid bimodular NRPS from SrfA and Tyc origin *in vitro*.[154] The six tested constructs comprised reassembly points between the T-C, C-A, A-T or T-TE domains and all catalysed the formation of Asp-Phe. They concluded that all strategies can be used for production of novel peptides; however, differences in enzyme activity suggested a rearrangement at the T-C linker and a preservation of the C-A interface. The influence of the C-A didomain as well as C domain specificity was e.g. investigated in the pyoverdine-producing NRPS (Pvd) from *P. aeruginosa*.[155] Substitution of the Thr-specific A domain led to WT pyoverdine *in vivo* although non-Thr-specific A domains from different *Pseudomonas* strains were used. Calcott *et al.* concluded that the adjacent C domain exhibits stronger acceptor substrate selectivity than the introduced A domain exhibits native substrate selectivity. In contrast, substitutions including this C domain led to the production of Lys- or Ser-containing pyoverdine but also several truncated peptides.[155] Expanding the exchanged didomain to T-C-A did furthermore not lead to a

higher yield.[156] However, the team around Ackerley reviewed in a current study their domain and module substitutions. Focusing on partially shuffling of regions within the C domain in order to narrow down the C domain's specificity, they identified that substitution of the A domain with the preceding C-A linker allowed incorporation of the non-native AA compared to substitutions of different C domain regions or the A domain alone.[157] Based on bioinformatic data suggesting that A domains may have evolved separately from C domains and the region harbouring their recombination site has low potential for structural disturbance substitution, they concluded that the C domain specificity is not mediated by the C domain but by the C-A linker. When applying their findings to other engineered PvdD constructs, they observed an increased yield compared to their previous study[155] as well as new derivatives. Furthermore, the authors transferred the C-A linker and following A domain substitution in the system of Belshaw *et al.* which was fundamental for introducing the hypothesis of C domain specificity and enabled production of diketopiperazines.[89]

In an prominent study, the NRPS of cyclic lipopeptide antibiotic daptomycin (Dpt) from *S. roseosporus* was subjected to multiple module substitutions *in vivo*.[158] First, the C-A-T modules at position 8 and 11 of Dpt were exchanged among each other, leading to a Ser or Ala-containing daptomycin at the respective position. A heterologous exchange of these positions against an Asn-specific module from a related NRPS (A54145 from *S. fradiae*)[159] with as well as without downstream E domain also resulted in the production of the expected peptides. Furthermore, the modules 8 to 11 (D-Ala-L-Asp-L-Gly-D-Ser) from Dpt were successfully replaced by the modules 8 to 11 (D-Lys-*O*-methyl-L-Asp-Gly-D-Asn) from A54145. Due to missing tailoring enzymes in *S. roseosporus*, non-methylated L-Asp was incorporated. Finally, these module substitutions were combined with exchange of the last Dpt subunit containing two modules. In total, Nguyen *et al.* generated a library of novel lipopeptides with partly WT activity against *S. aureus* but also noted a loss of production titre accompanied with an increasing number of module substitutions.[158]

All above mentioned approaches were performed in bacterial systems. Module exchanges have also been tested in iterative fungal NRPS systems to change the depsipeptide structures of enniatin and beauvericin.[160] In an additional work on these systems, Süssmuth and co-workers altered the ring size by terminal domain exchanges and introduced multiple

module substitutions without loss in production titre.[161] Furthermore, some hybrid compounds showed increased antiparasitic activity or were also be modified in their backbone methylation by deletion of M domains. Engineering of NRPSs on these modifying domains is even proven to convert a naturally uninterrupted A domain into a bifunctional A domain with retaining the adenylation and methylation function as their origins.[162]

1.4   Overview and aim of the thesis

The scope of this PhD thesis is the engineering of NRPSs in order to produce non-natural natural products. This will be achieved with the elaboration of two engineering approaches that provide guidelines for the modification of these enzymes and their subsequent utilization as well as of terminal R domains for reductive release of the peptides. Based on the findings of R domain-containing NRPSs, further *in silico* insights for the characterisation of those megasynthetases will be gained.

The current chapter 1 introduces NRPs, their biosynthesis as well as further details of the underlying biosynthetic machinery and how NRPSs can be engineered.

In chapter 2, all contributed publications are listed. These include "*De novo* design and engineering of non-ribosomal peptide synthetases" and "Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains" introducing two novel NRPS engineering approaches, as well as "Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains" focusing on aldehyde-generating R domains within identified wild type and engineered NRPSs.

Chapter 3 comprises additional results that were achieved within this research project but are not part of the listed publications. First, the use of R domains could be applied to other engineered NRPS systems in order to generate peptide aldehydes. Second, the study on an identified minimal NRPS is expanded to more *Xenorhabdus* strains.

The discussion in chapter 4 will compare both presented NRPS engineering approaches and also review current literature referring to both concepts. Additionally, the functional role of aldehyde-containing peptides will be demonstrated as well as a revised NRP

biosynthesis by a constructed recombinant NRPS system. Regarding the minimal NRPS from *Xenorhabdus*, insights into the non-canonical active site residues will be given and a classification into three subtypes will be postulated.

The PhD thesis finishes with the quotation of references, an attachment including all publications and supporting information, a curriculum vitae of the doctoral candidate, the list of publications and record of conferences as well as the declaration.

2 Publications

2.1 *De novo* design and engineering of non-ribosomal peptide synthetases

Kenan A. J. Bozhüyük[1], Florian Fleischhacker[1], Annabell Linck[1], Frank Wesche[1], Andreas Tietze[1], Claus-Peter Niesert[2] and Helge B. Bode[1,3] *

[1]   Merck Stiftungsprofessur für Molekulare Biotechnologie, Fachbereich Biowissenschaften, Goethe Universität Frankfurt, Max-von-Laue-Strasse 9, 60438 Frankfurt am Main, Germany.

[2]   Performance Materials/Process Technologies, Merck KGaA, Frankfurter Strasse 250, 64293 Darmstadt, Germany.

[3]   Buchmann Institute for Molecular Life Sciences (BMLS), Goethe Universität Frankfurt, Max-von-Laue-Strasse 15, 60438 Frankfurt am Main, Germany.

*   e-mail: h.bode@bio.uni-frankfurt.de

Attachments: Declaration on the contribution of the authors and the publication including supplementary information.

2.2 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

Attachments: Declaration on the contribution of the authors and the publication including supplementary information.

2.3 Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains

Andreas Tietze[1], Yan-Ni Shi[1], Max Kronenwerth[1], Helge B. Bode[1,2,3*]

[1] Fachbereich Biowissenschaften, Molekulare Biotechnologie, Goethe-Universität Frankfurt, Frankfurt am Main 60438, Germany.

[2] Buchmann Institute for Molecular Life Sciences (BMLS), Goethe-Universität Frankfurt, Frankfurt am Main 60438, Germany.

[3] Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt

* Corresponding author: h.bode@bio.uni-frankfurt.de

3 Additional results

3.1 R domains in engineered NRPSs

In the publication "Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains" (chapter 2.3) it has been shown that the R domain of tilivalline-producing synthetase (Xtv) from *X. eapokensis* DL20[27] processes different substrates and releases the respective peptides as aldehyde intermediates.[165]

3.1.1 Substitution of termination domain

The XtvB_R was tested to replace the TE domain in pyrrolizixenamide A-producing NRPS (PxaA) from *X. stockiae* DSM 17904. The bimodular PxaA is involved in the production of bacterial pyrrolizidine alkaloids and a monooxygenase (PxaB) processes the PxaA-derived intermediates **1a**, **1b** and **1c** (Fig. 8) via Baeyer-Villiger oxidation, hydrolysis and decarboxylation to pyrrolizixenamide A.[166] An exchange of the PxaA_TE against XtvB_R *C*-terminal of α-helix 4 of PxaA_T2 (NRPS-1) led to the production of **2a**, **2b** and **2c** after heterologous expression in *E. coli* DH10::*mtaA* as detected by high resolution (HR)-high performance liquid chromatography coupled mass spectrometry (HPLC-MS) analysis (Fig. 8A; Supplementary Fig. 1.1). This suggests the successful replacement of the termination domains.

Stable isotope labeling in combination with high-resolution MS, isolation and subsequent nuclear magnetic resonance (NMR) spectroscopy of **2a** (NMR measured and analysed by Yi-Ming Shi, Goethe-university Frankfurt, Supplementary Figs. 1.2 – 1.8, Supplementary Tab. 1.4) confirmed a structure which is similar to the PxaA products apart from a hydroxy instead of a ketone group and differing in fatty acid chain length among the derivatives (Fig. 8B). Absolute quantification revealed a production titre of 8.9 ± 2.3 mg/L for **2a**, 2.1 ± 0.8 mg/L for **2b** and 4.2 ± 0.8 mg/L for **2c**.
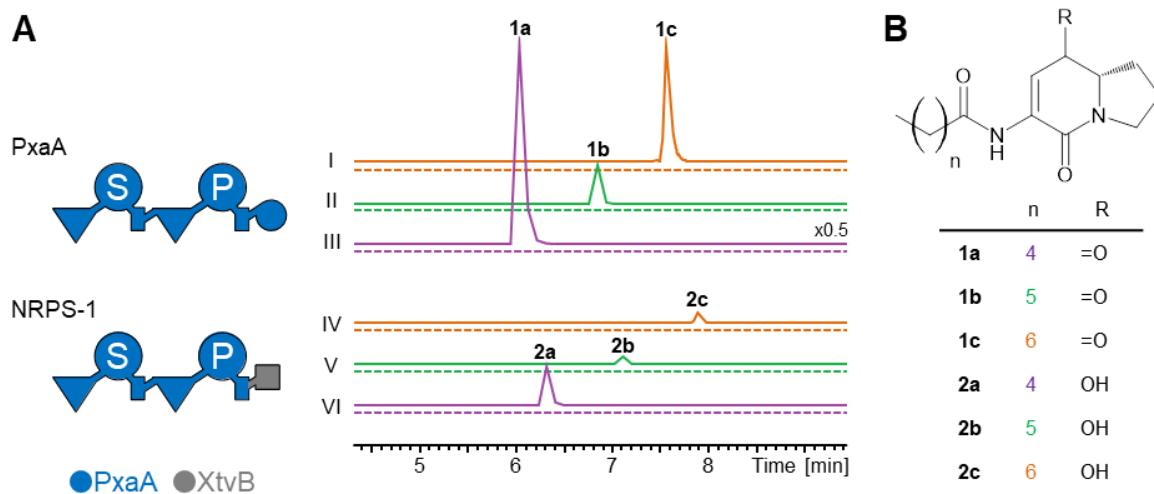
Figure 8. Exchange of the TE domain in PxaA against the XtvB_R domain. A. Production of **1a**, **1b** and **1c** by PxaA[166] as well as **2a**, **2b** and **2c** produced by NRPS-1 in *E. coli* DH10B::*mtaA* (induced: continuous line; non-induced: dashed line). EICs of (I) **1c** *(m/z* [M+H]$^+$ = 293.18), (II) **1b** *(m/z* [M+H]$^+$ = 279.17), (III) **1a** *(m/z* [M+H]$^+$ = 265.15; y-axis decreased by factor 0.5), (IV) **2a** *(m/z* [M+H]$^+$ = 267.17), (V) **2b** *(m/z* [M+H]$^+$ = 281.18) and (VI) **2c** *(m/z* [M+H]$^+$ = 295.20). The colours of the chromatograms are according to the length n of the fatty acid side chain. For domain assignment see Figs. 2 and 5. **B.** Structure of **1a**, **1b**, **1c**, **2a**, **2b** and **2c**.

3.1.2 *In vivo* production of a peptide aldehyde

In combination with the established exchange unit (XU) NRPS engineering approach,[163] an aldehyde-releasing NRPS was generated that should produce a linear lipopeptide with a *C*-terminal aldehyde. For this, the terminal XU including GxpS_A2T2 (*P. laumondii*) and XtvB_R (*X. eapokensis*)[165] was fused with the butyric acid-L-Pro activating XU1 of xenoamicin-producing synthetase (XabS; *X. doucetiae*)[25] and L-Ala activating XU2 of kolossin-producing synthetase (KolS; *P. laumondii*)[167] to generate NRPS-2. Upon heterologous expression in *E. coli* DH10B::*mtaA,* the expected lipopeptide **3a** was detected with software-based HR-HPLC-MS analysis (Figure 9A, Supplementary Fig. 1.9) and verified with stable isotope labelling and a synthetic standard (Supplementary Figs. 1.9 and 1.10).

To prove the existence of a free aldehyde in **3a**, the expression of NRPS-2 was performed in presence of the aldehyde capture reagent *O*-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine (PFBHA). The derivatization product **3b** was detected by HR-HPLC-MS verifying the peptide aldehyde (Figure 9B).
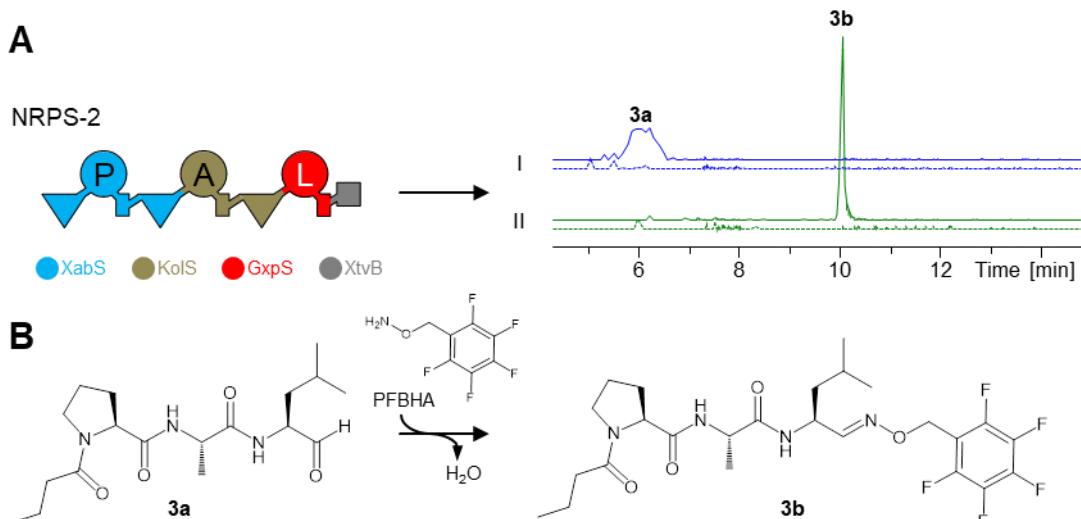
Figure 9. *In vivo* **production and detection of a peptide aldehyde. A.** Production of **3a** by NRPS-2 in *E. coli* DH10B::*mtaA* (induced: continuous line; non-induced: dashed line) (I) without and (II) with PFBHA. EICs of **3a** (*m/z* [M+H]$^+$ = 354.24) and **3b** (*m/z* [M+H]$^+$ = 549.25). For domain assignment see Figs. 2 and 5. **B.** Structure of **3a** and derivatization with PFBHA to yield **3b**.

3.2   Investigation of ATReds in *Xenorhabdus*

In the publication "Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains" (chapter 2.3), the production of the pyrazine **4a** in *X. indica* DSM 17382 has been assigned to the gene *xind01729* which encodes a minimal NRPS consisting of an A, T and R domain and is therefore called ATRed$_{xind01729}$.[165] AntiSMASH analysis[168] of *Xenorhabdus* WT genomes revealed additional 35 putative ATRed coding sequences (*atred*) in 20 strains in total (Tab. 1).

Table 1. Overview of ATRed encoding gene in *Xenorhabdus*. Organism and its abbreviation as well as ATRed encoding gene. For further details, please refer to Supplementary Tab. 2.2.

organism	gene	organism	gene
X. khoisanae DSM 25463	03561	*X. cabanillasii* JM26	01329
	03948		03628
Xenorhabdus sp. KK7.4	01108	*X. indica* DSM 17382	00627
	02190		01729
X. bovienii SS-2004	00464	*Xenorhabdus* sp. PB62.4	01459
X. cabanillasii DSM 17905	01493	*X. innexi* DSM 16336	00707
	03579		02671
X. poinarii DSM 4768	02758		02976
X. nematophila ATCC 19061	00646	*X. szentirmaii* DSM 16338	01262
	01475		03484
	01561	*X. mauleonii* DSM 17908	04014
X. miraniensis DSM 17902	01976		04297
X. vietnamensis DSM 22392	00828	*X. szentirmaii* US	00630
	03245		03375
Xenorhabdus sp. KJ12.1	01708	*X. kozodoi* DSM 17907	00716
	02365	*X. budapestensis* DSM 16342	02951
X. stockiae DSM 17904	02049		03352
	03518	*X. hominickii* DSM 17903	01101

3.2.1 Active site residues of R domains from ATReds

Clustal Omega alignment of the R domains of the ATReds with the structurally characterised R domains of myxalamid-producing synthetase (MxaA) from *S. aurantiaca*[116] as well as probable peptide synthetase (Nrp) from *M. tuberculosis*[118] revealed that the Lys of the catalytic triad of the SDR superfamily (Thr214-Tyr249-Lys253 in MxaA_R) is substituted by a Gln in six ATReds of six different *Xenorhabdus* strains including ATRed$_{xind01729}$ (ATReds of *xind01729*, *xcabDSM03579*, *xnem01561*, *xvie03245*, *xcabJM01329* and *xbud02951*; Supplementary Fig. 2.1). With exception of *X. budapestensis*, five of these six *Xenorhabdus* WT strains are capable of the biosynthesis of **4a** as detected by HR-HPLC-MS analysis (Fig. 10A, Supplementary Fig. 2.2). In addition, *X. innexi* produced a pyrazine with Ile (**4b**) instead of Phe-residues, although it

does only encode ATReds with the canonical catalytic triad (Supplementary Fig. 2.2 and 2.3). All other *Xenorhabdus* WT strains with a putative *atred* did not show production of a pyrazine in the HR-HPLC-MS analysis. This suggests that the ATReds with the unusual catalytic triad Thr-Tyr-Gln might be pivotal for the biosynthesis of **4a**.
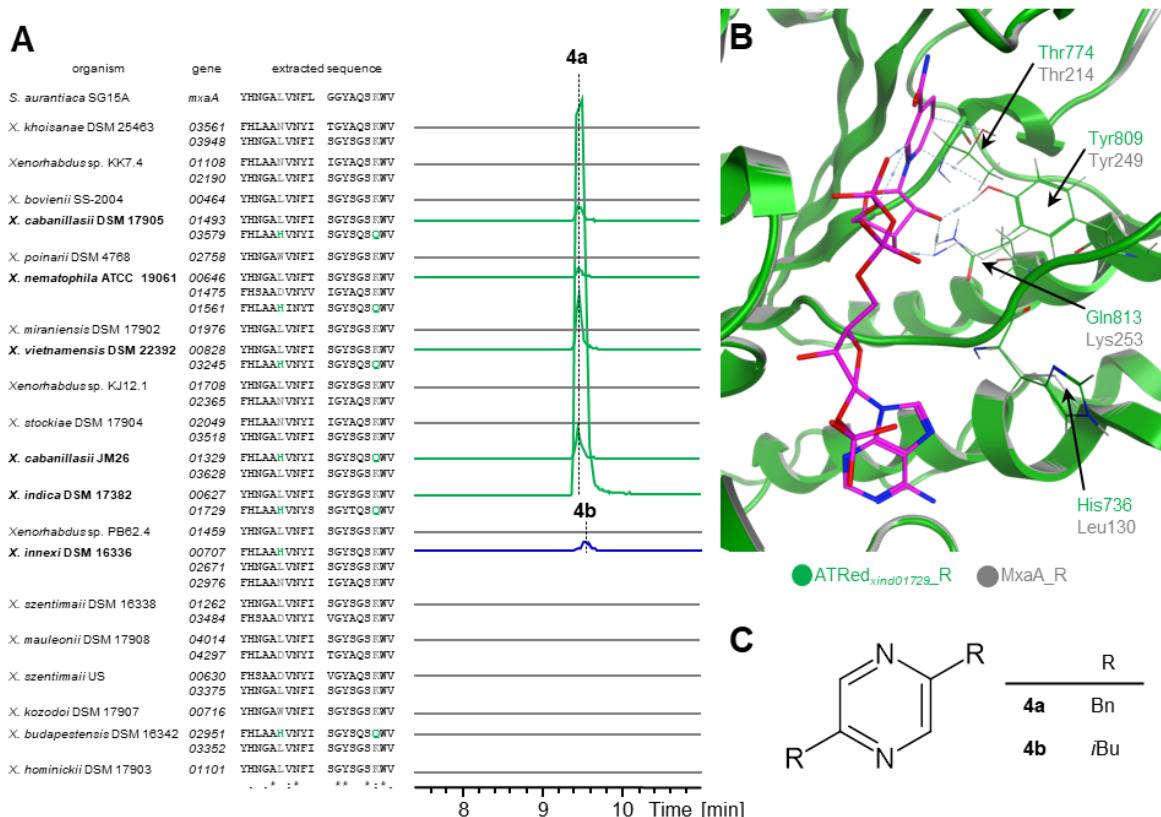
Figure 10. ATRed-dependent pyrazine production in *Xenorhabdus*. A. HR-HPLC-MS analysis (green: **4a**, EIC *m/z* $[M+H]^+$ = 261.14; blue: **4b**, EIC *m/z* $[M+H]^+$ = 193.17 and grey: no production of any pyrazines observed) of 20 *Xenorhabdus* WT strains and the extracted Clustal Omega alignment surrounding His736 and Gln813 in ATRed$_{xind01729}$ (presence of His and Gln is highlighted in green, otherwise grey) of 36 ATRed sequences as well as MxaA from *S. aurantiaca*. Strains with a detected pyrazine production are highlighted in bold. **B.** R domain active site of ribbon diagram of homology model (RMSD = 0.8 Å) of ATRed$_{xind01729}$ from *X. indica* (green) based on MxaA_R from *S. aurantiaca* (grey; PDB-ID 4U7W)[116]. Important residues and their respective nomenclature are highlighted by arrows and the NADPH cofactor is shown in pink. **C.** Structure of **4a** and **4b**. Abbreviations: Bn, benzyl; *i*Bu, *iso*-Butyl.

In order to gain further insights into the unusual catalytic triad, a homology model of the R domain from ATRed$_{xind01729}$ was calculated (Fig. 10B, Supplementary Fig. 2.4). This model bases on the crystal structure of MxaA_R from *S. aurantiaca* (PDB-ID 4U7W)[116] including the NADPH cofactor and has a root-mean-square deviation (RMSD) of 0.8 Å. Here, the catalytic triad of ATRed$_{xind01729}$ (Thr774-Tyr809-Gln813) is in close orientation to the NADPH cofactor like in the template structure. The only major difference between ATRed$_{xind01729}$ and MxaA_R within the active site was the presence of His736 instead of

Leu130 respectively. In the protein alignment of 36 ATReds, the appearance of this His-residue in the active site correlates in all cases with the Gln-residue in their catalytic triad (i.e. capable of the biosynthesis of **4a** with exception of *X. budapestensis*). His736 is absent in ATReds of *Xenorhabdus* strains that do not produce **4a** (Supplementary Fig. 2.1). In addition, **4b**-producing *X. innexi* encodes one ATRed with only the His-residue (His740 in ATRed$_{xinn00707}$) but not the Gln-residue (canonical Lys817 instead). These findings indicate that both residues, His736 and Gln813 in ATRed$_{xind01729}$, represent a different subtype of active site and might play an important role for successful biosynthesis of NRPs by ATReds in *Xenorhabdus*.

3.2.2 Clustering of ATReds in *Xenorhabdus*

Next, the AA sequences of all 36 ATReds from 20 *Xenorhabdus* strains were analysed for their similarity. Along with **4a**-producing ATRed$_{xind01729}$, all five ATReds with the identified unusual active site residues (His/Gln) of **4a**-producing *Xenorhabdus* strains (*xcabDSM03579*, *xnem01561*, *xvie03245* and *xcabJM01329*) and non-**4a**-producing *X. budapestensis* (*xbud02951*) as well as ATRed$_{xinn00707}$ from **4b**-producing *X. innexi* have a pairwise identity of at least 68 % (Fig. 11, Supplementary Fig. 2.5). This cluster of seven ATReds is distinct from all other examined ATReds. Those remaining ATReds can be divided in two clusters of 20 and 9 ATReds with high pairwise identity (at least 74 %) each. The ATReds of the three clusters share a pairwise similarity with ATReds from other clusters of below 48 % and differ over the full length protein. Notably, no ATReds of the same species containing two or three ATReds were found within the same cluster and with exception of ATRed$_{02758}$ from *X. poinarii*, exclusively ATReds of the largest cluster have been found to be associated with an MLP encoding gene (*mlp*) in the genome.
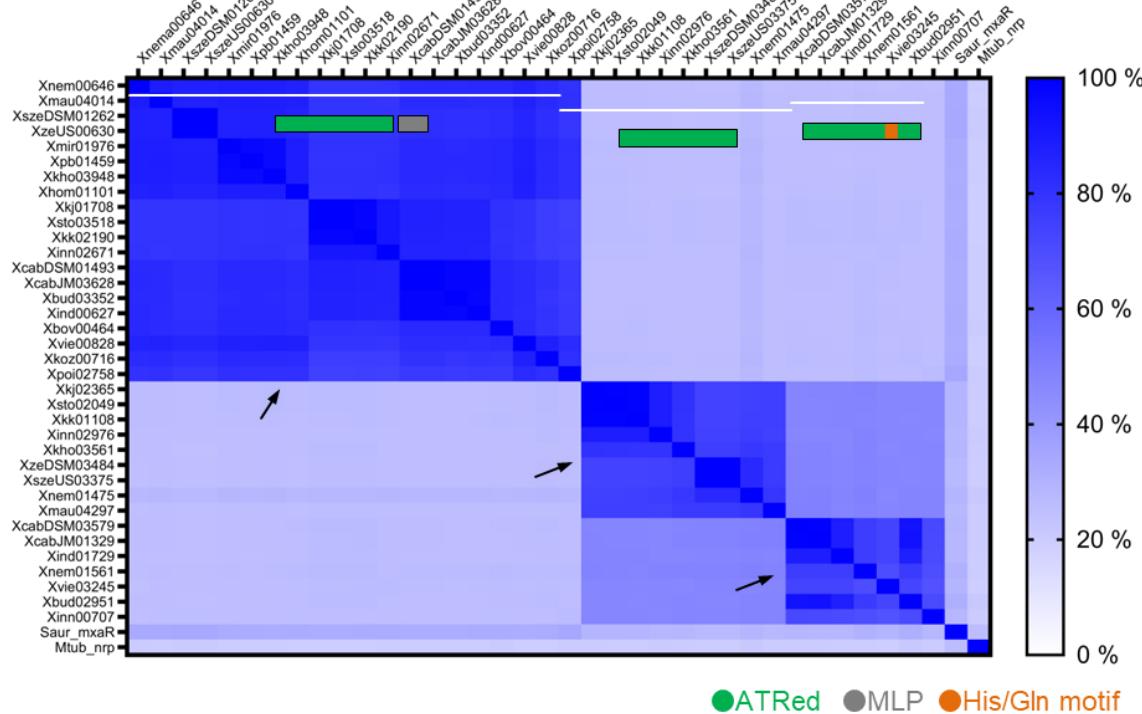
Figure 11. Clustering of ATReds in *Xenorhabdus*. Heatmap (white, 0% pairwise identity; blue, 100 % pairwise identity) of the Clustal Omega alignment of 36 ATRed AA sequences from 20 *Xenorhabdus* WT strains, MxaA_R and Nrp_R. The three clusters are indicated by arrows and the ATRed abbreviations are further specified in Supplementary Tab. 2.2. Corresponding schemes of ATReds (green) with an MLP (grey) or His/Gln motif (orange) are assigned to the single ATReds by white lines.

4 Discussion

4.1 The XU and XUC system for the engineering of NRPSs

Pioneered by the Marahiel group in 1995,[131] there has been a range of attempts on NRPS engineering using different approaches (see chapter 1.3).[16,132] However, only little success was reported on modifying the biosynthesis of these pharmaceutically relevant mega-synthetases. It resulted in a drop of yield compared to the WT and impaired enzymes and did not establish general applicable guidelines.

4.1.1 Comparison of the XU and XUC system

This study comprises the concept of exchange unit (XU)[163,169–171] and exchange unit condensation domain (XUC)[164,172] which provide simple, efficient and reproducible strategies for the production of *de novo* or modified NRPs by engineered NRPSs (Fig. 12). Both vary from the universal definition of an NRPS module

NRPS engineering using the XU concept follows three easy-to-follow rules. First, tridomains of A-T-C(/E) are used as building blocks and are called XU. Initiation XUs therefore may include a C_{start} domain whereas a termination XU ends with the termination domain (see chapter 1.2.5). Second, the WNATE motif defines the fusion point between the XUs and divides the C-A linker in a 22 AA long *N*-terminal part and a ten AA long *C*-terminal part. These two parts are structurally unrelated with the longer *N*-terminal part being involved in C-A interactions and the shorter *C*-terminal part associating only with the A domain. Third, the C domains acceptor site specificity has to be respected, i.e. the processed AA of the downstream A domain of a XU in the engineered NRPS system coincides with that of the XU of its origin. With this in hand, three WT NRPS were reconstructed, restoring a good production yield of up to 88 % compared to the WT, NRPs were modified by the AA configuration or AA substitution and even *de novo* peptides were created.[163,169–171]
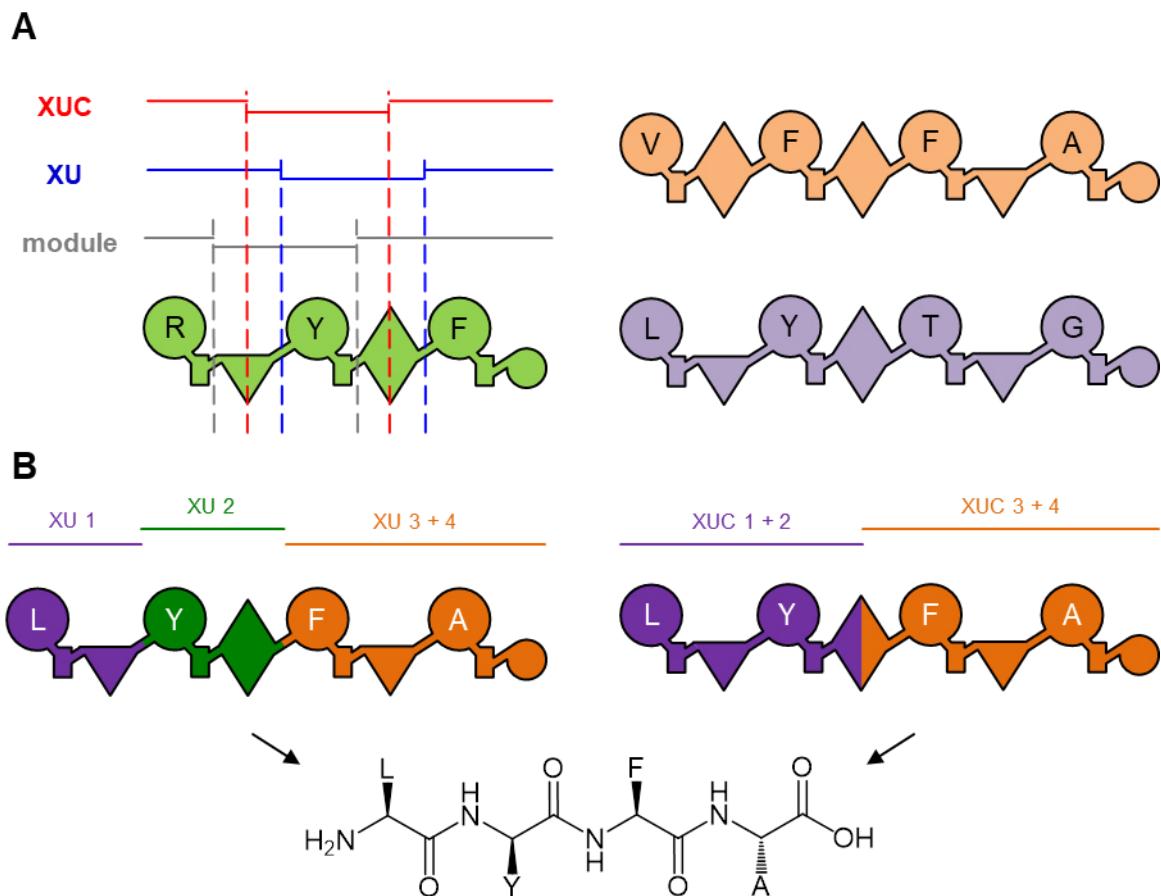
Figure 12. NRPS engineering using the XU and XUC system. A. Schematic representation of the fusion points of the XU[163] (blue) and XUC[164] (red) system compared to the canonical module definition (grey). Three hypothetical template NRPS (green, orange and purple) are shown. **B.** Production of a hypothetical linear tetrapeptide by engineered NRPSs using both strategies. For the XU-based NRPS, three building blocks from three origins have to be taken. The A domain specificity is indicated by one letter AA code within the A domains. For domain assignment see Figs. 2, 5 and 7.

The C domain specificity at the acceptor site depicts the major drawback of the XU concept.[88,89] The substitution of one building block that differs by the incorporated AA requires the exchange of at least two XU of the targeted NRPS, as exemplified in Fig. 12B. Even if this would not be limited by a sufficiently large number of building blocks, the increased number of artificial interfaces between the domains leads to a drop in peptide yields as also observed earlier.[158] This issue has been addressed by the development of the XUC concept.[164,172] Here, the catalytic units are defined as A-T didomains with flanking C_{Dsub} and C_{Asub} and the fusion point is located within the 4 AA long linker region which connects the two C subdomains.[173] Strikingly, XUCs comprise both C subdomains that interact with the 4'-PPant-bound AA activated by the same XUC, namely as acceptor substrate on the *N*-terminal C_{Asub} and as donor substrate on the *C*-terminal C_{Dsub} within one XUC. As a consequence, the C domain specificity is "integrated" within one XUC

enabling the more versatile combination possibilities of XUCs. In numbers, only one tenth of XUCs compared to XUs are necessary to provide a pool of building blocks for any peptide based on the 20 proteinogenic AAs.[164,172]

4.1.2 Placement within current literature

In the past few years, novel insights into NRPSs touching the XU and XUC approaches have been gained. Steiniger *et al.* reported the first *in vivo* combination of linear and iterative NRPSs from fungi by incorporation of C-A-M-T, A-T and C_{Asub}-A-T building blocks from cyclosporine-producing NRPS in enniatin- as well as bassianolide-producing NRPS.[174] Although, they used a fusion point for XU reassembly *C*-terminal to the C domain, their results confirmed the consideration of C domain specificity at the acceptor site as postulated by the XU strategy and also at the donor site in C-A-M-T swaps. Otherwise, no functional construct was observed. Interestingly, the same constructs of the A-M-T and C-A-M-T swaps were impeded in seven out of eight cases when a C subdomain fusion point allegedly to the XUC strategy was used. They argued that on the one hand fungal systems might respond different to bacterial NRPS engineering strategies (and *vice versa*). This would be supported by our findings, since even engineered NRPSs consisting of *Bacillus* and *Photorhabdus/Xenorhabdus* XUCs led to truncated NRPs due to non-functional interactions between the domains of different genera.[164] On the other hand, a possible disturbed integrity of the hybrid C domains with the floor loop and latch crossing from the C_{Asub} to the non-native C_{Dsub} led to non-functional NRPSs. However, in an earlier work, Steiniger *et al.* targeted the C_{term} domain of the same system.[161] The heterologous intra-domain interface by substitution of either the C_{Dsub} or C_{Asub} with related C_{term} subdomains did not hamper the biosynthesis. The authors did not mention that the hybrid C domains are derived from a $^{D}C_L$ (for D-2-hydroxyisovalerate) and a $^{L}C_L$ subtype. This issue of structural intra-domain interaction of hybrid C domains has already been reported for the incompatibility of C and C/E domains[164,172] and is likely to explain why their C subdomain exchanges did not work. This problem should be pursued once more structures of appropriate C domains are solved.

Impressive structural contributions to the understanding of NRPSs have already been achieved by the groups of Schmeing and Gulick (see chapter 1.2.6) but the issue of how C domain acceptor site specificity is mediated remained unclear.[70,79] The XUC concept

was shown to circumvent this limitation by dividing the C domain into its *N*-terminal C_{Dsub} ("left" subdomain) and *C*-terminal C_{Asub} ("right" subdomain), formally keeping the C domain's donor and acceptor subdomain with the C domain's donor and acceptor substrate within an XUC, respectively.[164,172] However, structural data of C domains interacting with their donor substrate bound to the *N*-terminal T domain (T_D) or the acceptor substrate bound to the *C*-terminal T domain (T_A) show that the interaction occurs from the "back" and "front" of the C domain and not from the "left" and "right". Both subdomains are therefore involved in domain interaction, regardless of being a hybrid or wild type C domain (Fig. 13). In the work of Reimer *et al.*, multiple T_D-C structures were captured in the condensation state for the first time and they noted slightly different residue-level contacts by van der Waals interactions within each structure.[85] The authors highlighted the importance of this interaction for the overall NRPS structure since it is the only point where neighbouring modules have to coordinate within the flexible and not constant overall architecture (see chapter 1.2.6).[85] Interestingly, six of seven C domain residues that were shown by mutational studies to influence this T_D-C interaction are located within the C_{Asub} and therefore non-native in XUC-based NRPSs. The GxpS comprising an XU (A-T-C) from the bicornutin-producing synthetase (BicA)[175] and therefore a native T_D-C_{Asub} interface showed no production whereas the same construct with a non-native T_D-C_{Asub} interface due to an A-T-C_{Dsub} exchange showed over 200 % production compared to the WT.[164] In addition, the successful generation of a peptide-library by randomization of XUCs with multiple shuffling of non-native T_D-C_{Asub} interfaces implicates that maintaining this interaction is not the only decisive factor regarding NRPS functionality. At the acceptor site of C domains, different studies revealed hydrophobic interactions and hydrogen bonds of the T_A with helices of the C_{Dsub} and predominantly C_{Asub}.[85,124,126] The interplay between the C domain's acceptor site with the following XU is thematised in one of the three rules of the XU approach or even predominantly maintained within XUCs. C domain specificity might be embedded in the C_{Asub} but the underlying structural basis was still not deduced. Here, partial swapping of C domain regions would help to understand this issue.

Discussion
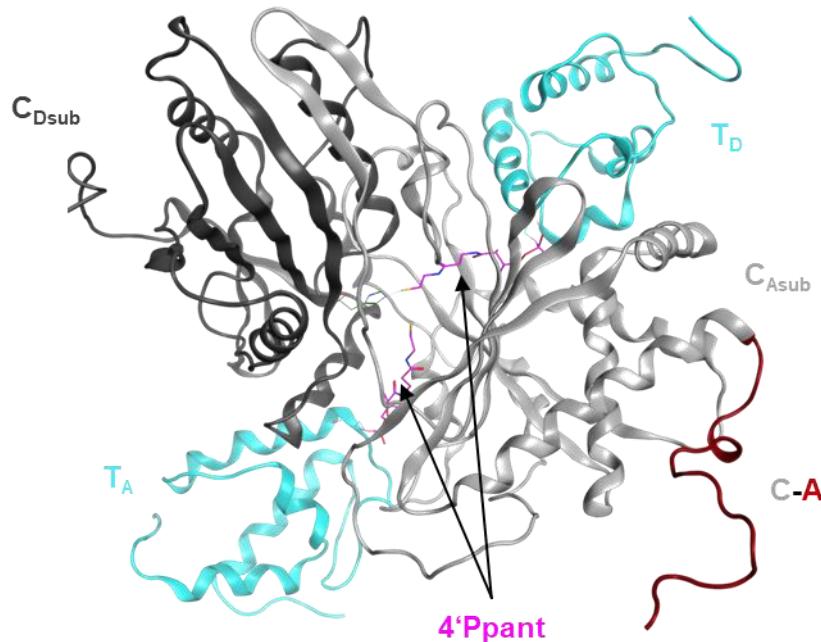
Figure 13. Model of a C domain in its condensation state with both T domain-bound 4'-PPant moieties. Superposed ribbon diagrams of the T_D domain with bound 4'-PPant moiety of LgrA in *B. parabrevis* (PDB-ID: 6MFY)[85] and the C and T_A domain with 4'-PPant moiety of the termination module of AB3403-NRPS in *A. baumannii* (PDB-ID: 4ZXI).[126] The T domains are in light blue, C_{Dsub} and C_{Asub} in dark and light grey respectively, the catalytic His-residue of the C domain in green, the 4'-PPant moieties in purple and the C-A linker[157] in red. The view is from top onto the C domain compared to Fig. 4. Based on [85,126] and processed with MOE 2016 (Chemical Computing Group).

A current study picks up this strategy when Calcott *et al.* reviewed their domain and module substitutions in PvdD from *P. aeruginosa*.[157] In short, they propose that C domain specificity is less important than discussed by the XU and XUC concept[163,164,170–172] or others[88,89,155] and specificity is embedded within the C-A linker (see chapter 1.3.3). However, their suggested recombination site *C*-terminal the C domain (Fig. 13) has already been described by Yan *et al.* in *M. xanthus*[176] and prior when engineering the GxpS with building blocks from *Photorhabdus* and *Xenorhabdus*.[163,170] Our results led to slightly reduced peptide production compared to the fusion point at the WNATE motif. Here, the GxpS could be reconstructed with XUs that do not respect to the C domain specificity and use the hybrid C-A linker[163,170,171] as well as the fusion point suggested by Calcott *et al.*[157] in order to verify their assumption.

Recently, over 39 000 module-connecting T-C linkers have been analysed with regard to their sequence and their adjacent AA substrates pairs of the A domains.[177] Farag *et al.* identified a striking relationship for more than 92 % of them suggesting that this region is specific to both adjacent A domains and specificity might be embedded within this region.

With referring to the XU approach in a retrospective analysis, they hypothesized that the reason for non-functional XU exchanges which disregard the third rule of the XU concept are incompatible intermodular interactions by the T-C linker and not due to C domain specificity. Apart from misquoting our results,[163,169,171] their assumption bases on *in silico* analysis and should be verified by experimental data. Furthermore, it is not apparent from existing structural data of modular NRPSs how the T-C linker can interact with its downstream A domain.[85] In addition, the authors emphasized that incompatibility of T-C linkers are the reason for a decreased yield compared to the wild type by engineered NRPSs even if these do consider the C domain's specificity.[163,169–171] In accordance with structural data suggesting the importance of T-C interaction,[85] the striking relationship of this linker region suggests that a rearrangement of naturally optimised interactions between modules and domains, is likely to explain decreasing production titre upon increasing number of substitutions i.e. number of non-native interfaces.[132,158,163,169–171]

4.2   R domains for peptide release

Overall structural elements of NRPs, like being cyclic, linear and *C*-terminal variations are mediated by terminal domains for peptide release (see chapter 1.2.4). Aldehyde-generating R domains were consequently addressed in terms of NRPS engineering. For TE domains, it has been shown before that internal T domains of an elongation module are unable to interact with the termination domain unless they were subjected to directed evolution.[178] Our results indicate that R domains can be introduced within an elongation module for premature release of peptides.[165] However, only few constructs were capable of the biosynthesis of pyrazines indicating that domain interaction might limit the general applicability.

Based on structural modelling experiments, TE and R domains have been shown to share a high conservation of the overall architecture as well as the location of active site-residues.[118] This finding by Chhabra *et al.* raised the question whether a 4'-PPant moiety of upstream T domain is able to interact with both terminal domains regardless of being a TE or R domain. An exchange of the PxaA_TE against XtvB_R resulted in the production of a 5,6-bicyclic compound verifying that the R domain can also replace the TE domain in an NRPS assembly-line *in vivo*. Here, the same fusion point was used that already led to functional NRPSs in our earlier experiments.[163,165,171] As one could expect,

the released 5,6-bicyclic compounds by the TE and R domain-terminated NRPS slightly differ i.e. by a ketone or a hydroxy group (Fig. 8). In summary, these results show that modification of NRPs can also be achieved by addressing the termination domain.

4.2.1 Deciphering the mechanism of PxaA

The occurrence of a ketone or a hydroxy group at the carbon atom 4 in the products of PxaA WT and NRPS-1 (Supplementary Fig. 1.3) can be explained by the mechanism of peptide release in PxaA as proposed by Schimming et al.[166] and the different substrates of TE and R domains (see chapter 1.2.5). In PxaA, dehydroalanine (DHA) cyclizes with the TE domain-bound ester to form tautomerized **1a** resulting in a ketone group (Fig. 14).[166] This is mediated by nucleophilic attack of the alkene of DHA onto the thioester. In contrast to this, NRPS-1 produces an aldehyde which results in a hydroxy group upon cyclization i.e. nucleophilic attack of the alkene of DHA onto the aldehyde. It is worth mentioning that in this nucleophilic addition an alkene acts as nucleophile for carbon-carbon bond formation.

Such biosynthesis requires the dehydration of serine to DHA in order to provide the nucleophile. This was proposed to be conducted by an unusual TE domain[166] as also observed in brabantamide A-producing synthetase (BraB) in *Pseudomonas* strain SH-C52.[179] However, the shown result implicates that this reaction is catalysed by another domain since NRPS-1 does not contain the TE domain. Pairwise identities of a Clustal Omega alignment revealed that the C2 domain of PxaA shares low similarities to other C domains of the $^{L}C_{L}$, $^{D}C_{L}$, dual C/E and Cy subtype (Supplementary Fig. 1.11 and Supplementary Tab. 1.5) and therefore might likely take an unusual role in the biosynthesis. Gaudelli et al. described the dehydration of serine to DHA in the biosynthesis of nocardicin A (Noc).[180] This is proposed to be mediated by the C5 domain of NocB. Here, an additional third His790-residue directly upstream of the catalytic His-motif acts as catalytic acid/base and abstracts the hydrogen of the α-carbon of T4 domain-bound donor serine. DHA formation is achieved by β-elimination of hydroxide from the seryl residue and is followed by nucleophilic attack of the α-amino group of T5 domain-bound acceptor substrate *p*-hydroxylphenylglycine onto the β-carbon of DHA. Finally, transfer from the T4 to T5 domain and formation of a β-lactam ring is achieved by an additional nucleophilic attack of the α-amino group of *p*-hydroxylphenylglycine onto the thioester.

Compared to PxaA, the role of DHA during biosynthesis differs. In NocB, DHA acts as electrophile for the α-amino group of the acceptor substrate, whereas in PxaA it serves as nucleophile onto the thioester of the acceptor substrate.
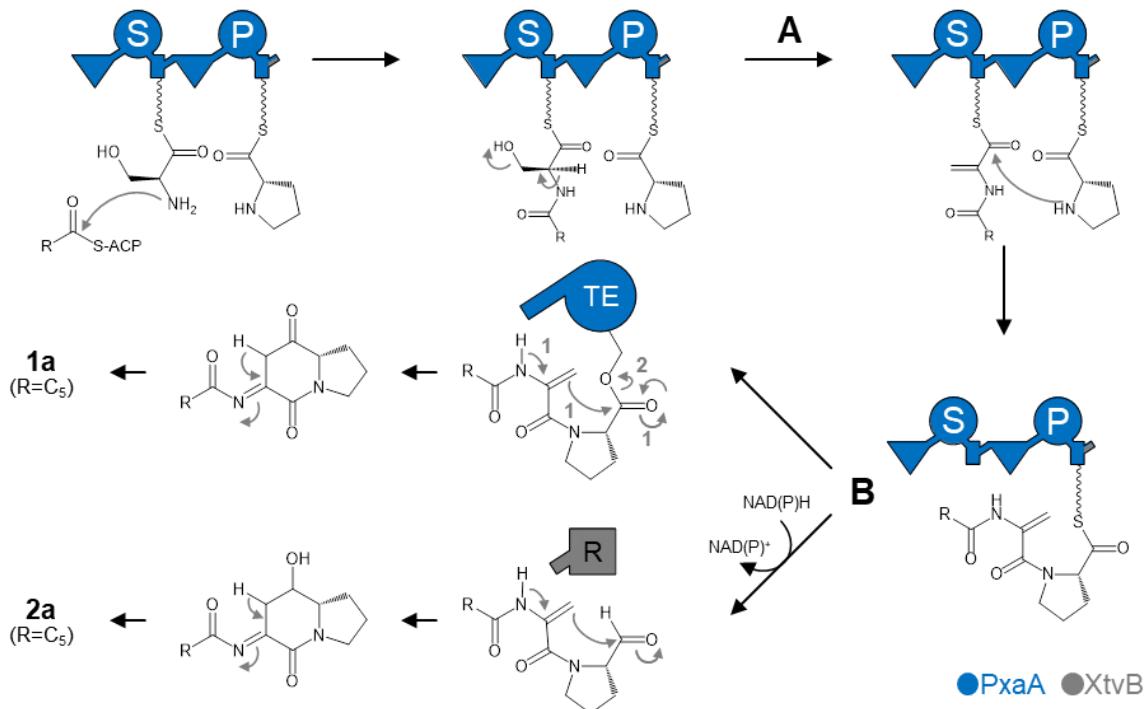
Figure 14. Biosynthesis of 1a and 2a by PxaA and NRPS-1 respectively.[166] Standard NRPS biochemistry attaches the ACP-bound fatty acid to serine on the T1 domain of PxaA. **A.** His1343 of the C2 domain likely catalyses the dehydration of serine to DHA. **B.** For the TE domain-based pathway in PxaA, the peptide is transferred to the termination domain and the side chain of DHA acts as a nucleophile to (1) mediate cyclization and (2) release from the NRPS creating a ketone group which arises from the carbonyl group of proline. This is followed by imine-enamine tautomerism to yield **1a**.[166] For the R domain-based pathway in NRPS-1, the peptide is reductively released by consumption of NAD(P)H and the side chain of DHA acts as a nucleophile to mediate cyclization creating a hydroxy group which arises from the aldehyde group of proline. Subsequent imine-enamine tautomerism yields **2a**. For domain assignment see Figs. 2 and 5.

However, PxaA_C2 does not harbour a third His prior to the canonical His-motif of C domains (Supplementary Fig. 1.12A). In the sequence alignment, PxaA_C2 alternatively exhibits an additional His-residue which does not occur in other C domains (Supplementary Fig. 1.12B). Notably, the same holds true for BraB_C2. In a homology model with LgrA_C2 (PDB-ID: 6MFY),[85] His1343 of PxaA is located within the floor loop at the substrate tunnel and therefore in a possible catalytic competent position (Supplementary Fig. 1.12C). Although the RMSD value is not significant due to the low sequence similarity to other C domains of PxaA_C2, this suggests that His1343 might act as the catalytical base for DHA formation in PxaA (His1379 in BraB, respectively). This has to be further investigated by mutational studies. It should be mentioned that the

moment of DHA formation in PxaA therefore rather occurs on the T1 domain-bound aminoacyl thioester before condensation with the second substrate,[180] than on the T2 domain-bound dipeptidyl thioester before transfer on the TE domain as proposed by Schimming *et al.*[166]

4.2.2 Peptide aldehydes as proteasome inhibitors

Natural products containing *C*-terminal warheads like β-lactone in belactosin A,[181] epoxyketone in epoxomicin[182] or aldehyde in fellutamide B[122] (see chapter 1.2.5) have been attributed with proteasome inhibition properties.[183] By reversible binding of the α-hydroxy group of the proteasome's active site residue Thr1 in one β-subunit, the peptide aldehyde fellutamide B exhibits an IC_{50} value in a nanomolar range against *M. tuberculosis*,[184] which triggers one of the world's deadliest infection diseases.[4]

Our results have shown the functional integration of an aldehyde-producing R domain within engineered NRPSs. Hitherto, freestanding aldehydes were never preserved within in the final molecule due to intramolecular reaction with nucleophiles (Fig. 8).[165] Upon XU-based exchanges covering all nucleophiles within the peptide, the free aldehyde was detected (Fig. 9). Since the production of the heterologous expression has not been optimized, chemical synthesis might allow structural data upon crystallization together with the proteasome. This potential proteasome inhibitory bioactivity has to be pursued in ongoing experiments. Based on structural data, the peptide could then be further modified since the P3 and P1 position as well as the aliphatic tail are known for fellutamide B to mediate preference to the active site.[185] As already applied in earlier studies, this can improve selectivity towards different proteasomal β-subunits in different organisms and therefore the activity to act as a species or tissue specific proteasome inhibitor.[186,187]

4.3   The ATRed subtype of minimal NRPSs

36 putative *atred*s have been identified in 20 different *Xenorhabdus* strains proposing the ATReds as a widespread representative of minimal NRPSs among *Xenorhabdus*. For *X. indica*, the predominant production of **4a** has been assigned to the ATRed$_{xind01729}$ which was proven by promoter exchange and subsequent HPLC-MS analysis.[165] The biosynthesis of pyrazines was furthermore observed in five other *Xenorhabdus* strains (Fig. 10).

4.3.1 Investigation of the active site

A Clustal Omega alignment revealed the replacement of the active site Lys against a Gln in ATRed$_{xind01729}$ at position 813 (all AA positions mentioned in this chapter are referred to ATRed$_{xind01729}$ unless stated otherwise). Reflecting this finding in the wider context of ATReds of all examined *Xenorhabdus* strains, exclusively those encoding an ATRed with Lys813 are capable of the biosynthesis of **4a**. In addition, these ATReds do also contain a His736 which is located at the active site and - with one exception – is not present in ATReds where no pyrazine production was observed by the respective WT strain. Importantly, no other residues which significantly deviate from the active of SDR enzyme MxaA_R, were found. Due to the fact that pyrazine production in strains with ATReds without His736 and Gln813 (i.e. the canonical catalytic triad) was not observed, one has to further evaluate the impact of these two residues. Here, the single and double point mutation His736Leu and Gln813Lys in ATRed$_{xind01729}$ and *vice versa* in a Leu736 and Lys813 containing ATRed will gain further insights.

In the canonical active site of R domains of the SDR superfamily, Tyr809 acts as the catalytic base (2e⁻ reduction) and is also required for high-affinity binding of NAD(P)H.[188] The substrate is stabilized by Thr774. Lys813 would form hydrogen bonds with NAD(P)H and lower the Tyr's pK$_a$ value to promote the proton relay system.[189] This is supported by a water molecule and a carbonyl backbone of a conserved Asn750 which also stabilizes the position of Lys813 and is therefore part of an extended catalytic tetrad.[190] Here, it is suggested that the substitution Lys813Gln still maintains the hydrogen bonding ability to the hydroxy groups of the ribose moiety of NAD(P)H (Fig. 15).[191] The Leu736His substitution is additionally involved in the proton relay system. Such role has already been shown in the ketoreductase (KR) 7 – another member of the SDR superfamily[192] - of the

simocyclinone D8-producing PKS (SimC7) from *S. antibioticus*.[193] Here, His interacts with the 2`-hydroxy group of the ribose moiety of NAD(P)H and a carbonyl backbone. In the structure of SimC7, the α-carbon of His is located 6.9 Å away from this hydroxy group[193] while the distance in the homology model of ATRed$_{xind01729}$_R between the α-carbon of His736 and the 3`-hydroxy group was also measured with 7.1 Å. This indicates that His736 is in sufficient proximity to the cofactor. Although the homology model positions the side chain of His736 not towards the cofactor, His736 is aligned in a structurally undefined region of the template MxaA_R and the true orientation might enable its function in the proton relay system. Structural data from X-ray crystallography of ATRed$_{xind01729}$_R would support this hypothesis and this should be addressed in the future. Otherwise, a π-π stacking interaction with the adenine moiety of NAD(P)H could also be possible and contribute to the overall functionality of the R domain. In this scenario, a water molecule would bypass the proton relay system from 2`-hydroxy group of NAD(P)H to Asn750.
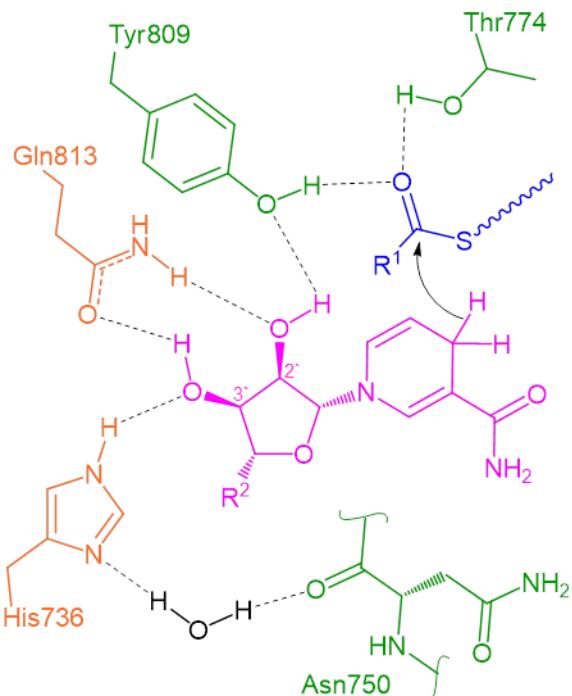
Figure 15. Proposed Gln/His-catalysed reduction of thioesters. NAD(P)H-derived hydride transfer onto the 4'-PPant-attached thioester and the corresponding proton relay system. Residues of the R domain are highlighted in green, Gln813 and His736 in orange, the substrate in blue and the cofactor in pink. Abbreviations are R^1 = peptide and R^2 = adenosine (3` phosphate-) ribose pyrophosphate moiety of NAD(P)H.

4.3.2 Classification of the ATReds

A heatmap of pairwise identity of a multiple alignment with all ATReds from *Xenorhabdus* unveiled three distinct clusters of ATReds (Fig. 11, Supplementary Fig. 2.5). According to this, it can be postulated that ATReds from *Xenorhabdus* are divided into three subtypes (Fig. 16). The largest cluster contains only ATReds with a canonical catalytic triad of R domains and with exception of *xpoi02758*, all respective *atred*s are associated with an *mlp*. Here, these are named as subtype 1 of ATReds. Subtype 2 ATReds differ only by the absence on an *mlp* from subtype 1. Finally, the third subtype harbours all ATReds with His736 and Gln813 or His736 only (for ATRed$_{xinn00707}$).
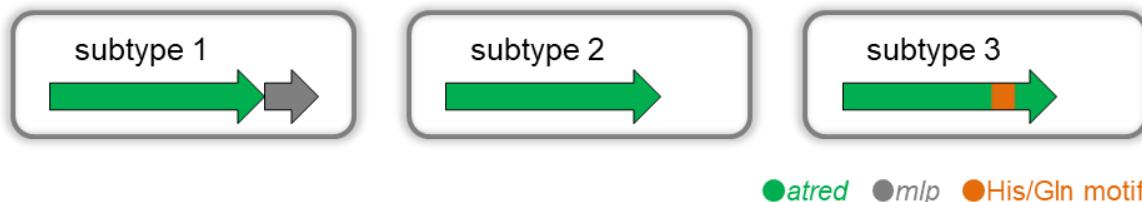
Figure 16. The three ATRed subtypes in *Xenorhabdus*. Schematic overview and classification of *atred*s (green) according to the presence of an *mlp* (grey) or the His/Gln motif (orange).

The strains differ by the number of encoded ATReds (e.g. one *atred* in *X. miraniensis*, two *atred*s in *X. vietnamensis* or three *atred*s in *X. nematophila*) and there is not more than one *atred* of each subtype present in each strain. However, pyrazine production was only observed in *Xenorhabdus* strains with *atred*s of subtype 3 and because of this, it can be assumed that the pyrazine production in *Xenorhabdus* strains is related to the subtype 3 ATReds. These are namely ATRed$_{xcabDSM03579}$ in *X. cabanillasii* DSM 17905, ATRed$_{xnem01561}$ in *X. nematophila ATCC 19061*, ATRed$_{xvie03245}$ in X. vietnamensis and ATRed$_{xcabJM01329}$ in *X. cabanillasii* JM26 as well as ATRed$_{xind01729}$ in *X. indica* DSM 17382, respectively. However, this should be verified by promoter exchange or deletion mutants as conducted in *X. indica*.[165]

In general, non-ribosomal R domains belong to the "extended" family of SDR enzymes. Those have been categorized in five families based on preserved residues within the coenzyme-binding site and active site region.[194] "Extended" SDRs are defined by a [ST]GxxGxxG cofactor binding motif, an Yxx[AST]K active site motif and a less conserved *C*-terminal extension as seen e.g. in the structure of MxaA_R[116] and homology model of ATRed$_{xind01729}$_R (Supplementary Fig. 2.4). Other SDR families are

Discussion

e.g. "classical" including CARs[194,195] or "complex" like KRs.[192] The latter ones have been reported to share an YxxxN active motif and a Tyr-based mechanism of reduction in an NAD(P)H-dependent manner as observed for ATRed$_{xind01729}$_R with Gln instead of Asn. Despite the different substrates, the role of this Asn in "complex" SDR is, however, not equivalent to the active site Gln of pyrazine-producing ATReds. As e.g. the crystal structure of the KR domain of 6-deoxyerythronolide B-producing PKS (DEBS) in *S. erythraea*[196] illustrates, DEBS_Asn1817 forms a hydrogen bond with the carbonyl backbone of DEBS_Tyr1813 and the interaction with the cofactor is facilitated by another Lys-residue resulting in the canonical catalytic triad which is just contributed by distinct parts of the enzyme scaffold.[192] Thus, the present case of Gln acting as part of the catalytic triad/proton relay system, represents to our knowledge a new active site motif within the families of SDRs.

It should be noted that within the third subtype of ATReds, two deviations occur. First, the ATRed$_{xinn00707}$ in **4b**-producing *X. innexi* does contain a His736 but a canonical Lys instead of the discussed Gln813 for subtype 3 ATReds. However, this would not exclude the proposed mechanism with His736 contributing to the proton relay system (Fig. 15) since Lys only substitutes Gln in order to fulfil its published role in the catalytic triad. Second, despite encoding *xbud02951*, *X. budapestensis* does not show production of a pyrazine as observed for other strains with subtype 3 ATReds. One explanation could be e.g. the absence of a transcriptional activator.[197] The experimental setup of all 20 tested *Xenorhabdus* strains was the same and no optimization on cultivation conditions was performed. The use of different media might stimulate the production of natural products by simulating the natural environment of *Xenorhabdus*[198] as e.g. observed for the 13 and 64 fold-increased tilivalline production in *X. eapokensis* in Schneider's or SF-900 insect medium respectively compared to LB medium.[27] Beside this ecological approach, the exchange of the native promoter against an arabinose-inducible promoter would address a "silent" gene.[199,200]

Nevertheless, it remains unclear why all strains with only ATReds of subtype 1 and/or 2 did not show production of pyrazines. Apart from the active site, the ATReds of the three classes share less similarity among each other (Supplementary Fig. 2.5) indicating that rather multiple and overall intramolecular interactions than only single active site residues

might contribute to the enzyme function. More relevant is the fact that pyrazines can be volatile widespread in nature.[201,202] In contrast to Phe-substituted pyrazine core structure **4a**, smaller substituents would lower their vapour pressure. In this case, the established extraction and detection method would not capture the whole spectrum of natural products. This can be addressed by using a dodecane or isopropyl myristate overlay, an *in situ* two-phase extraction or closed-loop stripping apparatus and subsequent gas chromatography-MS as reported for heterologous expression in *E. coli* or biosynthesis of volatiles in *Myxococcus*.[203–205]

Pyrazine-based compounds have been attributed with functions in bacteria as e.g. quorum sensing (autoinducer 3,5-dimethylpyrazine-2-ol in *V. cholera*)[206], antimicrobial activity (2,5-bis(1-methylethyl)-pyrazine produced by *Paenibacillus* sp. AD87 upon co-cultivation with *Burkholderia* sp. AD24)[207] or pheromone activity for symbiotic lifestyle (ant-associated *S. marcescens* 3B2).[208] Natural products with such properties fit to the chemical diversity of *Xenorhabdus* secondary metabolism.[209] In short, entomopathogenic *Xenorhabdus* spp. symbiotically colonizes the gut of the nematode *Steinernema* spp. which infests insect larvae and releases the bacteria until both re-associate and emerge from the cadaver.[198] The metabolites produced by *Xenorhabdus* fulfil crucial functions like disabling the insect immune system, killing the insect, defence against food competitors, support of nematode development and acting in cell-cell communication during the organization of mutualism and pathogenesis.[210] In this light, ATReds might contribute to their host`s complex lifecycle, although the true function of **4a**, **4b** as well as subtype 1 and 2 ATReds is unknown up to date. The widespread existence of three different ATReds with not more than one of each subtype being present per strain supports this assumption.

5   References

1.   Fleming, A. Nobel Lecture, Physiology or Medicine (1945).

2.   Rossiter, S. E., Fletcher, M. H. & Wuest, W. M. Natural Products as Platforms To Overcome Antibiotic Resistance. *Chem. Rev.* **117,** 12415–12474 (2017).

3.   Schäberle, T. F. & Hack, I. M. Overcoming the current deadlock in antibiotic research. *Trends Microbiol.* **22,** 165–167 (2014).

4.   U.S. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States* (2019).

5.   Sharma, A. Antimicrobial resistance. No action today, no cure tomorrow. *Indian J. Med. Microbiol.* **29,** 91–92 (2011).

6.   Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19,** 56–66 (2019).

7.   O'Neill, J. Tackling drug-resistant infections globally: final report and recommendations. *The Review on Antimicrobial Resistance* (2016).

8.   Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20,** 533–534 (2020).

9.   Nicola, M. *et al.* The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *Int. J. Surg.* **78,** 185–193 (2020).

10.   Cragg, G. M. & Newman, D. J. Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830,** 3670–3695 (2013).

11.   Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J. Antibiot.* **65,** 385–395 (2012).

12.   Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661 (2016).

13.   Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41,** 185–201 (2014).

14.   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 Discovery* **14,** 111–129 (2015).

15. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11,** 660–670 (2015).

16. Bozhüyük, K. A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51,** 88–96 (2019).

17. Katz, L. & Baltz, R. H. Natural product discovery. Past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43,** 155–176 (2016).

18. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed.* **56,** 3770–3821 (2017).

19. Mach, B., Reich, E. & Tatum, E. L. Separation of the biosynthesis of the antibiotic polypeptide tyrocidine from protein biosynthesis. *Proc. Natl. Acad. Sci. U.S.A* **50,** 175–181 (1963).

20. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U.S.A* **111,** 9259–9264 (2014).

21. Shou, Q. *et al.* A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat. Chem. Biol.* **12,** 770–772 (2016).

22. Richardt, A. *et al.* Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in *Drosophila*. *J. Biol. Chem.* **278,** 41160–41166 (2003).

23. Walsh, C. T. Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **303,** 1805–1810 (2004).

24. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis. Approaches to new antibiotics. *Chem. Rev.* **105,** 715–738 (2005).

25. Zhou, Q. *et al.* Structure and biosynthesis of xenoamicins from entomopathogenic *Xenorhabdus*. *Chemistry* **19,** 16772–16779 (2013).

26. May, J. J., Wendrich, T. M. & Marahiel, M. A. The dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J. Biol. Chem.* **276,** 7209–7217 (2001).

References

27.   Wolff, H. & Bode, H. B. The benzodiazepine-like natural product tilivalline is produced by the entomopathogenic bacterium *Xenorhabdus eapokensis*. *PloS one* **13,** e0194297 (2018).

28.   Chen, Y., McClure, R. A., Zheng, Y., Thomson, R. J. & Kelleher, N. L. Proteomics guided discovery of flavopeptins. Anti-proliferative aldehydes synthesized by a reductase domain-containing non-ribosomal peptide synthetase. *J. Am. Chem. Soc.* **135,** 10449–10456 (2013).

29.   Hubbard, B. K. & Walsh, C. T. Vancomycin Assembly: Nature's Way. *Angew. Chem. Int. Ed.* **42,** 730–765 (2003).

30.   Reimer, D. *et al.* Rhabdopeptides as insect-specific virulence factors from entomopathogenic bacteria. *ChemBioChem* **14,** 1991–1997 (2013).

31.   Xu, Y. *et al.* Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of *Beauveria bassiana*. *Fungal Genet. Biol.* **46,** 353–364 (2009).

32.   Gunsior, M. *et al.* The biosynthetic gene cluster for a monocyclic beta-lactam antibiotic, nocardicin A. *Chem. Biol.* **11,** 927–938 (2004).

33.   Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535,** 511–516 (2016).

34.   Gehring, A. G., Bradley, K. A. & Walsh, C. T. Enterobactin Biosynthesis in *Escherichia coli*: Isochorismate Lyase (EntB) Is a Bifunctional Enzyme That Is Phosphopantetheinylated by EntD and Then Acylated by EntE Using ATP and 2,3-Dihydroxybenzoate. *Biochemistry* **36,** 8495–8503 (1997).

35.   Christiansen, G., Fastner, J., Erhard, M., Börner, T. & Dittmann, E. Microcystin biosynthesis in *planktothrix*. Genes, evolution, and manipulation. *J. Bacteriol.* **185,** 564–572 (2003).

36.   Lawen, A. Biosynthesis of cyclosporins and other natural peptidyl prolyl cis/trans isomerase inhibitors. *Biochim. Biophys. Acta* **1850,** 2111–2120 (2015).

37.   Shen, B. *et al.* Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. *J. Nat. Prod.* **65,** 422–431 (2002).

38.   Gonsior, M. *et al.* Biosynthesis of the Peptide Antibiotic Feglymycin by a Linear Nonribosomal Peptide Synthetase Mechanism. *ChemBioChem* **16,** 2610–2614 (2015).

39.   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).

40.   Gevers, W., Kleinkauf, H. & Lipmann, F. The activation of amino acids for biosynthesis of gramicidin S. *Proc. Natl. Acad. Sci. U.S.A.* **60,** 269–276 (1968).

41.   Gevers, W., Kleinkauf, H. & Lipmann, F. Peptidyl transfer in gramicidin S biosynthesis from enzyme-bound thioester intermediates. *Proc. Natl. Acad. Sci. U.S.A.* **63,** 1335–1342 (1969).

42.   Kleinkauf, H., Gevers, W. & Lipmann, F. Interrelation between activation and polymerization in gramicidin S biosynthesis. *Proc Natl Acad Sci USA* **62,** 226–233 (1969).

43.   Lee, S. G., Roskoski, R., Bauer, K. & Lipmann, F. Purification of the polyenzymes responsible for tyrocidine synthesis and their dissociation into subunits. *Biochemistry* **12,** 398–405 (1973).

44.   Kleinkauf, H., Gevers, W., Roskoski, R. & Lipmann, F. Enzyme-bound phosphopantetheine in tyrocidine biosynthesis. *Biochem. Biophys. Res. Commun.* **41,** 1218–1222 (1970).

45.   Gilhuus-Moe, C. C., Kristensen, T., Bredesen, J. E., Zimmer, T.-L. & Laland, S. G. The presence and possible role of phosphopantothenic acid in gramicidin S synthetase. *FEBS Lett.* **7,** 287–290 (1970).

46.   Lipmann, F. Nonribosomal polypeptide synthesis on polyenzyme templates. *Acc. Chem. Res.* **6,** 361–367 (1973).

47.   Kurahashi, K. Biosynthesis of Small Peptides. *Annu. Rev. Biochem.* **43,** 445–459 (1974).

48.   Laland, S. G & Zimmer, T. L. The protein thiotemplate mechanism of synthesis for the peptide antibiotics produced by *Bacillus brevis*. *Essays Biochem.* **9,** 31–57 (1973).

References

49. Stein, T. *et al.* The Multiple Carrier Model of Nonribosomal Peptide Biosynthesis at Modular Multienzymatic Templates. *J. Biol. Chem.* **271,** 15428–15435 (1996).

50. Roskoski, R., Kleinkauf, H. & Gevers, W. & Lipmann, F. Isolation of enzyme-bound peptide intermediates in tyrocidine biosynthesis. *Biochemistry* **9,** 4846–4851 (1970).

51. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases A list of abbreviations can be found at the end of the text. *ChemBioChem* **3,** 490–504 (2002).

52. Kessler, N., Schuhmann, H., Morneweg, S., Linne, U. & Marahiel, M. A. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J. Biol. Chem.* **279,** 7413–7419 (2004).

53. 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).

54. Hoyer, K. M., Mahlert, C. & Marahiel, M. A. The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* **14,** 13–22 (2007).

55. Felnagle, E. A., Rondon, M. R., Berti, A. D., Crosby, H. A. & Thomas, M. G. Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. *Appl. Environ. Microbiol.* **73,** 4162–4170 (2007).

56. Sun, W.-W., Guo, C.-J. & Wang, C. C. C. Characterization of the product of a nonribosomal peptide synthetase-like (NRPS-like) gene using the doxycycline dependent Tet-on system in *Aspergillus terreus. Fungal Genet. Biol.* **89,** 84–88 (2016).

57. Fernández-Martínez, L. T. *et al.* New insights into chloramphenicol biosynthesis in *Streptomyces venezuelae* ATCC 10712. *Antimicrob. Agents Chemother.* **58,** 7441–7450 (2014).

58. 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).

59. Zhang, W., Ntai, I., Kelleher, N. L. & Walsh, C. T. tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 12249–12253 (2011).

60. Gulick, A. M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4,** 811–827 (2009).

61. Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H. & Zocher, R. Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *J. Biol. Chem.* **269,** 12710–12714 (1994).

62. Bučević-Popović, V., Sprung, M., Soldo, B. & Pavela-Vrančič, M. The A9 core sequence from NRPS adenylation domain is relevant for thioester formation. *ChemBioChem* **13,** 1913–1920 (2012).

63. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16,** 4174–4183 (1997).

64. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493–505 (1999).

65. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7,** 211–224 (2000).

66. Blin, K. *et al.* antiSMASH 5.0. Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47,** W81-W87 (2019).

67. Yonus, H. *et al.* Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J. Biol. Chem.* **283,** 32484–32491 (2008).

References

68. Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51,** 3252–3263 (2012).

69. Tan, X.-F. *et al.* Structure of the adenylation-peptidyl carrier protein didomain of the *Microcystis aeruginosa* microcystin synthetase McyG. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **71,** 873–881 (2015).

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

71. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **291,** 22559–22571 (2016).

72. Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288,** 1991–2003 (2013).

73. Boll, B., Taubitz, T. & Heide, L. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem.* **286,** 36281–36290 (2011).

74. Crosby, J. & Crump, M. P. The structural role of the carrier protein--active controller or passive carrier. *Nat. Prod. Rep.* **29,** 1111–1137 (2012).

75. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A. & Holak, T. A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8,** 407–418 (2000).

76. Koglin, A. *et al.* Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science* **312,** 273–276 (2006).

77. Reuter, K., Mofid, M. R., Marahiel, M. A. & Ficner, R. Crystal structure of the surfactin synthetase-activating enzyme Sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamiliy. *EMBO J.* **18,** 6823–6831 (1999).

78. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: Catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31,** 61–108 (2014).

79. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: Discovery, dissection and diversity. *Biochim. Biophys. Acta* **1865,** 1587–1604 (2017).

80. Crécy-Lagard, V. de, Marlière, P. & Saurin, W. Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C R Acad. Sci. III* **318,** 927–936 (1995).

81. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **273,** 22773–22781 (1998).

82. Bergendahl, V., Linne, U. & Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269,** 620–629 (2002).

83. 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).

84. Bloudoff, K., Alonzo, D. A. & Schmeing, T. M. Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **23,** 331–339 (2016).

85. Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366,** 6466 (2019).

86. 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).

87. 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).

88. Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.* **7,** 765–772 (2000).

89. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science* **284,** 486–489 (1999).

References

90.   Clugston, S. L., Sieber, S. A., Marahiel, M. A. & Walsh, C. T. Chirality of Peptide Bond-Forming Condensation Domains in Nonribosomal Peptide Synthetases: The C_5 Domain of Tyrocidine Synthetase Is a ^DC_L Catalyst. *Biochemistry* **42,** 12095–12104 (2003).

91.   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).

92.   Velasco, A. *et al.* Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2. Designing new cytotoxic compounds. *Mol. Microbiol.* **56,** 144–154 (2005).

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

94.   Kronenwerth, M. *et al.* Characterisation of taxlllaids A-G; natural products from *Xenorhabdus indica*. *Chemistry* **20,** 17478–17487 (2014).

95.   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).

96.   Crawford, J. M., Portmann, C., Kontnik, R., Walsh, C. T. & Clardy, J. NRPS substrate promiscuity diversifies the xenematides. *Org. Lett.* **13,** 5144–5147 (2011).

97.   Samel, S. A., Czodrowski, P. & Essen, L.-O. Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **70,** 1442–1452 (2014).

98.   Luo, L., Burkart, M. D., Stachelhaus, T. & Walsh, C. T. Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. *J. Am. Chem. Soc* **123,** 11208–11218 (2001).

99.   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).

100.   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).

101. McMahon, M. D., Rush, J. S. & Thomas, M. G. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194,** 2809–2818 (2012).

102. Duerfahrt, T., Eppelmann, K., Müller, R. & Marahiel, M. A. Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem. Biol.* **11,** 261–271 (2004).

103. Patel, H. M. & Walsh, C. T. *In vitro* reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin. Characterization of backbone tailoring thiazoline reductase and *N*-methyltransferase activities. *Biochemistry* **40,** 9023–9031 (2001).

104. Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis. Thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42,** 9722–9730 (2003).

105. Fujimori, D. G. *et al.* Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 16498–16503 (2007).

106. Pfeifer, B. A., Wang, C. C. C., Walsh, C. T. & Khosla, C. Biosynthesis of Yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host. *Appl. Environ. Microbiol.* **69,** 6698–6702 (2003).

107. Al-Mestarihi, A. H. *et al.* Adenylation and S-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis. *J. Am. Chem. Soc.* **136,** 17350–17354 (2014).

108. Müller, S. *et al.* Paenilamicin. Structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. *Angew. Chem. Int. Ed.* **53,** 10821–10825 (2014).

109. Shi, R. *et al.* Structure and function of the glycopeptide *N*-methyltransferase MtfA, a tool for the biosynthesis of modified glycopeptide antibiotics. *Chem. Biol.* **16,** 401–410 (2009).

110. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27,** 255–278 (2010).

111. Bruner, S. D. *et al.* Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE. *Structure* **10,** 301–310 (2002).

References

112. Lombó, F. *et al.* Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *streptomyces* species. *ChemBioChem* **7,** 366–376 (2006).

113. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24,** 735–749 (2007).

114. Schwarzer, D., Mootz, H., D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14083–14088 (2002).

115. Koglin, A. *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* **454,** 907–911 (2008).

116. Barajas, J. F. *et al.* Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chem. Biol.* **22,** 1018–1029 (2015).

117. Mullowney, M. W., McClure, R. A., Robey, M. T., Kelleher, N. L. & Thomson, R. J. Natural products from thioester reductase containing biosynthetic pathways. *Nat. Prod. Rep.* **35,** 847–878 (2018).

118. Chhabra, A. *et al.* Nonprocessive 2 + 2e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A* **109,** 5681–5686 (2012).

119. Silkowski, B., Nordsiek, G., Kunze, G., Blöcker, H. & Müller, R. Novel features in a combined polyketide synthase/non-ribosomalpeptide synthetase : the myxalamid biosynthetic gene cluster of themyxobacterium *Stigmatella aurantiaca* Sga15. *Chem. Biol.* **8,** 59–69 (2001).

120. Gahloth, D. *et al.* Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. *Nat. Chem. Biol.* **13,** 975–981 (2017).

121. Becker, J. E., Moore, R. E. & Moore, B. S. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster. Molecular basis for imine macrocyclization. *Gene* **325,** 35–42 (2004).

122. Yeh, H.-H. *et al.* Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in *Aspergillus nidulans* Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor. *ACS Chem. Biol.* **11,** 2275–2284 (2016).

123. Guo, C.-J. *et al.* Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. *Cell* **168,** 517-526 (2017).

124. 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).

125. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Chem. Biol.* **20,** 234–240 (2010).

126. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529,** 235–238 (2016).

127. Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **25,** 783-793 (2017).

128. Reimer, J. M., Aloise, M. N., Harrison, P. M. & Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529,** 239–242 (2016).

129. Reimer, J. M., Haque, A. S., Tarry, M. J. & Schmeing, T. M. Piecing together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.* **49,** 104–113 (2018).

130. Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33,** 136–140 (2016).

131. 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).

132. Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35,** 1210–1228 (2018).

133. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33,** 317–347 (2016).

134. 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).

References

135. 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).

136. Fischbach, M. A., Walsh, C. T. & Clardy, J. The evolution ofgene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U.S.A* **105,** 4601–4608 (2008).

137. Alanjary, M., Cano-Prieto, C., Gross, H. & Medema, M. H. Computer-aided re-engineering of nonribosomal peptide and polyketide biosynthetic assembly lines. *Nat. Prod. Rep.* **36,** 1249–1261 (2019).

138. Moran, S., Rai, D. K., Clark, B. R. & Murphy, C. D. Precursor-directed biosynthesis of fluorinated iturin A in *Bacillus* spp. *Org. Biomol. Chem.* **7,** 644–646 (2009).

139. Hojati, Z. *et al.* Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from *Streptomyces coelicolor. Chem. Biol.* **9,** 1175–1187 (2002).

140. Deb Roy, A., 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).

141. Lewis, R. A. *et al.* Active site modification of the β-ketoacyl-ACP synthase FabF3 of *Streptomyces coelicolor* affects the fatty acid chain length of the CDA lipopeptides. *Chem. Commun.* **47,** 1860–1862 (2011).

142. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41,** 9718–9726 (2002).

143. 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).

144. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angew. Chem. Int. Ed.* **53,** 10105–10108 (2014).

145. Hein, C. D., Liu, X.-M. & Wang, D. Click chemistry, a powerful tool for pharmaceutical sciences. *Pharm. Res.* **25,** 2216–2230 (2008).

146. 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. *Chemistry* **22,** 639–645 (2016).

147. Zhang, K. *et al.* Engineering the substrate specificity of the DhbE adenylation domain by yeast cell surface display. *Chem. Biol.* **20,** 92–101 (2013).

148. Villiers, B. & Hollfelder, F. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chem. Biol.* **18,** 1290–1299 (2011).

149. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10,** 282–287 (2018).

150. Cole, M. F. & Gaucher, E. A. Utilizing natural diversity to evolve protein function. Applications towards thermostability. *Curr. Opin. Chem. Biol.* **15,** 399–406 (2011).

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

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

153. Schneider, A., Stachelhaus, T. & Marahiel, M. A. Targeted alteration of the substrate specicity of peptide synthetases by rational module swapping. *Mol Gen Genet* **257,** 308–318 (1998).

154. 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).

155. 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).

156. Calcott, M. J. & Ackerley, D. F. Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.* **15,** 162–174 (2015).

157. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *bioRxiv*; 10.1101/2020.02.28.970632 (2020).

References

158. Nguyen, K. T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 17462–17467 (2006).

159. Miao, V. *et al.* The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J. Ind. Microbiol. Biotechnol.* **33,** 129–140 (2006).

160. Zobel, S. *et al.* Reprogramming the Biosynthesis of Cyclodepsipeptide Synthetases to Obtain New Enniatins and Beauvericins. *ChemBioChem* **17,** 283–287 (2016).

161. Steiniger, C. *et al.* Harnessing fungal nonribosomal cyclodepsipeptide synthetases for mechanistic insights and tailored engineering. *Chem. Sci.* **8,** 7834–7843 (2017).

162. Lundy, T. A., Mori, S. & Garneau-Tsodikova, S. Engineering Bifunctional Enzymes Capable of Adenylating and Selectively Methylating the Side Chain or Core of Amino Acids. *ACS Synth. Biol* **7,** 399–404 (2018).

163. Bozhüyük, K. A. J. *et al. De novo* design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10,** 275–281 (2018).

164. Bozhüyük, K. A. J. *et al.* Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11,** 653–661 (2019).

165. Tietze, A., Shi, Y.-N., Kronenwerth, M. & Bode, H. B. Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains. *ChemBioChem*; 10.1002/cbic.202000176 (2020).

166. Schimming, O. *et al.* Structure, Biosynthesis, and Occurrence of Bacterial Pyrrolizidine Alkaloids. *Angew. Chem. Int. Ed.* **54,** 12702–12705 (2015).

167. 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).

168. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** 237-243 (2015).

169. Tietze, A. Reprogrammierung von nichtribosomalen Peptidsynthetasen aus entomopathogenen Bakterien. Masterarbeit. Goethe-Universität Frankfurt am Main (2016).

170. Fleischhacker, F. Functional characterization and reprogramming of nonribosomal peptide synthetases from entomopathogenic bacteria. Dissertation. Goethe-Universität Frankfurt am Main (2017).

171. Bozhüyük, K. A. J. Reprogramming Non-ribosomal Peptide Synthetases. Dissertation. Goethe-Universität Frankfurt am Main (2016).

172. Linck, A. Mass spectrometric characterisation of natural products and reprogramming of non-ribosomal peptide synthetases from entomopathogenic bacteria. Dissertation. Goethe-Universität Frankfurt am Main (2018).

173. 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).

174. Steiniger, C., Hoffmann, S. & Süssmuth, R. D. Probing Exchange Units for Combining Iterative and Linear Fungal Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **26,** 1526–1534 (2019).

175. 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).

176. Yan, F. *et al.* Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9,** 7510–7519 (2018).

177. Farag, S. *et al.* Inter-Modular Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides. *Bioinformatics* **35,** 3584–3591 (2019).

178. Owen, J. G., Calcott, M. J., Robins, K. J. & Ackerley, D. F. Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem. Biol.* **23,** 1395–1406 (2016).

179. Schmidt, Y. *et al.* Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*. *ChemBioChem* **15,** 259–266 (2014).

180. Gaudelli, N. M., Long, D. H. & Townsend, C. A. β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520,** 383–387 (2015).

181. Wolf, F. *et al.* Biosynthesis of the β-Lactone Proteasome Inhibitors Belactosin and Cystargolide. *Angew. Chem. Int. Ed.* **56,** 6665–6668 (2017).

182. Schorn, M. *et al.* Genetic basis for the biosynthesis of the pharmaceutically important class of epoxyketone proteasome inhibitors. *ACS Chem. Biol.* **9,** 301–309 (2014).

References

183. Kisselev, A. F., van der Linden, W. A. & Overkleeft, H. S. Proteasome inhibitors. An expanding army attacking a unique target. *Chem. Biol.* **19,** 99–115 (2012).

184. Lin, G., Li, D., Chidawanyika, T., Nathan, C. & Li, H. Fellutamide B is a potent inhibitor of the *Mycobacterium tuberculosis* proteasome. *Arch. Biochem. Biophys.* **501,** 214–220 (2010).

185. Hines, J., Groll, M., Fahnestoc, M. & Crews, C. M. Proteasome Inhibition by Fellutamide B Induces Nerve Growth Factor Synthesis. *Chem. Biol.* **15,** 501–512 (2008).

186. Li, H. *et al.* Structure- and function-based design of Plasmodium-selective proteasome inhibitors. *Nature* **530,** 233–236 (2016).

187. Totaro, K. A. *et al.* Rational Design of Selective and Bioactive Inhibitors of the *Mycobacterium tuberculosis* Proteasome. *ACS Infect. Dis.* **3,** 176–181 (2017).

188. Haque, A. S. *et al.* Delineating the reaction mechanism of reductase domains of Nonribosomal Peptide Synthetases from mycobacteria. *J. Struct. Biol.* **187,** 207–214 (2014).

189. Benach, J., Atrian, S., Gonzàlez-Duarte, R. & Ladenstein, R. The Catalytic Reaction and Inhibition Mechanism of *Drosophila* Alcohol Dehydrogenase: Observation of an Enzyme-bound NAD-ketone Adduct at 1.4 Å Resolution by X-ray Crystallography. *J. Mol. Biol.* **289,** 335–355 (1999).

190. Filling, C. *et al.* Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J. Biol. Chem.* **277,** 25677–25684 (2002).

191. Rhys, N. H., Soper, A. K. & Dougan, L. The hydrogen-bonding ability of the amino acid glutamine revealed by neutron diffraction experiments. *J. Phys. Chem. B* **116,** 13308–13319 (2012).

192. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65,** 3895–3906 (2008).

193. Schäfer, M., Stevenson, C. E. M., Wilkinson, B., Lawson, D. M. & Buttner, M. J. Substrate-Assisted Catalysis in Polyketide Reduction Proceeds via a Phenolate Intermediate. *Cell Chem. Biol.* **23,** 1091–1097 (2016).

194. Kallberg, Y., Oppermann, U., Jörnvall, H. & Persson, B. Short-chain dehydrogenases/reductases (SDRs). *Eur. J. Biochem.* **269,** 4409–4417 (2002).

195. Gahloth, D., Aleku, G. A. & Leys, D. Carboxylic acid reductase: Structure and mechanism. *J. Biotechnol.* **307,** 107–113 (2020).

196. Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **14,** 737–748 (2006).

197. Bervoets, I. & Charlier, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiol. Rev.* **43,** 304–339 (2019).

198. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: Two roads to the same destination. *Mol. Microbiol.* **64,** 260–268 (2007).

199. Bode, E. *et al.* Simple "on-demand" production of bioactive natural products. *ChemBioChem* **16,** 1115–1119 (2015).

200. Bode, E. *et al.* Promoter Activation in Δhfq Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angew. Chem. Int. Ed.* **58,** 18957–18963 (2019).

201. Schulz, S. & Dickschat, J. S. Bacterial volatiles: The smell of small organisms. *Nat. Prod. Rep.* **24,** 814–842 (2007).

202. Müller, R. & Rappert, S. Pyrazines: Occurrence, formation and biodegradation. *Appl. Microbiol. Biotechnol.* **85,** 1315–1320 (2010).

203. Liu, W. *et al.* Engineering *Escherichia coli* for high-yield geraniol production with biotransformation of geranyl acetate to geraniol under fed-batch culture. *Biotechnol. Biofuels* **9,** 58 (2016).

204. Kim, E.-M., Eom, J.-H., Um, Y., Kim, Y. & Woo, H. M. Microbial Synthesis of Myrcene by Metabolically Engineered *Escherichia coli*. *J. Agric. Food. Chem.* **63,** 4606–4612 (2015).

205. Dickschat, J. S., Wenzel, S. C., Bode, H. B., Müller, R. & Schulz, S. Biosynthesis of volatiles by the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **5,** 778–787 (2004).

References

206. Papenfort, K. *et al.* A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nat. Chem. Biol.* **13,** 551–557 (2017).

207. Janssens, T. K. S., Tyc, O., Besselink, H., Boer, W. de & Garbeva, P. Biological activities associated with the volatile compound 2,5-bis(1-methylethyl)-pyrazine. *FEMS Microbiol. Lett.* **366,** fnz023 (2019).

208. Silva-Junior, E. A. *et al.* Pyrazines from bacteria and ants: convergent chemistry within an ecological niche. *Sci. Rep.* **8,** 2595–2601 (2018).

209. Shi, Y.-M. & Bode, H. B. Chemical language and warfare of bacterial natural products in bacteria-nematode-insect interactions. *Nat. Prod. Rep.* **35,** 309–335 (2018).

210. Tobias, N. J., Shi, Y.-M. & Bode, H. B. Refining the Natural Product Repertoire in Entomopathogenic Bacteria. *Trends Microbiol.* **26,** 833–840 (2018).

211. Hanahan, D. Studies on Transformation of *Escherichia coli* with Plasmids. *J. Mol. Biol.* **166,** 557–580 (1983).

212. 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).

213. 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).

214. Wichard, T., Poulet, S. A. & Pohnert, G. Determination and quantification of α,β,γ,δ-unsaturated aldehydes as pentafluorobenzyl-oxime derivates in diatom cultures and natural phytoplankton populations: Application in marine field studies. *J. Chromatogr. B* **814,** 155–161 (2005).

215. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47,** 636 - 641 (2019).

216. Schilling, N. A. *et al.* Synthetic Lugdunin Analogues Reveal Essential Structural Motifs for Antimicrobial Action and Proton Translocation Capability. *Angew. Chem. Int. Ed.* **58,** 9234–9238 (2019).

217. Tobias, N. J. *et al.* Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.* **2,** 1676–1685 (2017).

218. Thanwisai, A. *et al.* Diversity of *Xenorhabdus* and *Photorhabdus* spp. and their symbiotic entomopathogenic nematodes from Thailand. *PloS one* **7,** e43835 (2012).

219. Cai, X. *et al.* Biosynthesis of the Antibiotic Nematophin and Its Elongated Derivatives in Entomopathogenic Bacteria. *Org. Lett.* **19,** 806–809 (2017).

6   Attachments

6.1   *De novo* design and engineering of non-ribosomal peptide synthetases

6.1.1 Erklärung zu den Autorenanteilen an der Publikation

Status: published

Name der Zeitschrift: *Nat. Chem.* **10,** 275–281 (2018)[163]

Autoren: Kenan A. J. Bozhüyük (KAJB), Florian Fleischhacker (FF),
Annabell Linck (AL), Frank Wesche (FW), Andreas Tietze
(AT), Claus-Peter Niesert (CPN) und Helge B. Bode (HBB)

Erläuterung: AT: Diese Arbeit wurde während der Doktorarbeit erarbeitet.
AT‡: Diese Arbeit wurde bereits in der Masterarbeit von AT
„Reprogrammierung von nichtribosomalen Peptidsynthetasen
aus entomopathogenen Bakterien" (2016) Goethe-Universität
Frankfurt am Main, als Prüfungsleistung eingereicht, ist aber
Teil der Publikation.

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

KAJB (45 %), FF (20 %), AL (2.5 %), AT (2.5 %), HBB (30 %)

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Klonierung von Plasmiden: KAJB (10 %), FF (10 %), AL (4 %), AT (4 %), AT‡ (2 %);
Heterologe Expression: KAJB (5 %), FF (5 %), AL (2 %), AT (2 %), AT‡ (1 %);
HPLC-MS: KAJB (5 %), FF (5 %), AL (2 %), AT (2 %), AT‡ (1 %); Peptidisolation:
KAJB (5 %), FF (5 %), AL (10 %), FW (2 %); Peptidquantifizierung: KAJB (2 %),
FF (2 %), AL (2 %), AT (2 %); Chemische Synthese: FW (10 %)

(3) zur Erstellung der Datensammlung und Abbildungen

Sequenzalignment und Strukturanalyse: KAJB (20 %), FF (10 %); Domänen und
Modultausch in AmbS: KAJB (20 %), FF (10 %); Workflow XU Konzept: KAJB (5 %);
Reprogrammierung XtpS und GxpS: KAJB (10 %), FF (5 %); Produktion nicht-natürlicher

Peptide: KAJB (4 %), AL (3 %), AT (3 %); C Domäne als Terminationsdomäne: KAJB (10 %)

(4) zur Analyse und Interpretation der Daten

Sequenzalignment und Strukturanalyse: KAJB (10 %), FF (5 %), CPN (5 %), HBB (10 %); Domänen und Modultausch in AmbS: KAJB (20 %), FF (8 %), AT (2 %); Workflow XU-Konzept: KAJB (4 %); Reprogrammierung XtpS und GxpS: KAJB (5 %), FF (5 %), AT (1 %); Produktion nicht-natürlicher Peptide: KAJB (5 %), AL (5 %), AT (5 %); C Domäne als Terminationsdomäne: KAJB (7 %), FF (3 %)

(5) zum Verfassen des Manuskriptes

KAJB (60 %), HBB (40 %)

Ort/Datum	Unterschrift des Promovierenden

Ort/Datum	Unterschrift des Betreuers

6.1.2 Publication

De novo design and engineering of non-ribosomal peptide synthetases

Kenan A. J. Bozhüyük[1], Florian Fleischhacker[1], Annabell Linck[1], Frank Wesche[1], Andreas Tietze[1], Claus-Peter Niesert[2] and Helge B. Bode[1,3]*

Peptides derived from non-ribosomal peptide synthetases (NRPSs) represent an important class of pharmaceutically relevant drugs. Methods to generate novel non-ribosomal peptides or to modify peptide natural products in an easy and predictable way are therefore of great interest. However, although the overall modular structure of NRPSs suggests the possibility of adjusting domain specificity and selectivity, only a few examples have been reported and these usually show a severe drop in production titre. Here we report a new strategy for the modification of NRPSs that uses defined exchange units (XUs) and not modules as functional units. XUs are fused at specific positions that connect the condensation and adenylation domains and respect the original specificity of the downstream module to enable the production of the desired peptides. We also present the use of internal condensation domains as an alternative to other peptide-chain-releasing domains for the production of cyclic peptides.

Non-ribosomal peptide synthetases (NRPSs) are multimodular enzymes or enzyme complexes from bacteria and fungi that are capable of producing a large variety of natural products, several of which are used clinically (for example, cyclosporin, vancomycin and daptomycin)[1,2]. The chemical diversity of non-ribosomal peptides (NRPs) not only relies on the incorporation of non-proteinogenic amino acids, fatty acids, β-amino acids or α-hydroxy acids as building blocks[3], but also on the formation of cyclic and branched cyclic peptides or depsipeptides[4] that additionally can be modified after peptide formation via glycosylation and other modifications[2].

Despite the chemical diversity of the produced NRPs, all NRPSs use a multiple-carrier thiotemplate mechanism. They harbour a modular architecture in which every module consists of different catalytic domains connected via defined linker regions[2,5]. One module is responsible for the selection, activation, processing and connection of a specific amino acid to a second amino acid and, with a few exceptions[6,7], one NRPS protein consists of multiple modules. A minimal NRPS elongation module consists of an adenylation (A) domain for the amino acid selection and activation, a 4′-phosphopantetheinylated thiolation (T) domain as the amino acid carrier and a condensation (C) domain for peptide-bond formation. Additional modifying domains, such as epimerization (E) or *N*-methyltransferase (MT) domains are commonly inserted in *cis* in one or more modules, whereas other enzymes, such as halogenases or oxygenases, work in *trans*[3]. Release of the mature peptide from the NRPS is catalysed by a thioesterase (TE) domain or a terminal C domain (C_{term}) by hydrolysis or cyclization via intramolecular nucleophiles[4,8].

The basic mechanism of most NRPS domains is well understood from detailed *in vitro* experiments using NRPS model systems or individual domains and several efforts tried to transfer this knowledge to the *in vivo* production of NRP analogues[9]. Besides precursor-directed biosynthesis and mutasynthesis[10], engineering of the biosynthetic machinery itself is the most-promising option for the construction of novel NRPs, including the replacement of individual

A or A-T didomains, swapping C-A didomains, complete modules (C-A-T tridomains) and T-C-A tridomains[9,11], altering individual amino acids in the A domain binding pocket and thus the 'specificity-conferring code'[12,13], swapping subdomains of A domains[14], deletions[15] and the insertion[16] of modules.

Although all of the mentioned studies had problems with a decrease in peptide-production yield with the titre progressively decreasing with the number of modifications in the NRPS system, *de novo* NRPS construction is currently not practicable using state-of-the-art methods[9,17]. One possible explanation for this is that not only A but also C domains show a strong stereo and significant side-chain selectivity of the connected amino acids[18]. Moreover, it is clear from NRPS structural biology that a NRPS is highly dynamic and probably requires several different protein–protein interactions to be maintained[5]. Linker sequences that connect the different domains or modules in *cis* (one protein) and docking domains[19] that mediate interactions between two NRPS subunits are both crucial for protein–protein interaction[20].

As none of the current methods show reproducible guidelines to create whole new NRPSs, our goal was to establish a general strategy to generate new (artificial) NRPSs[17]. Such a method would make it possible to establish a NRPS system from a library of NRPS components and to produce new peptides, peptide derivatives, functionalizable peptides and peptide-based compounds.

Results and discussion

Verification of the state-of-the-art methods. To test the state-of-the-art methods, we decided to alter the NRPS AmbS, responsible for the production of ambactin (**1**) (Supplementary Table 1), originally described from *Xenorhabdus*, but functionally also in *E. coli* (Supplementary Fig. 1)[21]. However, exchange of the entire (C-A-T, C/E-A-T) or partial (A-T) domains against homologous fragments from other NRPSs from *Xenorhabdus* or *Photorhabdus* (Fig. 1a) via the yeast assembly of different PCR fragments[21] did not result in the production of the desired or other peptides

---

[1]Merck Stiftungsprofessur für Molekulare Biotechnologie, Fachbereich Biowissenschaften, Goethe Universität Frankfurt, Max-von-Laue-Strasse 9, 60438 Frankfurt am Main, Germany. [2]Performance Materials/Process Technologies, Merck KGaA, Frankfurter Strasse 250, 64293 Darmstadt, Germany. [3]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe Universität Frankfurt, Max-von-Laue-Strasse 15, 60438 Frankfurt am Main, Germany. *e-mail: h.bode@bio.uni-frankfurt.de
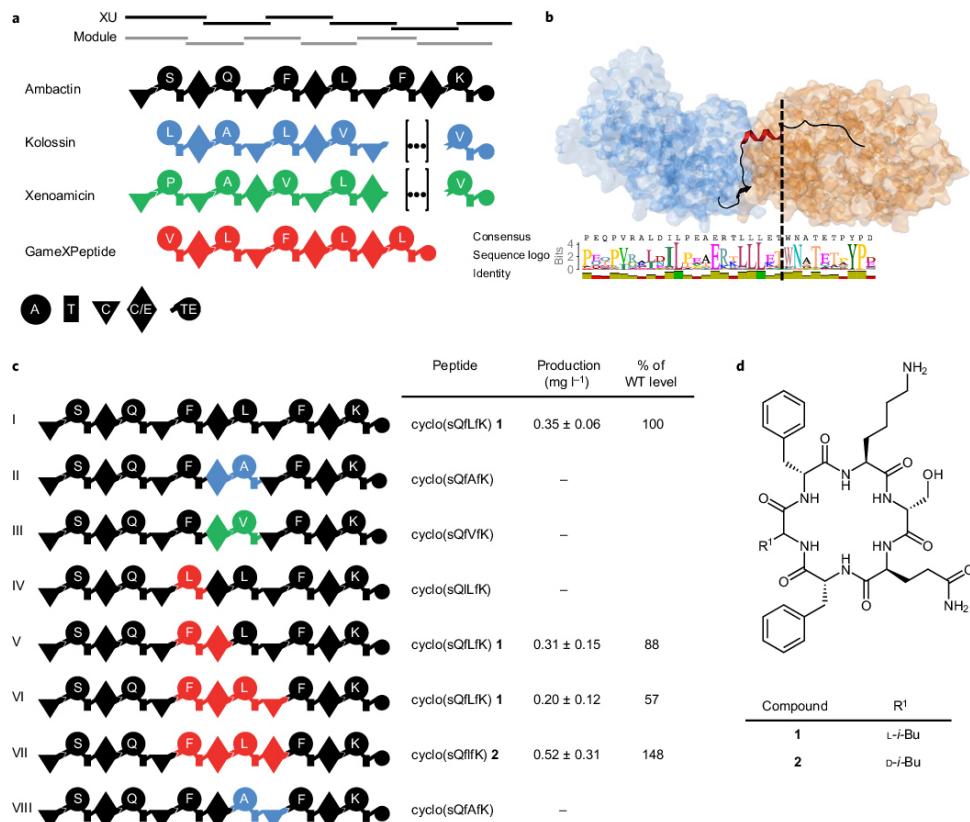
Figure 1 | Domain and module swaps in the ambactin-producing NRPS AmbS. a, Schematic representation of NRPSs used as building blocks in the experiments. Modules and XUs are highlighted for AmbS and specificities are assigned for all of the A domains. The symbols used for domain assignment are: A, large circle; T, rectangle; C, triangle; C/E, diamond; TE, C-terminal small circle. **b**, C-A didomain excised from the SrfA-C crystal structure (Protein Database ID: 2VSQ)[22] with the C-A linker depicted in a ribbon representation (top). C domain, blue; A domain, orange. C-A linker sequence logo of linkers excised from *Photorhabdus* and *Xenorhabdus* NRPSs (bottom). Dashed line shows the used fusion point of the C-A hybrid linker. **c**, Generated AmbS derivatives (I–VIII) and corresponding peptide yields as obtained from triplicate experiments. **d**, Structures of ambactin derivatives **1** and **2**.

(Supplementary Tables 2–4 give the strains, plasmids and oligonucleotides used in this work).

Identification of a new interdomain fusion point. From sequence alignments of interdomain linker regions from NRPSs found in *Photorhabdus* and *Xenorhabdus* (Supplementary Fig. 2), as well as from reviewing available structural data[22-25], the C-A linker was identified as an ideal target to achieve a successful NRPS redesign[22]. Compared with the 32 amino acid long and conserved (on average 44% sequence identity) C-A linker, the shorter A-T (~15 amino acids) and T-C (~18 amino acids) linkers are highly variable (on average only 23% sequence identity) and form a multitude of specific interactions with both domains involved during the NRPS catalytic cycle[18,23]. The C and A domain–domain interactions are formed by 'weak' hydrophobic interactions between both domains, and the small C-terminal subunit of the A domain has to adopt a second conformation during one catalytic cycle of

peptide elongation[26]. The second conformation[27] (Supplementary Fig. 3) is necessary to form the A-T didomain interface and suggests some kind of structural flexibility of the C-A domain interface as well as of the interaction-mediating linker. Moreover, the C-A linker can be divided into two structurally unrelated parts. The first 22 N-terminal amino acids mediate the C-domain–A-domain interactions, of which eight amino acids form a helical structure, which is mainly associated with the C domain (Fig. 1b and Supplementary Fig. 3). The C-terminal ten amino acids starting in *Photorhabdus* with the consensus motif WNATE (Fig. 1b and Supplementary Fig. 2) are only associated with the downstream A domain and form no secondary structure (Supplementary Fig. 3). This position was predicted to be the ideal fusion point for A-T-C tridomains because (1) no secondary structure interferes with the A or C domains, (2) the fusion point is located within a conformationally flexible loop, (3) no domain–domain interactions are disrupted and (4) no C-domain–A-domain

Attachments
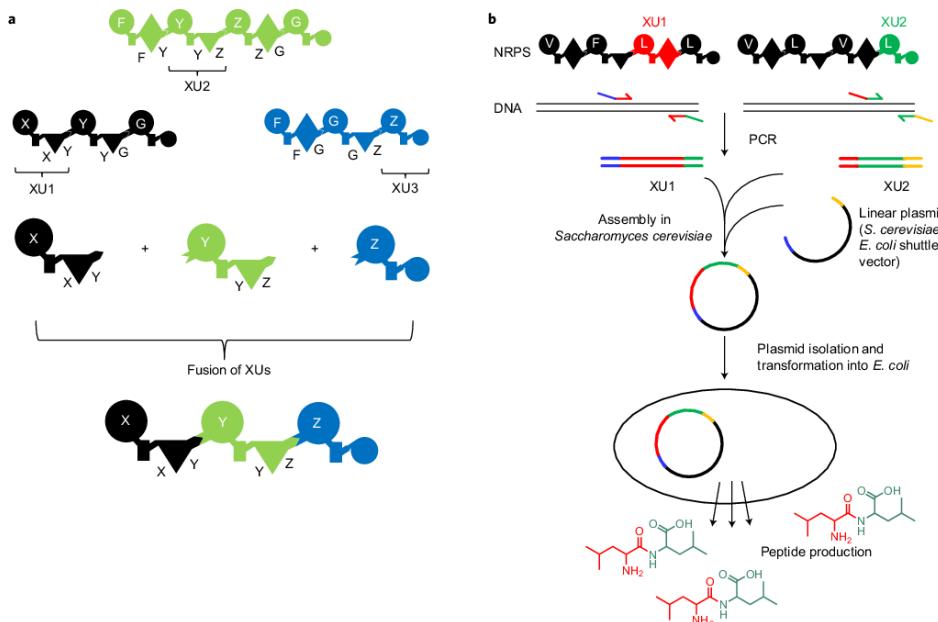
Figure 2 | XU-concept workflow. a, Schematic representation of NRPS recombination rules using XUs, taking into account the C-domain specificities. The symbols used are as in Fig. 1. Amino acid specificities of the A and the C or C/E domains are indicated by capital letters. **b**, Schematic overview of the generation of novel peptides exemplified for a dipeptide using the XU concept as applied in this work.

interface that mediates interactions is present within the second part of the C-A linker (Fig. 1b). The assumed C-A interface flexibility contradicts the common view that C-A domain interactions remain invariant during the catalytic cycle and should not be separated to reprogram NRPSs[28]. In addition, the recently elucidated structures of two distinct conformations of *holo*-NRPSs (EntF, AB3403) suggest that the C-A domain platform may be more dynamic than previously proposed[24]. Thus, tridomains that consist of A-T-C or A-T-C/E were chosen as the exchange units (XUs) (Supplementary Fig. 4). A-T-C tridomains have been mentioned once before to construct artificial NRPSs[29]; however, there the fusion point was located within the α-helical structure of the C-A linker, which resulted in impaired enzymes and decreased product yields.

XU assembly results in functional NRPSs. The XU concept (Fig. 2) is based on the following three rules: (1) A-T-C or A-T-C/E are used as XUs. (2) The specificity of the downstream C domain must be respected. This means that XU1 can only fuse with an XU2 that has the same amino acid specificity as that of the downstream natural XU1–XU2 arrangement. (3) XUs are fused in the C-A linker at the conserved WNATE sequence. When these three rules were applied to exchange the Phe-specific XU3 against a Phe-specific XU from the GxpS NRPS, almost no loss in productivity was observed (Fig. 1c V). Similarly, exchange of AmbS XU3–XU4 against the two building blocks XU3 and XU2 from GxpS, respectively, resulted also in the production of **1** in 57% yield (Fig. 1c VI) compared with the wild-type (WT) construct (Fig. 1c I). The new derivative **2** having a D-Leu at position 4 that results from an additional E domain in XU4 was

produced even better when XU3–XU4 was exchanged against XU3–XU4 from the GxpS NRPS (Fig. 1c VII). To the contrary, the use of an XU specific for Ala and Ala-Leu condensation, and therefore not respecting the downstream C domain specificity, did not produce any peptide (Fig. 1c VIII). The structures of all of the derivatives were confirmed by tandem mass spectroscopy (MS/MS) analysis and by comparison with synthetic standards (Supplementary Fig. 5).

As the construction of a functional artificial NRPS with more than two modules has not been shown yet, we first tried the reconstruction of naturally occurring NRPSs to validate the XU concept and to compare production titres of natural NRPSs with artificially assembled ones. Such NRPSs are ideal for the validation of the XU concept because peptide-releasing TE domains are available that could otherwise prevent the formation/release of the desired peptides because of substrate incompatibilities.

As a starting point, the xenotetrapeptide (**3**)-producing NRPS XtpS (Fig. 3a I) was reconstructed from fragments of the natural NRPS GxpS (refs 30,31), KolS (ref. 32) (Supplementary Fig. 6) and the natural terminal XtpS XU A-T–TE (Fig. 3a II) or the XtpS TE alone (Fig. 3a III), both of which led to the production of **3** in almost 50% yield compared with the original XtpS (ref. 33), as confirmed by MS/MS analysis and a comparison of the retention times (Supplementary Fig. 7). Next, the GameXPeptide-producing NRPS GxpS (ref. 30) was reconstructed from 2–5 XUs from up to four different NRPSs (GxpS, GarS, KolS and XtpS (Supplementary Fig. 6)), as shown in Fig. 3b II–IV and Supplementary Figs 8 and 9a. Although the production titre inversely correlated with the number of XUs, the use of a C-A linker derived exclusively from the downstream A domain (Fig. 3b

72

ARTICLES
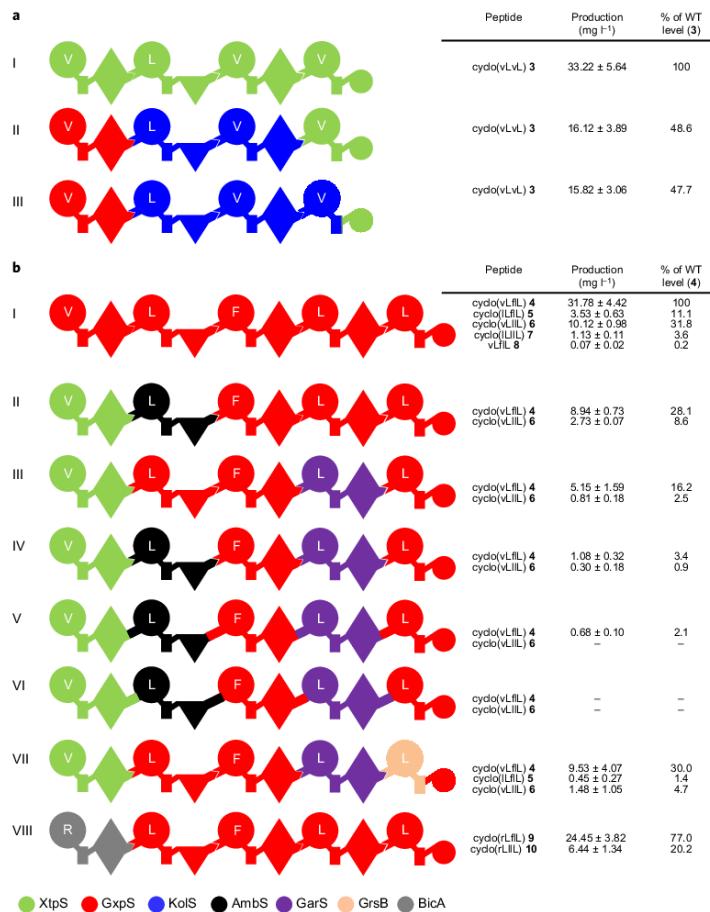
Figure 3 | *De novo* design of XtpS and GxpS NRPSs for xenotetrapeptide and GameXPeptide production. To reveal the general applicability of the XU concept and to verify the benefit of the identified fusion point, several homologue and non-homologue domain alterations were performed. **a**, Redesign of the WT NRPS XtpS (I). Recombined XtpS derivatives (II and III) and corresponding relative peptide levels compared with the level of the original XtpS (I). Production titres revealed that T–TE didomains can be separated and recombined. **b**, Step-by-step conversion of GxpS (I). Recombined GxpS derivatives (II–VIII) and corresponding amounts of GameXPeptide derivatives **4–10** as determined in triplicate. Not respecting the novel fusion point (Fig. 1b) results in a severe drop in peptide titres (V and VI). The colour code of the NRPSs used as building blocks is depicted at the bottom of the figure (Supplementary Fig. 6 gives details). The domain assignment is as in Fig. 1.

V) or exclusively from the C domain (Fig. 3b VI) resulted in an even lower production titre or no peptide at all, respectively.

One possible explanation for the reduced production titres might be the non-natural C-A interface that is made of hydrophobic and ionic interactions. As no structural data for a C-A interface from a *Photorhabdus* NRPS is available, a sequence-based comparison of the original and newly generated artificial C-A interface regarding the amino acid identified in the SrfA-C–C-A interface[22] was performed (Supplementary Fig. 10). In all cases, a drop in identity compared with the original interface was observed that somewhat correlated with the drop in the production titre, especially when multiple XUs were assembled (Supplementary Fig. 11).

GameXPeptides A–C (**4–6**) were also produced when XU5 was exchanged against XU4 from the gramicidin-producing NRPS GrsB from *Bacillus subtilis* (Fig. 3b VII), as expected from the use of the nonspecific modules 1 and 3 in GxpS (ref. 31). Similarly, the formal exchange of the nonspecific XU1 from GxpS (Val/Leu) against the Val-specific XU1 from XtpS led to the exclusive production of **4** and **6** without the production of **5** and **7** (Fig. 3b II–IV), which indicates that the XU concept can also be used to increase product specificity and reduce side products. Additionally, two Arg-containing GameXPeptide derivatives (**9** and **10**) were produced (Supplementary Fig. 9b) from the formal exchange of XU1 against XU1 from the bicornutin-producing
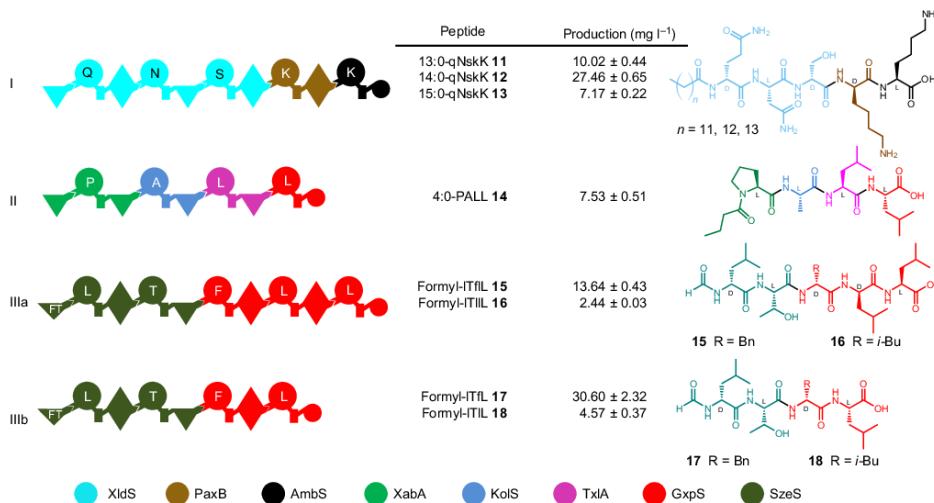
Figure 4 | Production of novel peptides. Combining XUs from different organisms resulted in the production of eight new peptides and proved the general applicability of XUs. I–III generated peptides **11**–**18** and their corresponding amounts are given as determined in triplicate. The colour code of the NRPSs used as building blocks is depicted at the bottom of the figure (Supplementary Fig. 6 gives details). The domain assignment is as in Fig. 1 plus FT (formyl transferase, N-terminal triangle).
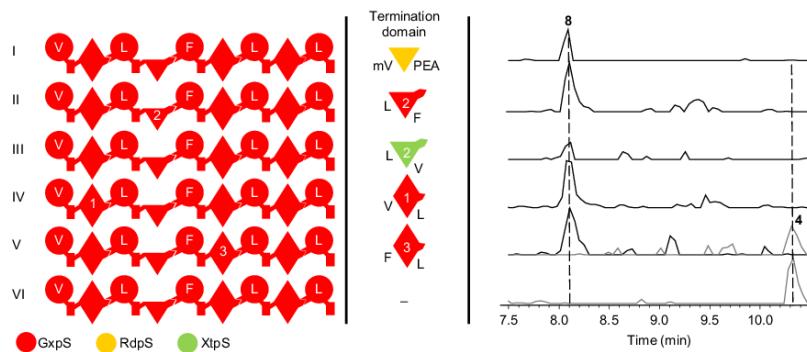
Figure 5 | Internal C domains as alternative termination domains. GxpS derivatives that show different covalently connected termination domains (I–V) or no termination domain (VI) (left) and HPLC/MS traces of **4** (m/z [M + H]$^+$ = 586.4) and **8** (m/z [M + H]$^+$ = 604.4) (right). GxpS with a C$_{term}$ from RdpC (I; mV (N-methyl valine), PEA (phenylethylamine)), an internal C domain from GxpS (II) or XtpS (III), internal C/E domains from GxpS (IV and V) and GxpS without a TE domain (VI). C-domain specificities and the origin of the used domains are indicated. Peptide amounts for **8** were detected in the range 0.09–0.40 mg l^{-1} and for **4** in the range 0.15–0.36 mg l^{-1} (all HPLC/MS traces have the same scale). The colour code of the NRPSs from *Photorhabdus* and *Xenorhabdus* used as building blocks is shown at the bottom (Supplementary Fig. 6 gives details). The domain assignment is as in Fig. 1.

NRPS BicA (ref. 34) (Fig. 3b VIII) and their structures were confirmed by chemical synthesis (Supplementary Fig. 12).

De novo construction of artificial NRPSs. Three additional and non-natural peptide types were generated (Fig. 4) using the XU concept by the formal fusion of 2–4 fragments from up to four different natural NRPSs that were used as the construction material (Supplementary Fig. 6). The expected lipopeptides **11**–**14** were produced in yields of 7–27 mg l^{-1} (Fig. 4 I and II) and were

structurally confirmed by chemical synthesis (Supplementary Figs 12–14). They differ only in the acyl moiety used as the starter unit and originating from the *E. coli* fatty-acid pool, as also observed in the original xenolindicins[21]. The formal fusion of XU1–XU2 from the szentiamide-producing NRPS SzeS and XU3–XU4–XU5 from GxpS resulted in the expected pentamodular NRPS that produced the formylated peptides **15** and **16** (Fig. 4 IIIa), only differing in Leu or Phe at position 3 from the relaxed substrate specificity of XU3 from GxpS (refs 30,31). However, most probably during the

ARTICLES
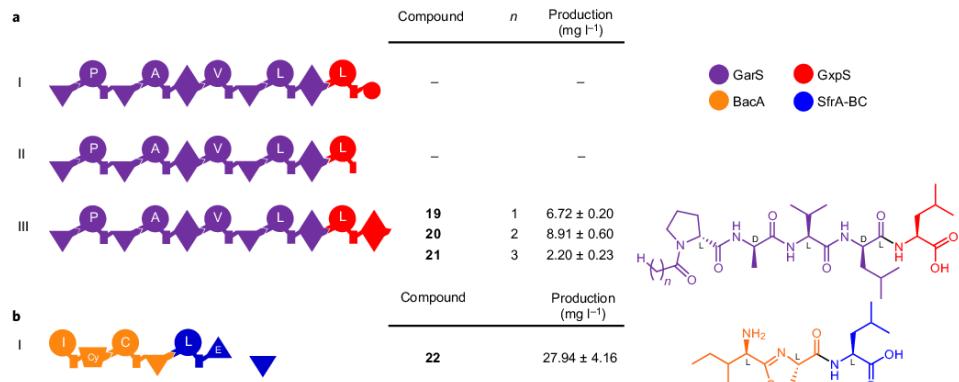
Figure 6 | Production of novel peptides by applying internal C domains as termination domains. a, Schematic representation of three tailor-made NRPSs with different peptide-releasing domains and structures of the lipopeptides **19**–**21**. I, TE domain; II, T domain; III, internal C/E domain from XU4 of GxpS. **b**, Schematic representation of a tailor-made NRPS designed from *Bacillus* XUs and the structure of the thiazoline-containing peptide **22** using XU6 from SrfA-BC. The colour code of the NRPSs used as building blocks is shown at the bottom (Supplementary Fig. 6 gives details). The domain assignment is as in Fig. 1 plus Cy (trapezium) and E (inverted triangle).

yeast assembly a tetramodular NRPS without XU4 was detected and produced the formylated peptides **17** and **18** (Fig. 4 IIIb) because of a high sequence similarity between C/E4 and C/E5 of GxpS, which allows an additional homologous recombination event and a subsequent deletion of XU4 (Supplementary Fig. 15), as confirmed by sequencing. The structures of **15** and **17** were confirmed by chemical synthesis (Supplementary Figs 16 and 17).

Avoiding TE-domain substrate specificity. The two limiting factors of the XU concept to produce novel peptides are the strict specificity of downstream C domains and often peptide-specific TE domains. To address the latter problem, the use of C or C/E domains as termination domains similar to fungal and some bacterial NRPSs[35] was tested, allowing the production of linear **8** but also cyclic **4** (Fig. 5 I–V) in very low yields of 0.09–0.4 mg l^{-1} compared with the TE-containing constructs (Fig. 3b). No epimerization of the terminal Leu was observed when a C/E domain was used as the termination domain (Fig. 5 IV and V). Even the complete loss of any termination domain resulted in the formation of **4** (Fig. 5 VI), which indicates some autocatalytic mechanism that has to be studied in more detail in the future. Although no general rules for the formation of cyclic or linear peptides using C or C/E domains can be concluded from these results, they show the applicability of internal C and C/E domains as termination domains, especially for the production of linear peptides that can be easily fused to the NRPS-encoding genes using the XU concept. Using the internal XU4 of GxpS (with a C/E domain) as terminal XU5 following XU1–XU4 from the GarS NRPS, the production of novel lipopeptides (**19**–**21**) was possible in good yield, which were not produced with XU5 from GxpS or without a termination domain (Fig. 6a and Supplementary Fig. 18). Again, the differences between **19**, **20** and **21** are only in the acyl moiety that is derived from the fatty-acid pool of *E. coli* (see also Fig. 4 I). The use of a C/E domain as the termination domain showed no epimerization of the terminal Leu, as described for the production of GameXPeptide (Fig. 5). Similarly, when an E and stand-alone C domain were used as the termination XU for a tripeptide that results from the formal fusion of XU1 and XU2 from the bacitracin-producing NRPS BacA with XU6 from the surfactin-producing NRPS SrfA-ABC

(Fig. 6b and Supplementary Fig. 19), no terminal epimerization was observed.

Conclusion

Despite intensive research during the past two decades and the increasing knowledge of NRPSs, it has been very difficult to swap domains and/or modules to result in active NRPSs, not to mention the construction of complete NRPSs *de novo*. As all NRPSs exhibit homologous structures and C-A domain linkers, the XU concept in combination with the knowledge about internal C domains as alternative termination domains should enable the production of several new cyclic or linear peptides. In principle, the number of possible peptides is only dependent on the identification of the suitable XU in a natural NRPS system and here hundreds of NRPS or NRPS/polyketide synthase hybrids with known substrate specificities have been described[36,37], with even more awaiting their identification in microbial genomes. Here we have shown the assembly of XUs from *Photorhabdus*, *Xenorhabdus*, *Bacillus* and combinations thereof. Although the XU concept seems to be universal, for its application and generation of modified peptides from other microorganisms, such as *Streptomyces*, *Pseudomonas* or myxobacteria, we suggest to use XUs from evolutionary-related organisms, as their underlying genes often have a much higher G-C content. As the current XU concept relies on a homologous assembly in yeast it might also allow the production of unexpected peptides through unpredicted homologous recombinations (Fig. 4 III) and thus in an additional layer of peptide diversification. With the potential automation of the cloning and heterologous expression steps being part of the XU concept, and especially with the co-expression of biosynthesis pathways for unusual amino acids, the biotechnological production of peptides might be more economical and faster than chemical synthesis once a large XU database is available. For compounds **11**, **14**, **15**, **17** and **22**, we have tested the number of *E. coli* clones that produce the desired compounds and could detect between 6.7 and 100% positive clones (Supplementary Materials 3) using our assembly pipeline (Fig. 2). At worst, 15 clones had to be analysed to find the expected producer. Although the production titres are often still decreased compared with the WT level, the XU concept shows good production titres even when multiple parts were assembled (Fig. 3),

Attachments

and when new peptides were generated (Figs 4 and 6). Furthermore, internal C or C/E domains can be used for peptide release or cyclization (Figs 5 and 6), which avoids the specificity of the TE domains. Our data suggest that, even without further optimization, the amount of most peptides is high enough to perform (automated) bioactivity assays with crude extracts and result in the potential identification of novel lead compounds in the future. One can probably further increase the titres using random mutagenesis or directed evolution via error-prone PCR, as previously described[38], and combine this with classical strain optimization, often applied to industrial natural product producers. Moreover, the XU concept might also be suitable to increase production specificity, as shown for the GameXPeptides (Fig. 3b), and thus to minimize the formation of undesired side products that might be responsible for side effects and thus need to be removed during the production process, which requires time and infrastructure.

Data availability statement. All data reported here, including all HPLC-MS data, alignments, materials and methods, sequences for oligonucleotides, plasmids and strains generated are described in the Supplementary Information and also available from the corresponding author on request.

References

1. Felnagle, E. A. *et al.* Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211 (2008).
2. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
3. Walsh, C. T. The chemical versatility of natural-product assembly lines. *Acc. Chem. Res.* **41**, 4–10 (2008).
4. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24**, 735–715 (2007).
5. Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140 (2016).
6. Cai, X. *et al.* Biosynthesis of the antibiotic nematophin and its elongated derivatives in entomopathogenic bacteria. *Org. Lett.* **19**, 806–809 (2017).
7. Cai, X. *et al.* Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nat. Chem.* **9**, 379–386 (2017).
8. Gao, X. *et al.* Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.* **8**, 823–830 (2012).
9. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347 (2016).
10. Weist, S. & Süssmuth, R. D. Mutational biosynthesis—a tool for the generation of structural diversity in the biosynthesis of antibiotics. *Appl. Microbiol. Biotechnol.* **68**, 141–150 (2005).
11. 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).
12. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
13. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
14. Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–648 (2015).
15. Mootz, H. D. *et al.* Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **124**, 10980–10981 (2002).
16. Butz, D. *et al.* Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem* **9**, 1195–1200 (2008).
17. Kries, H. Biosynthetic engineering of nonribosomal peptide synthetases. *J. Pept. Sci.* **22**, 564–570 (2016).
18. 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).
19. 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).
20. Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A. & Gulick, A. M. Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* **82**, 2691–2702 (2014).
21. 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).
22. 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).
23. Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188–198 (2012).
24. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
25. Liu, Y., Zheng, T. & Bruner, S. D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–1488 (2011).
26. Gulick, A. M. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827 (2009).
27. Tan, X.-F. *et al.* Structure of the adenylation-peptidyl carrier protein didomain of the *Microcystis aeruginosa* microcystin synthetase McyG. *Acta. Crystallogr. D* **71**, 873–881 (2015).
28. Strieker, M., Tanovic, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240 (2010).
29. Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P. J. L. M. & Marahiel, M. A. Construction of hybrid peptide synthetases for the production of α-L-aspartyl-L-phenylalanine, a precursor for the high-intensity sweetener aspartame. *Eur. J. Biochem.* **270**, 4555–4563 (2003).
30. 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).
31. 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).
32. 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).
33. Kegler, C. *et al.* Rapid determination of the amino acid configuration of xenotetrapeptide. *ChemBioChem* **15**, 826–828 (2014).
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. Haynes, S. W., Ames, B. D., Gao, X., Tang, Y. & Walsh, C. T. Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites. *Biochemistry* **50**, 5668–5679 (2011).
36. Flissi, A. *et al.* Norine, the knowledge-base dedicated to non-ribosomal peptides, is now open to crowdsourcing. *Nucleic Acids Res.* **44**, D1113–D1118 (2016).
37. Medema, M. H. *et al.* Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
38. Fischbach, M. A., Lai, J. R., Roche, E. D., Walsh, C. T. & Liu, D. R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl Acad. Sci. USA* **104**, 11951–11956 (2007).

Author contributions

K.A.J.B. and H.B.B. conceived and designed the experiments. K.A.J.B., F.F., A.L. and A.T. performed the experiments and analysed the data together with H.B.B. and C.-P.N. F.W. performed the chemical synthesis of all of the peptides. K.A.J.B. and H.B.B. wrote the paper. All of the authors discussed the results and commented on the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to H.B.B.

Competing financial interests

The authors declare no competing financial interests.

76

6.1.3 Supplementary information

De novo design and engineering of non-ribosomal peptide synthetases

Supplementary data

Kenan A. J. Bozhüyük, Florian Fleischhacker, Annabell Linck, Frank Wesche, Andreas Tietze, Claus-Peter Niesert, Helge B. Bode[†]

[†]   Kenan A. J. Bozhüyük, Florian Fleischhacker, Annabell Linck, Andreas Tietze, Frank Wesche, Helge B. Bode

Merck Stiftungsprofessur für Molekulare Biotechnologie

Fachbereich Biowissenschaften, Goethe Universität Frankfurt

Max-von-Laue-Str. 9, 60438 Frankfurt am Main, 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

Dr. Claus-Peter Niesert, Performance Materials/Process Technologies, Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, German.

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

Attachments

2

78

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, 10g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl). Solid media contained 1% (w/v) agar. *S. cerevisiae* strain CEN.PK 113-7D 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 2% (w/v) agar. Ampicillin (100 μg/ml), kanamycin (50 μg/ml) and G418 (200 μg/ml) were used as selection markers. All strains were cultivated at 30°C.

2. Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus* and *Photorhabdus* strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. Polymerase chain reaction (PCR) was performed with oligonucleotides obtained from Sigma-Aldrich and Eurofins Genomics (Tab. 4). Fragments for yeast homologues recombination were amplified with homology arms (40 – 80 bp) in a two-step PCR program using Phire Hot Start II DNA polymerase (Thermo Scientific) and for all other applications Phusion Hot Start II High-Fidelity DNA polymerase (Thermo Scientific) was used. Both polymerases were used according to the manufacturers' instructions. DNA purification was performed using MinElute PCR Purification Kit (Qiagen). Plasmid isolation from *E. coli* was done by alkaline lysis.

3. Overlap extension PCR-yeast homologous recombination (ExRec)

Transformation of yeast cells was done according to the protocols from Gietz and Schiestl[1,2]. 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. The DNA insert encoding the shortened construct

4

responsible for the production of compounds **17** and **18** (Fig. 4 III b) was sequenced in order to identify the missing sequence (see Supplementary Figure 15).

For compound **11** 25 *E. coli* clones were analyzed by HPLC/MS in 5 pools of 5 clones each. While all pools showed the production of **11**, a detailed analysis of one pool showed that 4 of these 5 clones indeed produced **11** (cloning efficiency 80%). For compound **14** 4 from 5x5 clones showed production, with one out of five showing production upon detailed analysis (20% cloning efficiency). For compounds **15** and **17** one of 15 clones tested showed production (6.7 % cloning efficiency). For compound **22** 5 from 5x5 clone pools showed production with all five cloned being confirmed producers (100% cloning efficiency).

4. 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 containing 50 µg/mL kanamycin or 100 µg/mL ampicillin, respectively. 100 µl of an overnight culture were used for inoculation of 10 ml cultures, containing 0.5 mg/ml L-arabinose or 2 mM IPTG, and 2 % (v/v) XAD-16. 50 µg/mL kanamycin or 100 ampicillin µg/mL were used as selection markers. After incubation for 48-72 h at 22°C and 30°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[3,4]. 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 a Dionex Ultimate 3000 RSLC Coupled to a Bruker micro-TOF-Q II equipped with an ESI Source set to positive ionization mode. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

5

82

5. Peptide quantification

All peptides were quantified using a calibration curve of synthetic **1** (for quantification of **1-2**), **3**, **4** (for quantification of **4-7**), **8**, **9** (for quantification of **9** and **10**), **12** (for quantification of **11-13**), **14**, **15** (for quantification of **15** and **16**), **17** (for quantification of **17** and **18**), **19** (for quantification of **19-21**) and **22** using HPLC/MS measurements as described above. Triplicates of all experiments were measured.

6. Chemical synthesis

6.1 Synthesis of linear peptides

The linear sequences were synthesized on preloaded resins (H-AA-2ClTrt PS resin, Sigma Aldrich, Germany) on a 25 μM scale with a Syro Wave peptide synthesizer (Biotage, Sweden) by using standard Fmoc/*t*-Bu chemistry. Fmoc-amino acids were purchased from Carbolution Chemicals (Germany), Iris Biotech (Germany) or Bachem (Switz). Therefore, the resin was placed in a plastic reactor vessel with a Teflon frit and an amount of 6 eq. of amino acid derivative (c = 0.2 M) was activated *in situ* at room temperature with 6 eq. of *O*-(6-chlorobenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HCTU, Carl Roth, Germany, c = 0.6 M) in dimethylformamide (DMF, Carl Roth, Germany) in the presence of 12 eq. *N,N*-diisopropylethylamine (DIPEA, Iris Biotech, c = 2.4 M) in *N*-methylpyrrolidone (NMP, Iris Biotech) for 50 min. Fmoc-protecting groups were removed with a solution of 40 % piperidine (Iris Biotech) in NMP (v/v %) for 5 min and followed by a second deprotection step with 20 % piperidine in NMP (v/v %) for 10 min. After each coupling and deprotection step, the resin was washed with NMP (4 ×). After addition of the final amino acid and deprotection step, the resin was washed with NMP (5 ×), DMF (5 ×) and DCM (5 ×).

6.2 Formylation

Formylation of the free N-terminus was performed as previously described by Nollmann *et al.*[5]. Therefor the resin was treated with 5 eq. 4-nitrophenylformate (Sigma Aldrich) and 3 eq. N-methylmorpholine (Sigma Aldrich) at 4 °C (c = 12.5 μmol/mL) overnight. Afterwards the resin was washed with NMP (5 ×), DMF (5 ×) and DCM (5 ×).

6.3 Acylation

The free N-terminus was acylated by treating the peptidyl resin with 10 eq. fatty acid, 10 eq. O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU, Carbolution Chemicals) and 20 eq. DIPEA in DMF (c = 40 μmol/mL) overnight. Afterwards the resin was washed with NMP (5 ×), DMF (5 ×) and DCM (5 ×).

6.4 Cyclization

For subsequent cyclization, the protected peptide was cleaved with 20 % hexafluoroisopropanol (HFIP, Carbolution Chemicals) for 1 h. The resin was removed by filtration and the cleavage cocktail was evaporated. The residue was dissolved in DMF (c = 1 mM) and cyclized in solution (micro wave, Discover CEM system, 25 W, 75 °C, 20 min) using 2 eq. HATU, 2 eq. 1-hydroxy-7-azabenzotriazole (HOAt, Carbolution Chemicals) and 4 eq. DIPEA. After evaporation under reduced pressure, the residue was dissolved in DCM and the solution was washed with saturated $NaHCO_3$. The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*.

6.5 Synthesis of "short bacitracin"

2-[1′(S)-(tert-Butyloxycarbonylamino)-2′(R)-methylbutyl]-4(R)-carboxy-Δ^2-thiazoline was synthesized according to S. Nozaki, I. Muramatsu[6] and J. Lee *et al.*[7] Afterwards 2.2 eq. thiazoline derivative were coupled to 50 μmol H-DLEU-2ClTrt PS resin using 2 eq. HATU,

7

2 eq. HOAt and 4 eq. DIPEA in DMF (c = 50 µmol/mL) overnight. The resin was washed with DMF (5 ×) and DCM (5 ×).

6.6 Cleavage or total deprotection

For total deprotection or cleavage 0.5 mL 95 % trifluoroacetic acid (TFA, Iris Biotech) and 2.5 % triisopropylsilane (TIS, Sigma Aldrich) in water were added to the cyclized peptide or peptidyl resin and the mixture was agitated for at least 1 h at room temperature. The resin was removed by filtration and washed twice with TFA. Then the cleavage cocktail was evaporated. Either the cyclized or linear peptide was dissolved in MeOH in order to purify it by semi-preparative HPLC–MS (Waters Purification TM System, Waters Corporation, USA; Jupiter Proteo, Phenomenex, Germany). The purity was determined by RP-UPLC coupled with ESI–MS.

8

Supplementary Tables

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

Compound	MS detected $[M+\mathrm{H}]^+$ (* = $[M+2\mathrm{H}]^+$)	MS calculated $[M+\mathrm{H}]^+$ (* = $[M+2\mathrm{H}]^+$)	Molecular formular	Δppm	Reference
1	751.4124	751.4137	$C_{38}H_{54}N_8O_8$	1.8	8
2	751.4119	751.4137	$C_{38}H_{54}N_8O_8$	2.5	
3	411.2961	411.2965	$C_{21}H_{38}N_4O_4$	1.1	9
4	586.3952	586.3962	$C_{32}H_{51}N_5O_5$	1.9	10
5	600.4103	600.4119	$C_{33}H_{53}N_5O_5$	2.7	10
6	552.4106	552.4119	$C_{29}H_{53}N_5O_5$	2.4	10
7	566.4259	566.4275	$C_{30}H_{55}N_5O_5$	3.0	10
8	604.4054	604.4069	$C_{32}H_{53}N_5O_6$	2.5	
9	643.4285	643.4289	$C_{33}H_{54}N_8O_5$	0.8	
10	609.4441	609.4446	$C_{30}H_{56}N_8O_5$	0.9	
11	400.7661*	400.7656*	$C_{37}H_{69}N_9O_{10}$	-1.0	
12	407.7736*	407.7734*	$C_{38}H_{71}N_9O_{10}$	-0.9	
13	414.7817*	414.7812*	$C_{39}H_{73}N_9O_{10}$	-1.0	
14	483.3170	484.3255	$C_{24}H_{43}N_4O_6$	1.5	
15	634.3792	634.3810	$C_{32}H_{51}N_5O_8$	1.8	
16	600.3947	600.3966	$C_{29}H_{53}N_5O_8$	2.0	
17	521.2962	521.2969	$C_{26}H_{40}N_4O_7$	1.4	
18	487.3122	487.3126	$C_{23}H_{42}N_4O_7$	0.9	
19	554.3539	554.3548	$C_{27}H_{47}N_5O_7$	1.7	
20	568.3694	568.3704	$C_{28}H_{49}N_5O_7$	2.0	
21	582.3846	582.3861	$C_{29}H_{51}N_5O_7$	2.6	
22	330.1845	330.1845	$C_{15}H_{27}N_3O_3S$	0.2	

9

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

Strain	Genotype	Reference
E. coli DH10B	F_ mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lac*Z Δ, M15, Δ *lac*X74 *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 Δ entD	8
S. cerevisiae CEN. PK 113-7D	MATα, *MAL2-8ᶜ*, SUC2	Euroscarf
P. asymbiotica ATCC 43949		ATCC
P. luminescens TTO1		DSMZ
X. bovienii SS2004		DSMZ
X. budapestensis DSM 16342		DSMZ
X. doucetiae DSM 17909		DSMZ
X. indica DSM 17382		DSMZ
X. miraniensis DSM 17902		DSMZ
X. nematophila ATCC 19061		ATCC
X. stockiae DSM 17904		DSMZ
X. szentirmaii DSM 16338		DSMZ
B. brevis ATCC 999		M. A. Marahiel / ATCC
B. licheniformis ATCC 10716		M. A. Marahiel / ATCC
B. subtilis ATCC 21332		M. A. Marahiel / ATCC
E. coli DH10B::mtaA pFF62A_*ambS*	*E. coli* DH10B::mtaA pFF62A_*ambS*, Amp^R	8
E. coli DH10B::mtaA pFF62A_*ambS*_gxpS-A4T4	*E. coli* DH10B::mtaA pFF62A_*ambS*_gxpS_A4T4, Amp^R	this work
E. coli DH10B::mtaA pFF62A_*ambS*_gxpS-A3T3C4	*E. coli* DH10B::mtaA pFF62A_*ambS*_gxpS_A3T3C4, Amp^R	this work
E. coli DH10B::mtaA pFF62A_*ambS*_kolS_C2A2T2	*E. coli* DH10B::mtaA pFF62A_*ambS*_kolS_C2A2T2, Amp^R	this work
E. coli DH10B::mtaA pFF62A_*ambS*_kolS_A2T2C3	*E. coli* DH10B::mtaA pFF62A_*ambS*_kolS_A2T2C3, Amp^R	this work
E. coli DH10B::mtaA pFF62A_*ambS*_xabB_C3A3T3	*E. coli* DH10B::mtaA pFF62A_*ambS*_xabB_C3A3T3, Amp^R	this work
E. coli DH10B::mtaA pFF1_*ambS*_gxpS_A3T3C4_A4T4C5	pFF62A_*ambS*_gxpS_A3T3C4A4T4C5, Kan^R	this work
E. coli DH10B::mtaA pFF1_*ambS*_gxpS_A3T3C4_A2T2C3	pFF62A_*ambS*_gxpS_A3T3C4A2T2C3, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*	*E. coli* DH10B::mtaA pFF1_*gxpS-de-novo*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS-variant_1*	*E. coli* DH10B::mtaA pFF1_*gxpS-variant_1*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS-variant_2*	*E. coli* DH10B::mtaA pFF1_*gxpS-variant_2*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS-*	*E. coli* DH10B::mtaA pFF1_*gxpS-variant_3*,	this work

10

86

variant_3	Kan^R	
E. coli DH10B::mtaA pFF1_*gxpS-variant_4*	*E. coli* DH10B::mtaA pFF1_ *gxpS-variant_4*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS-variant_5*	*E. coli* DH10B::mtaA pFF1_ *gxpS-variant_5*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_bicA-A1T2C2	*E. coli* DH10B::mtaA pFF1_ *gxpS*_bicA-A1T2C2, Kan^R	this work
E. coli DH10B::mtaA pFF1_*xtpS*	*E. coli* DH10B::mtaA pFF1_*xtpS*, Kan^R	this work
E. coli DH10B::mtaA pFF1_xtpS-*de-novo_1*	*E. coli* DH10B::mtaA pFF1_*xtpS-de-novo*, Kan^R	this work
E. coli DH10B::mtaA pFF1_xtpS-*de-novo_2*	*E. coli* DH10B::mtaA pFF1_*xtpS-de-novo*, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_grsB_A4T4	*E. coli* DH10B::mtaA pFF1_*gxpS*_grsB_A4T4, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_C$_{term}$	*E. coli* DH10B::mtaA pFF1_*gxpS*_C$_{term}$, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_C2$_{int}$	*E. coli* DH10B::mtaA pFF1_*gxpS*_C2$_{int}$, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_C/E1$_{int}$	*E. coli* DH10B::mtaA pFF1_*gxpS*_C/E1$_{int}$, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_C/E3$_{int}$	*E. coli* DH10B::mtaA pFF1_*gxpS*_C/E3$_{int}$, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS*_T	*E. coli* DH10B::mtaA pFF1_*gxpS*_T, Kan^R	this work
E. coli DH10B::mtaA pFF1_*garS_gxpS*_T	*E. coli* DH10B::mtaA pFF1_*garS_gxpS*_T, Kan^R	this work
E. coli DH10B::mtaA pFF1_*garS_gxpS*_TE	*E. coli* DH10B::mtaA pFF1_*garS_gxpS*_mit_TE, Kan^R	this work
E. coli DH10B::mtaA pFF1_*garS_gxpS*_C/E3$_{int}$	*E. coli* DH10B::mtaA pFF1_*garS_gxpS*_C/E3$_{int}$, Kan^R	this work
E. coli DH10B::mtaA pFF1_*gxpS_xcn1*_C2$_{int}$	*E. coli* DH10B::mtaA pFF1_*gxpS_xcn1*_C2, Kan^R	this work
E. coli DH10B::mtaA pFF1_2A_*xldS_paxB_ambS*	*E. coli* DH10B::mtaA pFF1_2A_*xldS_paxB_ambS*, Kan^R	this work
E. coli DH10B::mtaA pFF1_13A_*xabABC_kolS_txlA_gxpS*	*E. coli* DH10B::mtaA pFF1_13A_*xabABC_kolS_txlA_gxpS*, Kan^R	this work
E. coli DH10B::mtaA pFF1_22A_*szeS_gxpS*	*E. coli* DH10B::mtaA pFF1_22A_*szeS_gxpS*, Kan^R	this work
E. coli DH10B::mtaA pFF1_1B_*bacA_srfA*-BC	*E. coli* DH10B::mtaA pFF1_1B_*bacA_srfA*-BC, Kan^R	this work

11

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

Plasmid	Genotype	Reference
pFF1_Ypet	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	this work
pFF62A	2μ ori, kanMX4, T7lac promoter, MCS, pBR322 ori, Amp^R	8
pFF62A_*ambS*	2μ ori, kanMX4, T7lac promoter, *ambS*, pBR322 ori, Amp^R	8
pFF62A_*ambS*_gxpS-A4T4	2μ ori, kanMX4, T7lac promoter, *ambS*_C1A1T1C2A2T2C3-*gxpS*_A4T4-*ambS*_C4A4T4C5A5T5C6A6T6TE, pBR322 ori, Amp^R	this work
pFF62A_*ambS*_gxpS_A3T3C4	2μ ori, kanMX4, T7lac promoter, *ambS*_C1A1T1C2A2T2C3-*gxpS*_A3T3C4-*ambS*_A4T4C5A5T5C6A6T6TE, pBR322 ori, Amp^R	this work
pFF62A_*ambS*_kolS_C2A2T2	2μ ori, kanMX4, T7lac promoter, *ambS*_C1A1T1C2A2T2C3A3T3-*kolS*_C2A2T2-*ambS*_C5A5T5C6A6T6TE, pBR322 ori, Amp^R	this work
pFF62A_*ambS*_kolS_A2T2C3	2μ ori, kanMX4, T7lac promoter, *ambS*_C1A1T1C2A2T2C3A3T3C4-*kolS*_A2T2C3-*ambS*_A5T5C6A6T6TE, pBR322 ori, Amp^R	this work
pFF62A_*ambS*_xabB_C3A3T3	2μ ori, kanMX4, T7lac promoter, *ambS*_C1A1T1C2A2T2C3A3T3-*xabB*_C3A3T3-*ambS*_C5A5T5C6A6T6TE, pBR322 ori, Amp^R	this work
pFF1_*ambS*_gxpS_A3T3C4_A4T4C5	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *ambS*_C1A1T1C2A2T2C3-*gxpS*_A3T3C4A4T4C5-*ambS*_A5T5C6A6T6TE	this work
pFF1_*ambS*_gxpS_A3T3C4_A2T2C3	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *ambS*_C1A1T1C2A2T2C3--*gxpS*_A3T3C4_A2T2C3-*ambS*_A5T5C6A6T6TE	this work
pFF1_*gxpS*	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*	this work
pFF1_gxpS-*variant*-1	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*_A1T1C2-*ambS*_A4T4C5-*gxpS*_A3T3C4A4T4C5A5T5TE	this work
pFF1_gxpS-*variant*-2	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*_A1T1C2-*gxpS*_A2T2C3A3T3C4-*garS*_A4T4C5-*gxpS*_A5T5TE	this work
pFF1_gxpS-*variant*-3	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*_A1T1C2-*ambS*_A4T4C5-*gxpS*_A3T3C4-*garS*_A4T4C5-*gxpS*_A5T5TE	this work
pFF1_gxpS-*variant*-4	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-	this work

12

88

	Flag, Kan^R, *xtpS*_A1T1C2-*ambS*_A4T4C5-*gxpS*_A3T3C4-*garS*_A4T4C5-*gxpS*_A5T5TE	
pFF1_*gxpS*-*variant*-5	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*_A1T1C2-*ambS*_A4T4C5-*gxpS*_A3T3C4-*garS*_A4T4C5-*gxpS*_A5T5TE	this work
pFF1_*gxpS*_bicA-A1T2C2	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bicA*_A1T1C2-*gxpS*_A2T2C3A3T3C4A4T4C5A5T5TE	this work
pFF1_xtpS	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*	this work
pFF1_xtpS-*de-novo*-1	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1C2-*kolS*_A14T14C15A15T15C16-*xtpS*_A4T4TE	this work
pFF1_xtpS-*de-novo*-2	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1C2-*kolS*_A14T14C15A15T15C16A16T16-*xtpS*_TE	this work
pFF1_*gxpS*_*grsB*	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T1C2A2T2C3A3T3C4A4T4C5-*grsB*_A4T4-*gxpS*_TE	this work
pFF1_*gxpS*_C_{term}	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *rdpC* (from base 4597 to 5997) was inserted downstream of *gxpS* (from base 1 to 14625)	this work
pFF1_*gxpS*_C2_{int}	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* from *P. luminescens* ATCC 43949 (from base 5177 to 6637) was inserted downstream of *gxpS* from *P. asymbiotica* ATCC 43949 (from base 1 to14625)	this work
pFF1_*gxpS*_C/E1_{int}	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* from *P. asymbiotica* ATCC 43949 (from base 1831 to 3336) was inserted downstream of *gxpS* from *P. asymbiotica* ATCC 43949 (from base 1 to 14625)	this work
pFF1_*gxpS*_C/E3_{int}	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *rdpC* from *P. asymbiotica* ATCC 43949 (from base 8200 to 9705) was inserted downstream of *gxpS from P. asymbiotica* ATCC 43949 (from base 1 to 14625)	this work
pFF1_*gxpS*_T	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* from *P. asymbiotica* ATCC 43949 (from base 1 to 14784)	this work
pFF1_*garS*_*gxpS*_T	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* from *P. luminescens* TTO1 (from base 13034 to 14944) was inserted downstream of *garS* (from base 1 to 14241)	this work

13

pFF1_*garS_gxpS*_TE	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* from *P. luminescens* TTO1 (from base 13034 to 15699) was inserted downstream of *garS* (from base 1 to 14241)	this work
pFF1_*garS_gxpS*_C/E3$_{int}$	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS* (nt) from *P. luminescens* TTO1 (from base 13034 to 14785) followed by *gxpS from P. asymbiotica* ATCC 43949 (from base 8200 to 9705) was inserted downstream of *garS* (from base 1 to 14241)	this work
pFF1_*gxpS_xcn1*_C2$_{int}$	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *rdpC* from *P. asymbiotica* ATCC 43949 (from base 5119 to 6567) was inserted downstream of *gxpS from P. asymbiotica* ATCC 43949 (from base 1 to 14625)	this work
pFF1_2A_*xldS_paxB_ambS*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *xldS*_C1A1T1C/E2A2T2C3A3T3C/E4-*paxB*_A2T2C3-*ambS*_A6T6TE	this work
pFF1_13A_*xabABC_kolS_txlA_gxpS*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *xabABC*_C1A1T1C2-*kolS*_A2T2C3-*txlA*_A3T3C4-*gxpS*_A5T5TE	this work
pFF1_22A_*szeS_gxpS*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *szeS*_FtA1T1C/E2A2T2C3-*gxpS*_A3T3C/E4A4T4C/E5A5T5TE	this work
pFF1_1B_*bacA_srfA-BC*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *srfA_BC* from *B. subtilis* ATCC 21332 (*srfA-B*: from base 7588 to 10752; *srfA-C* from base 1 to 1410) was inserted downstream of *bacA from B. licheniformis* ATCC 10716 (from base 1 to 6339)	this work

14

90

Supplementary Table 4. Oligonucleotides used in this work. Correlations to figures from the main text corresponding to the plasmids are represented in brackets.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pFF62A_ambS_kolS_C2A2T2 (Fig. 1c - II)	pFF62A_ambS F1 rev	TGTCGATAAACTGCTTTCTCACATTTGGTTGGCAGAGTCATCACCTTATCATTTTCATATGTGTTCCTGTGTGAAATTGTTATC	pFF62A_ambS
	pFF62A_ambS F3 for	AATAGAATTGTTTTAGCTCAAGTGCTCAATACGGCATTAGCAAGCAATGATTATACTTAGCCAATTCGCCGAACCGTA	
	ambS F1 for	ATGAAAAATGATAAGGTGATGACTCTG	X. miraniensis DSM 17902
	ambS F1 rev	AGTTTGTGCAAGCCCGGA	
	ambS F2 C1 for	TCTGTTCCAGACGCGCGGTATTGTCCGGGCTTGCACAAACTGTAGGGCAGTCACAGGCGGT	P. luminescens TTO1
	ambS F2 C1 rev	TCTGAGCCAGAAACCACAAACGTTGTTGAGCAAATGACAGCGGCAATTGGTGATCCTCAC	X. miraniensis DSM 17902
	ambS F3 for	CTGTCATTTGCTCAACAACG	
	ambS F4 rev	CTAAGTATAATCATTGCTTGCTAATGC	
pFF62A_ambS_xab_C3A3T3 (Fig. 1c - III)	FF_ambS F2 C5 rev	GGTATCCACCGATATTCTCAATCTGAGCCAGAAACCACAAACGTTGTTGAGCAAATGACAGAGGCAGGGCGGGTGACGGT	X. doucetiae DSM 17909
	FF_ambS F2 C5 for	CGCACGTTGGCCGTGCGCGATCTGTTCCAGACGCGCCGGTATTGTCCGGGCTTGCACAAACTATTAAAGGGGATATCGATAGCTTTATCG	
pFF62A_ambS_gxpS-A4T4 (Fig. 1c - IV)	KB_pFF62A_ambS_AT-P1	CAAGCGCTTGATGCGCTATTTAC	pFF62A_ambS
	KB_pFF62A_ambS_AT-P2	TGGTTTTCATGTATCAGCGCCGTGGCATCCGGGGTCTTTTCCACCTGCTGTTCAAACAGTTGATGAATGCACAACCGCTCAGGATACGATGTTGCTGTCC	
	KB_pFF62A_ambS_AT-P3	GAGCGGTTGTGCATTCATC	P. luminescens TTO1
	KB_pFF62A_ambS_AT-P4	TGAATATTGGCCAGCCCGCCCGGCACTTGCCGGATGATATGGTCAATCTCTGGCTGGGTAAGGTCAATTAAGGGCAACATTTCTGGCGTTAACACCGTGG,	
	KB_pFF62A_ambS_AT	ATGTTGCCCTTAATTGACC	pFF62A_ambS

Construct	Primer	Sequence	Template
	-P5	TATCGGTAGACGCTACTGGCG	
KB_pFF62A_ambS_AT -P6		CAAGCGCCTTGATGCGCTATTTAC	pFF62A_ambS
pFF62A_ambS_gxpS_A3T3C4 (Fig. 1c-V)	KB-ambS-Swap-P1	CTCAGAGGATGACAAAATATCGATAGCCGTGACAGGTGTTGAGGATCAGCGACCATCGCCTG CAAGACTGCGTGGAGATATC	
	KB-ambS-Swap-P2	CCTCAACAACCTGTCACGGC GGTTTCTCTACTTGCTGTTCAAACAACTGATGGATACATAACTGCTCAGGATAGGCTGTTTCAG	P. luminescens TTO1
	KB-ambS-Swap-P3		
	KB-ambS-Swap-P4 AT_19	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAAATGATAAGGTGATGACTCTG	pFF62A_ambS_gxpS _A3T3C4
pFF1_ambS_gxpS_A3T3C4_A2T2C3 (Fig. 1c - VI)	FF_216	AGTTTTTAACAACAATGTGCGTTC	
	AT_23	TCCGGCGTTTTTCCTGCCTGTAGTTCAAATACCTGATGAATACACAATGGGTCAGGATACGATG TTGCTGTCCCATTCCAATTTCCAGTAATAACTCCCGCTC	P. luminescens TT01
	AT_24	CGCTTGAGCAGGCGTCCGATATGCCGGTACAGCAGTTAGACATTCTGCCGGCAACTGAACGCA CATTGTTGTTAAAAACTTGGAACGCGACAGAAACC	
	AT_20	TGGAATGGGACAGCAACAT	pFF62A_ambS_gxpS _A3T3C4
	AT_21	AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCC AGCGGATCCGGCGCGCCCTAAGTATAATCATTGCTTGCTAATGCC	
pFF1_ambS_gxpS_A3T3C4_A4T4C5 (Fig. 1c - VII)	AT_19	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAAATGATAAGGTGATGACTCTG	pFF62A_ambS_gxpS _A3T3C4
	FF_216	AGTTTTTAACAACAATGTGCGTTC	
	AT_22	TCCGGCGTTTTTCCTGCCTGTAGTTCAAATACCTGATGAATACACAATGGGTCAGGATACGATG TTGCTGTCCCATTCCAGACTTTCAGTAACAGTTTCCTTTCTACTGC	P. luminescens TT01
	AT_25	CGCTTGAGCAGGCGTCCGATATGCCGGTACAGCAGTTAGACATTCTGCCGGCAACTGAACGCA	

Construct	Primer	Sequence	Template
	AT_20	CATTGTTGTTAAAAACTTGGAACGCCACTGAAACAG TGGAATGGGACAGCAACAT	pFF62A_ambS_gxpS _A3T3C4
	AT_21	AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGGAGCC AGCGGATCCGGCGCCTAAGTATAATCATTGCTTGCTAATGCC	
pFF62A_ambS_kolS _A2T2C3 (Fig. 1c - VIII)	pFF62A_ambS F1 rev	TGTCGATAAACTTGCTTTCTCACATTTGGTTGGCAGAGTCATCACCTTATCATTTTTCATATGTGT TTCCTGTGTGAAATTGTTATC	pFF62A_ambS
	pFF62A_ambS F3 for	AATAGAATTGTTTTAGCTCAAGTGCTCAATACGGCATTAGCAAGCAATGATTATACTTAGCCAAT TCGCCGAACCGTA	
	ambS F1 for	ATGAAAAATGATAAGGTGATGACTCTG	X. miraniensis DSM 17902
	ambS F1 rev	AGTTTGTGCAAGCCCGGA	
	ambS F2 C5 for	CGCACGTTGGCCGTGCGCGATCTGTTCCAGACGCCGGTATTGTCCGGGCTTGCACAAACTATT AAAGGGGATATCGATAGCTTTATCG	X. doucetiae DSM 17909
	ambS F2 C5 rev	GGTATCACCGATATTCTCAATCTGAGCCAGAAACCACAAACGTTGTTGAGCAAATGACAGAGGC AGGGGCGGGGTGACGGT	
	ambS F3 for	CTGTCATTTGCTCAACAACG	X. miraniensis DSM 17902
	ambS F4 rev	CTAAGTATAATCATTGCTTGCTAATGC	
pFF1_xtpS-de-novo- 1 (Fig. 3a - II)	KB-Pep1-P1	CGGATCCTACCTGACGCGTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	P. luminescens TTO1
	KB-Pep1-P2	TACACCAGCGCCGTAGCATCCGGGTTTTTCTCCACTTGTTCAAATAACTGATGGACACATA ACTGCTTAGGGTAGACGGTTTCTGTCGCGTTCCACGTTTC	
	KB-Pep1-P3	GTCTACCCTAAGCAGTTATGTGTCC	P. luminescens TTO1
	KB-Pep2-P4	GCTCTCCAGCAACAGCTTCTTCTCAG	
	KB-Pep2-P5	CGCTTGAACAGAAGTGGAGACACCGATACGGTTGCTGAATATCTTGCCGGAAGCTGAGAAGA AGCTGTTGCTGGAGAGCTGGAATGCCACTGAAACCGTATC	X. nematophila ATCC 19061
	KB-Pep1-P6	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG	

Attachments

Construct	Primer	Sequence	Organism
pFF1_xtpS-de-novo-2 (Fig. 3a - III)	KB-Pep1-P1	CGGAGCCAGCGGATCCCAGCGCCTCCACTTCGCAATTC	P. luminescens TTO1
	KB-Pep1-P2	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	
	KB-Pep1-P3	TACACCAGCGCCGTAGCATCCGGGTTTTCTCCACTTGTTGTTCAAATAACTGATGGACACATA ACTGCTTAGGGTAGACGGTTTCTGTCGCGGTTCCACGTTTC	P. luminescens TTO1
	KB-Pep4-P4	GTCTACCCTAAGCAGTTATGTGTCC TGCGCTATTTCTCGGTTGAGACAG	
	KB-Pep4-P5	ATCTGTTTCAATTCCCTGTTCTGTCGAACTGGCCGCGAAAATGACATCAGATAAGCTGTCTCAA CCGAGAAATAGCGCAGTACCGGTACGACCCGATGG	X. nematophila ATCC 19061
	KB-Pep1-P6	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGGCCAGAACCAGCAG	
pFF1_gxpS (Fig. 3b - I)	KB_pFF1-YPet_plu3263_P1	CGGAGCCAGCGGATCCCAGCGCCTCCACTTCGCAATTC	P. luminescens TTO1
	KB_pFF1-YPet_plu3263_P2	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGC	
	KB_pFF1-YPet_plu3263_P3	TATAATCAGGCAACATCGCG	P. luminescens TTO1
	KB_pFF1-YPet_plu3263_P4	CGAAGCGGATAGTGGACTGG	
	KB_pFF1-YPet_plu3263_P5	AGGCGTCAACGGAAAAGCATC GTGAACATCTGAGTGGAT	P. luminescens TTO1
	KB_pFF1-YPet_plu3263_P6	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGGCCAGAACCAGCAG	
pFF1_gxpS-variant-1 (Fig. 3b - II)	FF_029	CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAATTC CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT	
	FF_212	AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG GGTTTTCAGCAATAACGTGC	X. nematophila HGB081
	FF_213	CCCTTGAACTGGCACCGGAAACACCGGTACGGGCACTCAATATTTTGCCTGCATCAGAACGCA CGTTATTGCTGAAAACCTGGAATGCGACAGAGACC	X. miraniensis DSM 17902
	FF_214	TCCGGGAGTCTTCTCTATCTGTTGTTCAAACAACTGATGGACACATACCTGAGTAGGATACGGTT	

Construct	Primer	Sequence (5'→3')	Template organism
	FF_215	CTTCGGTCGCATTCCACGTTTCCAGCAATAACGG	P. luminescens TT01
	FF_220	TGGAATGCGACCGAAGAAC	
		AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG	
		CGGAGCCAGCGGATCCTTACAGCGCCTCCGCTTCAC	
pFF1_gxpS-variant-2 (Fig. 3b - III)	FF_029	CGGATCCTACCTGACGCGTTTTATCGCAACTCTCTACTGTTCTCCATACCCGTTTTTTTGGGCT	X. nematophila HGB081
		AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG	
	FF_212	GGTTTTCAGCAATAACGTGC	
	FF_221	CCCTTGAACTGGCACCGGAAACACCGGTACGGGCACTCAATATTTTGCCTGCATCAGAACGCA	P. luminescens TT01
		CGTTATTGCTGAAAACCTGGAAACGCGACAGAAACC	
	FF_216	AGTTTTTAACAACAATGTGCGTTC	
	FF_217	CGCTTGAGCAGGCGGTCCGATATGCCGGTACAGCAGTTAGACATTCTGCCGGCAACTGAACGCA	X. bovienii SS2004
		CATTGTTGTTAAAAACTTTCAATGCTACCCAAGCCGA	
	FF_218	TCAGAATTTTTCGCCACTTGCTGTTCAAACAGATGAATACAACGAGCCGGGGATAGGCCG	
		TTTGCGGGCCATTGCCCTCTACCCAGCAGTTGTTGTCGC	
	FF_219	GGCAATGGCCCGCAAAC	P. luminescens TT01
	FF_220	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG	
		CGGAGCCAGCGGATCCTTACAGCGCCTCCGCTTCAC	
pFF1_gxpS-variant-3 (Fig. 3b - IV)	gxpS A1/A2 M1 fw YHR	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT	X. nematophila ATCC 19061
		AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG	
	GxpS A3/A4 F1A rv	GGTTTTCAGCAATAACGTGC	
	GxpS A3/A4 F2A fw	CCCTTGAACTGGCACCGGAAACACCGGTACGGGCACTCAATATTTTGCCTGCATCAGAACGCA	X. miraniensis DSM 17902
	GxpS A3/A4 F2A rv	CGTTATTGCTGAAAACCTGGAAACGCGACAGAGACC	
		TCCGGAGTCTTCTCTATCGTGTTGTTCAAACAACTGATGGACAACATACCTGAGTAGGATACGGTT	
	GxpS A3/A4 F3 fw	CTTCGGTCGCATTCCACGTTTCCAGCAATAACCGG	P. luminescens TT01
		TGGAATGCGACCGAAGAAC	
	GxpS A3/A4 F3 rv	AGTTTTTAACAACAATGTGCGTTC	
	GxpS A3/A4 F4 fw	CGCTTGAGCAGGCGGTCCGATATGCCGGTACAGCAGTTAGACATTCTGCCGGCAACTGAACGCA	X. bovienii SS-2004
		CATTGTTGTTAAAAACTTTCAATGCTACCCAAGCCGA	
	GxpS A3/A4 F4 rv	TCAGAATTTTTCGCCACTTGCTGTTCAAACAGATGATGAATACAACAGAGCCGGGGATAGGCCG	

19

	Primer	Sequence	Organism
	GxpS A3/A4 F5 fw	TTTGCGGGGCCATTGCCCTCTACCAGCAGTTGTTGTGCGGCAATGGCCCGCAAAC	*P. luminescens* TTO1
	GxpS A3/A4 F5 rv	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTACAGCGCCTCCGCTTCAC	
pFF1_gxpS-variant-4 (Fig. 3b - V)	FF_29	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG	*X. nematophila* ATCC 19061
	FF_30	CACGTTTTTAACAGCAGTGTCCGCTCAGTTTCCGGCAGAATTTCTAATGCCCTTATTGGCCGGTCCGGTGCCTGTTCCAGGGCATCGGCCAGACTGTC	
	FF_31	CTGCTCAGGTGGTGGTGGAACCGTTTGATTCAGAAAGAATATGCGGTTATATGCAGCAGGCGCTGGACAGTCTGGCCGATGCCCTGGAACAGGCACCGGAC	*X. miraniensis* DSM 17902
	FF_32	GCATTCCAATTTCCAGTAATAACTCCCGCTCAGAGGATGACAAAATATCGATAGCCGTGACAGGTTGTTGAGGCTGGTTGACCATCGCCTGCAAGACTG	
	FF_33	ATTATGCCACGGCGCTGTGTTTGATAAAACGGCCATTGAACGGCAGGTGGGGTACCTGCACGCAGTCTTGCAGGCGATGGTCAACCAGCCTCAACAACCTGTC	*P. luminescens* TTO1
	FF_34	TCGGTCAGCAACAGTTGCCGCTCCGAGTCCGGCAGCATCGACAGACCTGCAACAGTTTGTGTCGCATCGGCAGCCATCGCTGCCAACGCCTGTTGCATA	
	FF_35	CTGATCTAGGGCTGACGGCTCAAGTGGTGCAACCATTCGATCCAGAACGGATATGTGGCTATATGCAACAGGCGTTGGAGCAGCGATGGCTGCCGATGCG	*X. bovienii* SS-2004
	FF_36	CCGACTTTCAGTAACAGTTTCCTTTCTACTGCCGGTAAAATTTCCAACCGCTGAACCGGGGTTTCCGGGGCGTGTTCCAGGGCATCGACCAGGACCGCTGA	
	FF_37	CCGCTCAGGCCGTAACCGGAATCGTGCCGTCTCGCATAACCGCTTATCTGGTTACCGCCATCAGCGGTCTGGTCGATGCCGTGGAACACGCCCCGGAAAC	*P. luminescens* TTO1
	FF_38	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAATTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAAT	
pFF1_gxpS-variant-5 (Fig. 3b - VI)	FF_29	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG	*X. nematophila* ATCC 19061
	FF_39	GTTTGCCCTTCGTATATTAAAGCTGAGCATCCGGGGGTTTTCTCTACTTGCTGTTCAAACAACTGATGGATACATAACTGTTCAGGATAGGGGGTTTCAA	
	FF_40	GGGCACTCAATATTTTGCCTGCATCAGAGAACGCACGTTATTGCTGAAAACCTGAGCACGGTTGA	*X. miraniensis* DSM

Name	Sequence	Source
FF_41	AACCCCCTATCCTGAACAGTTATGTATCCATCAGTTGTTTG CTGAGGCGTCTGGTTTTCATAGATTACCGCTATGGCGTCCGGAGTCTTCTCTATCTGTTGTTCAA ACAACTGATGGACACACAATGGGTCAGGATACGATG	17902
FF_42	TGGCTATGCTGGCCCCGGCAGAGCGCCGGTTATTGCTGGAAACGTGAATGGGACAGCAACA TCGTATCCTGACCCATTGTGTCCATCAGTTGTTTGAACAAC	P. luminescens TTO1
FF_43	TAGCTGAGTGTCTGATCCTCAAACACCACGGCAGTGGCATCAGGGGGCATGTGCCGCTTGGGCT TCGAATAATTGGTGGATGCACAACCGCTCAGGATAGG	
FF_44	ACATTCTGCCGGCAACTGAACGCACATTGTTGTTAAAAACTTGGAACGCCACTGAAACAGCCTA TCCTGAGCGGTTGTGCATCCACCAATTATTCGAAGCC	X. bovienii SS-2004
FF_45	TAGCTGAGAGCCTGTTCTTCATAGACCAGTGCCGTAGCCTCAGAATTTTCGCCACTTGCTGTT CAAACAGATGATGAATTAGTGCATCCTGTGGGAAATC	
FF_46	CCATCTTACCGCCGACAGAGCGACAACAACTGCTGGTAGAGTTCAACGCCACCCAGACGGATT TCCCACAGGATGCACTAATTCATCATCTGTTTGAACAGCA	P. luminescens TTO1
FF_38	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAATTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCGCCTCCGCTTCACAAT	
pFF1_gxpS_grsB (Fig. 3b - VII)		
KB_pFF1-YPet_plu3263_P1	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGC	P. luminescens TTO1
KB_pFF1-YPet_plu3263_P2	TATAATCAGGCAACATCGCG	
KB_pFF1-YPet_plu3263_P3	CGAAGCGGATAGTGGACTGG	P. luminescens TTO1
KB_pFF1-YPet_plu3263_P4	AGGGCGTCAACGGAAAAGCATC	
KB_pFF1-YPet_plu3263_P5	GTGAACATCTGAGTGGAT	P. luminescens TTO1
KB-Gib-P8	GGGATAGGCCGTTTGCGGG	
KB-Gib-P3	TCAGATTCGCTGTCTCAATC	pFF1-Ypet
KB_pFF1-YPet_plu3263_P6	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCGCCTCCGCTTCACAATTC	

21

Construct	ID	Sequence	Species
	KB-P31	TTCAGCGGGTTGGAAATTTTACCGGCAGTAGAAAGGAAACTGTTACTGAAAGTCGGCAATGGCC CGCAAACGGCCTATCCCCAGAATCAAACAATACAGGAATTG	*B. brevis* ATCC 999
	KB-P32	GGCACGAAGAAGAGCGGCGGTTCGGTTCCATCCGGTCGCACCGGTATGGCGCGCTGCTTTGCGA TTGAGACAGCGAATCTGATGTAATATATTCCGCTATTTTCATAATAG	
pFF1_gxpS_bicA-A1T2C2 (Fig. 3b - VIII)	ML20	AGATTAGCGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTT TTGGGCTAACAGGAGGAGAATTCCATGAAAGATAAACATTGCTACAGTG	*X. budapestensis* DSM 16342
	ML18	TCTGGCGTTTTCTCCGCTTGTTGTTCAATCAACTGATGAATACAGCGTTGTTCAGGATAGGGGG TTTCTGTCGCCGTTCCAGGTGCTTAATAAACAGTGTACGTTC	
	ML19	TGGAACGGCGACAGAAACC	*P. luminescens* TT01
	ML10	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAACCCAGCGGATCCGGCGCGCCTTACAGCGCCTCCGCTTC	
pFF1_2A_xldS_paxB_ambS (Fig. 4 - I)	AT-2.1	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCTGAAACTTTGGAACTATAAAATGAATATGACACG	*X. indica* DSM 17382
	AT-2.2	TCAGGTGAGCGTAACACCTGTTGTTCGAAGAGTTGTTGCAGTAACAGATCCTGAGGGTAAGCG ACTTCTGTGGTGCTGAAGTTCACCAGTAGCTGCTGACGTTC	
	AT-2.3	TTCAGCACCACAGAAGTCGC	*X. stockiae* DSM 17904
	AT-2.4	TTCCGTCAGGAGGCTGTGTG	
	AT-2.5	CCATAGTGGCGGATGAAACACACAACAAGTCGCCGATTTAGCGCTGCTTACACCACAGCAACGCA CACAGCTCCTGACGAATGGAATCCGACAGCAACAG	*X. miraniensis* DSM 17902
	AT-2.7	AACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGCGGAGCCAGC GGATCCGGCGCGCCTTACAGATCCTGCGGGGTAAGCG	
pFF1_13A_xabABC_kolS_txlA_gxpS (Fig. 4 - II)	AL10-1	CTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAGAATTCCATGCCTATGTCATG CAATGGTATTAA	*X. doucetiae* DSM 17909
	AL13-2	ATCCACCAGCAGTTGTTGTCG	
	AL13-3	TCGCCTGCCGATGTTGTCAGCGTCGGAGCGACAACAACTGCTGGTGGATTGGAATGCAACCGC AACC	*P. luminescens* TTO1

23

	Primer	Sequence	Source
	AL13-4	GTTGCGAGGGTTTTGTCCTGCGGGTACGGTACGTCGGTCTGGTTCCAAGTTTCCAATAACAACTTGCGCTC	
	AL13-5	TGGAACCAGACCGACGTACC	X. bovienii SS-2004
	AL13-6	GCGGTGCAGCAGGGTATG	
	AL13-7	GGATTTGACGTATTGTCCGCGCAGGAGCGCCATACCCTGCTGCACCGCGCGGCAATGGCCCGCAAAC	P. luminescens TTO1
	AL13-8	ATGAACTCGCCAGAACCAGCAGCGGGAGCCAGCGCGGATCCGGCGCGCGCCTTACAGCGCGCCTCCGCTTCAC	
pFF1_22A_szeS_gxpS (Fig. 4 - III)	KB22-1	CGGATCCTACCTGACGCGTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGGTAGTATTGCTAAAAAGG	X. szertirmaii DSM 16338
	KB22-6N	GTGCTCTGCTTCATTGACACC	
	KB22-7N	CCCTGCAACAGGTCATTAACC	X. szertirmaii DSM 16338
	KB22-2	TCCGGAGTCTTCTCTATCTGTTGTTCAAACAACTGATGGACACATACCTGAGTAGGATACGGTTCTTCGGTCGCATTCCAGCTTTCCAGCAATAACCG	
	KB22-3N	TGGAATGCGACCGAAG	P. luminescens TTO1
	KB22-8N	CGATATTGACGTGGGTTAAAACG	
	KB22-9N	GCATCGTCGAACAGGTACCG	P. luminescens TTO1
	KB22-4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTACAGCGCCTCCGCTTC	
pFF1_gxpS_Cterm (Fig. 5 - I)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGAGAGCATCGTGAG	P. asymbiotica ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	P. asymbiotica ATCC 43949
	KB_Pau-P4	GATTTCTGCTACCAGTTCAGCC	
	KB-Rdp3-FW	ATTTGCACATTGAATAATCTGTTCCAATTCCCTGTGTTGGCTGAACTGGTAGCAGAAATCCGTAGCGCTCAAGACCATG	X. nematophila ATCC 19061

Construct	Primer	Sequence	Organism
	KB-Rdp3-RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGGATCCGTC ATAAAAGTAACTGATATTTTC	P. asymbiotica ATCC 43949
pFF1_gxpS_C2int (Fig. 5 - II)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAA AGAGAGCATCGTGAG	P. asymbiotica ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	P. asymbiotica ATCC 43949
	KB_Pau-P4	GATTTCTGCTACCAGTTCAGCC	
	KB-PluC2-FW	ATTTGCACATTGAATAAATCTGTTCCAATTCCCTGTGTTGGCTGAACTGGTAGCAGAAATCTGCG CACAGATCTGTGCAC	P. luminescens TTO1
	KB-PluC2-RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGGAGCCAGCGGGATCCATG GACACATACCTGAGTAGG	
pFF1_gxpS_xcn1_C 2int (Fig. 5 - III)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAA AGAGAGCATCGTGAG	P. asymbiotica ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	P. asymbiotica ATCC 43949
	KB_Pau-P4	GATTTCTGCTACCAGTTCAGCC	
	KB_XcnC2_FW	ATTTGCACATTGAATAATCTGTTCCAATTCCCTGTGTTGGCTGAACTGGTAGCAGAAATCTGCGT ACAACGTCATGCG	X. nematophila ATCC 19061
	KB_XcnC2_RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGGAGCCAGCGGGATCCATG AATACATAAACGATTCAGG	
pFF1_gxpS_C/E1int (Fig. 5 - IV)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAA AGAGAGCATCGTGAG	P. asymbiotica ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	P. asymbiotica ATCC 43949
	KB_Pau-P4	GATTTCTGCTACCAGTTCAGCC	

|		Sequence	Organism
	KB-Pau-CE1-FW	ATTTGCACATTGAATAATCTGTTCCAATTCCCTGTGTTGGCTGAACTGGTAGCAGAAATCGAGC ACCATCAGTCTTTCG	*P. asymbiotica* ATCC 43949
	KB-Pau-CE1-RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCGGAGCCAGCGGATCCATG GATACACAACGAATCAGG	*P. asymbiotica* ATCC 43949
pFF1_gxpS_C/E3int (Fig. 5 - V)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAA AGAGAGCATCGTGAG	*P. asymbiotica* ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	*P. asymbiotica* ATCC 43949
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	
	KB_Pau-P4	GATTTCTGCTACCAGTTCAGCC	
	KB-Pau-CE3-FW	ATTTGCACATTGAATAATCTGTTCCAATTCCCTGTGTTGGCTGAACTGGTAGCAGAAATCGAGC AACATCGTGAAATCAG	*P. asymbiotica* ATCC 43949
	KB-Pau-CE3-RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCGGAGCCAGCGGATCCATG AATGCACAATTGGTCAG	
pFF1_gxpS_T (Fig. 5 - VI)	KB_Pau-P1	TTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAA AGAGAGCATCGTGAG	*P. asymbiotica* ATCC 43949
	KB_Pau-P2	ATAATGCCACAGGCGACCTG	*P. asymbiotica* ATCC 43949
	KB_Pau-P3	ATACGTCTGGCTCTACCGG	
	KB-Pau-TE-RV	AAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTAA CGCATAAATCGGGTAATC	
pFF1_garS_gxpS_T E (Fig. 6a - I)	LH 6 P1	CGGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGCCTATGCTTGTCATGCAATCGTATC	*X. bovienii* SS-2004
	LH 6 P2	GTTGCGCCAGTGCTAACG	
	LH 6 P3	CGTCTGGGTGTCAGTCCG	*X. bovienii* SS-2004
	LH 6 P4	CTCTACCAGCAGTTGTTGTCGC	
	LH 6 P5	CCCTGACCCGAGATCCGCAACAATTGATCCGGGATGTATCCATCTTACCGCCGACAGAGCGAC AACAACTGCTGGTAGAGGGCAATGGCCCGCAAACG	*P. luminescens* TTO1
	LH 6 P6	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG	

| Construct	Primer	Sequence	Organism
pFF1_garS_gxpS_T (Fig. 6a - II)	LH 6 P1	CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCACAATTC CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT	X. bovienii SS-2004
	LH 6 P2	AACAGGAGGAATTCCATGCCTATGTCATGCAATCGTATC GTTGCGCCAGTGCTAACG	X. bovienii SS-2004
	LH 6 P3	CGTCTGGGTGTCAGTCCG	
	LH 6 P4	CTCTACCAGCAGTTGTTGTCGC	
	LH 6 P5	CCCTGACCCGAGATCCGCAACAATTGATCCGGGATGTATCCATCTTACCGCCGACAGAGCGAC AACAACTGCTGGTAGAGGGCAATGGCCCGCAAACG	P. luminescens TTO1
	LH 6 P7	AGAATCGGAACAACACCGGTAAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGGATCCTAGCGCATAAAATCGGGTAATCC	
pFF1_garS_gxpS_C/ E3int (Fig. 6a - III)	LH 6 P1	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCATGCCTATGTCATGCAATCGTATC	X. bovienii SS-2004
	LH 6 P2	GTTGCGCCAGTGCTAACG	X. bovienii SS-2004
	LH 6 P3	CGTCTGGGTGTCAGTCCG	
	LH 6 P4	CTCTACCAGCAGTTGTTGTCGC	
	LH 6 P5	CCCTGACCCGAGATCCGCAACAATTGATCCGGGATGTATCCATCTTACCGCCGACAGAGCGAC AACAACTGCTGGTAGAGGGCAATGGCCCGCAAACG	P. luminescens TTO1
	LH 6 P8	AACGGTAACATCGCCGGCGTCAGTACAACCGTATCCAGTGTAATGCTGTTGTCAGGCACCCTG ATTTCACGATGTTGCTCGATCTCTGCCACCAGTTCCG	
	LH 3 P13	GAGCAACATCGTGAAATCAG	P. asymbiotica ATCC 43949
	LH 3 P14	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGGATCCATGAATGCACAATTGGTCAG	
pFF1_1B_bacA_srfA-BC (Fig. 6b)	AT-1.1	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCATGGTTGCTAAAACATTCATTAGAAAATG	B. licheniformis ATCC 10716
	AT-1.2	TCTTTGTGCGGCTGGACAGTCTCTTCGAATAGCTGATGAACCGTTTTGTCTGTCGGCACAGGCA GCGCTTTGCCCTTCCAGGACTCTAAAAGTGTCCGTTTTTC	
	AT-1.3	TGGAAGGGCAAAGCGC	B. subtilis ATCC 21332

AT-1.5

AACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGGAGCCAGC

GGATCCGGCGCGCCTTAGAACCAATACGTCAGAGGCTTTG

Supplementary Figures
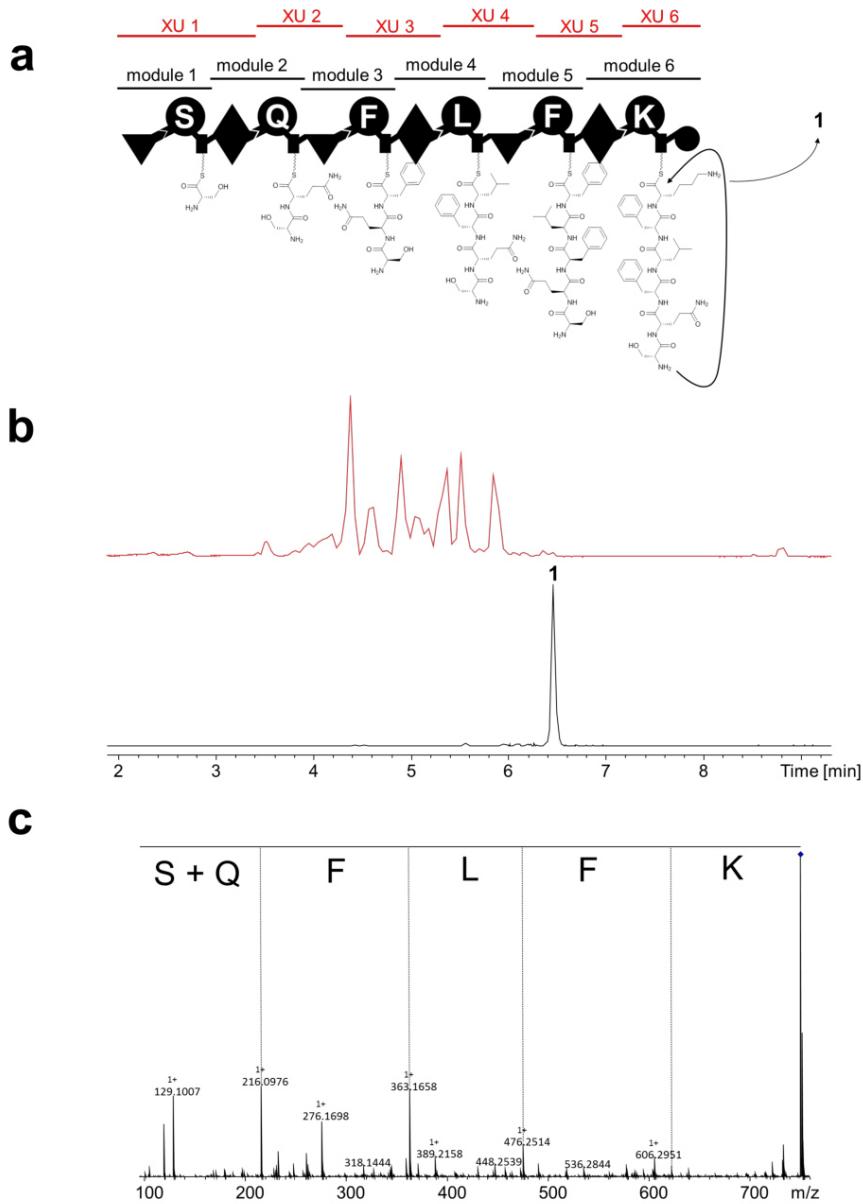
Supplementary Figure 1. Heterologous production of ambactin in E. coli DH10B::mtaA. Schematic representation of the AmbS assembly line (a). Base peak chromatogram (red) and extracted ion chromatogram (black) of **1** (m/z $[M$+H$]^+$ = 751.4) (**b**). MS-MS spectra of **1** (**c**).

28

a

PaxC	2	PEHAIEAIDIISQAERHQLLENFNDTAVALPQ
PaxC	3	PGQSIRSISLPPAECIQVLEDFNDTDLDVPQ
AmbS	2	PDMPIRALEIIPETERTLLLTTWNATETAYPE
AmbS	4	PDRPIRALEIIPETERTLLLKTWNATETAYPE
GxpS	3	SDMPVQQLDILPATERTLLLKTWNATETAYPE
XtpS	3	PDMPVQQLEILPETERTLLLKTWNATETVYPE
BicA	4	PAMPVRELDILPQAERTLLLTTWNNTETPYPP
BicA	3	PAMPVRALDILPEAERTLLLTTWNNTETPYPA
HctAB	1	PDMPVRALDILPEAERTLLLTTWNNTETPYPE
HctAB	3	PEMPVRALEIIPEAERTLLLNIWNGTKTAYPE
HctAB	2	PETPVRQLVILPEAECTLLLKTWNATEAAYPE
KolS	5	PETPVRTLEIIPEAERTLLLKTWNETEIAYSD
BicA	1	PETPVRQLNIIPEAERTLLLSTWNNTETAYPE
GxpS	1	PETPVRALNVLPASERTLLLETWNATETPYPE
XtpS	1	PETPVRALNIIPASERTLLLKTWSTVETPYPE
KolS	12	PETPVRALNIIPEVERRLLLKTWNATETVYPK
KolS	10	PEAPVRLLNILPASEKKLLLESWNATQAPYPD
KolS	14	SETPIRLLNIIPEAEKRLLLLAWNATQALYPD
KolS	8	PETPVRLLNIILPEAEKRLLLATWNATQAPYPD
KolS	3	PETPIRALNILPAAEKRLLLETWNATQAPYPD
KolS	7	PEIPVRTLNILPTAEKKLLLETWNATETSYPD
GxpS	4	PETPVQRLEIIPAVERKLLLKVGNGPQTAYPG
AmbS	6	PETSIQTLEIIPEIERKQLLETWNPTATEYPD
HctAB	4	AQQRIGDIDILTPAERRLLLETWNNTETPYPE
BicA	2	AQRRIGDIDILTSAERQLLLKTWNATETPYPD
KolS	6	PQQPVGKIDILTVAERKLLLETWNATQCRYPD
KolS	13	PQQPVGKIDILTVEERKLLLESWNATQTRYPD
KolS	2	PQQPVGKIDILTVTERKLLLETWNATKTSYPD
KolS	11	PQQTVGKIDILTVAERKLLLETRNVIETRYPK
KolS	4	PQQPVGKIDILTVAERKLLLETWSMTETSYPD
KolS	9	PQQPVGKIDILTVAERKLLLETWNATEMPYPA
AmbS	5	LQQPVGKMAMLAPAERRLLLETWNGTATSYPD
AmbS	3	LQQPVGKMAMLAPAERRLLLETWNGTATSYPD
GxpS	2	PQQPVTAIDILSSSERELLLENWNATEEPYPT
XtpS	2	PQQPVAVIDILSSTERTLLLKTWNATETVYPE
PaxC	1	DSLRVEDLPLLQPQQRAHLLQDFNDTTLSYPP

b

AmbS	4	APDEKAFAHQVYEAPQ
AmbS	2	APDEKAFARQAYEAPQ
AmbS	5	VFDNKALVNQVYEAPK
AmbS	3	VPGEKALVRQIYEAPQ
PaxC	3	APDESAMITQGYEAPI
BicA	1	AFGEEAFVRQAYEAPQ
AmbS	6	VPGEEAFVRQAYETPQ
KolS	9	VPGEEDFARQIYAAPL
KolS	4	VPGEEDFARQIYAVPS
KolS	11	APGEEDFARQIYAAPS
KolS	2	APGEEDFARQIYAAPS
KolS	13	VPGEEAFVRQIYAAPS
GxpS	2	APGEDAFARQIYVAPQ
XtpS	2	APGEEAFARQIYVAPQ
GxpS	3	APGEDAFARQAYQAPQ
KolS	12	APGEDAFARQVYAAPQ
KolS	5	APGADAFARQVYAAPQ
GxpS	4	APGEDAFARQIYAAPQ
KolS	8	VPGENAFARQVYAAPQ
XtpS	1	VPGEAAFARQVYAAPQ
GxpS	1	APGENAFARQVYAAPQ
KolS	10	APGVEAFARQVYAAPQ
HctAB	3	SPDSKAFARQTYAAPQ
HctAB	2	APDSEAFARQIYAAPQ
XtpS	4	IPDNEAFARQVYAAPQ
XtpS	3	IPDNEAFARQVYAAPQ
AmbS	1	EPDSTAYQRQYYQSPQ
GxpS	5	VPDQNAFARQIYEAPQ
HctAB	4	APDGDAFARQAYEAPQ
BicA	4	APDGEAFARQAYEAPQ
BicA	5	APDGEAFARQIYEAPH
KolS	6	APDEEAFSRQIYEAPQ
PaxC	1	APDRTAIVSREYEAPQ
PaxC	2	APDSAATVSREFEAPQ
HctAB	1	APNEEAFVRRTYEAPQ
BicA	2	APDNQAFARRAYEAPQ

c

PaxC	1	VLDVPPNLILEESQVITP
PaxC	2	ELAVPPNLIHADSQIITP
AmbS	1	TFAIPPNLIGMECERITP
KolS	14	AVVVPANVITPATTVLTP
KolS	12	AVVVPANVITPVTTILTP
GxpS	4	TEKVPANVITPATTVLTP
KolS	1	AVIVPANIMTPAITALTP
KolS	7	PVVVPANVITPATTALTP
KolS	5	AVVVPENVITPATTALTP
KolS	8	EVVVPTNVITPVTTALTP
GxpS	1	IVTVPPNVITPATTALTP
XtpS	1	VVTVPPNFITATTTALTP
AmbS	5	SVEVPPNVITPATTALTP
AmbS	3	TMVVPPNVITPATTALTS
KolS	10	VVAVPTNVITSATTKLTP
HctAB	3	AVEVPPNAITPETQHLTP
HctAB	2	AIEVPPNVITPETQHLTP
BicA	4	TIEIPPNVITPATQQLTP
BicA	1	TIEIPPNVITPVTQQLTP
BicA	3	VIEVPPNVITPATQQLTP
HctAB	1	AIEVPPNVITPAIQQLTP
GxpS	3	EISVPDNGITADTTVLTP
XtpS	3	EMSVPDNVITPDAAVLTP
HctAB	4	RFDQQSSTLPPIMPVSRE
HctAB	5	RFDQQGSVLPPIMPVSRE
BicA	2	RIDKQDNVSAIMPVSRE
KolS	6	QLGEQNNSVSAMLPISRE
AmbS	2	RFNERDNRRPAILPIPRD
KolS	4	RFDESGEALPAIIPLSRE
KolS	11	RFDESGEALPAIIPLSRE
KolS	2	QFDESGAALPAIVPLSRE
KolS	9	QFGESDRILPAIVPLSRE
AmbS	4	RLNKKESIQPVIVPISRE
KolS	13	RLNEQSSKLPVIIPLSGE
GxpS	2	QRNTGSDKLPEIRSISRD
XtpS	2	QRHANSDQLPEIRPISRD

Supplementary Figure 2. Sequence alignments of 36 selected NRPS linker sequences each. All linkers are from *Photorhabdus* and *Xenorhabdus*. (**a**) C-A and C/E-A linker sequences; the second part of the proposed hybrid linker is highlighted red. Analysis of overall 121 linker sequences revealed a sequence identity of 43.6 %. (**b**) T-C and T-C/E linker sequences. Analysis of overall 91 linker sequences revealed a sequence identity of 23.3 %. (**c**) A-T linker sequences. Analysis of overall 137 linker sequences revealed a sequence identity of 23.7 %.

105
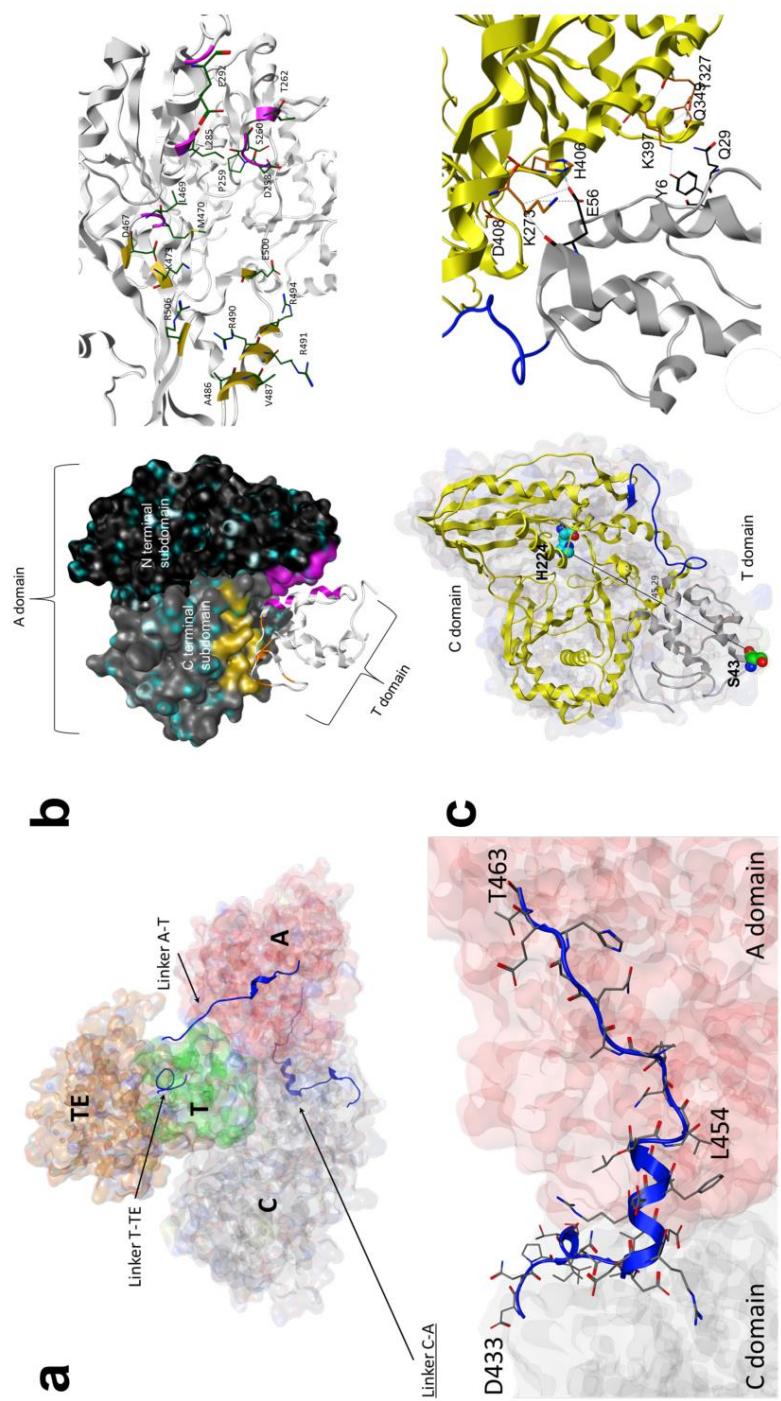
Supplementary Figure 3. Interdomain linkers. (**a**) Interdomain linkers of the SrfA-C termination module[12]. Overall structure of the terminal synthetase SrfA-C (PDB-ID: 2VSQ) at 2.6 Å resolution comprising the condensation domain (C, gray), adenylation domain (A, red), thiolation domain (T, green), and thioesterase domain (TE, brown) with linkers in blue (top). Zoomed view of the C-A linker (**32** aa residues, from D433 to T463) connecting condensation (C) and adenylation (A) domains (bottom). L454 of SrfA-C represents the identified fusion point. (**b**) Interactions between adenylation and thiolation/carrier domains (PDB-ID: 3RG2)[13]. Shown are the interactions between the A domain (grey = C terminal subdomain; black = N terminal subdomain) and the T domain (ribbon presentation in silver) (left). The two observed interaction surfaces are highlighted in pink and yellow. A domain residues that contribute to the A-T didomain interface are shown in green (right). (**c**) Structural analysis of a T-C didomain from NRPS. Overall structure of the TycC5-6 PCP (grey)-C (yellow) didomain (PDB-ID: 2JGP)[14]. The linker region is highlighted in blue and the conserved catalytically active AA residues His224 (turquoise) and S43 (green) are in space filling representation (left). Overview of the PCP domain and its interactions with the C domain with the linker displayed in blue (right). Residues contributing to the domain-domain interface are highlighted in black (T domain) and orange (C domain).

Attachments
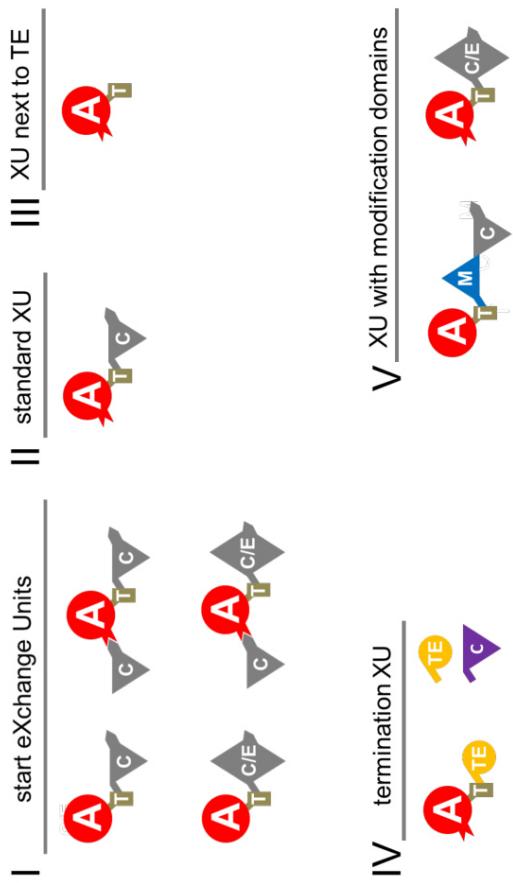
Supplementary Figure 4. The eXchange Unit (XU) concept. (**a**) Schematic representation of selected exchange units with all domains assigned: Adenylation (A, red), thiolation (T, light green), condensation (C, grey), condensation dual (C/E, grey), terminal condensation (C_{term}, purple), thioesterase (TE, yellow), modification (M, blue).

108

32
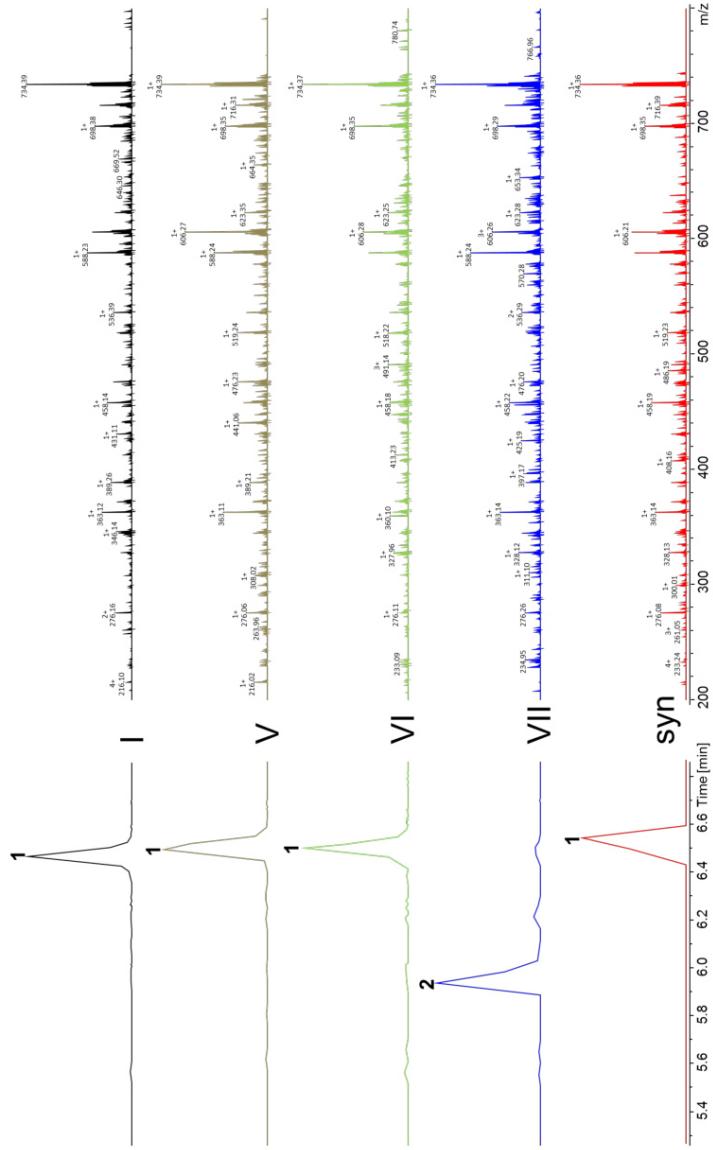
Supplementary Figure 5. HPLC/MS data of ambactin (**1**, *m/z* [*M*+H]⁺ = 751.4) and its derivative **2** (*m/z* [*M*+H]⁺ = 714.4), all from *E. coli* DH10B::mtaA extracts. Roman numbers refer to Fig. 1c. **I**: WT AmbS; **V**: Exchange of A3-T3-C/E4 against A3-T3-C/E4 (D-Phe) of GxpS; **VI**: Exchange of A3-T3-C/E4-A4-T4-C5 against A3-T3-C/E4-A2-T2-C3 (D-Phe-L-Leu) of GxpS; **VII**: Exchange of A3-T3-C/E4-A4-T4-C5 against A3-T3-C/E4-A4-T4-C/E5 (D-Phe-D-Leu) of GxpS; **syn:** HPLC/MS data of synthetic **1** (*m/z* [*M*+2H]⁺ = 751.4).
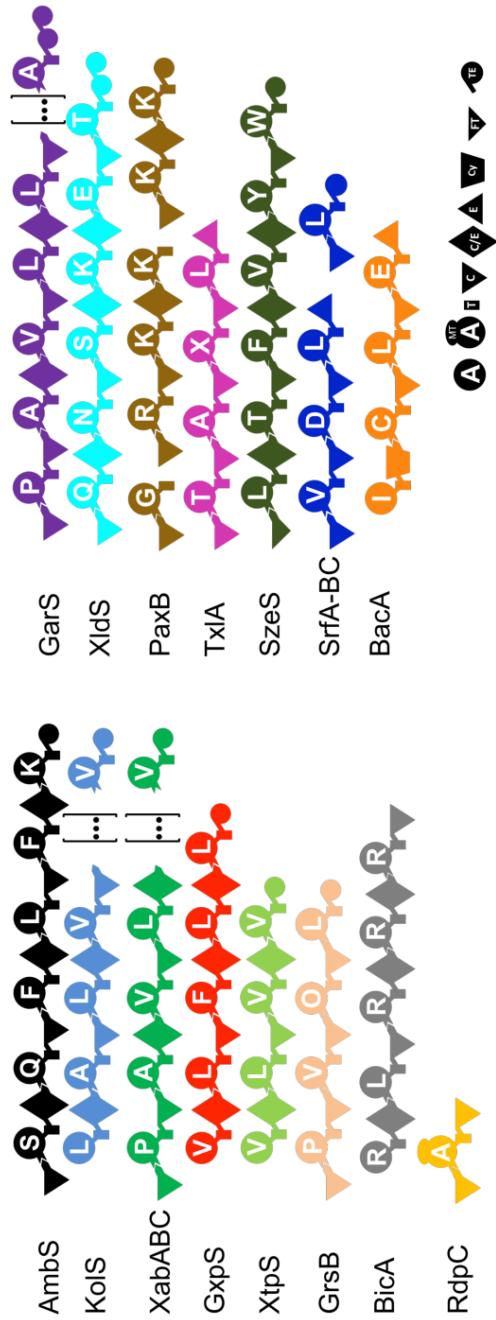
Supplementary Figure 6. Schematic overview of all NRPS used for XU generation. AmbS, KolS, XabABC, GxpS, XtpS, GrsB, BicA, RdpC have all been described previously. For GarS producing gargantuanin see Genbank accession number NC013892.1 (locus_tag: XBJ1_RS08370, XBJ1_RS08360 and XBJ1_RS19040 and XBJ1_RS19040. Except for SrfA-BC and BacA (both of *Bacillus* origin) all other NRPS originate from *Xenorhabdus* or *Photorhabdus* strains.
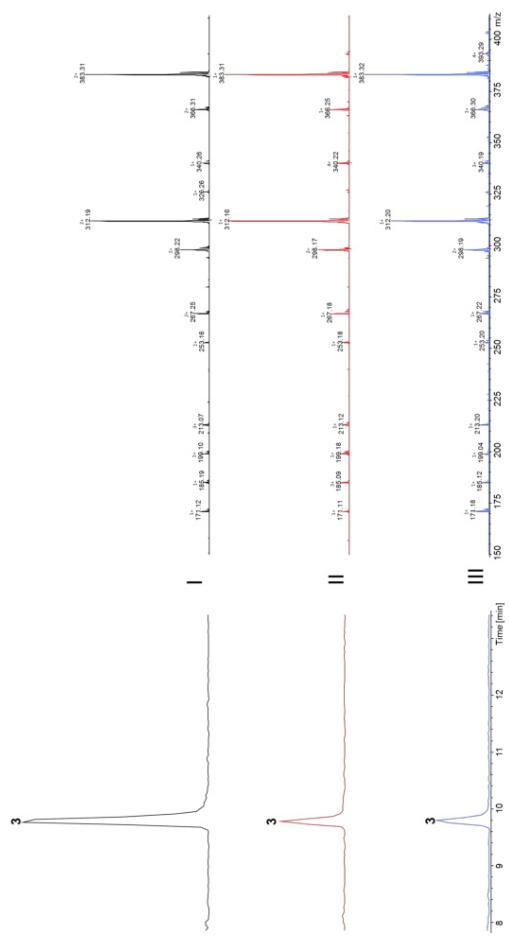
Supplementary Figure 7. HPLC/MS data of xenotetrapeptide (**3**, m/z [M+H]$^+$ = 411.3) produced in *E. coli* DH10B::mtaA. Roman numbers refer to Fig 2a.
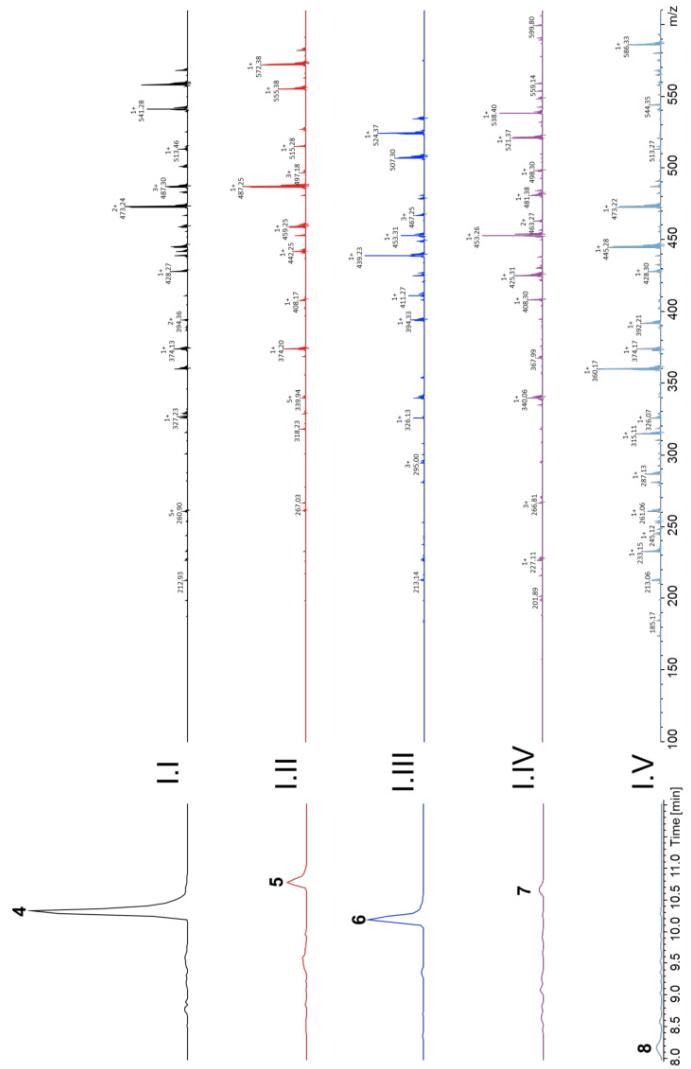
Supplementary Figure 8. HPLC/MS data of GameXPeptides A-D (**5-8**) and linear GameXPeptide A (**9**) produced in *E. coli* DH10B::mtaA. Roman numbers refer to Fig. 2b. (I.I) HPLC/MS data of **5** (*m/z* [M+H]⁺ = 586.4). (I.II) HPLC/MS data of **6** (*m/z* [M+H]⁺ = 600.4). (I.III) HPLC/MS data of **7** (*m/z* [M+H]⁺ = 522.4). (I.IV) HPLC/MS data of **8** (*m/z* [M+H]⁺ = 566.4). (I.V) HPLC/MS data of **9** (*m/z* [M+H]⁺ = 604.4).
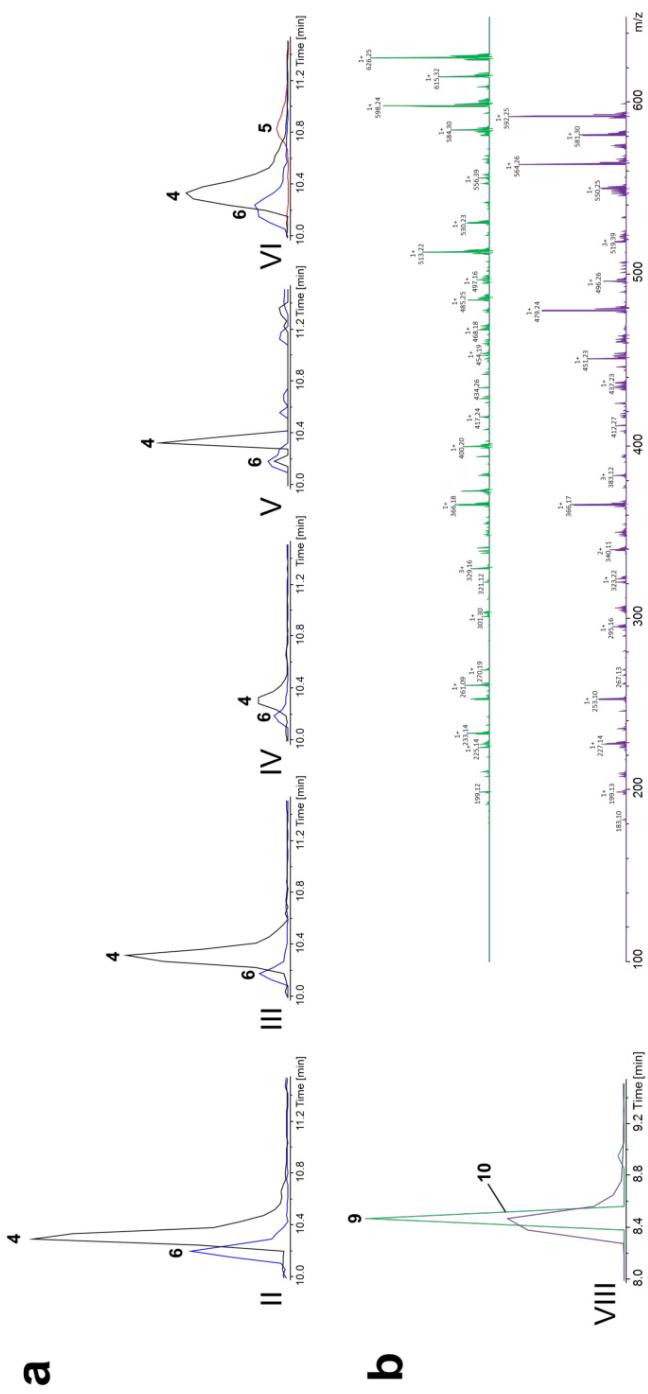
Supplementary Figure 9. Reprogramming GameXPeptide producing NRPS GxpS. Roman numbers refer to Fig. 2b. (**a**) Extracted ion chromatograms (EIC) of compounds **5**–**7**. II–IV: EICs of **5** and **7**. V: EICs of **5** and **7**, both with tenfold increased intensity. VI: EICs of **5**-**7**. (**b**) HPLC/MS data of new GameXPeptide derivatives **10** (*m/z* [*M*+H]⁺ = 643.4) and **11** (*m/z* [*M*+H]⁺ = 609.4).
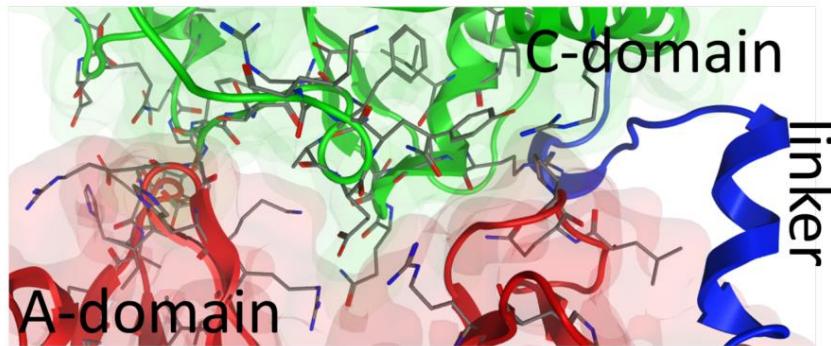
Supplementary Figure 10a. The C-A-Didomain interface of SrfA-C **(PDB-ID: 2VSQ_A.** Amino acids contributing to the interacting non-linker regions and included in the analysis are depicted in stick representation.
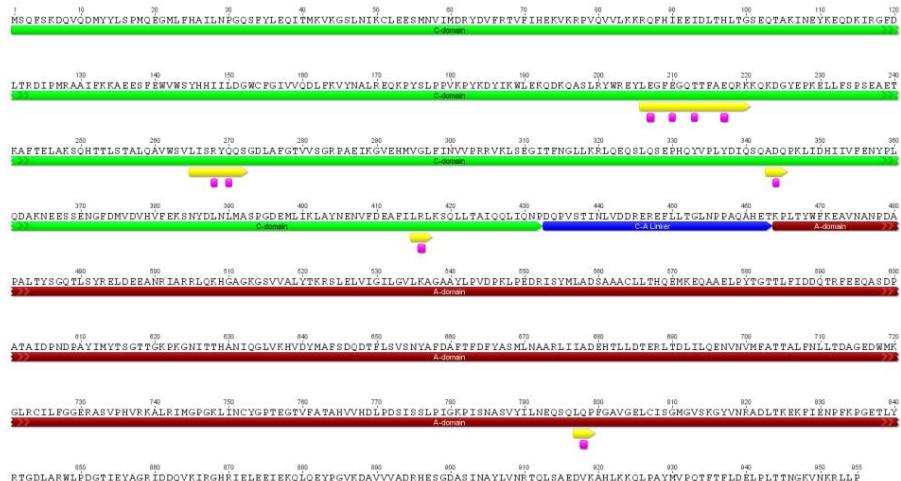
Supplementary Figure 10b. Primary sequence of the SrfA-C C-A didomain. C-domain (green), C-A linker (blue), A-domain (red), H-bond forming aa (pink), and aa included in the analysis (yellow, *cf.* Supplementary Fig. 10c).

Supplementary Figure 10c. Amino acids (aa) included in the Analysis. Depicted are the aa marked in yellow from Supplementary Fig. 10b.

38

114

a

Fig. 1c-V

Interface 1   Interface 2

Production in comparison to WT 88%

```
            1        10        20        30        40        50
            |        |         |         |         |         |
AmbS-C3A3   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
GxpS-C3A3   LADAPVLLELPIDRPLLSRLSGQVQPGYL-PVPLARYVKIRGFRIEDGQDKRL
GxpS-C4A4   MLAEVDEPTLLFGLAVLSRTSGQVESGYM-PVPLVRYIKIRGFRIEDGQDKRL
AmbS-C4A4   MLIGVSEPTLLFGIKILSRISGQVASGYM-PVPLARYVKIRSFRIEDGYDKRL
```
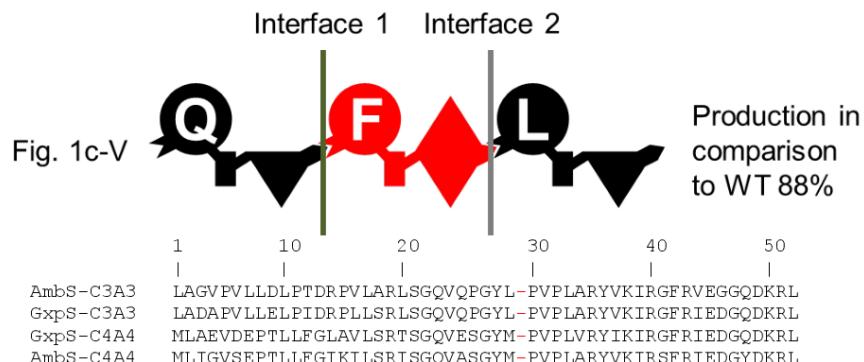
Interface 1: AmbS_A3 vs. GxpS_A3 -> 91.3% Identity
Interface 2: GxpS_A4 vs. AmbS_A4 -> 82.6% Identity

b

Fig. 1c-VI

Interface 1   Interface 2

Production in comparison to WT 57%

```
            1        10        20        30        40        50
            |        |         |         |         |         |
AmbS-C5A5   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
AmbS-C3A3   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
GxpS-C3A3   LADAPVLLELPIDRPLLSRLSGQVQPGYL-PVPLARYVKIRGFRIEDGQDKRL
```
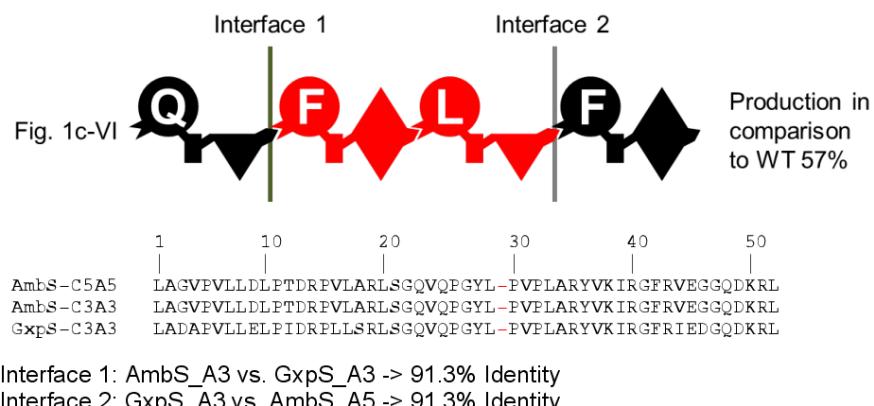
Interface 1: AmbS_A3 vs. GxpS_A3 -> 91.3% Identity
Interface 2: GxpS_A3 vs. AmbS_A5 -> 91.3% Identity

c
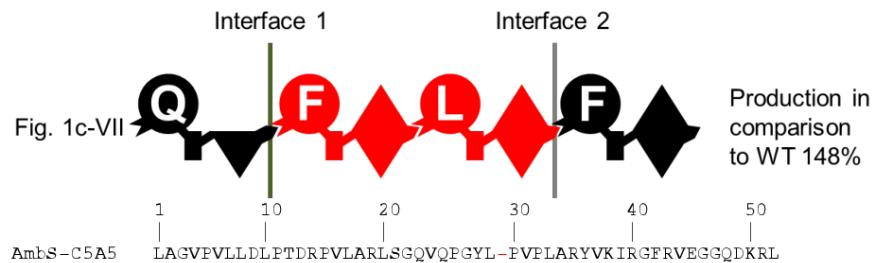
Fig. 1c-VII

Interface 1   Interface 2

Production in comparison to WT 148%

```
            1        10        20        30        40        50
            |        |         |         |         |         |
AmbS-C5A5   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
```

39

115

```
AmbS-C3A3    LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
GxpS-C3A3    LADAPVLLELPIDRPLLSRLSGQVQPGYL-PVPLARYVKIRGFRIEDGQDKRL
GxpS-C5A5    MLAEVDEPTLPFGLAVLSRTSGQVQGGYM-PVPLARYIKIRGFRIEDGQDKRL
```

Interface 1: AmbS_A3 vs. GxpS_A3 -> 91.3% Identity
Interface 2: GxpS_A5 vs. AmbS_A5 -> 87.0% Identity

c

Fig. 2a-II

Production in comparison to WT 48%

```
               1        10        20        30        40        50
               |        |         |         |         |         |
KolS-C13A13   MLAEVDEPTLPFGLTVLSRISGQVSGGYM-PVPLARYVKIRGFRIEDEQNKRL
  XtpS-C4A4   MLAEVDEPTLSFGLSVLSRSSGQVASGYM-PVPLARYVKIRGFRIEDGQDKRL
KolS-C15A15   MLAEVDQPTLPFGLTVLSRTSGQVQNDYM-PVPLARYIKIRGFRIEDGQDKRL
  GxpS-C2A2   MLAEVDEPTLPFGLTVLSRTSGQVQGGYM-PVPLVRYIKIRGFRIEDGQDKRL
```

Interface 1: GxpS_A2 vs. KolS_A13 -> 82.6% Identity
Interface 2: KolS_A15 vs. XtpS_A4 -> 95.7% Identity

d

Fig. 2b-II

Production in comparison to WT 28%

```
             1        10        20        30        40        50
             |        |         |         |         |         |
GxpS-C3A3   LADAPVLLELPIDRPLLSRLSGQVQPGYL-PVPLARYVKIRGFRIEDGQDKRL
AmbS-C5A5   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDKRL
XtpS-C2A2   MLAEVDEPTLPFGLTVLSHTSGQVQSGYM-PVPLVRYIKIRGFRIEDGQDKRL
AmbS-C4A4   MLIGVSEPTLLFGIKILSRISGQVASGYM-PVPLARYVKIRSFRIEDGYDKRL
```

Interface 1: XtpS_A2 vs. AmbS_A4 -> 82.6% Identity
Interface 2: AmbS_A5 vs. GxpS_A3 -> 91.3% Identity

40

116

e

Interface 1   Interface 2   Interface 3

Fig. 2b-III   V   L   F   L   L

Production in comparison to WT 16%

```
                1         10        20        30        40        50
                |         |         |         |         |         |
XtpS-C2A2   MLAEVDEPTLPFGLTVLSHTSGQVQSGYM-PVPLVRYIKIRGFRIEDGQDK-RL
GxpS-C4A4   MLAEVDEPTLLFGLAVLSRTSGQVESGYM-PVPLVRYIKIRGFRIEDGQDK-RL
GxpS-C2A2   MLAEVDEPTLPFGLTVLSRTSGQVQGGYM-PVPLVRYIKIRGFRIEDGQDK-RL
GxpS-C5A5   MLAEVDEPTLPFGLAVLSRTSGQVQGGYM-PVPLARYIKIRGFRIEDGQDK-RL
GarS-C5A5   QLADVDAPTAPFG--VLAQTSGHVA--YL-PVPLGRWVKLRGFRIEDELSHPRL
GarS-C4A4   LQNAPSLLTLPTDRPVLSRLSGQLQPGYL-PVPLGRWVKLRGFRIEDEPNHPRL
```

Interface 1: XtpS_A2 vs. GxpS_A2 -> 100.0% Identity
Interface 2: GxpS_A4 vs. GarS_A4 -> 62.5% Identity
Interface 3: GarS_A5 vs. GxpS_A5 -> 62.5% Identity

f

Interface 1   Interface 2   Interface 3   Interface 4

Fig. 2b-VI   V   L   F   L   L

Production in comparison to WT 3.2%

```
                1         10        20        30        40        50
                |         |         |         |         |         |
XtpS-C2A2   MLAEVDEPTLPFGLTVLSHTSGQVQSGYM-PVPLVRYIKIRGFRIEDGQDK-RL
GxpS-C2A2   MLAEVDEPTLPFGLTVLSRTSGQVQGGYM-PVPLVRYIKIRGFRIEDGQDK-RL
GxpS-C5A5   MLAEVDEPTLPFGLAVLSRTSGQVQGGYM-PVPLARYIKIRGFRIEDGQDK-RL
GxpS-C4A4   MLAEVDEPTLLFGLAVLSRTSGQVESGYM-PVPLVRYIKIRGFRIEDGQDK-RL
AmbS-C4A4   MLIGVSEPTLLFGIKILSRISGQVASGYM-PVPLARYVKIRSFRIEDGYDK-RL
GarS-C5A5   QLADVDAPTAPFG--VLAQTSGHVA—YL--PVPLGRWVKLRGFRIEDELSHPRL
GxpS-C3A3   LADAPVLLELPIDRPLLSRLSGQVQPGYL-PVPLARYVKIRGFRIEDGQDK-RL
AmbS-C5A5   LAGVPVLLDLPTDRPVLARLSGQVQPGYL-PVPLARYVKIRGFRVEGGQDK-RL
GarS-C4A4   LQNAPSLLTLPTDRPVLSRLSGQLQPGYL-PVPLGRWVKLRGFRIEDEPNHPRL
```

Interface 1: XtpS_A2 vs. AmbS_A4 -> 82.6% Identity
Interface 2: AmbS_A5 vs. GxpS_A3 -> 91.3% Identity
Interface 3: GxpS_A4 vs. GarS_A4 -> 62.5% Identity
Interface 4: GarS_A5 vs. GxpS_A5 -> 62.5% Identity

41

117

g

Interface 1

Fig. 2b-VII

Production in comparison to WT 30%

```
               1         10        20        30        40        50
               |         |         |         |         |         |
GxpS-C5A5    MLAEVDEPTLPFGLAVLSRTSGQVQGGYM-PVPLARYIKIRGFRIEDGQDKRL-
GrsB-C4A4    FAEDIPVLNLSTDYPLLHKYSGQLQVTHF-IQPLARWVKIRGIRIEDHMKEKYL
```

Interface 1: GxpS_A5 vs. GrsB_A4 -> 52.2% Identity

h

Interface 1

Fig. 2b-VIII

Production in comparison to WT 70%

```
               1         10        20        30        40        50
               |         |         |         |         |         |
BicA-C2A2    MLAEVEEPTLPFGLTVLSRVCGQVDRGYM-PVPLARYVKIRGFRIEEGQDRRL
GxpS-C2A2    MLAEVDEPTLPFGLTVLSRTSGQVQGGYM-PVPLVRYIKIRGFRIEDGQDKRL
```

Interface 1: BicA_A2 vs. GxpS_A2 -> 82.6% Identity

Supplementary Figure 11. Analysis of the non-linker C-A interface regions. Depicted are the interfaces and sequence homologies of the NRPS shown in Fig. 1c-V (**a**), 1c-VI (**b**), 2a-II (**c**), 2b-II (**d**), 2b-III (**e**), 2b-VI (**f**), 2b-VII (**g**), and Fig. 2b-VIII (**h**). (**Top**) Schematic representation of the recombinant NRPSs, forming artificial interfaces. (**Center**) Alignments of aa forming the naturally occurring C-A non-linker interface between the respective XUs. The red gap (-) separates the interface forming aa of the C- (left, 29 aa) and A-domain (right, 23 aa). Colour coded (green grey, turquoise, and red) are by recombination introduced artificial interfaces. (**Bottom**) The quality/homology of the recombinant C-A interfaces were

42

118

evaluated via measuring the sequence identity of the interface forming aa of the WT A-domain compared to the introduced A-domain.
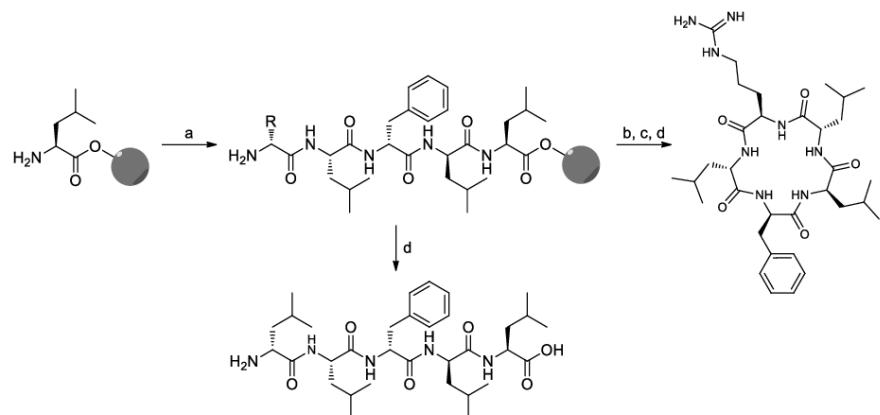
Supplementary Figure 12. Chemical synthesis as exemplarily shown for **8** and **9**. **a**, Fmoc-AA-OH (6 eq.), HCTU (6 eq.), DIPEA (12 eq.), NMP, 50 min, then piperidine/NMP. **b**, HFIP/ DCM (1:4), 1 h. **c**, HATU (2 eq.), HOAt (2 eq.), DIPEA (4 eq.), DMF, 25 W, 75 °C, 20 min. **d**, TFA/TIS/water (95:2.5:2.5), 1-2 h.

43

Attachments
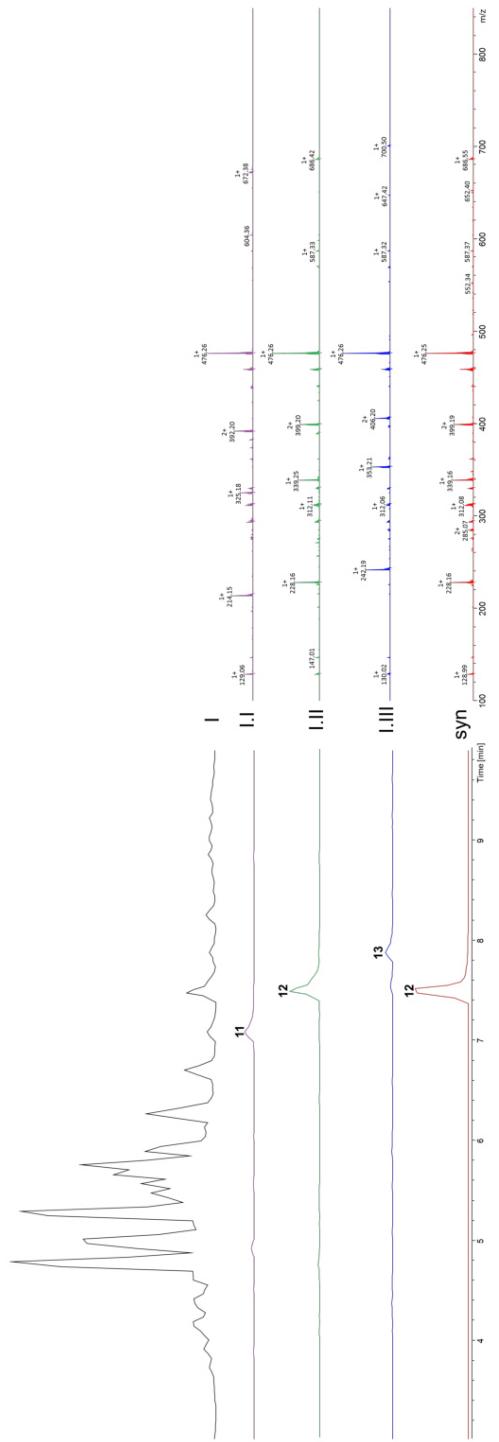
Supplementary Figure 13. HPLC/MS data of compounds **11-13** produced in E. coli DH10B::mtaA and synthetic **12**. Roman numbers refer to Fig. 3. (I) Base Peak Chromatogramm (BPC) of a methanolic XAD extract. (I.I) HPLC/MS data of **11** (m/z $[M+2H]^+$ = 400.8). (I.II) HPLC/MS data of **12** (m/z $[M+2H]^+$= 407.8). (I.III) HPLC/MS data of **13** (m/z $[M+2H]^+$= 414.8). (syn) HPLC/MS data of synthetic **12** (m/z $[M+2H]^+$ = 407.8).
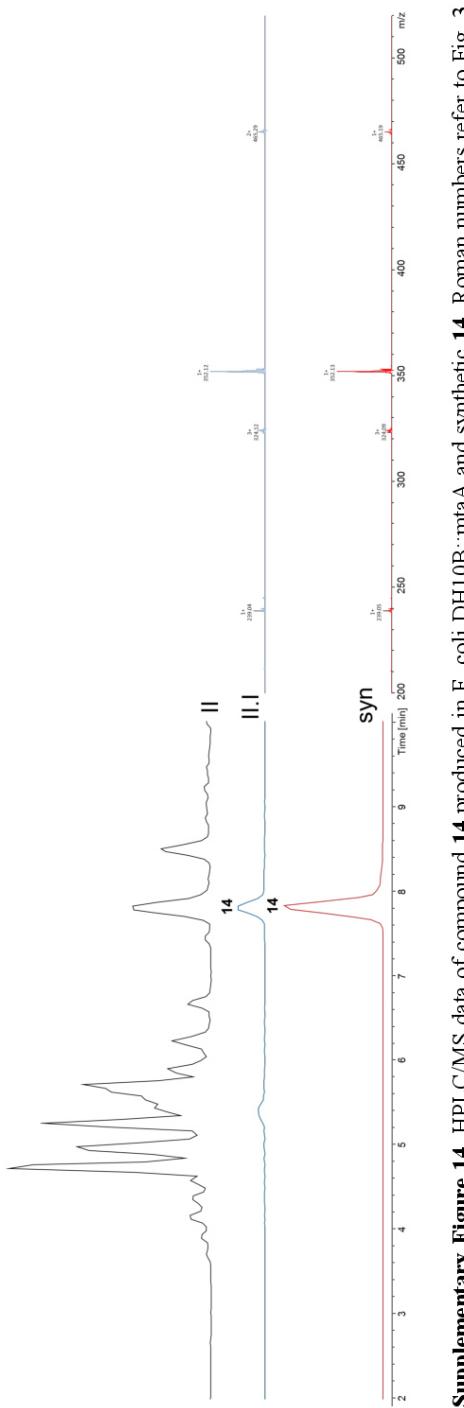
Supplementary Figure 14. HPLC/MS data of compound **14** produced in E. coli DH10B::mtaA and synthetic **14**. Roman numbers refer to Fig. 3. (II) BPC of a methanolic XAD extract. (II.I) HPLC/MS data of **14** (m/z $[M+H]^{+}$ = 484.4). (syn) HPLC/MS data of synthetic **14** (m/z $[M+H]^{+}$ = 484.4).

Supplementary Figure 15. Homologues recombination of C/E4 and C/E5 of GxpS. Roman numbers refer to Fig. 4. (**a**) Schematic representation of two recombinant NRPS, shown in Fig. 4. The regions of the homologues recombination event are indicated by black lines. The outcrossed region is depicted in black. (**b**) Sequence alignment of the C/E4/5 hybrid domain of pFF1_22A_szeS_gxpS, C/E4 domain of GxpS and C/E5 domain of GxpS. The site of recombination during yeast cloning is marked in **red**.

46
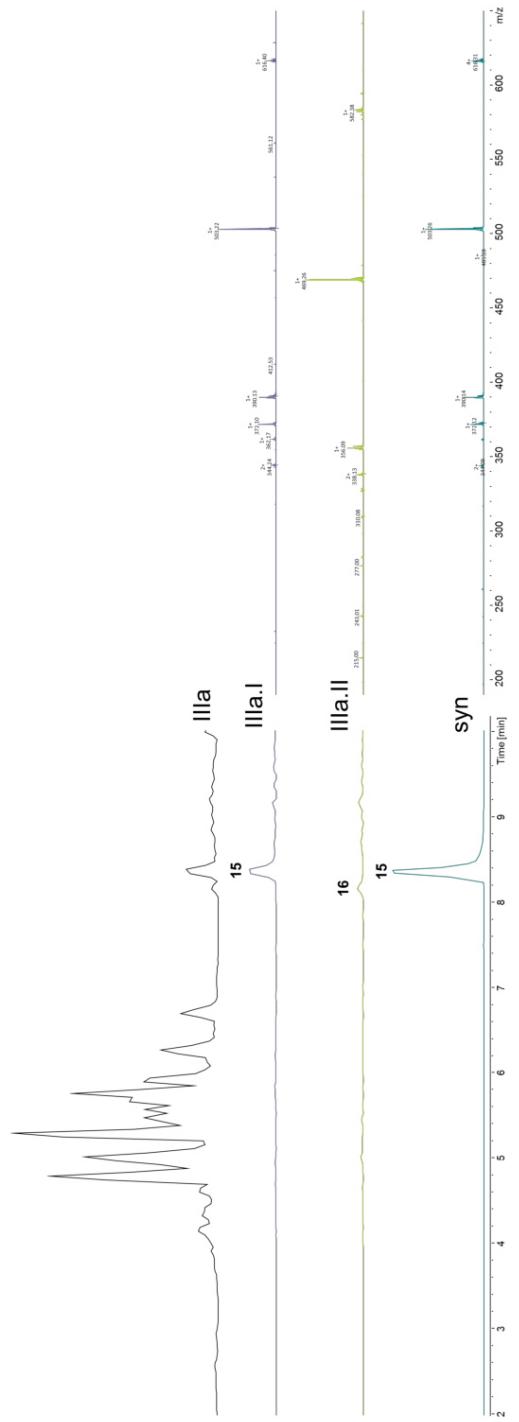
Supplementary Figure 16. HPLC/MS data of compounds **15** and **16** produced in E. coli DH10B::mtaA and synthetic **15**. Roman numbers refer to Fig. 3. (IIIa) BPC of a methanolic XAD extract. (IIIa.I) HPLC/MS data of **15** (m/z [M+H]$^+$ = 634.4). (IIIa.II) HPLC/MS data of **16** (m/z [M+H]$^+$ = 600.4). (syn) HPLC/MS data of synthetic **15** (m/z [M+H]$^+$ = 634.4).
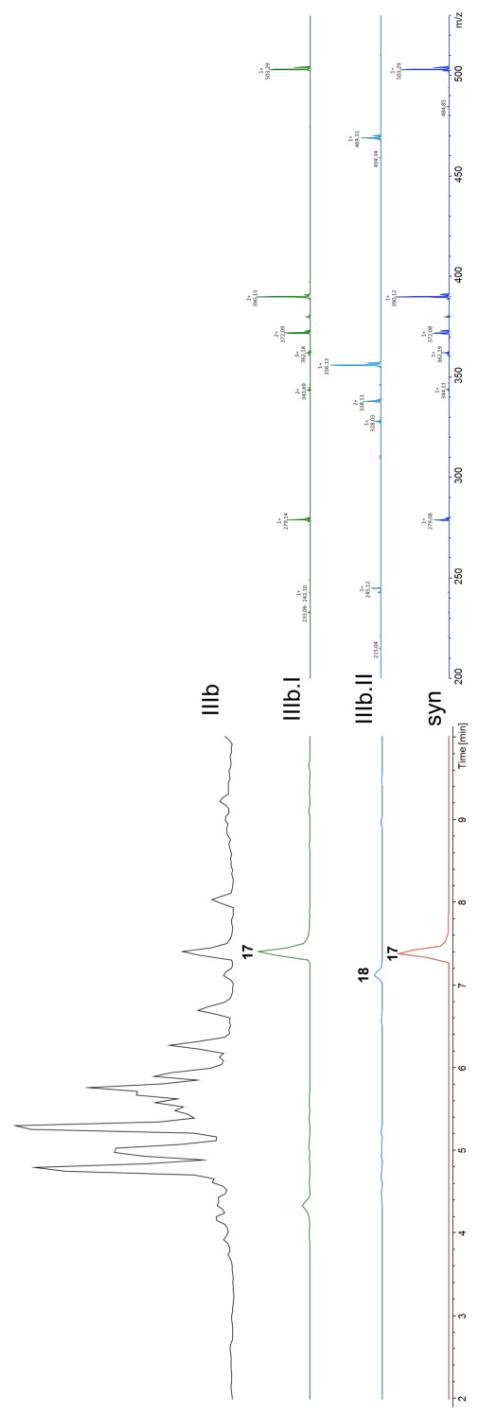
Supplementary Figure 17. HPLC/MS data of compounds **17** and **18** produced in *E. coli* DH10B::mtaA and synthetic **17**. Roman numbers refer to Fig. 3. (IIIb) BPC of a methanolic XAD extract. (IIIb.I) HPLC/MS data of **17** (m/z [M+H]$^+$ = 521.3). (IIIb.II) HPLC/MS data of **18** (m/z [M+H]$^+$ = 487.3). (syn) HPLC/MS data of synthetic **17** (m/z [M+H]$^+$ = 521.3).
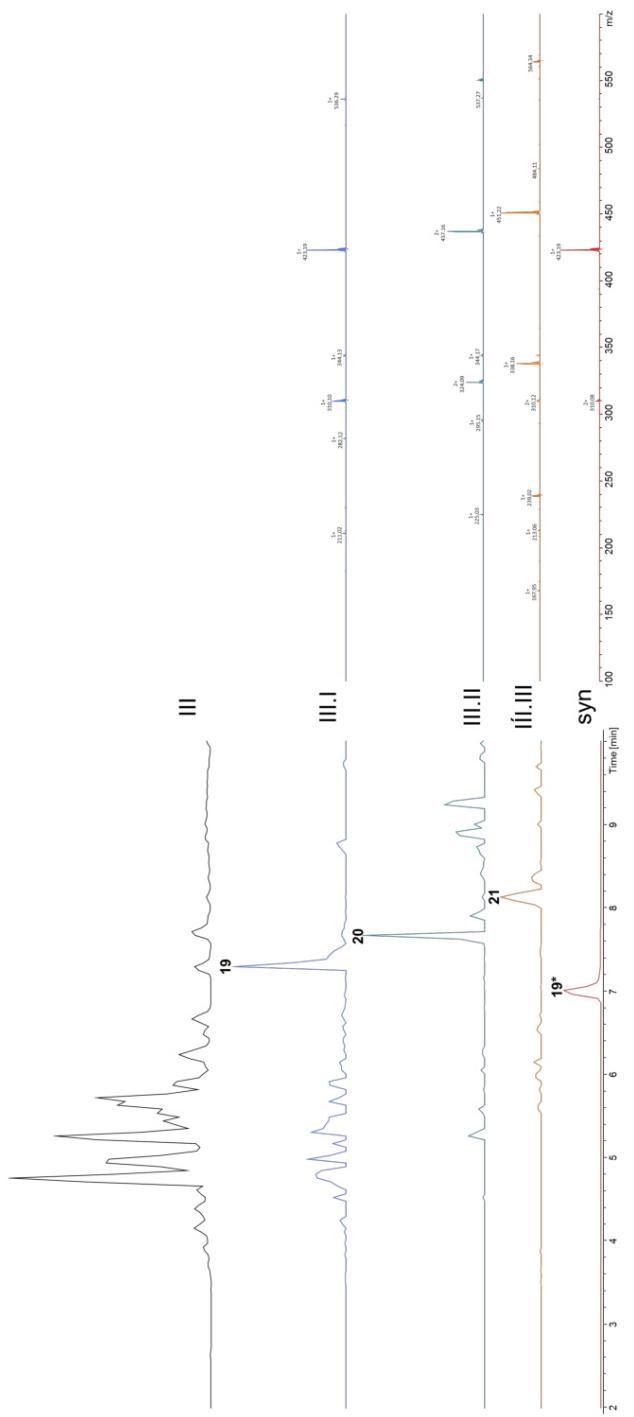
Supplementary Figure 18. HPLC/MS data of compounds **19-21** produced in E. coli DH10B::ntaA. Roman numbers refer to figure 5a. (III) BPC of a methanolic XAD extract. II-IV with tenfold increased intensity. (III.I) HPLC/MS data of **19** (m/z $[M+H]^+$ = 554.4). (III.II) HPLC/MS data of **20** (m/z $[M+H]^+$ = 568.4). (III.III) HPLC/MS data of **21** (m/z $[M+H]^+$ = 582.4). (syn) HPLC/MS data of synthetic **19*** showing a terminal D-Leu (m/z $[M+H]^+$ = 554.4).
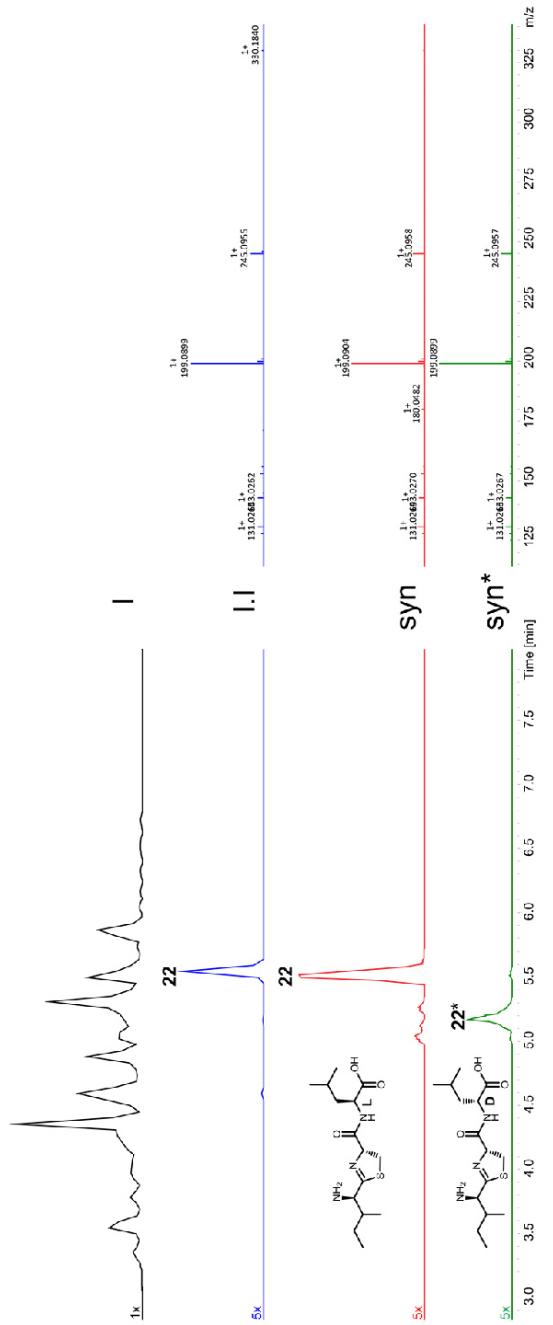
Supplementary Figure 19. HPLC/MS data of compound **22** produced in E. coli DH10B::mtaA and synthetic **22**. Roman numbers refer to figure 5b. (I) BPC of a methanolic XAD extract. (I.I) HPLC/MS data of **22** (m/z $[M+H]^+$ = 330.2). (syn) HPLC/MS data of synthetic **22** (m/z $[M+H]^+$ = 330.2). (syn*) HPLC/MS data of synthetic **22*** showing a terminal D-Leu (m/z $[M+H]^+$ = 330.2). The intensities of the chromatograms of I.I, syn and syn* are 5-fold increased.

50

References

1. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2,** 31–34 (2007).
2. 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).
3. 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).
4. Fuchs, S. W. *et al.* Formation of 1,3-cyclohexanediones and resorcinols catalyzed by a widely occurring ketosynthase. *Angew Chem Int Ed Engl* **52,** 4108–12–4112 (2013).
5. Nollmann, F. I. *et al.* A Photorhabdus natural product inhibits insect juvenile hormone epoxide hydrolase. *ChemBioChem* **16,** 766–771 (2015).
6. Nozaki, S. & Muramatsu, I. Convenient Synthesis of *N*-Protected Amino Acid Amides. *Bull Chem Soc Jpn* (1988).
7. Lee, J., Griffin, J. H. & Nicas, T. I. Solid-Phase Total Synthesis of Bacitracin A. *J Org Chem* **61,** 3983–3986 (1996).
8. 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 (2014).
9. Kegler, C. *et al.* Rapid Determination of the Amino Acid Configuration of Xenotetrapeptide. *ChemBioChem* **15,** 826 (2014).
10. Nollmann, F. I. *et al.* Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. *ChemBioChem* **16,** 205–8 (2014).
11. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166,** 557–580 (1983).
12. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321,** 659–63–663 (2008).
13. Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol* **19,** 188 (2012).
14. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci* **16,** 2093–2107 (2007).

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

6.2.1 Erklärung zu den Autorenanteilen an der Publikation

Status: published

Name der Zeitschrift: *Nat. Chem.* **11**, 653–661 (2019)[164]

Autoren: Kenan A. J. Bozhüyük (KAJB)*, Annabell Linck (AL)*, Andreas Tietze (AT)*, Janik Kranz (JK)*, Frank Wesche (FW), Sarah Nowak (SN), Florian Fleischhacker (FF), Yan-Ni Shi (YNS), Peter Grün (PG) und Helge B. Bode (HBB)

                               *gemeinsame Erstautorenschaft

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

KAJB (26 %), AL (20 %), AT (12 %), JK (12 %), HBB (30 %)

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Klonierung von Plasmiden: KAJB (5 %), AL (5 %), AT (3 %), JK (1 %), FF (2 %); Heterologe Expression: KAJB (3 %), AL (3 %), AT (3 %), JK (3 %), FF (3 %); Expression von His-markierten Proteinen: SN (2 %); Pyrophosphat-Assay: KAJB (2 %), SN (3 %), JK (1 %); HPLC-MS: KAJB (2 %), AL (3 %), AT (4 %), JK (4 %), FF (2 %); Homologie-Modell: KAJB (1 %); Peptidisolation: AL (3 %), AT (3 %), JK (6 %), YNS (2 %), PG (5 %), FW (2 %); Peptidquantifizierung: KAJB (1 %), AL (4 %), AT (6 %), JK (4 %); Chemische Synthese: FW (5 %); NMR Experimente: YNS (3 %)

(3) zur Erstellung der Datensammlung und Abbildungen

Sequenzalignment und Strukturanalyse: KAJB (3 %), AL (3 %), AT (2 %), JK (2 %); Verifizierung des XUC-Konzepts: KAJB (4 %), AL (5 %), AT (4 %), JK (4 %), FF (3 %); Fusion Gram-positiver und –negativer XUCs: AL (2 %); AT (4 %); JK (4 %); *in vitro*-Assay: KAJB (2 %), JK (3 %); Erweiterung der Starter XUCs: KAJB (2 %); Fütterungsexperimente mit nicht-natürlichen Aminosäuren: KAJB (2 %), AL (3 %), AT (6 %), JK (6 %), YNS (2 %); Erstellung einer Peptidbibliothek: KAJB (4 %),

AL (6 %), AT (3 %), FF (3 %); Isolierung und Strukturaufklärung von Peptiden: AL (2 %), AT (5 %), JK (6 %), YNS (5 %)

(4) zur Analyse und Interpretation der Daten

Sequenzalignment und Strukturanalyse: KAJB (5 %), AL (3 %), AT (2 %), JK (2 %); Verifizierung des XUC-Konzepts: KAJB (3 %), AL (5 %), AT (4 %), FF (4 %); Fusion Gram-positiver und –negativer XUCs: AL (2 %); AT (4 %); JK (4 %); *in vitro*-Assay: KAJB (4 %), JK (3 %), SN (4 %); Erweiterung der Starter XUCs: KAJB (2 %); Fütterungsexperimente mit nicht-natürlichen Aminosäuren: KAJB (2 %), AL (2 %), AT (4 %), JK (4 %); Erstellung einer Peptidbibliothek: KAJB (4 %), AL (6 %), AT (3 %), FF (3 %); Isolierung und Strukturaufklärung von Peptiden: AL (3 %), AT (4 %), JK (5 %), YNS (4 %), PG (5 %)

(5) zum Verfassen des Manuskriptes

KAJB (30 %), AL (10 %), AT (10 %), JK (10 %), HBB (40 %)

Ort/Datum	Unterschrift des Promovierenden

Ort/Datum	Unterschrift des Betreuers

6.2.2 Publication

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
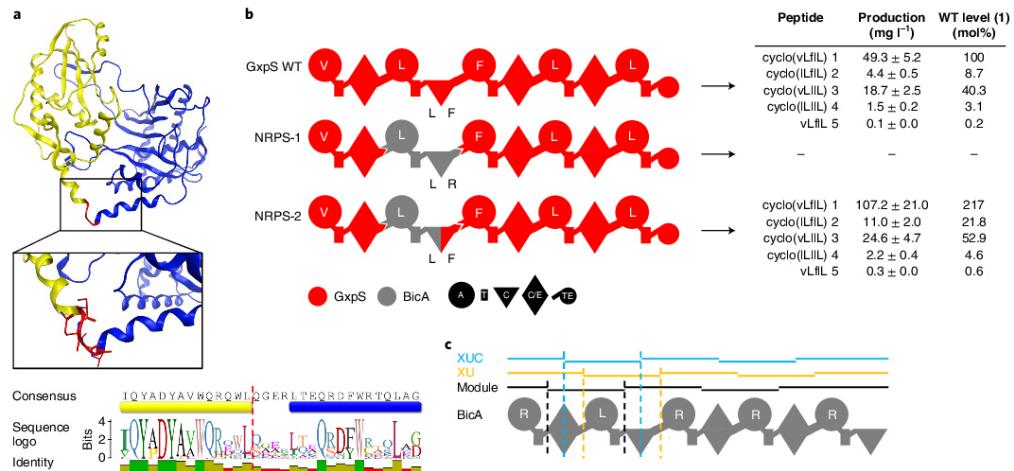
Fig. 1 | Modulation of C domain substrate specificity. a, The C domain excised from the T-C bidomain TycC 5–6 from tyrocidine syntethase (TycC) of *Brevibacillus brevis* (PDB ID: 2JGP). The C domains' N-terminal donor (yellow) and C-terminal acceptor site (blue) sub-domains are depicted in ribbon representation (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 mg l⁻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 mg l⁻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 mg l⁻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 mg l⁻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

ARTICLES
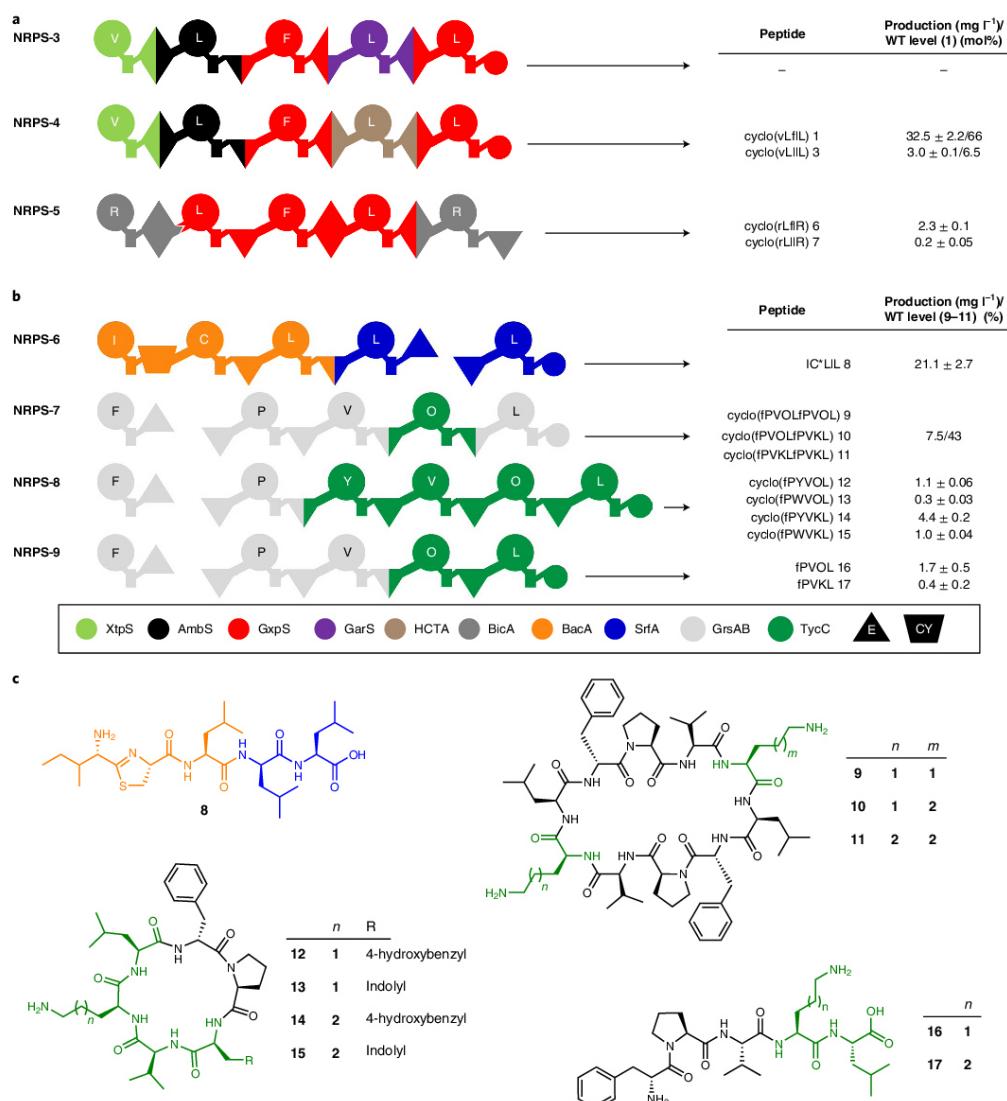
Fig. 2 | Design of recombinant NRPS for peptide production. a, The generated recombinant GxpS (**NRPS-3** to **-5**) and corresponding amounts of GameXPeptide derivatives **1**, **3**, **6** and **7** as determined in triplicates. **b**, Recombinant **NRPS-6** to **-9** using building blocks of Gram-positive 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 (¥-Y). When the Val-specific XUC3 of the xenotetrapeptide[55] (**26**) (Supplementary Fig. 18) producing NRPS (XtpS) from *X. nematophila* HGB081 was exchanged against

Attachments

	Peptide	Production (mg l^{-1}) WT level (1) (%)
NRPS-18	–	–
NRPS-19	cyclo(vLflL) 1	0.3 ± 0.03/0.6
	cyclo(vLlIL) 3	0.3 ± 0.04/0.7
NRPS-20	cyclo(vLflL) 1	0.4 ± 0.03/0.7
	cyclo(vLlIL) 3	0.4 ± 0.01/0.9
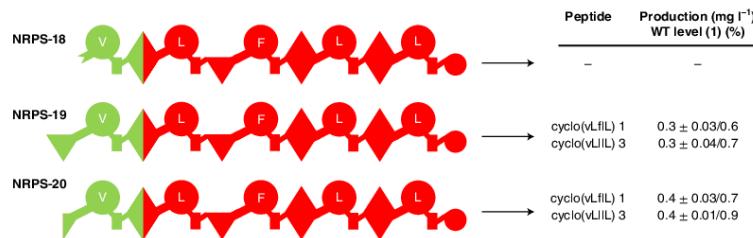
Fig. 3 | Elongation XUCs can be used as starting XUC. A schematic representation of recombinant GxpS (**NRPS-18** to -**20**) and the corresponding peptide yields obtained from triplicate experiments. For the peptide nomenclature, the standard AA one-letter code (with lowercase for D-AA) is used. See Fig. 1

XUC3 of GxpS, six new xenotetrapeptide derivatives (**29–34**) in yields of 0.17–106 mg l^{-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 mg l^{-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 recombind 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 mg l^{-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 mg l^{-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[19]. 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 mg l^{-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 mg l^{-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

134
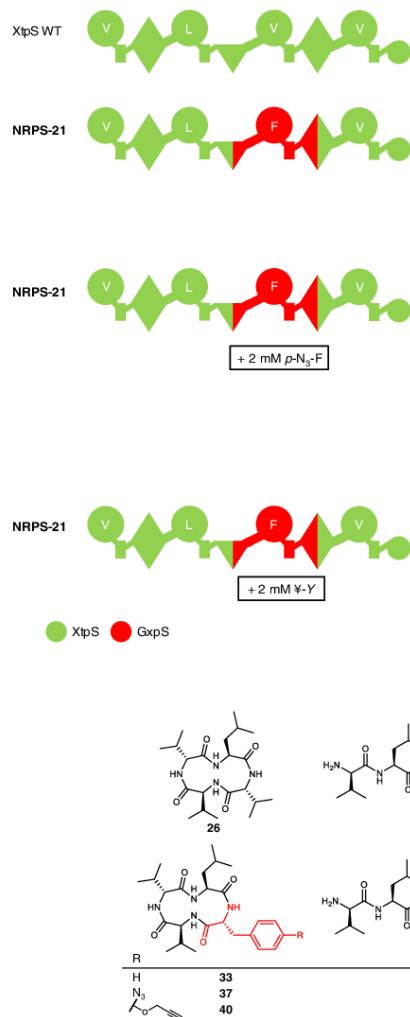
Peptide	Production (mg l⁻¹)	WT level (26) (mol%)
XtpS WT		
cyclo(vLvV) 26	134.1 ± 26.9	100
vLvV 27	0.4 ± 0.06	0.3
vLvV-OMe 28	1.3 ± 0.1	0.9
NRPS-21		
cyclo(vLvV) 26	7.5 ± 1.2	5.6
vLfV 29	5.6 ± 0.4	3.5
vLfV-OMe 30	24.4 ± 1.9	4.1
vLIV 31	0.2 ± 0.01	0.2
vLIV-OMe 32	0.6 ± 0.08	0.1
cyclo(vLfV) 33	106.6 ± 15.6	71.2
cyclo(vLIV) 34	80.1 ± 6.3	57.8
NRPS-21 + 2 mM p-N₃-F		
vL-[p-N₃-f]-V-OMe 35	46.0 ± 5.4	27.2
vL-[p-N₃-f]-V 36	283.5 ± 29.5	163.4
cyclo(vL-[p-N₃-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
NRPS-21 + 2 mM ¥-Y		
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

R			
H	33	29	30
N₃	37	36	35
(propargyl)	40	38	39

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.

a

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

b

Peptide	Production (mg l⁻¹)/ WT level (1) (mol%)
ILmL 45	0.8 ± 0.01
ILwL 46	4.4 ± 0.1
vLwL 47	1.4 ± 0.01
vLIV-OMe 32	0.7 ± 0.01
ILfL 48	9.9 ± 0.2
vLfV-OMe 30	2.2 ± 0.1
ILlL 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

c

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(ILwlL) 58	3.2 ± 0.1
cyclo(vLrlL) 42	4.2 ± 0.3

KolS   BicA   XeyS   XtpS   Xenolindicin-like
Pax   AmbS_miraniensis   AmbS_indica   GxpS

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.

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 synthesis 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 taxllaids 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).

Attachments

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

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

138

6.2.3 Supplementary information

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

140

2

3

142

144

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[TM] 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

7

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

8

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

10

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.

11

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	

12

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

13

152

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

14

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

Strain	Genotype	Reference
E. coli DH10B	F_ mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lac*ZΔ, M15, ΔlacX74 recA1 endA1 araD 139 Δ(ara, leu)7697 galU galK λrpsL (Strr) nupG	[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^H	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^H	[9]
E. coli DH10B::mtaA pFF1_*NRPS*_0	*E. coli* DH10B::mtaA pFF1_*NRPS*_0, Kan^H	[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^H	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^H	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^H	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^H	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^H	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

15

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	[9]
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*	[9]
pFF1_*xtpS*_WT	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *xtpS*	[9]
pFF1_NRPS_0	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bicA*-A1T1C2_*gxpS*-A2T2C3A3T3C4A4T4C5A5T5TE	[9]
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*-A1T1E2C2A2T2C3A3T3C$_{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

17

156

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

18

157

Supplementary Table 4. Oligonucleotides used in this work.

Plasmid	Oligonucleotide		Sequence (5'->3')	Template
pCOLA_gxpS_A3	pCOLA_Gib_A3 FW	Insert	CATCACCATCATCACCACCCTCAACAACCTGTCACGGC	P. luminescens TT01
	pCOLA_Gib_A3 RV	Insert	CAGCCTAGGTTAATTAAGCTGTTAAGTCAGATCAATCAGCGGCAAC	
	DUET_Gib_FW		CAGCTTAATTAACCTAGGCTG	pCOLADUET-1
	DUET_Gib_RV		GTGGTGATGATGGTGATG	
pFF1_grsTAB_WT	grs fw1		CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	grs rv1		GTCTTTCCATCCAACTGCAAC	
	grs fw2		CAGAAATCGAGATATTGTCTGAAG	A. migulanus ATCC9999
	grs rv2		ACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGA TCCGGCGGCCGGCCGAGCTCTTATTTTACTACAAATGTCCCTTGTAG	
pFF1_NRPS_1	KB-RT-6		TCATGAACTCGCCAGAACCAGCAGCGGAGCCGAGCCAGCGGATCCCAGCGGCCTCCGCTTCACAATTC	P. luminescens TT01
	KB-RT-7		TGGAATGCGACCGAAGAACC	
	KB-RT-8		CACATACCTGAGTAGGATACGGTTCTTCGGTCGCATTCCAAGTTTTCAGCAACAACTGGC	X. budapestensis DSM 16342
	KB-RT-9		TGTTTTGCCTGCATCGGAACCGCACGTTGTTGCTGGAAACGTGGAATACAACGGAAACTGC	
	KB-RT-10		CGTTTCCAGCAACAACG	P. luminescens TT01
	KB-RT-11		TTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG G	
pFF1_NRPS_2	KB-RT-6		TCATGAACTCGCCAGAACCAGCAGCGGAGCCGAGCCAGCGGATCCCAGCGGCCTCCGCTTCACAATTC	P. luminescens TT01
	KB-RT-7		TGGAATGCGACCGAAGAACC	
	KB-RT-15		AAGCCATTGACGCTGCCAG	X. budapestensis DSM 16342
	KB-RT-9		TGTTTTGCCTGCATCGGAACCGCACGTTGTTGCTGGAAACGTGGAATACAACGGAAACTGC	P. luminescens TT01
	KB-RT-10		CGTTTCCAGCAACAACG	
	KB-RT-11		TTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAG G	
	KB-RT-13		CACATACCTGAGTAGGATACGGTTCTTCGGTCGCATTCCAGTAATAACTCCCGCTC	P. luminescens TT01
	KB-RT-14		TCAATATCCTGATTATGCGGTCTGGCAGCGTCAATGGCTTTCAGGTGAAGGAGTACAGGC	
pFF1_NRPS_3	AL-GxpS-2-1		ACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTA AAAAGG	X. nematophila ATCC 19061
	AL-GxpS-2-2		CCCAATCAACATATCGGTAAAAAAGCGAGTATGTTCCATCTGGCTCACCCCCTGGTGGGCC	
	AL-GxpS-2-3		CCCGTACCTTTCCGTAATCTGGTCGCTCAGCGCCACCAGGGGGTGAGCCAGATGGAACATACT 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		CCGGTCCCGTTCCGCCATTTAGTGGCACAGGCTCGGCAGGGTGTCAAGGCGCGTGTTCTCACT	X. bovienii SS2004

	Primer	Sequence	Organism
	AL-GxpS-2-8	G CTCAGCCAACATTTCAGTAAAGAAACGGGTATGTTCAGCCTGACTCACGCTCAGGGTCTGGG	P. luminescens TT01
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGAATTCCATGAAAGATAGCATGGCTA	X. nematophila ATCC 19061
pFF1_NRPS_4	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATCTGGCTCACCCCCTGGTGGGCC AAAAGG	
	AL-GxpS-2-2	CCCAATCAACATATCGGTAAAAAGCGAGTATGTTCCATCTGGCTCACCCCCTGGTGGGCC	X. miraniensis DSM 17902
	AL-GxpS-2-3	CCCGTACCTTTCCGTAATCTGGTCGCTCAGCCCACCAGGGGGTGAGCCAGATGGAACATACT CG	
	AL-GxpS-2-4	CGTCCGACGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAAAACCACTGGCGTTGCC	P. luminescens TT01
	AL-GxpS-2-5	TCAGGTGAAGGAGTACAGGCAC	
	AL-GxpS-2-6	GACACCCTGCCGAGCC	
	AL-GxpS-2-11	GCCGGTCCCGTTCCGCCATTTAGTGGCACAGGCTCGGCCAGGGTGTCAGTCAGGAAGCCCACA CC	X. miraniensis DSM 17902
	AL-GxpS-2-12	CTCAGCCAACATTTCAGTAAAGAAACGGGTATGTTCAGCCTGACTCACGCTCAGGGTCTGGG	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	P. luminescens TT01
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCGCCTCGGTTCAC	
pFF1_NRPS_5	ML020	AGATTAGCGGATCCTACCTGACGCGCTTTTTATCGGCAACTCTCTACTGTTTCTCCATACCCGTTTTT TTGGGCTAACAGGAGGAATTCCATGAAAGATAACACATTGCTACACAGTG	pFF1_NRPS_0
	FF_305	CCAATAATCACTCTGTGCCTG	
	FF_306	GAACTGGTTGCACTTTACGC	
	FF_307	CATCCCTAACCGGGACTG	
	FF_308	CATACCTTGACGGACGAGGGCGGCAACCTGCCAGCGCCGGCCACCCTTCCGCAATCTGGTAGCG CAGTCCCGGTTAGGGATGAGTCAGGCAGCCCATACC	X. budapestensis DSM 16342
	FF_309	AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCC AGCGGATCCGGCGCGCCCTATTGCTCTGCTGATATCAGAA	
pFF1_NRPS_6	AT_105	CGGATCCTACCTGACGCGCTTTTTATCGGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	B. licheniformis ATCC 10716
	AT_107	GGCATGGCTATTTTCCCATTC	
	AT_157	GGAGCGCTGTATGAAGGGAAATCACTGAAGCCGTTCCATATTCAATATAAAGATTACGCAGAAT	B. subtilis MR 168
	AT_118	GGGAAAATAGCCATGCCCAAAAAGAACGGATGAAGGAGC AACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCC AGCGGATCCGGCGCGCCCTTATGAAACCGTTACGGTTTGTGTATTAAG	
pFF1_NRPS_7	FF_316	CGGATCCTACCTGACGCGCTTTTTATCGGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACACATATATTTTGC	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	

20

Construct	Primer	Sequence	Source
pFF1_NRPS_8	FF_316	TCGGGCGGCGCGAGCTCTTATTTTACTACAAATGTCCCTTGTAG CGGATCCTACCTGACGCGCTTTTTATCGGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	B. brevis ATCC 8185
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACC GAGGAATTTGCC	
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCGGCGCGCCTTATT TCAGGATGAACAGTTCTTGC	
pFF1_NRPS_9	FF_316	CGGATCCTACCTGACGCGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTTGC	A. migulanus ATCC9999
	FF_376	AAACCATTCGTTTTGCCATAC	B. brevis ATCC 8185
	FF_377	GAGCTTGCTGAACTGCATATTCAGTATAAAGATTTTGCTGTATGGCAAAACGAATGGTTTCAGTC GGATCGCTTCCAAAAAC	
	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 GATGTTCTGTCGTGGTCGACACCCTGCCGAGCC	X. nematophila HGB081
	ML_P6.1	ATCCTACCTGACGCGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCATGCGGGCAATGGTGAACC	
pFF1_NRPS_19	ML_P1	GACCAGACAGAACATCACCG	pFF1_gxpS_WT
	ML_P2	GGCCCAATCCTATACGCC	
	ML_P3	CTTACCAAGCGCCACAAGG	
	ML_P4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCA	
	ML_P5	TCGGTCAGCCCAAACGGTAATGTCGGTTCATCCACTTCTGCCAACATGTCGGTAAAGAATCGGT GATGTTCTGTCGTGGTCGACACCCTGCCGAGCC	X. nematophila HGB081
	ML_P6.3	ATCCTACCTGACGCGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCATGTCTGATGAAGGCGTGC	
pFF1_NRPS_20	ML_P1	GACCAGACAGAACATCACCG	pFF1_gxpS_WT
	ML_P2	GGCCCAATCCTATACGCC	
	ML_P3	CTTACCAAGCGCCACAAGG	
	ML_P4	AGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAG CGGAGCCAGCGGATCCCAGCGCCTCCGCTTCA	
	ML_P5	TCGGTCAGCCCAAACGGTAATGTCGGTTCATCCACTTCTGCCAACATGTCGGTAAAGAATCGGT GATGTTCTGTCGTGGTCGACACCCTGCCGAGCC	X. nematophila HGB081

21

	Name	Sequence	Organism
	ML_P6.2	ATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC GGGAGGAATTCCATGGCATTTACCGAAAAGATCTGCG	X. nematophila HGB081
pFF1_NRPS_21	ML_P7	CGGATCCTACCTGACGCGTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCT AACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAGGG	X. nematophila HGB081
	ML_P8	CTATCGGCAATTCAAGTAACACCGGTGCATCTGCCAACGTCCGACGCGCCAATAATCACTCTGTGC CTGTACTCCTTCACCTGAAAAATACCTGCCGCTGCC	
	ML_P9	GCTTGTCTGAATCAACAACCTGATCCGCTGCCGCCATTGACCATTCAATATCCTGATTATGCTG CCTGGCAGCGGCAGGTATTTTCAGGTGAAGGAGTACAGGC	
	ML_P10	GCACTTCCGACAATCCAAATGACAACGTTGGCTCATCTACCTCAGCCAACATATCGGTAAAGAA ACGGGTATGTTCTGCCTGACTGACACCCTGCCGAGCC	
	ML_P11	GGCTTGCCTCTTGGGGCAAATGGATAGCCTGCCTGCCGGTCCCGTTCCGCCATTTAGTGGC ACAGGCTCGGCAGAGGTGTCAGTCAGGCAGAACATACCCG	X. nematophila HGB081
	ML_P12	CCAGAATCGGAACAACACCGGTAAACAGTTCTTCACCTTTGCTCATGAACTCGCCAGAACCAGC AGCGGAGCCGAACGCGGATCCCAGCGCCTCCACTTCG	
pFF1_library_1	KB-AmbF-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGTTGAACGCTTACAAT CC	X. miraniensis DSM 17902
	KB-AmbF-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGATACCCAGCCGGGCCTTG	X. indica DSM 17382
	KB-AmbW-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCATCGGAACGGGGTACAAA TTC	
	KB-AmbW-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACCCCCATCCGTGCCTG	P. luminescens TT01
	KB-Thr1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGCACAGATACAGT CTC	
	KB-Thr2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Arg1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGG TGCAG	X. budapestensis DSM 16342
	KB-Arg2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	
	KB-Ser1	GTCTAAATCAACAGCCAGATCCGTTGTTGCCATTGACCATTCAATATCCTGATTATGCGGCTTG	X. szentirmaii DSM 16338
	KB-Ser2	GCAGCGGCAGGTATTTCAGGGTGACCGCCTGAC CCGCCAACCCAAATAGCAGCGTCGGTTCATCCACCTCAGCCAACATGTCAGTAAAGAAGCGA GTATGTTCTGCCTGACTCACACTCAGGATTTGAGCGATAAAG	
	KB-Lys1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGAGGTACTGGAA AAGC	X. nematophila HGB081
	KB-Lys2	CTCAGCCAACATGTCAGTAAAGAAGCGGAGTATGTTCTGCCTGACTTACACTGCGGGGTTTGGGC	
	AL-GxpS-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTA AAAAGG	pFF1_gxpS_WT
	AL-GxpS-2	AAATACCTGCCGCTGCC	
	AL-GxpS-3	AGTCAGGCAGAACATACTCGCTTCTTTAC	
	AL-GxpS-4	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCGAACGCGGATCCCAGCGCCTCCGCTTCAC	
pFF1_library_2	KB-xeyS-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGCTGCTCGTCTGGTCTTCCCCAACCAGGACTG	X. indica DSM 17382
	KB-xeyS-N	ACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAAACATGGCTAC AACG	
	KB-BicA-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCCATCCCCAACCAGGACTG	X. budapestensis DSM 16342

22

Primer	Sequence	Organism
KB-BicA-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTGG	X. miraniensis DSM 17902
KB-17902-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCAACGCCCAGCCGCGGGCTTGAGC	
KB-17902-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAAATGATAAGGTGATGACTCTGC	X. nematophila HGB081
KB-2022-C	TTCTGCCAACATGTCGGTAAAGAATCGGTGATGTTCTGTCTGGTCCACCCCCTGGTGGGCC	
KB-2022-N	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAGG	X. szentirmaii DSM 16338
KB-XLSer1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGACCGCCTGAC	
KB-XLSer2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACACTCAGGATTTGAGCGATAAAG	X. miraniensis DSM 17902
KB-AmbF-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCGGTTGAACGCTTACAATCC	
KB-AmbF-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGATACCCAGCCGGGCTTG	X. indica DSM 17382
KB-AmbW-1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCATCGGAACGGGTACAAATTC	
KB-AmbW-2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACCCCCATCCGTGCCTG	P. luminescens TT01
KB-Thr1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGCACAGATACAGTCTC	
KB-Thr2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	X. budapestensis DSM 16342
KB-Arg1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGGTGCAG	
KB-Arg2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	X. nematophila HGB081
KB-Lys1	ACCATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGAGGTACTGGAAAAGC	
KB-Lys2	CTCAGCCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTTACACCTGCGGGTTTGGGC	pFF1_gxpS_WT
AL-GxpS P3	AGTCAGGCAGAACATACTCGCTTCTTTAC	
AL-GxpS P4	TTTGCTCATGAAACTCGCCAGAACCAGCAGCGGAGCCCAGCGGATCCCAGCGCCTCCGCTTCAC	
KB-Lib3-1	GACCAGCACAGAACATCACCG	
KB-Lib3-2	AAATACCTGCCGCTGCC	
pFF1_library_3 KB-XL-X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACACTCAGGATTTGAGCG	X. szentirmaii DSM 16338
KB-XL-X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCAGGGTGACCGCCTGACC	
KB-XI-amb X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACTACCTAGACGTGCCTGTGC	X. indica DSM 17382
KB-XI-amb X3 FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTTCGGCAGAACGGATACAAATTC	
KB-Kol X3 - RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACCGCCCAACCGGACC	P. luminescens TT01
KB-Kol X3 - FW	ATTCAATATCCTGATTATGCGGCTTGGCAGCGGCAGGTATTTCGGCAGCACAGATACAGTC	
KB-Kol X2 RV	AAATACCTGCCGCTGCCAGCCGCATAATCAGGATATTGAATCGTCAACGGTAGCAACGG	P. luminescens TT01
KB-Kol X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGTCAGGGTCAGGCATACCCG	
KB-BicA X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTCACGCCCAACCGGACC	X. budapestensis

23

Primer	Sequence	Template
KB-BicA X3 FW	ATTCAATATCCTGATTATGCGGGCTTGGCAGCGGCAGCGGCAGGTATTTTCTGCTGATCGTATTCAGGTGC	DSM 16342
KB-Bb 2 RV	TCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGGATCCCAGCGCCTCCGCTTCACAATTC	pFF1_gxpS_WT
KB-Bb 2 FW	AGTCAGGCAGAACATACTCGC	
KB-Bb 1 RV	AACCCCCTGCCGAGCC	pFF1_gxpS_WT
KB-Bb 1 FW	TTCTCCCATACCCGTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	
KB-amb X2 RV	AAATACCTGCCGCTGCCAAGCCGCATAATCAGGATATTGAATTGCCAATGGTGGCAAGGG	X. indica DSM 17382
KB-amb X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGCCAGACAGAGCACACCCG	
KB-17902 X3 RV	CCAACATGTCAGTAAAGAAGGGAGTATGTTCTGCCTGACTGAGCCGGGGCTTGTGC	X. miraniensis DSM 17902
KB-17902 X3 FW	ATTCAATATCCTGATTATGCGGGCTTGGCAGCGGCAGGTATTTTCGGTTGAACGCTTACAATCC	
KB-2022 X3 RV	CCAACATGTCAGTAAAGAAGCGAGTATGTTCTGCCTGACTGACACCCTGCCGAGCC	X. nematophila HGB081
KB-2022 X3 FW	ATTCAATATCCTGATTATGCGGGCTTGGCAGCGGCAGGTATTTTCTGATGAAGGCGTGCAGG	
KB-2022 X2 RV	AAATACCTGCCGCTGCCAAGCCGCATAATCAGGATATTGAATGGTCAATGGCGGCAGCGCG	X. nematophila HGB081
KB-2022 X2 FW	ACCTTTCCGCAATCTGGTGGCTCAGGCTCGGCAGGGGGTTAGTCAGGAAGCGTACACGCG	

24

163

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[H]	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[H]	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[H]	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

26

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	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	B. brevis ATCC 8185
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTGGCATAATCGCTTGCTTCAGACCGA GGAATTTGCC	
	AL_C9_rv	AGCCAGCTTGGTCTGCC	
	AL_odl_fw	CCAGTACAAAGACTTTGCTGTGTGGCAGACCAAGCTGGCTGCTGATGGCATATTTGATGC	X. nematophila ATCC 19061
	AL_odl_rv	TGCAACCAGTACGCTTCCTGCTTGTGGAAGGCAGCCGACTGTAATTGTTGTCGCTGCCAG	B. brevis ATCC 8185
	AL_C10_fw	CAGTCGGCTGCCTTCC	
	AL_tycC_rv	CTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_NRPS_11	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	AL_xabdou_fw	CAATATAAAGACTTTACTGTTGGCATAATCGCTTGCTTCCGGAAGCTAAACTGACGG	X. doucetiae DSM 17909
	AL_xabdou_rv	CCGGGTCCAAAAATCCTCCTGTTTTGGAAGGCGATCCGACTGCAGCCTGTCACGCTGC	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	B. brevis ATCC 8185
	AL_tycC_rv	CTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_NRPS_12	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTGGCATAATCGCTTGCTTCAGACCGA GGAATTTGCC	B. brevis ATCC 8185
	AL_C8_rv	GAACAGTTCAGACTGCCAGAC	
	AL_txl_fw	CCATTACAAAGATTTCGCCGTCTGGCAGTCTGAACTGTTCCAGGGGGAGGTGGGGG	X. bovienii SS2004
	AL_txl_rv	GGGTCCAAAAATCCTCCTGTTTTGGAAGGCGATCCGACTGCAAATAGTCACGCTGCCATTT	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	B. brevis ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCTTTCAGGATGAACA GTTCTTGC	
pFF1_NRPS_13	FF_316	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAAC AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGCGATTATGCCAAAC	

27

Construct	Primer	Sequence (5'→3')	Template
	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	AGGAGGAATTCCAAGCAATTTATTCTTACATATATTTTGC	A. migulanus ATCC9999
	FF_373	AAGCAAGGCGATTATGCCAAAC	
	FF_374	GAACTTCCTACGTTAGGCATTCAATATAAAGACTTTACTGTTTGGCATAATCGCTTGCTTCAGACCGAGGAATTTGCC	B. brevis ATCC 8185
	AL_C8_rv	GAACAGTTCAGACTGCCAGAC	
	AL_xabnem_fw	CGTCCATTACAAAGATTTCGCCGTCTGGGCAGTCTGAACTGTTCCAGGGCAATGCCCTGAC	X. nematophila ATCC 19061
	AL_xabnem_rv	GGGTCCAAAAATCCTCCTGTTTTTGGAAGGCGGATCCGACTGCAAACATATCATGTTGCCAGACAG	
	AL_C9_fw	CAGTCGGATCGCTTCCAAAAAC	B. brevis ATCC 8185
	FF_375	CTTCACCTTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCGAGCGGATCCTTTCAGGATGAACAGTTCTTGC	
pFF1_NRPS_15	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	B. licheniformis ATCC 10716
	ALAT_1	CGTCCGACGCGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAATTAAATGTATGATTCCATTCCACATAATC	
	AT_99	TCAGGTGAAAGGAGTACAGGCAC	
	ABCR14_neu	GAAACGGGGTATGTTCAGCCTGAC	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	pFF1_gxpS_WT
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	pFF1_gxpS_WT
pFF1_NRPS_16	AT_105	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATG	B. licheniformis ATCC 10716
	ALAT_2	CGTCCGACGGCCAATAATCACTCTGTGCCTGTACTCCTTCACCTGAGGCATGGCTATTTTCCCATTC	
	AT_99	TCAGGTGAAAGGAGTACAGGCAC	
	ABCR14_neu	GAAACGGGGTATGTTCAGCCTGAC	
	AL-GxpS-2-9	AGTCAGGCTGAACATACCCG	pFF1_gxpS_WT
	AL-GxpS-2-10	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCAGCGCCTCCGCTTCAC	pFF1_gxpS_WT
pFF1_NRPS_17	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	pFF1_gxpS_WT
	AT_217	AAATAACCTGCCGCTGCC	
	ALAT_3	ACCATTCAATATCCTGATTATGCGGGCTTGGCAGCGGCAGGTATTTCAGTCGGATCGCTTCCAAAAAC	B. brevis ATCC 8185
	ALAT_4	TTTGCTCATGAACTCGCCAGAACCAGCAGCGGAGCCGAGCGGATCCTTTCAGGATGAACAGTTCTTGCAGG	

28

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**).

29

168
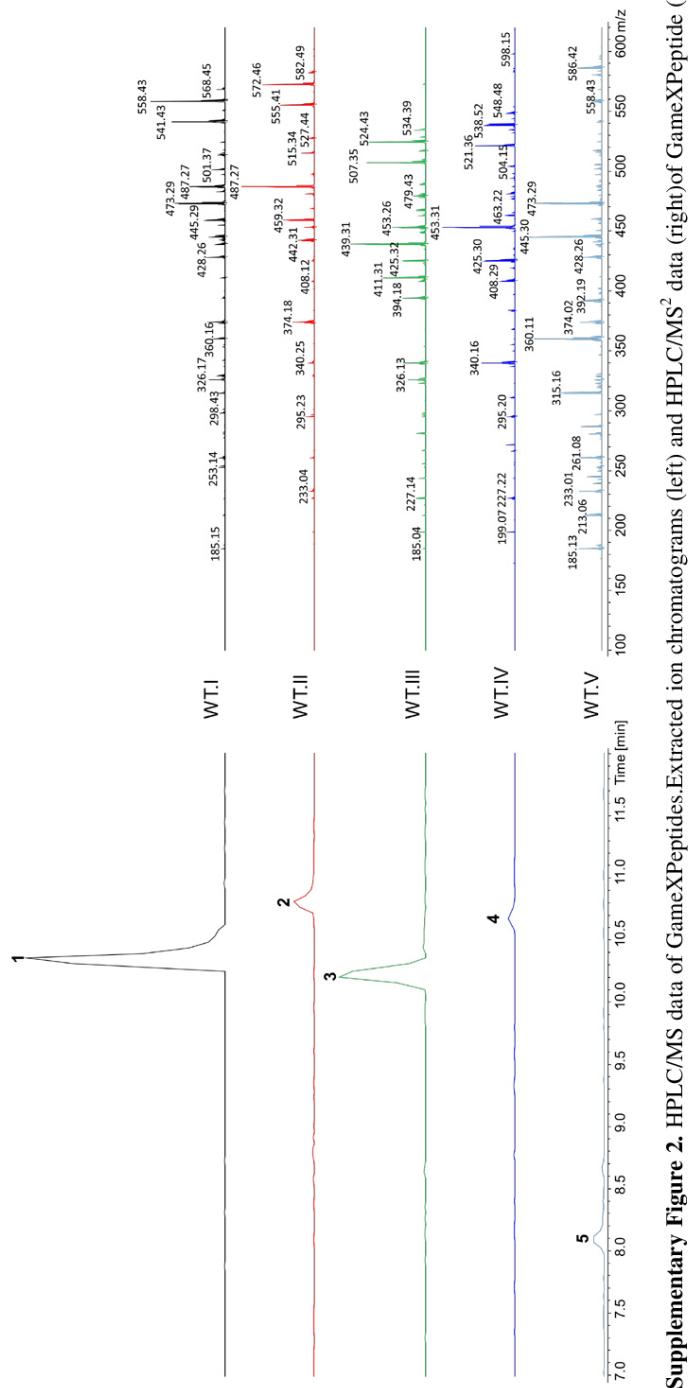
Supplementary Figure 2. HPLC/MS data of GameXPeptides.Extracted ion chromatograms (left) and HPLC/MS^2 data (right)of GameXPeptide (**1**, *m/z* [*M*+H]$^+$ = 586.4) and its derivatives**2** (*m/z* [*M*+H]$^+$ = 600.4), **3** (*m/z* [*M*+H]$^+$ = 552.4), **4** (*m/z* [*M*+H]$^+$ = 566.4) and the linear GameXPeptide**5** (*m/z* [*M*+H]$^+$ = 604.4) all from *E. coli* DH10B::mtaA extracts.

30

169
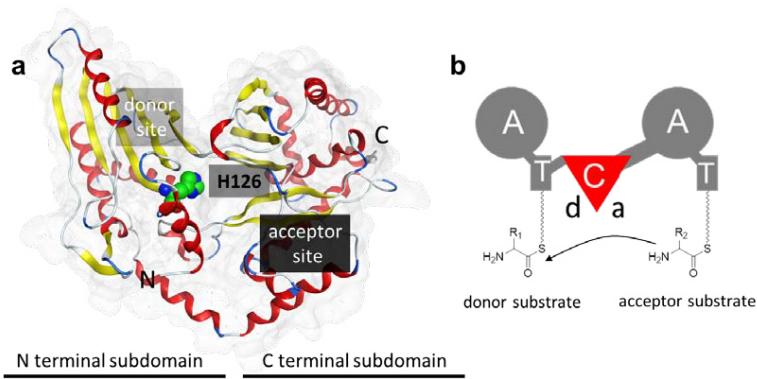
Supplementary Figure 3. Structure and mechanism of a C domain.(**a**) Crystal structureof VibH, a C domain of *V. cholera* vibriobactin synthethase (PDB-ID: 1L5A)[13],subdivided into N terminal subdomain (donor site) and C terminal subdomain (acceptor site). The catalytic center (H126) is highlighted 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).

31
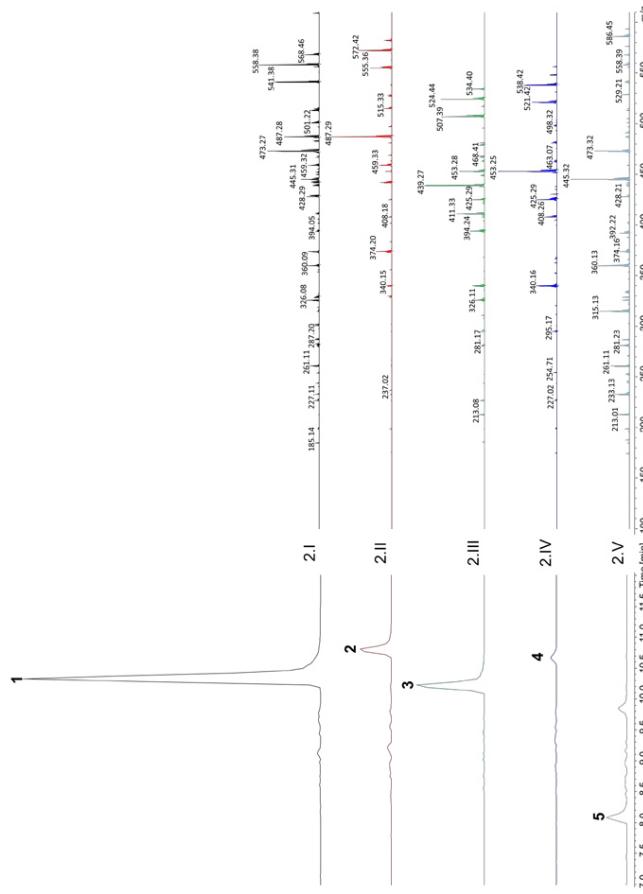
Supplementary Figure 4. HPLC/MS data of compounds **1-5** produced by NRPS-2 in *E. coli* DH10B::mtaA.Extracted ion chromatograms (left) and HPLC/MS² data (right) of GameXPeptide (**1**, *m/z* [*M*+H]⁺ = 586.4) and its derivatives **2** (*m/z* [*M*+H]⁺ = 600.4), **3** (*m/z* [*M*+H]⁺ = 552.4), **4** (*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.

32

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.

33
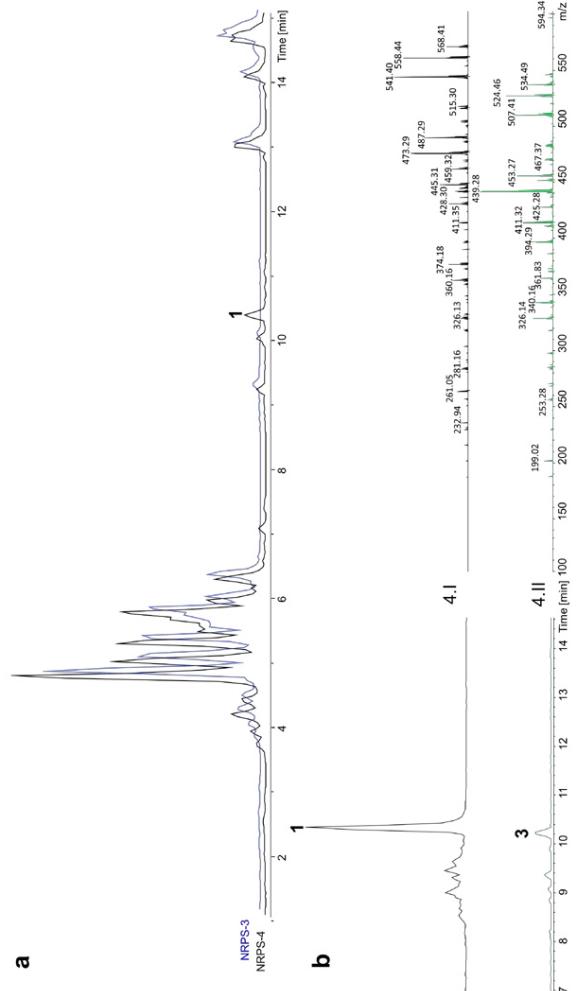
Supplementary Figure 6.HPLC/MS data of compounds **1** and **3** produced by NRPS-4 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-3 and NRPS-4 (Figure 2) in *E. coli* DH10B::mtaA. (**b**)EIC (left) and HPLC/MS[2] data (right) of GameXPeptide (**1**, *m/z* [*M*+H]⁺ = 586.4) and its derivative**3** (*m/z* [*M*+H]⁺ = 552.4) produced by NRPS-4 and (Figure 1b) in *E. coli* DH10B::mtaA.

Attachments
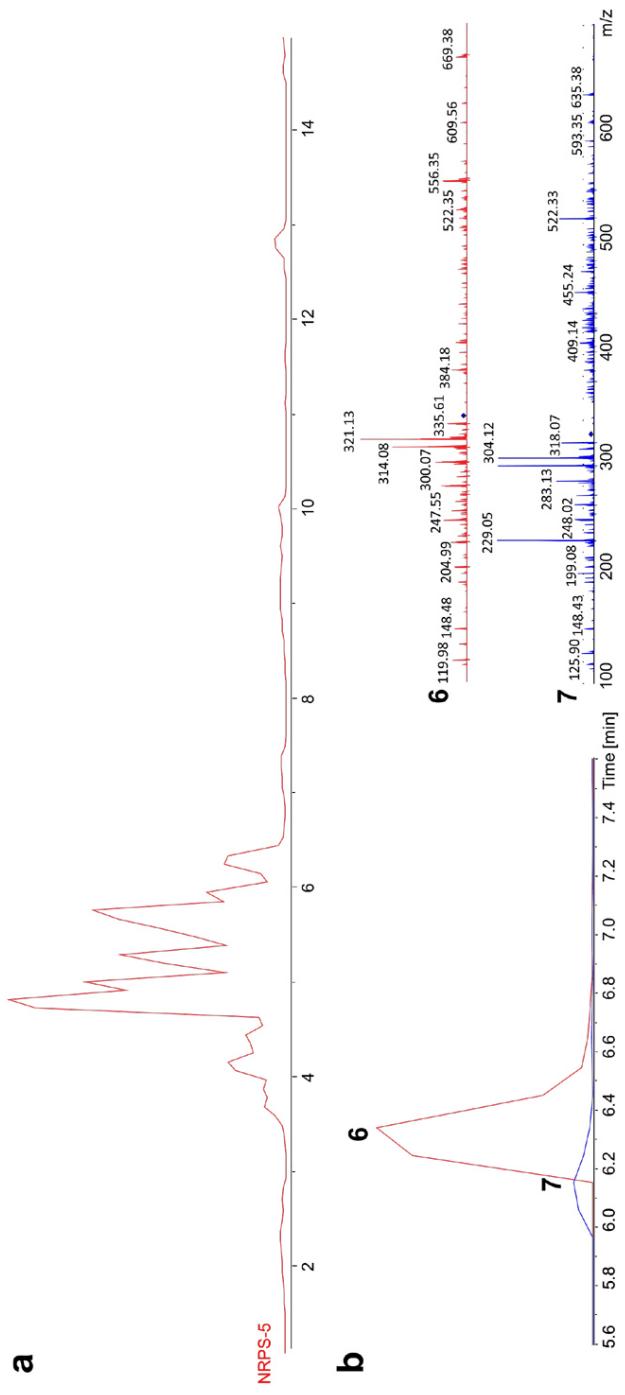
Supplementary Figure 7.HPLC/MS data of compounds **6** and **7** produced by NRPS-5 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-5 (Figure 2) in *E. coli* DH10B::mtaA. (**b**) EIC (left) and HPLC/MS^2 data (right) of arginine containing GameXPeptide **6**, *m/z* $[M+2H]^{2+}$ = 352.7 and its derivative **7** (*m/z* $[M+2H]^{2+}$ = 335.7).

174

35
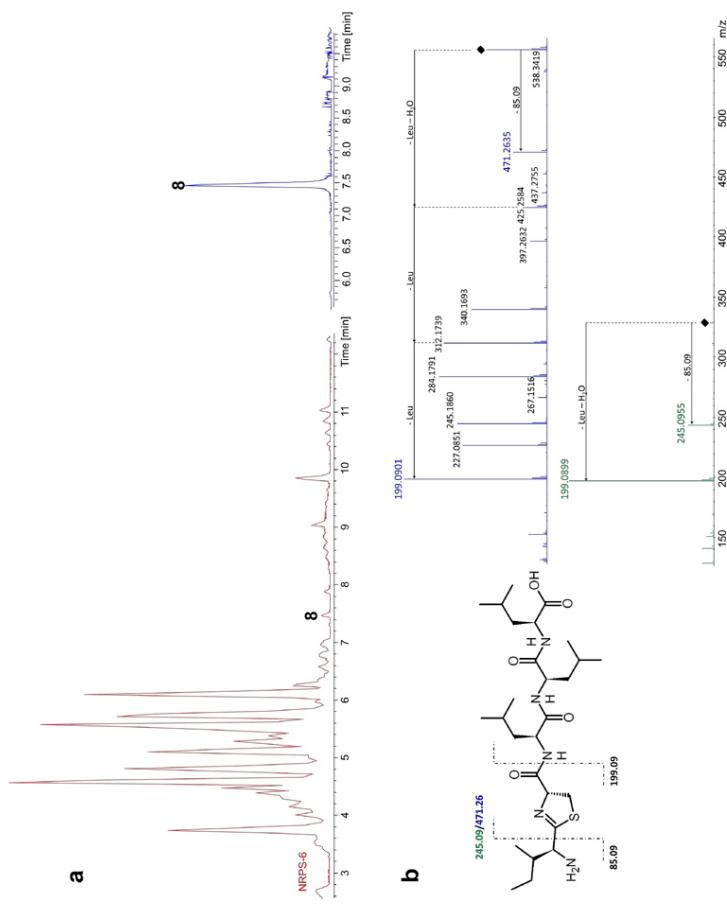
Supplementary Figure 8.HPLC/MS data of compound **8** produced by NRPS-6 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram (left) of production from NRPS-6 (Figure 2) in *E. coli* DH10B::mtaA and EIC (right) of **8** (*m/z* [*M*+H]⁺ = 556.35). (**b**) Structure, fragmentation and HPLC/HR-MS² data of **8**(blue) compared to Ile-Thiazoline-Leu (green; *m/z* [*M*+H]⁺ = 330.18)⁹.

36

175
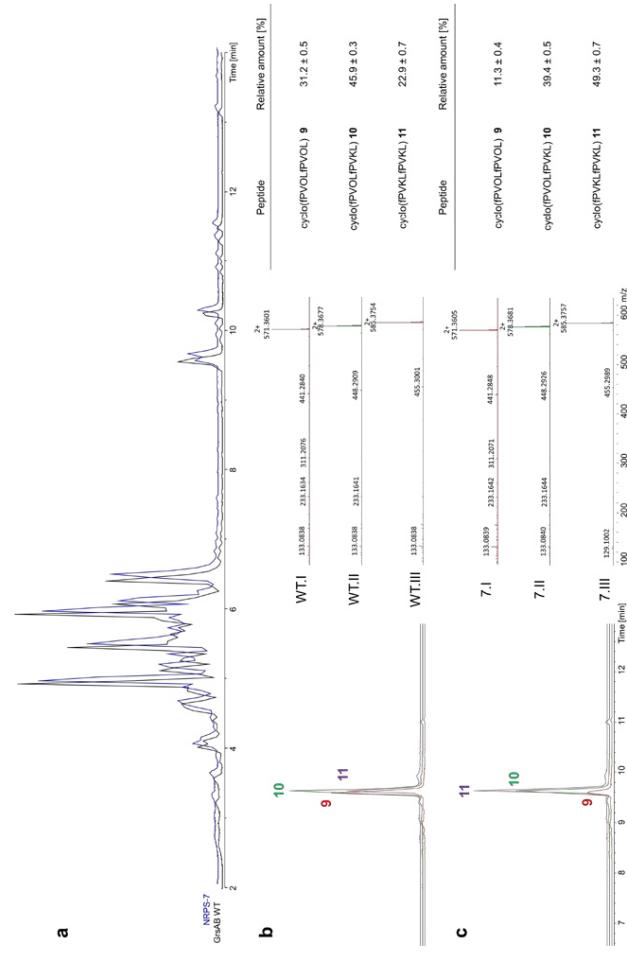
Supplementary Figure 9. HPLC/MS data of compounds **9-11** produced by GrsAB and NRPS-7 in *E. coli* DH10B::mtaA. (**a**) Basepeak chromatogram of production from NRPS-7 (Figure 2) and Gramicidin S-producing synthetase GrsAB in *E. coli* DH10B::mtaA. (**b**) overlaid EIC (left) and HR-HPLC/MS² data (middle) 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/MS² 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.

37
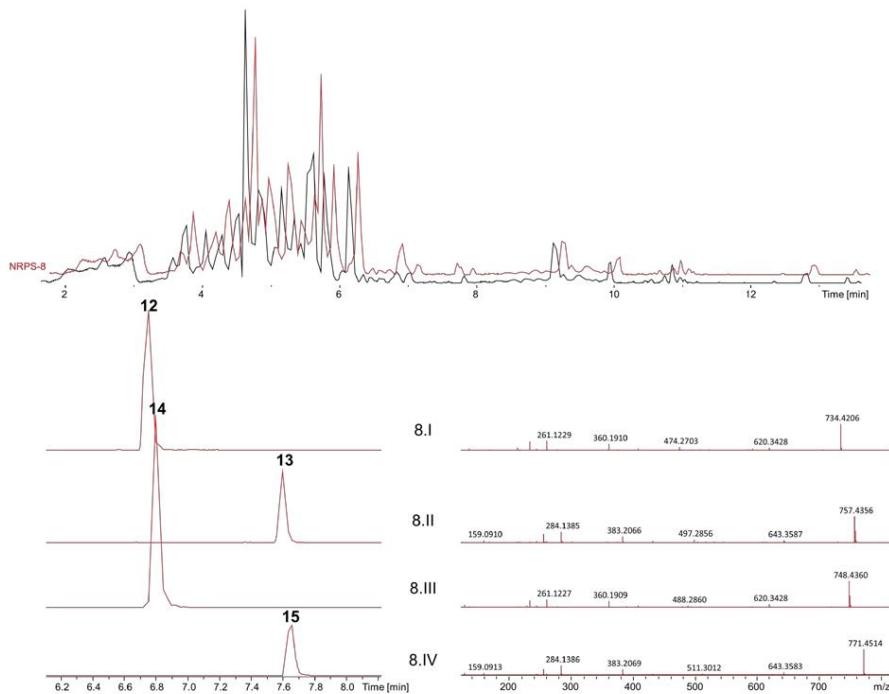
Supplementary Figure 10. HPLC/MS data of compounds **12-14** produced by NRPS-8 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-8 (Figure 2) in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/HR-MS[2] data (right) of **12** (m/z $[M+H]^+$= 734.420), **13** (m/z $[M+H]^+$ = 757.4396), **14** (m/z $[M+H]^+$ = 748.435) and **15** (m/z $[M+H]^+$ = 771.451).

38
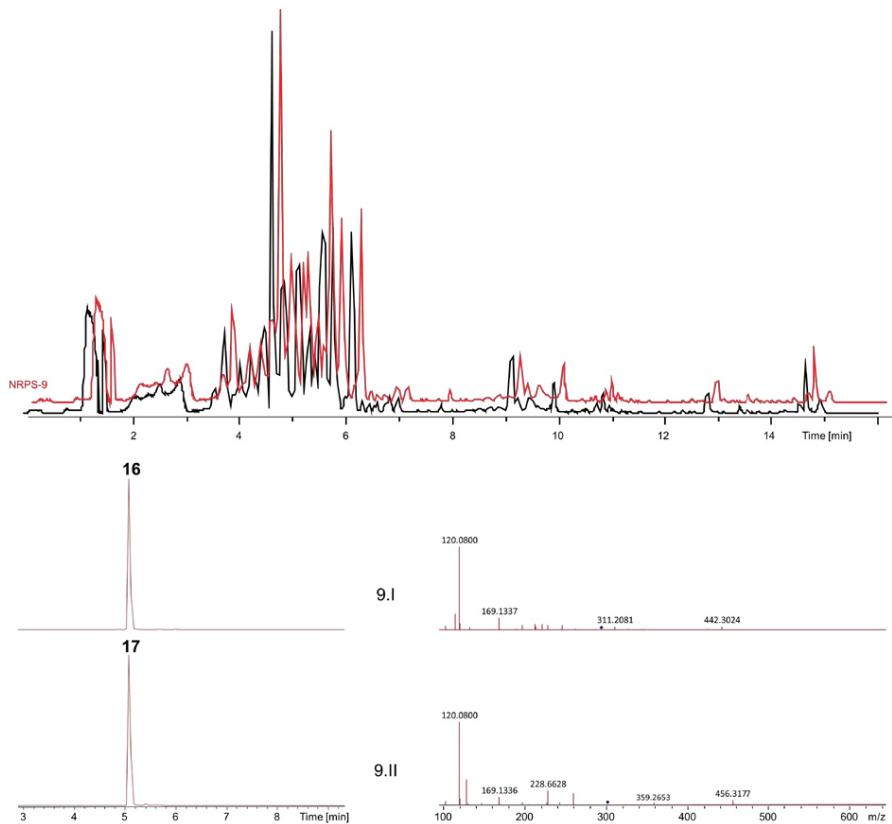
Supplementary Figure 11. HPLC/MS data of compounds **16** and **17** produced by NRPS-9 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-9 (Figure 2) in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/HR-MS[2] data (right) of **16** (*m/z* $[M+2H]^{2+}$ = 295.189) and **17** (*m/z* $[M+2H]^{2+}$ = 302.197).

39
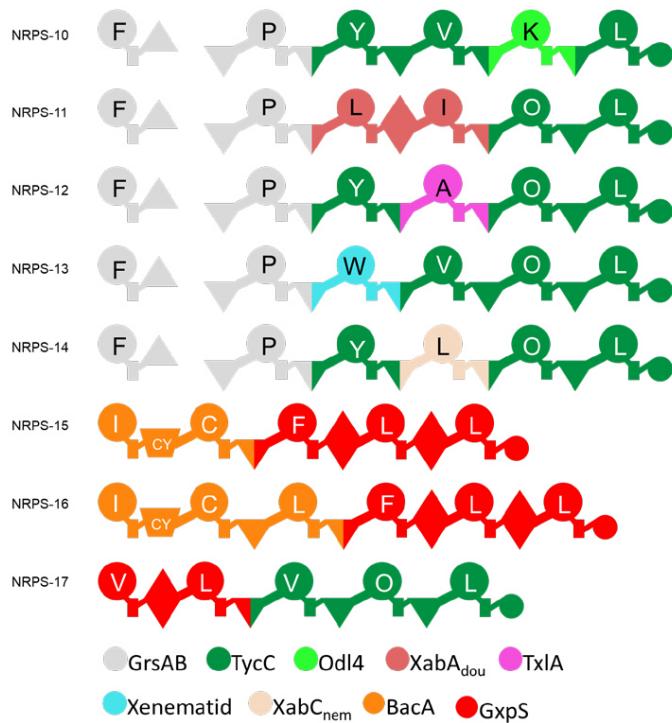
Supplementary Figure 12. Generated recombinant NRPS from Gram-positive and -negative origin. For assignment of domain symbols see Fig. 1; further symbols are E (epimerization; inverted triangle), CY (heterocyclization; trapezium). Bottom: Color code of NRPS used as building blocks (for details see Supplementary Figure 5).

40
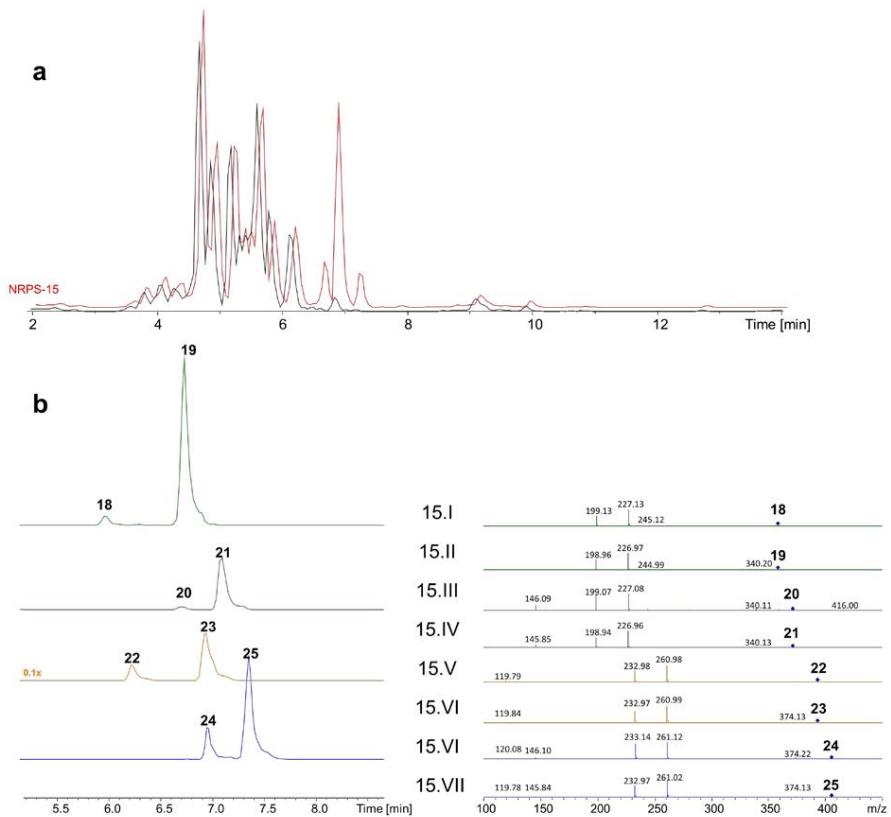
Supplementary Figure 13.HPLC/MS data of compounds **18-25** produced by NRPS-15 in *E. coli* DH10B::mtaA.(**a**) Basepeak chromatogram of production from NRPS-15 in *E. coli* DH10B::mtaA (red: induced, black non-induced). (**b**) EIC (left) and HPLC/MS^2 data (right) of **18** (*m/z* [*M*+H]$^+$ = 358.22), **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.

41
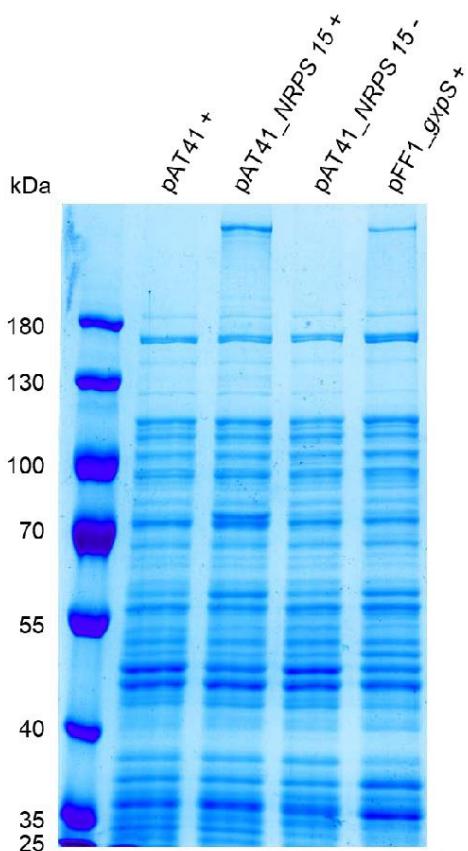
Supplementary Figure 14. SDS-PAGE assay of protein extracts of *E. coli* DH10B::mtaA harboring pAT41 with no insert (control), pAT41_*NRPS 15* and pFF1_*gxpS*. All samples marked with "**+**" were inoculated with 0.02% arabinose and "-" without arabinose. The expected molecular weight for NRPS 15 and GxpS is 572 kDa and 514 kDa, respectively.

42
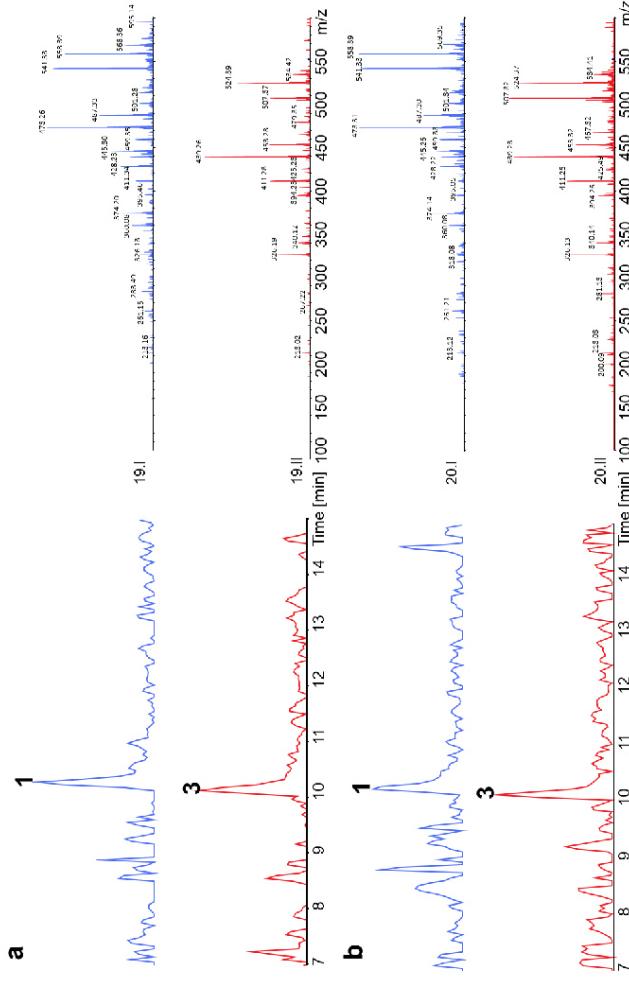
Supplementary Figure 15.HPLC/MS data of compounds **1** and **3** produced by NRPS-19 and -20 in *E. coli* DH10B::mtaA.(**a**) EIC (left) and HPLC/MS[2] data (right) of GameXPeptide **1**, *m/z* [*M*+H]⁺ = 586.4 and its derivative **3** (*m/z* [*M*+H]⁺ = 552.4) produced by NRPS-19 (Figure 3) in *E. coli* 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.

43
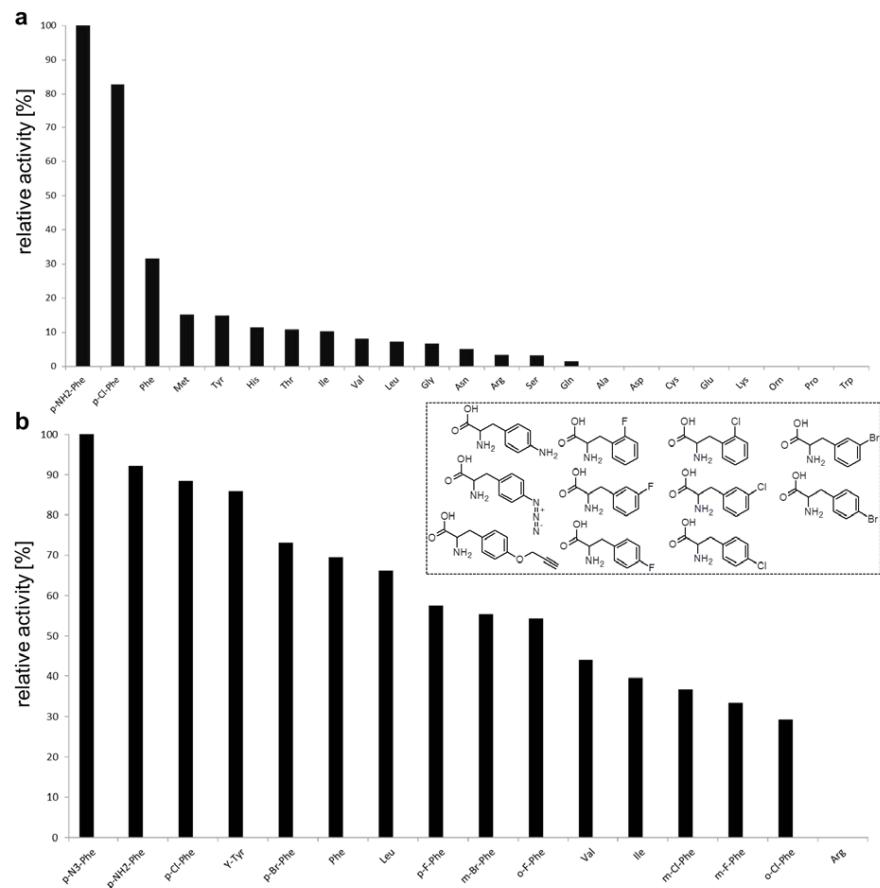
Supplementary Figure 16.*In vitro* adenylation activity of GxpS_A3.(**a**)Adenylation activity[3] of GxpS_A3tested with proteinogenic AAs and *para* substituted phenylalanine.The activities are calculated relative to the substrate with the highest activity (*p*-NH$_2$-Phe). (**b**) GxpS_A3 adenylation 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).

44
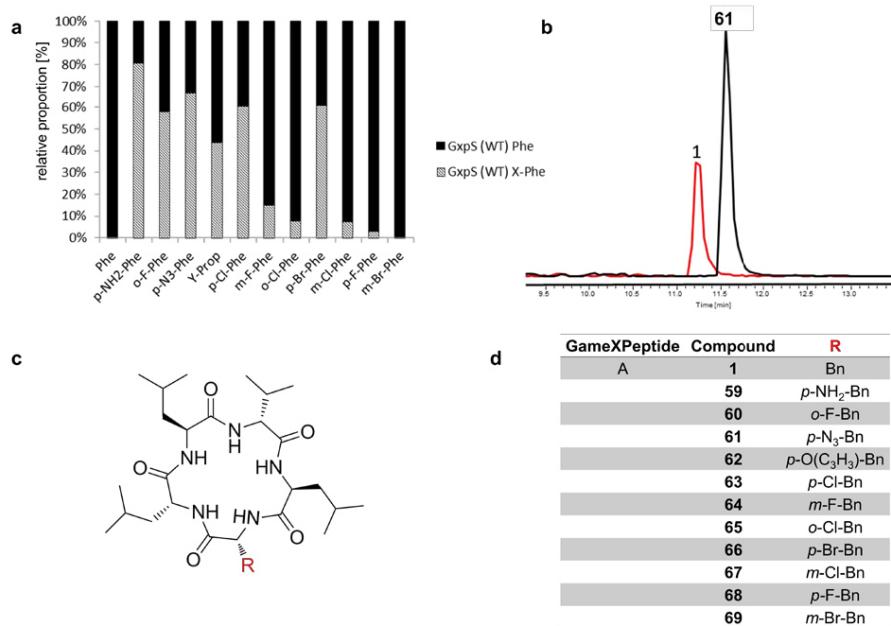
Supplementary Figure 17. *In vivo* characterization of GxpS in *E. coli* DH10B::mtaA. (**a**) *In vivo* feeding experiments with the GameXPeptide producing WT and mutant (W239S) GxpS. The relative proportions of the HPLC/MS detected signals of peptides (**1, 59-69**) are shown, according to the supplemented 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.

45
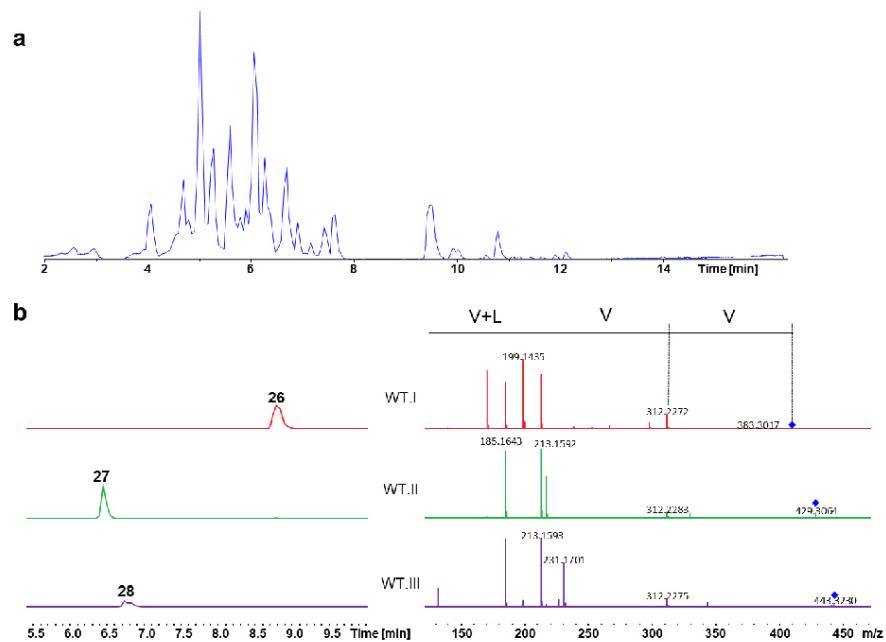
Supplementary Figure 18. Heterologous production of xenotetrapeptide in *E. coli* DH10B::mtaAand HPLC/HR-MS analysis. (**a**) Base peak chromatogram (blue) and (**b**) extracted ion chromatograms (left) of **26**(*m/z* [*M*+H]$^+$ = 411.29), **27** (*m/z* [*M*+H]$^+$ = 429.30), **28** (*m/z* [*M*+H]$^+$ = 443.32) and MS-MS spectra.

46
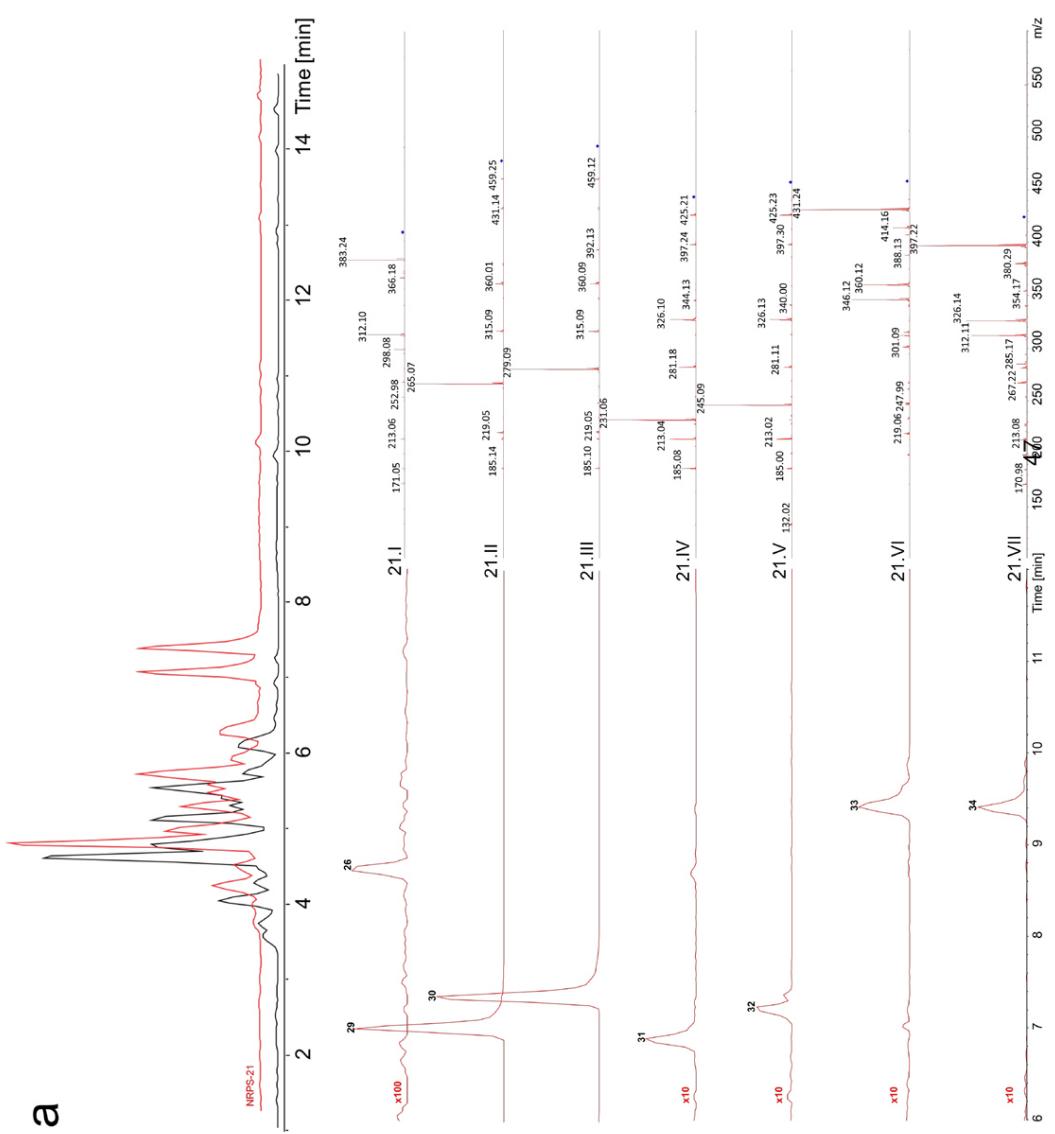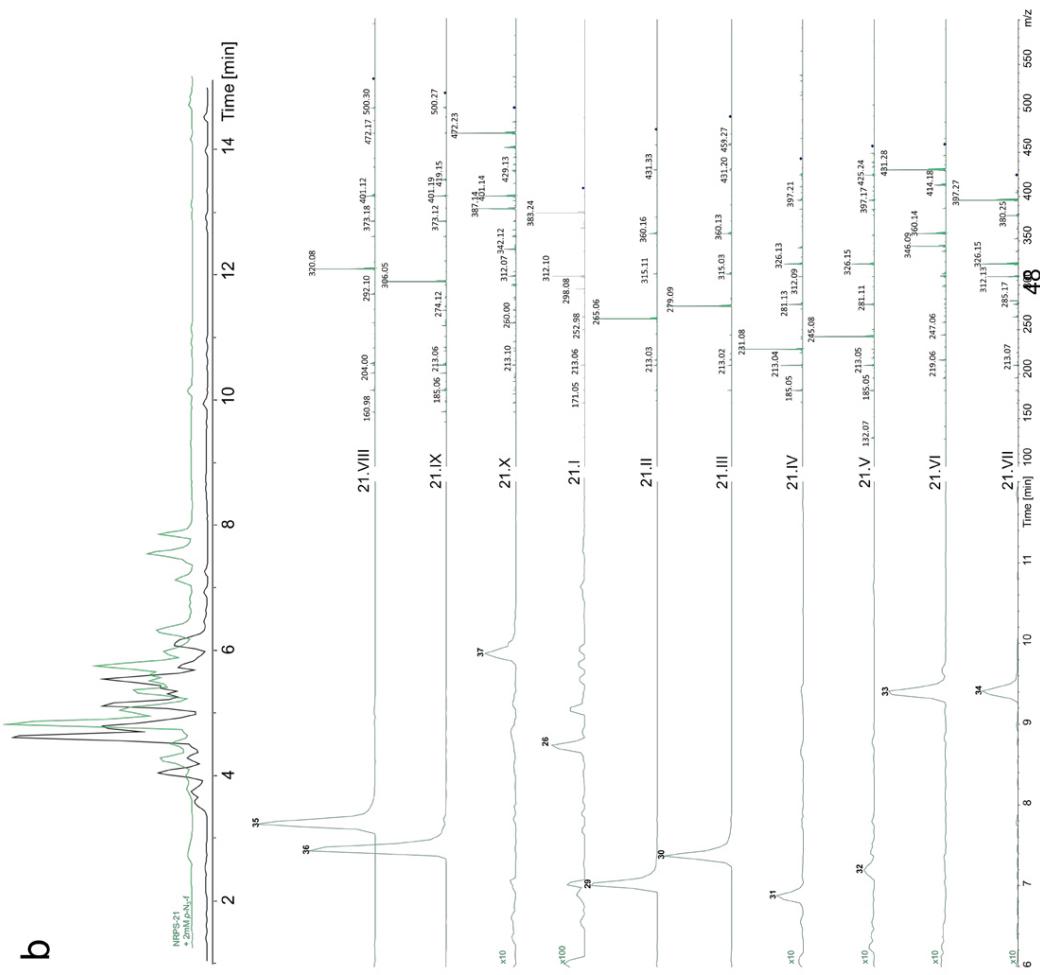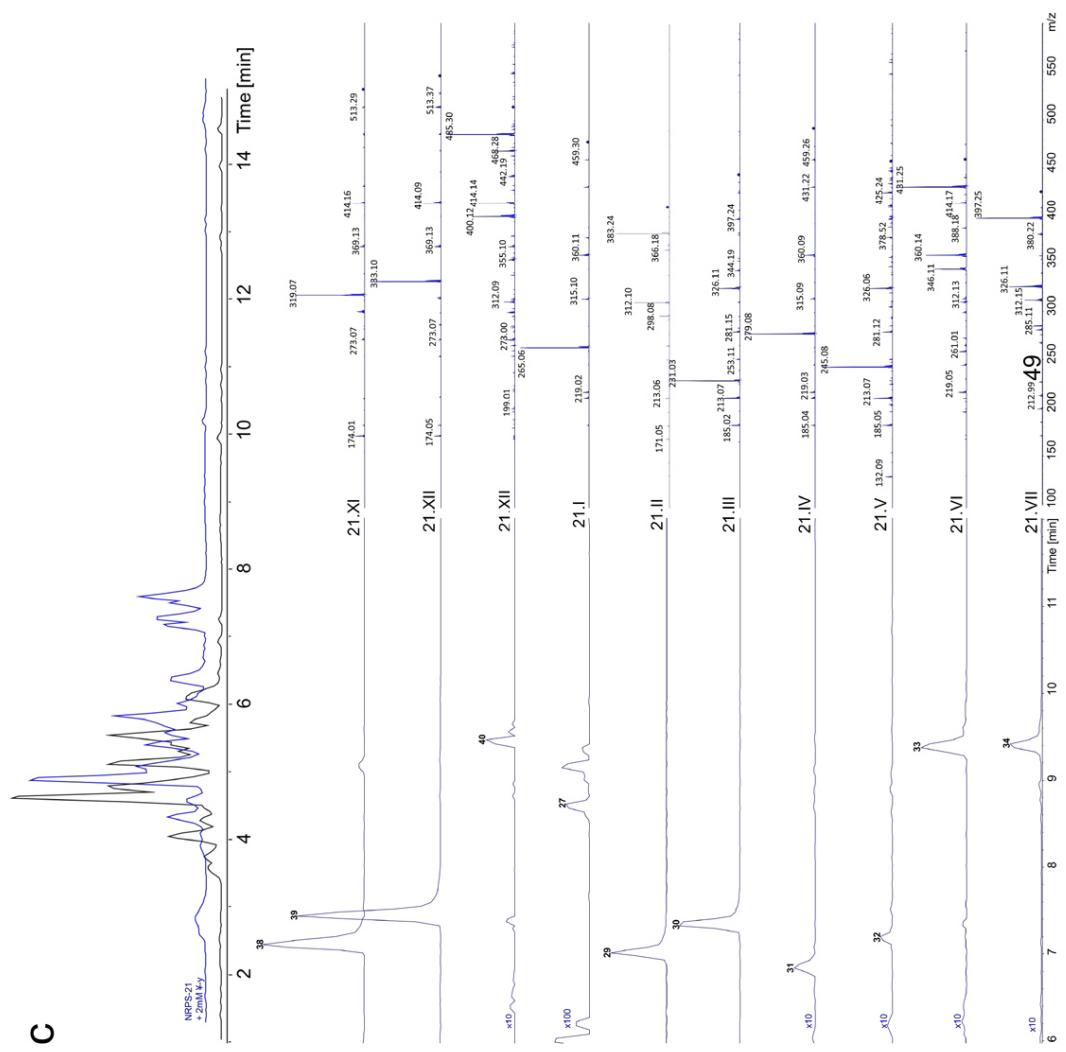
Supplementary Figure 19. HPLC/MS data of xenotetrapeptide derivatives produced by NRPS-21 in *E. coli* DH10B::mtaA.BPCs (top), EICs (left) and HPLC/MS^2 data (right). (**a**)**26** (*m/z* [*M*+H]$^+$ = 411.2)**29** (*m/z* [*M*+H]$^+$ = 477.2), **30** (*m/z* [*M*+H]$^+$ = 491.2),**31** (*m/z* [*M*+H]$^+$ = 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).

50

189
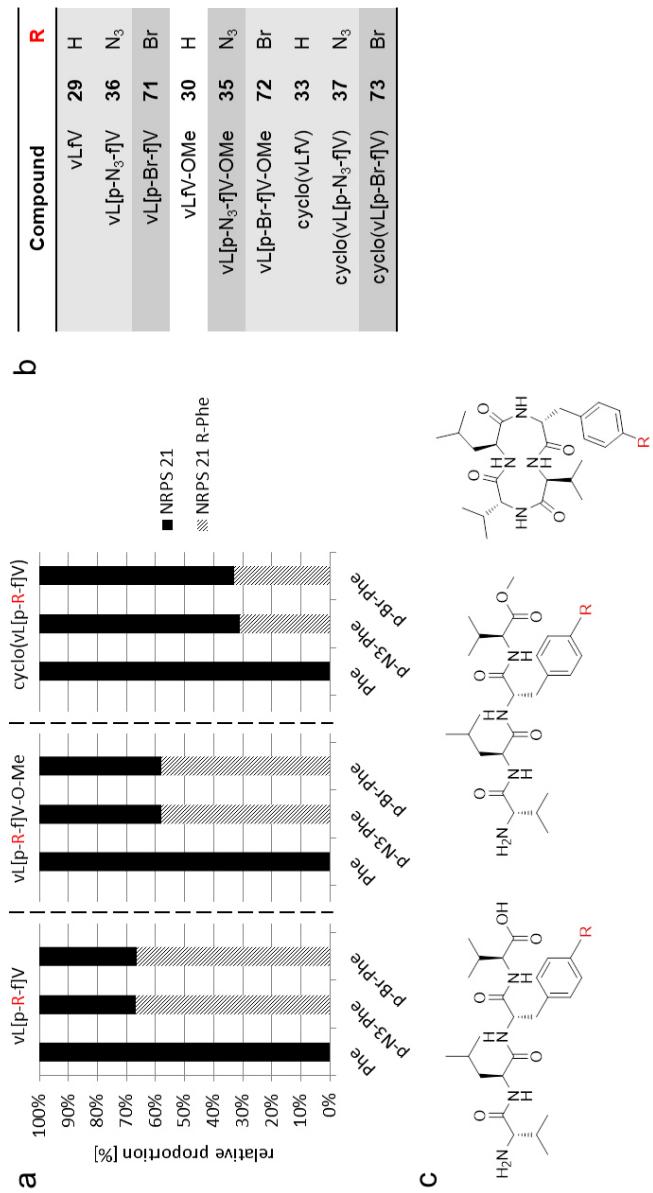
Supplementary Figure 20. *In vivo* characterization of NRPS-21 in *E. coli* DH10B::mtaA. (**a**) *In vivo* feeding experiments with NRPS-21 incorporating substituted Phe derivatives (R). The relative proportions of the HPLC/MS detected signals of peptides (**29, 30, 33, 35-37,71-73**) are shown, according 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.

51
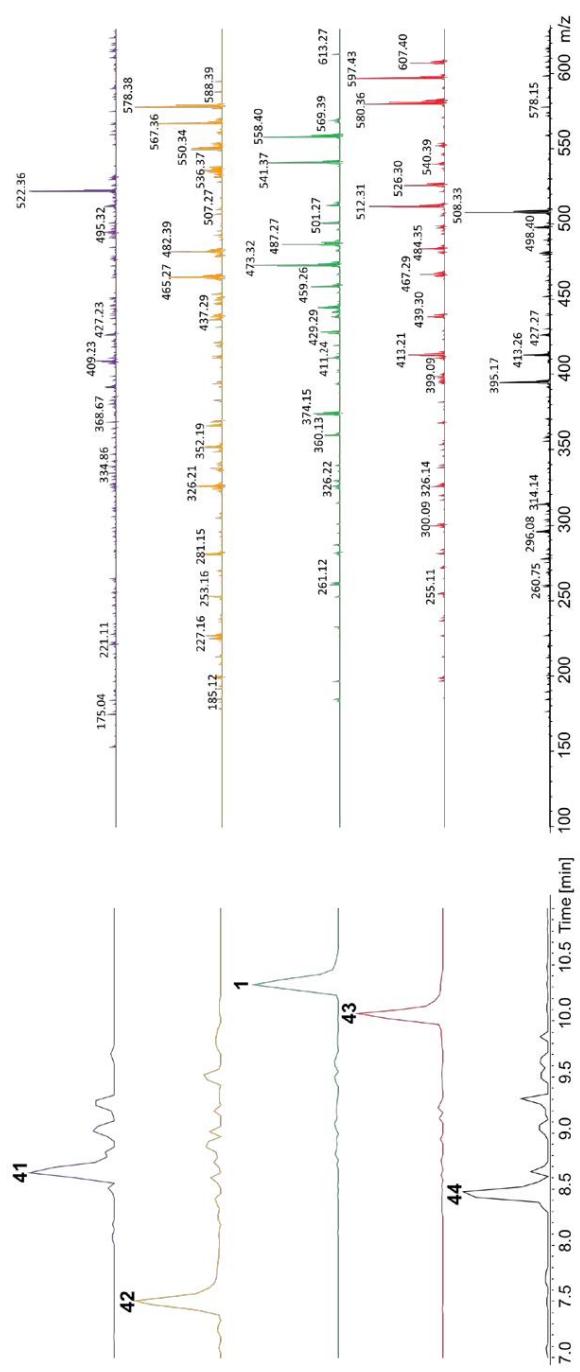
Supplementary Figure 21. Targeted randomization of GxpS at position three. EIC (left) and HPLC/MS² data (right) of GameXPeptide derivatives produced by library 1 (Figure 4)**41** (*m/z* [*M*+H]⁺ = 540.3),**42** (*m/z* [*M*+H]⁺ = 595.4),**1** (*m/z* [*M*+H]⁺ = 586.4)**43** (*m/z* [*M*+H]⁺ = 625.4)and **44** (*m/z* [*M*+H]⁺ = 526.3) in *E. coli* DH10B::mtaA.

c

52
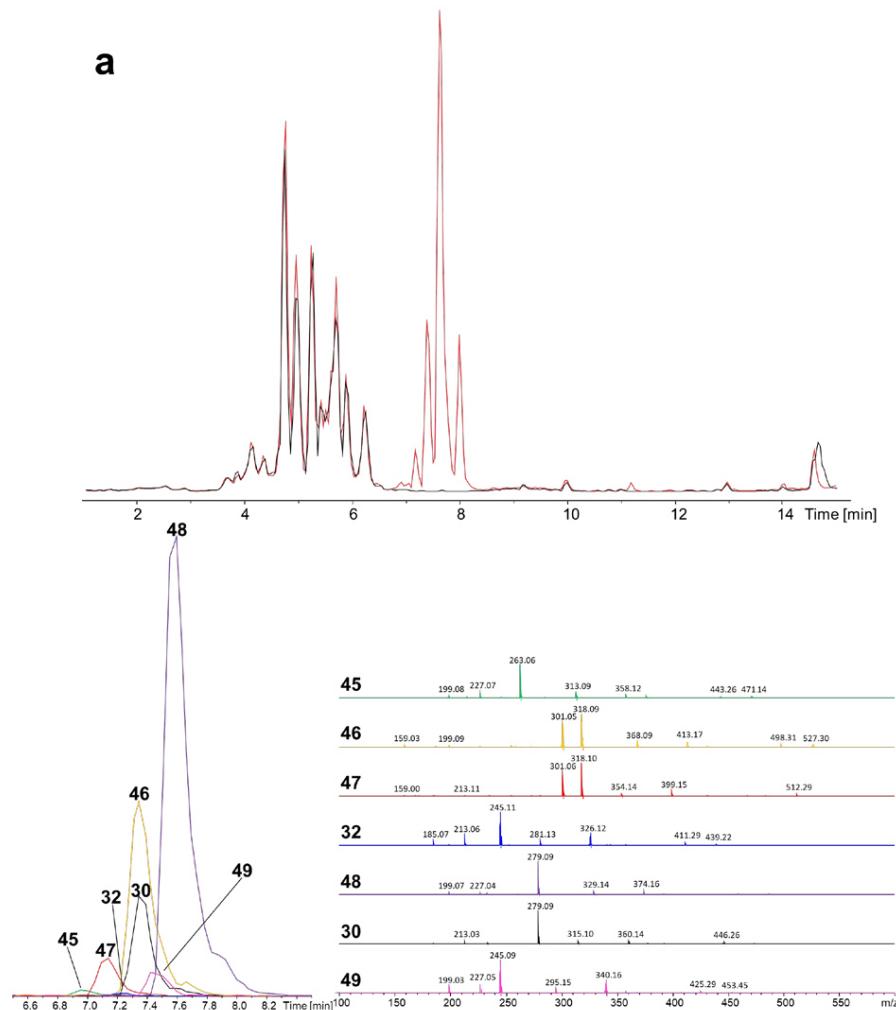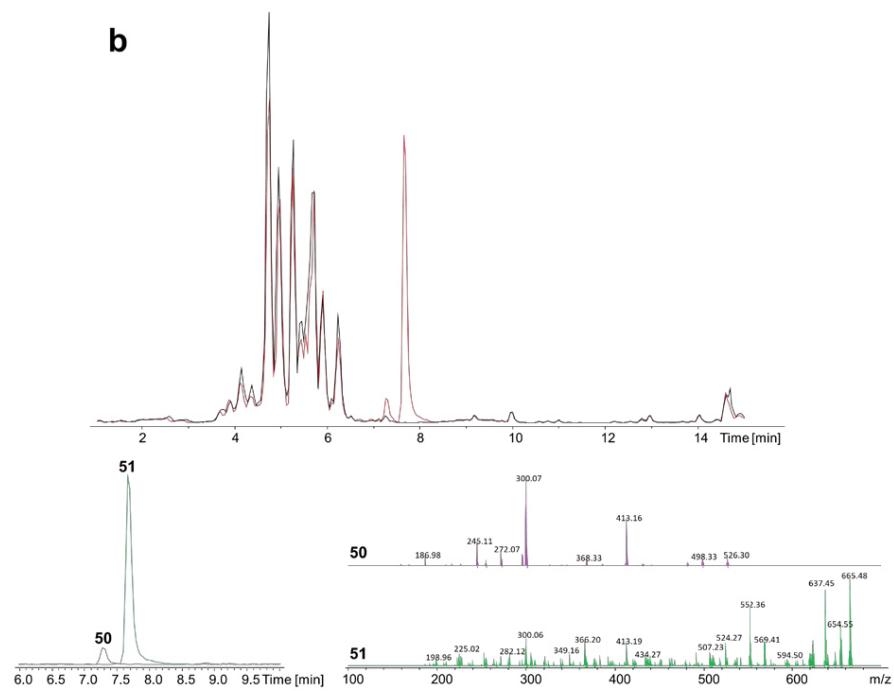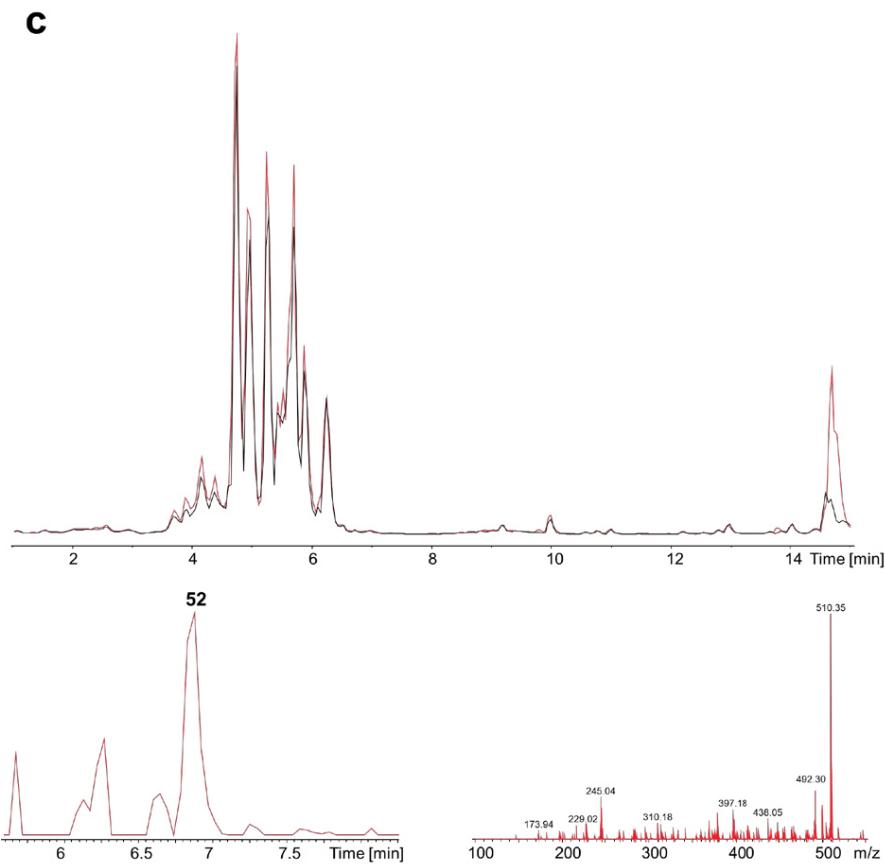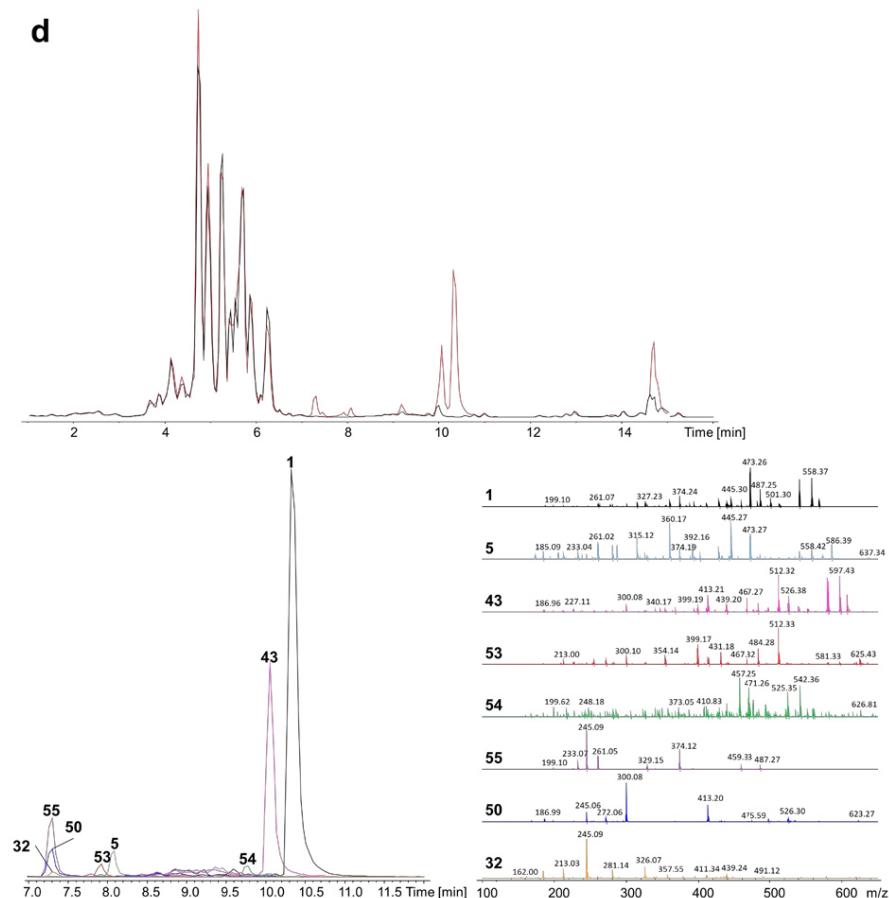
Supplementary Figure 22. The creation of a library of GxpS where position one and three were randomized. (**a**) Top: BPCs of production from Lib2_NRPS-1in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom:EICs (left) and HPLC/MS^2 data (right)of peptides produced by 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

56

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).
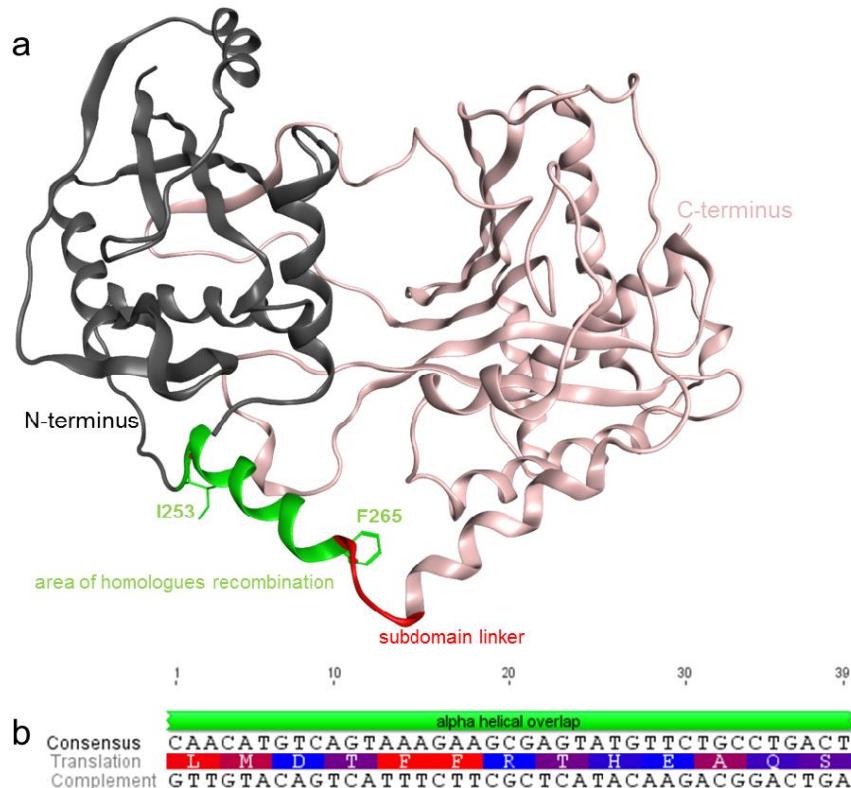
Supplementary Figure 23. Design of an artificial α5 helix.(**a**) Crystal structure of TycC6 (PDB-ID: 2JGP)[8], subdivided into N terminal subdomain (grey) and C terminal subdomain (light red). The subdomain linker is highlighted in red and the targeted area (I253 – F265) for homologous recombination in yeast is highlighted in green (39 nucleotides). (**b**) Consensus sequence used to generate library three (Figure 5c).

57

a

58

b
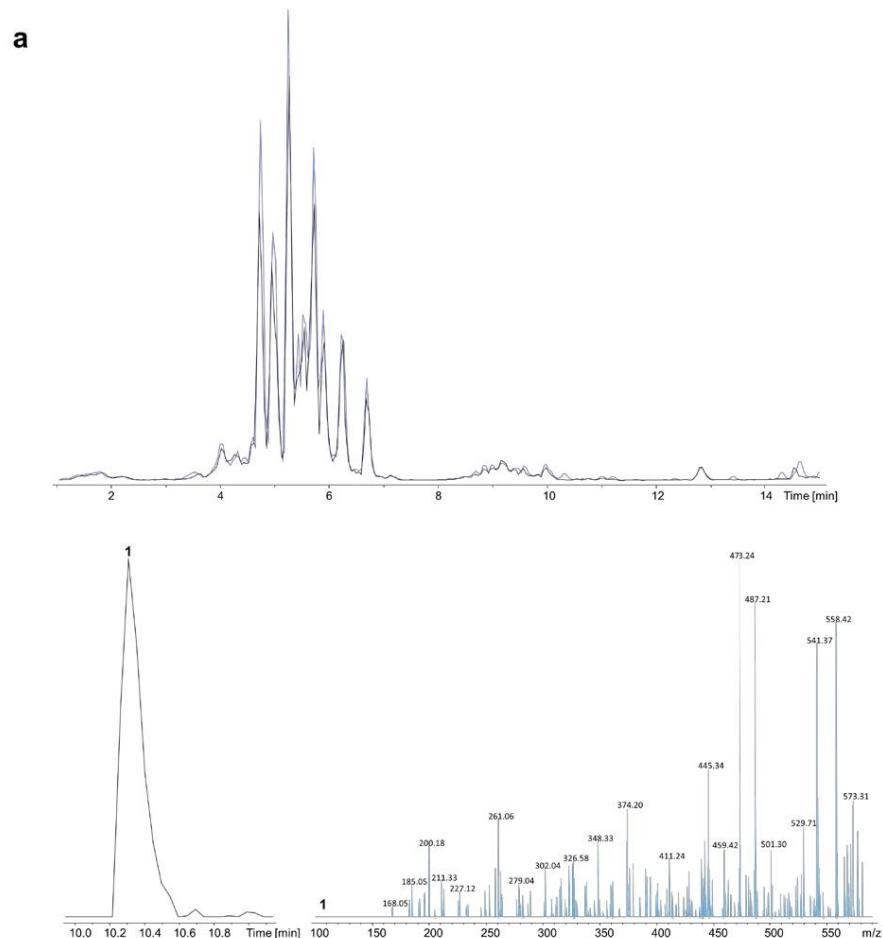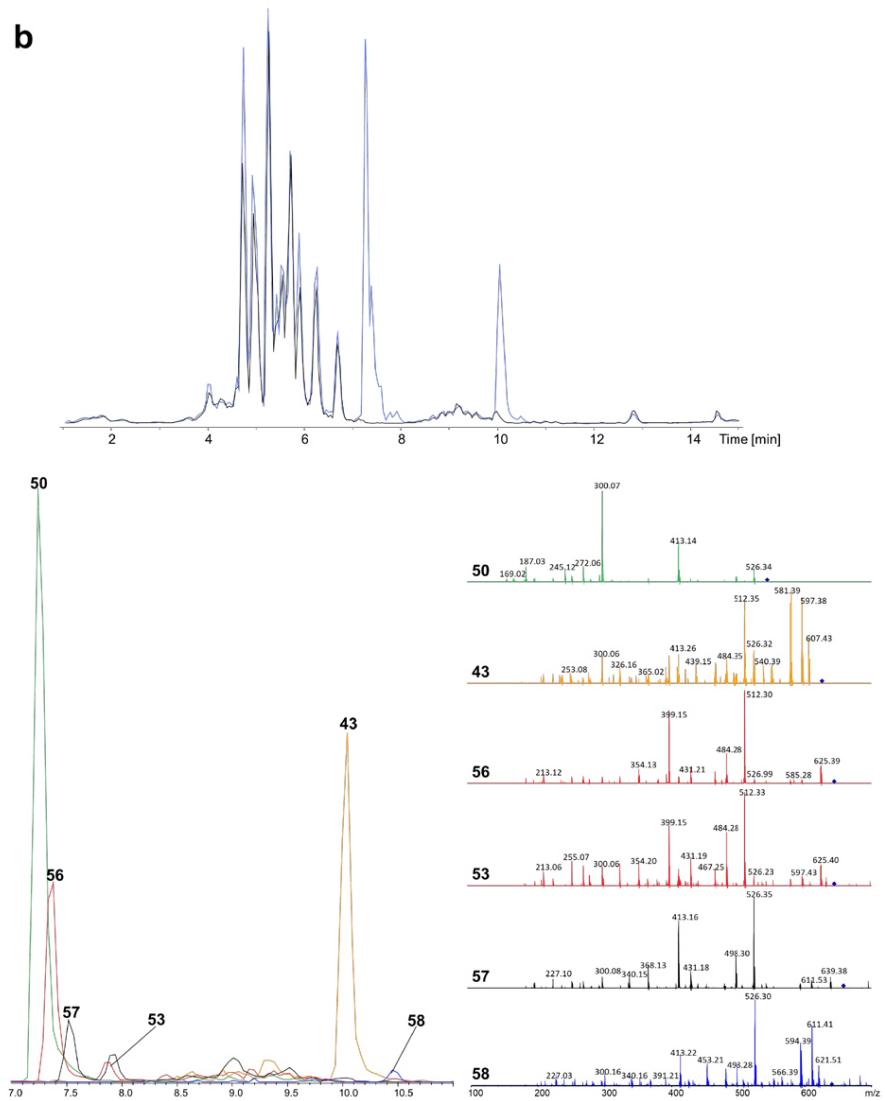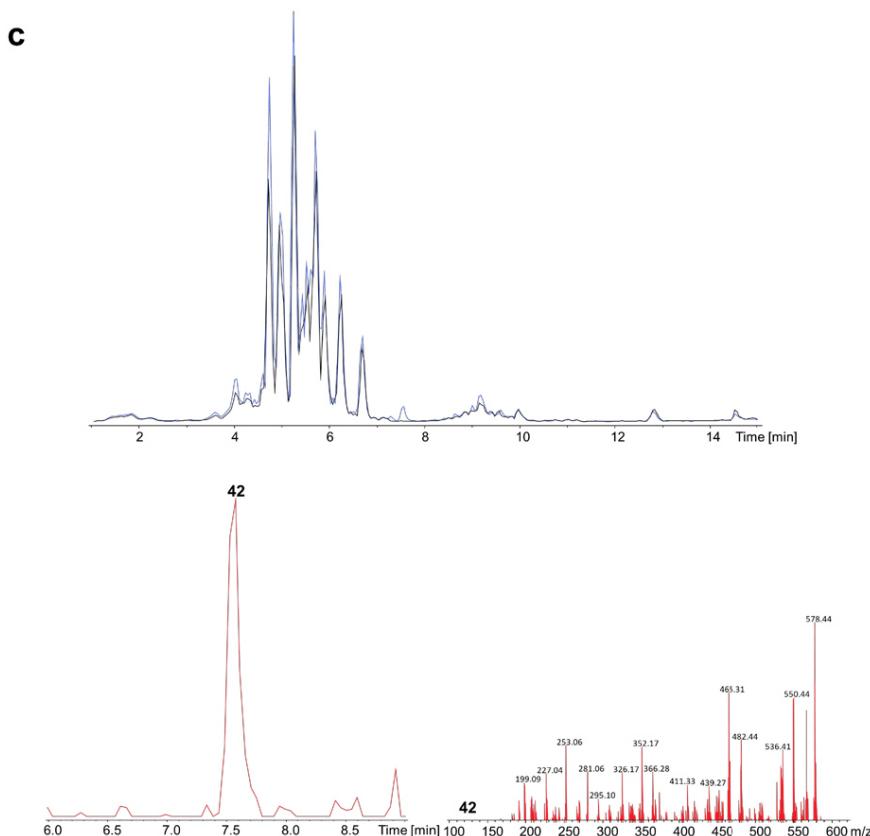
Supplementary Figure 24. The creation of a random library via an artificial α5 helix. (**a**) Top: BPCs of production from Lib3_NRPS-1 in *E. coli* DH10B::mtaA induced with L-arabinose (red) and not induced (black). Bottom: EICs (left) and HPLC/MS^2 data (right) of peptides produced by Lib3_NRPS-1 (**1**, *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).

60

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$_4$O$_5$, 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.
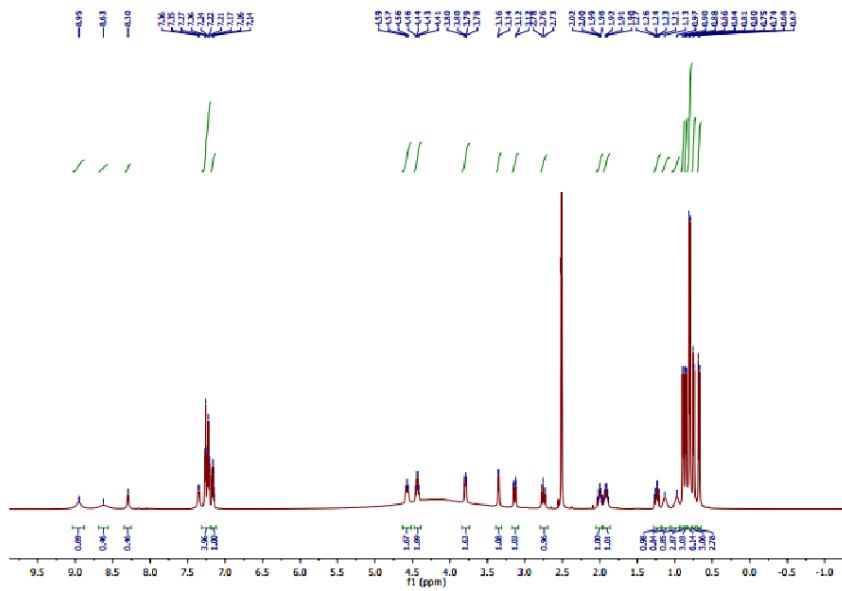
Supplementary Figure 25. ^{1}H NMR spectrum of compound **29**.

61

200
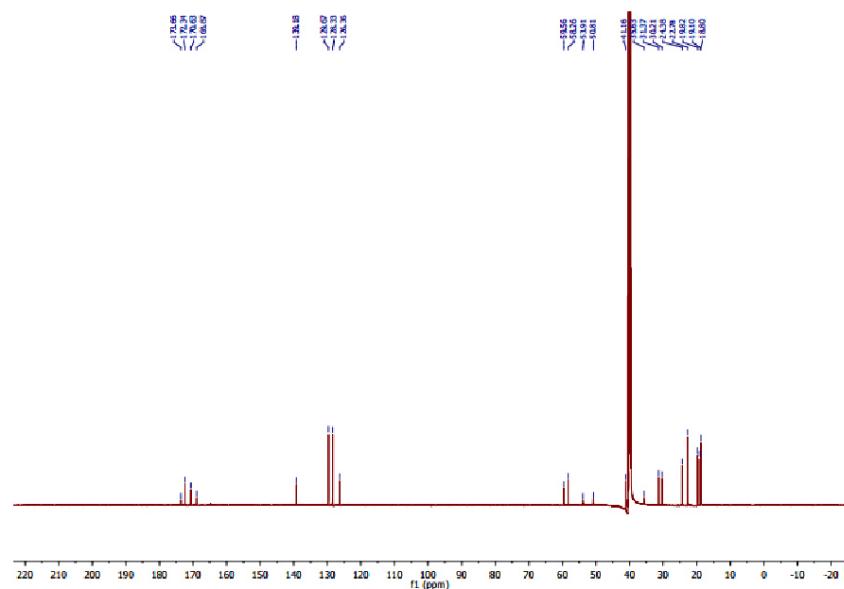
Supplementary Figure 26. ^{13}C NMR spectrum of compound **29**.
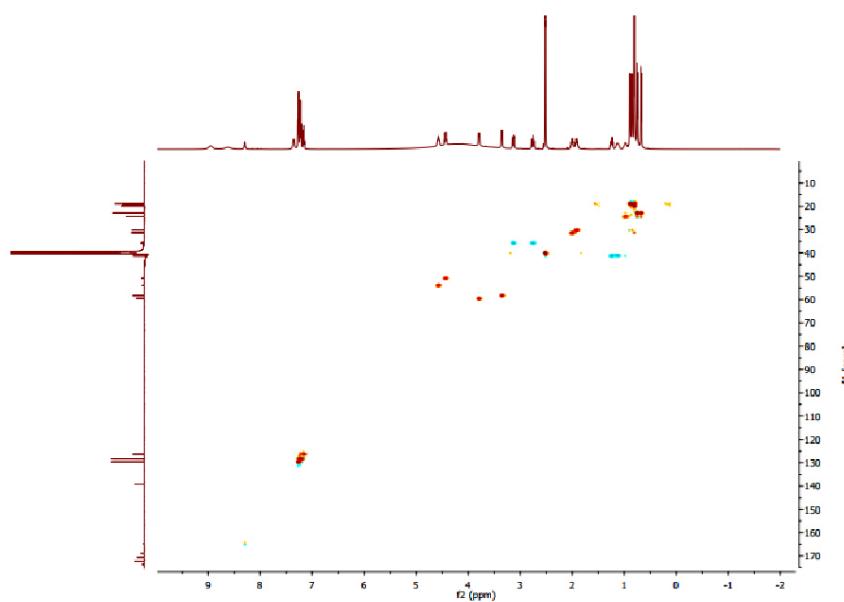
Supplementary Figure 27. HSQC spectrum of compound **29**.

62
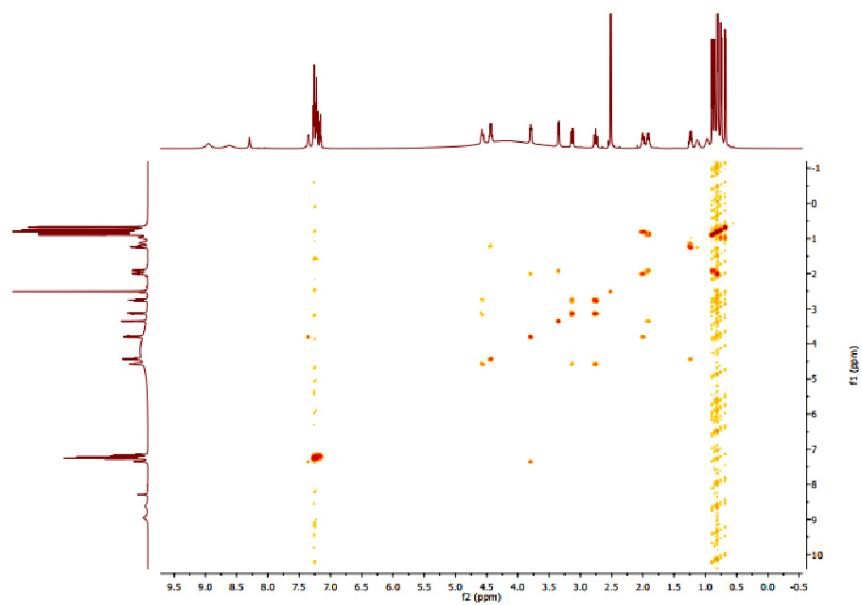
Supplementary Figure 28. COSY spectrum of compound **29.**
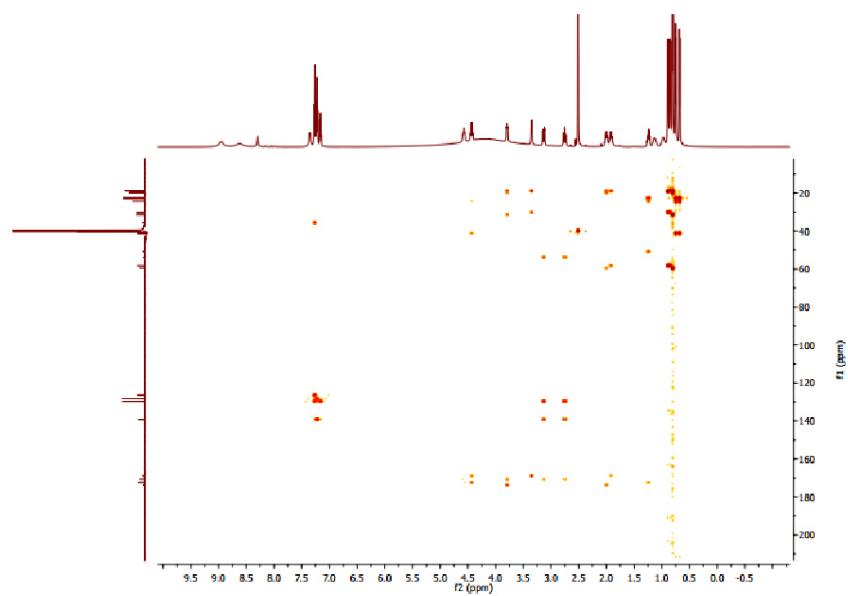
Supplementary Figure 29. HMBC spectrum of compound **29**.

63

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).¹H 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). ¹³C 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.
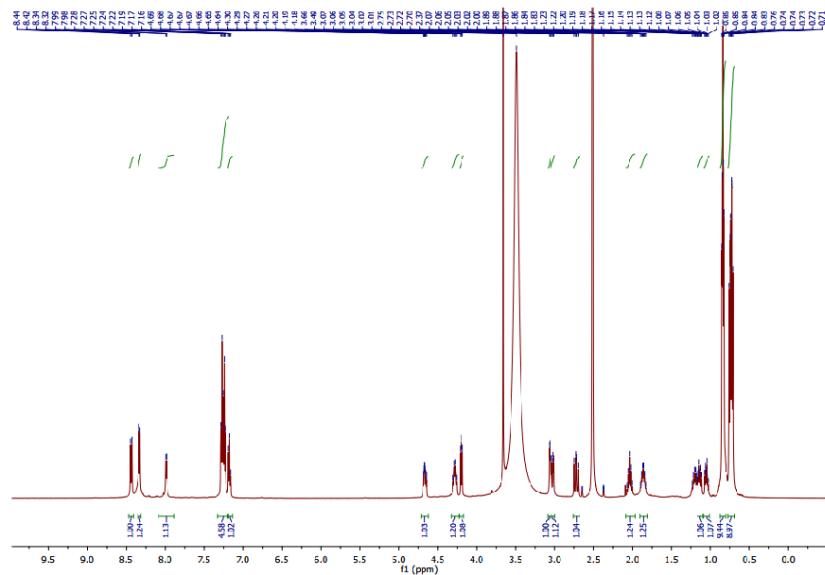
Supplementary Figure 30. ¹H NMR spectrum of compound **30**.
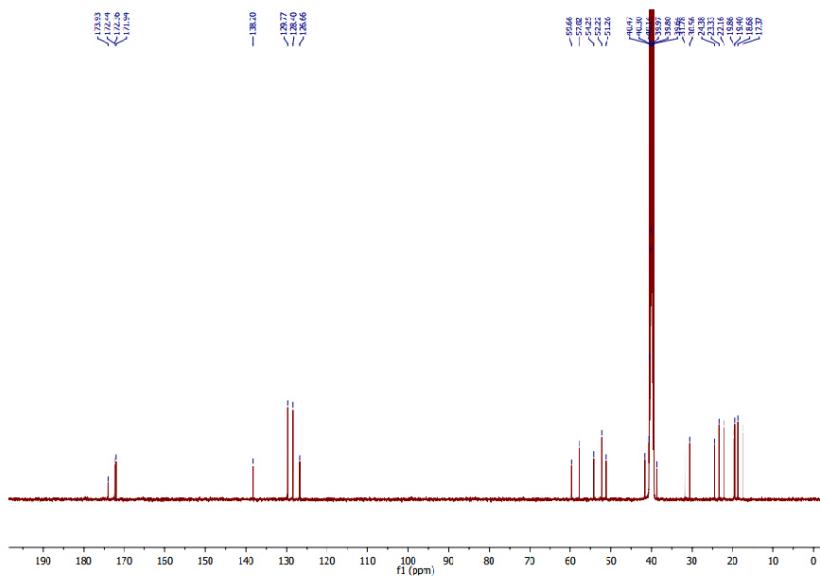
Supplementary Figure 31. ^{13}C NMR spectrum of compound **30**.
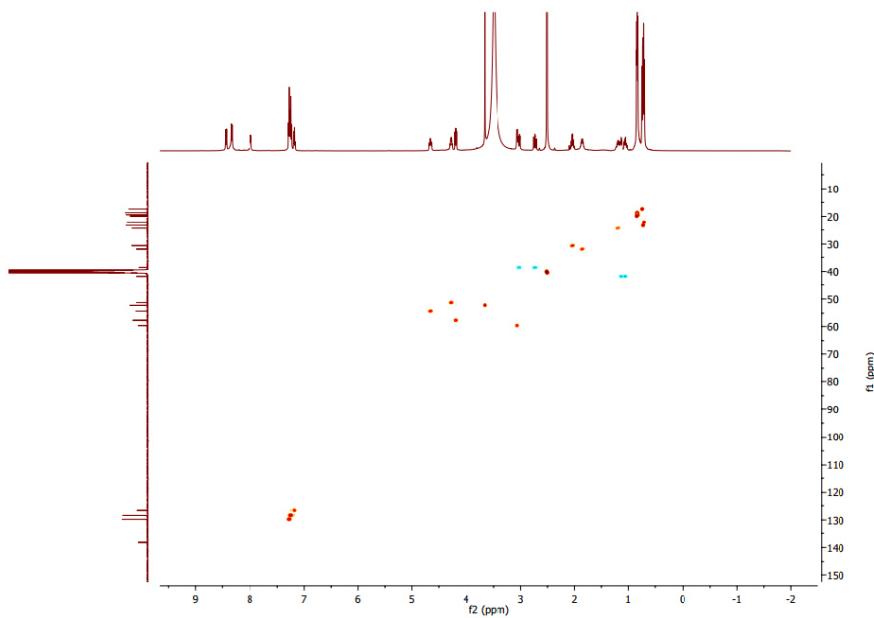
Supplementary Figure 32. HSQC spectrum of compound **30**.

65
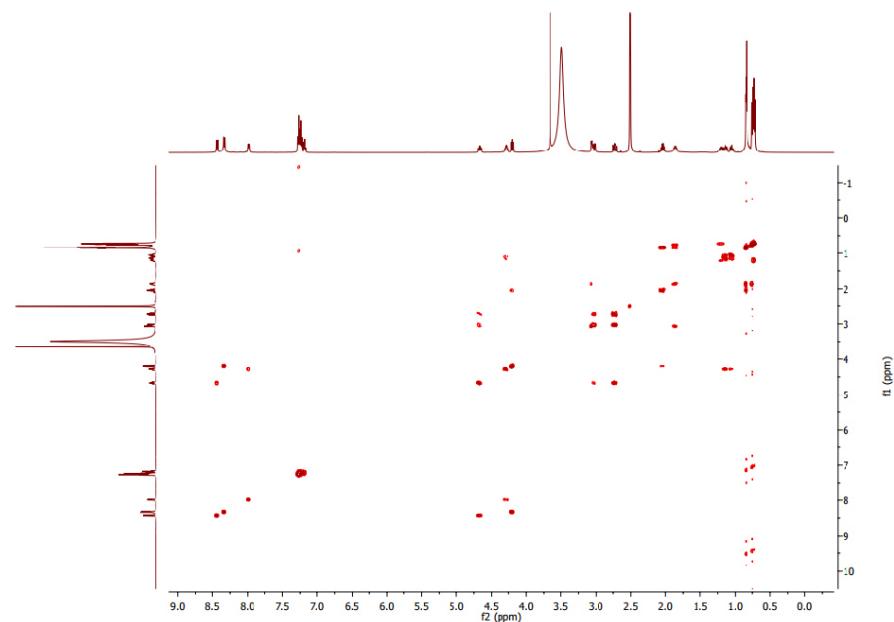
Supplementary Figure 33. COSY spectrum of compound **30**.
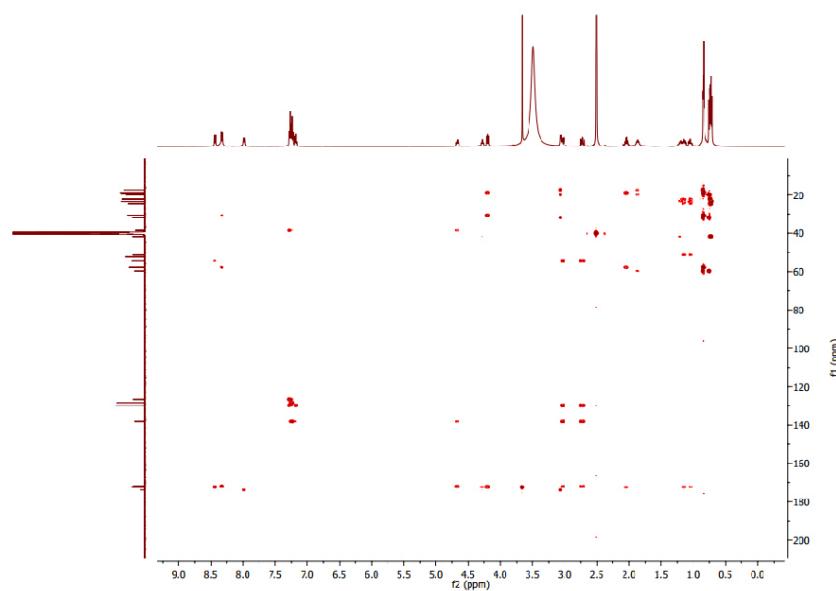
Supplementary Figure 34. HMBC spectrum of compound **30**.

66

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$_7$O$_5$, 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).
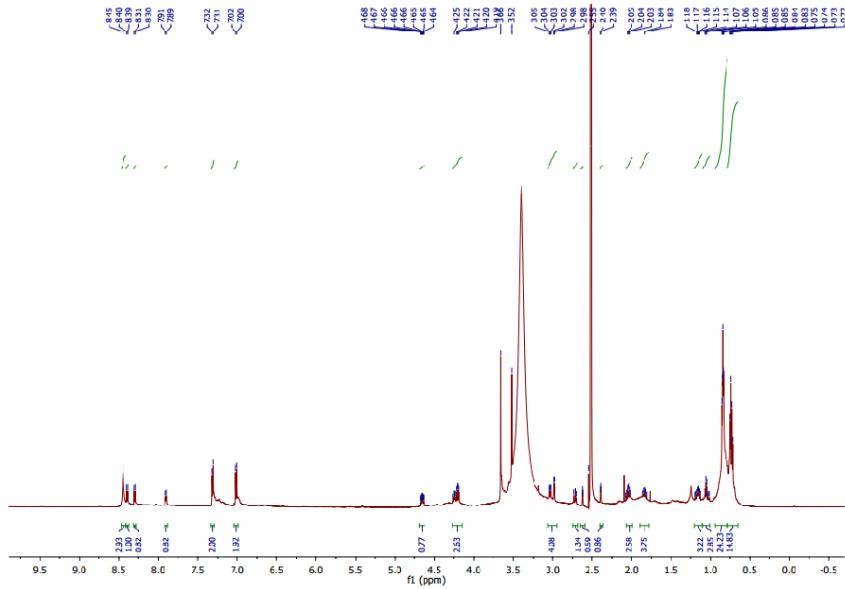
Supplementary Figure 35. ^{1}H NMR spectrum of compound **35**.

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). ¹H 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).
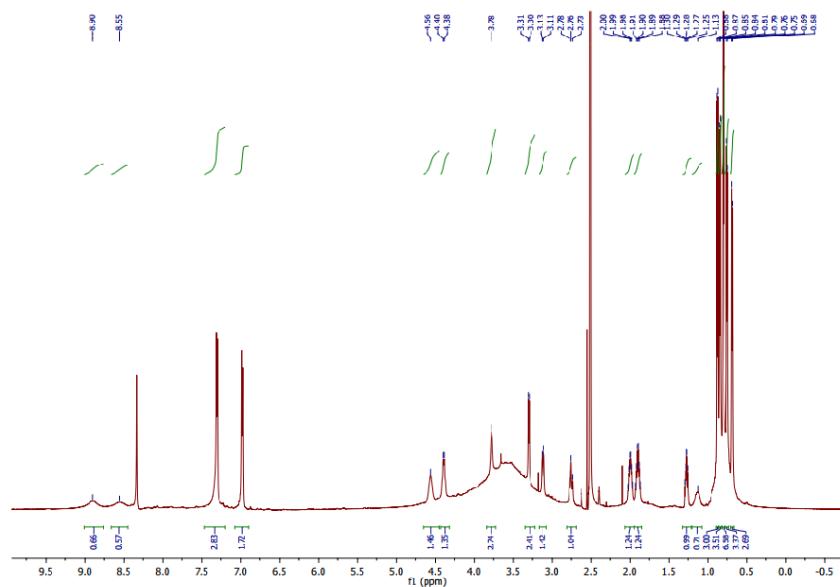
Supplementary Figure 36. ¹H NMR spectrum of compound **36**.
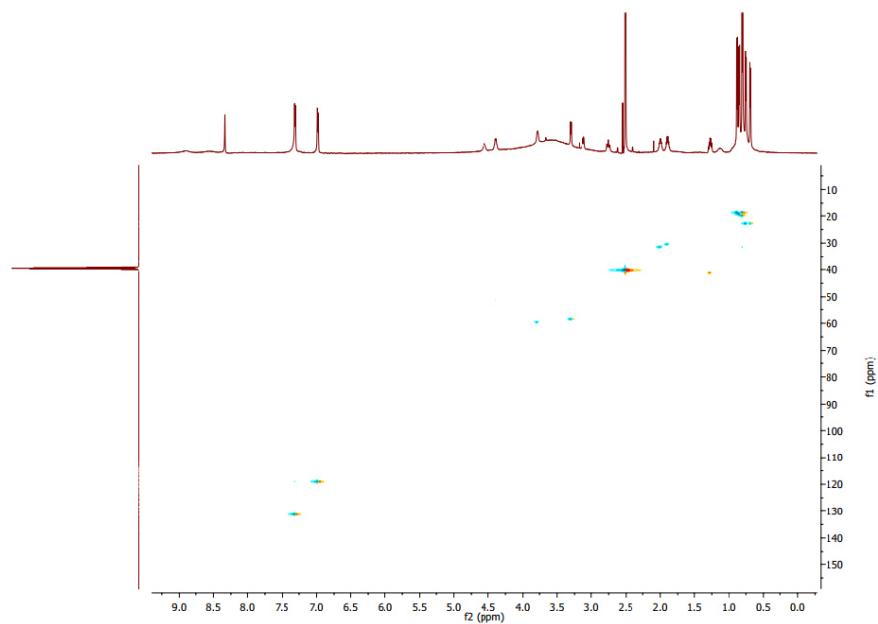
Supplementary Figure 37. HSQC spectrum of compound**36**.
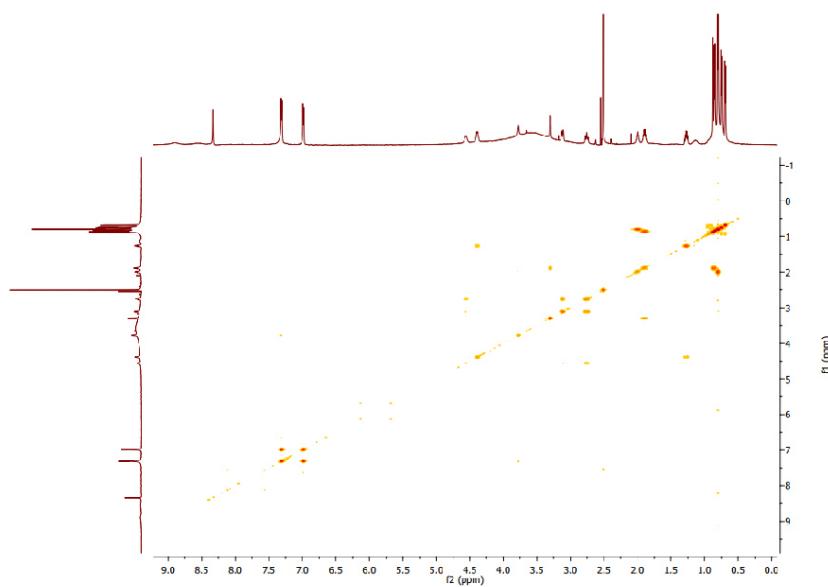
Supplementary Figure 38. COSY spectrum of compound **36**.

69
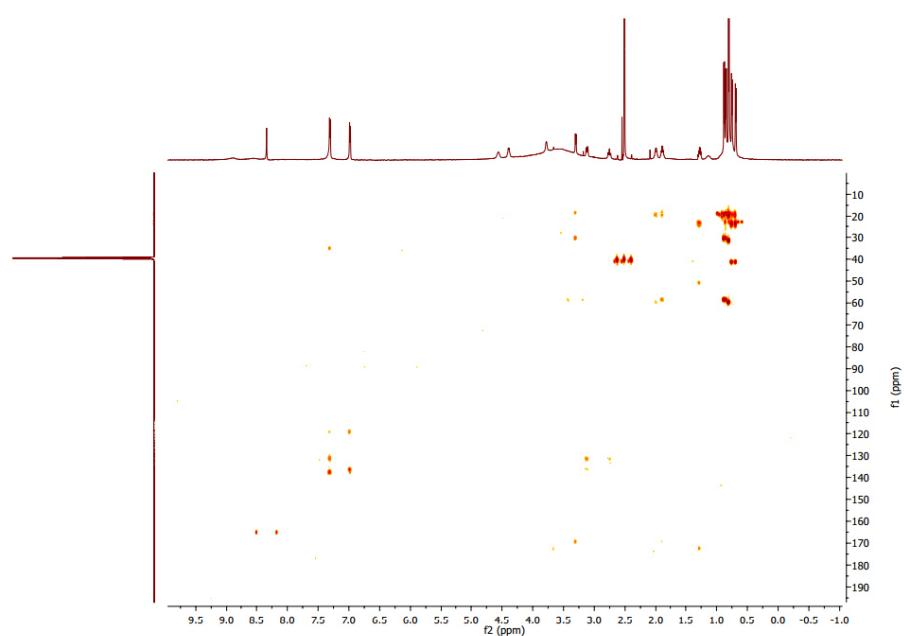
Supplementary Figure 39. HMBC spectrum of compound**36**.

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). ¹H 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).
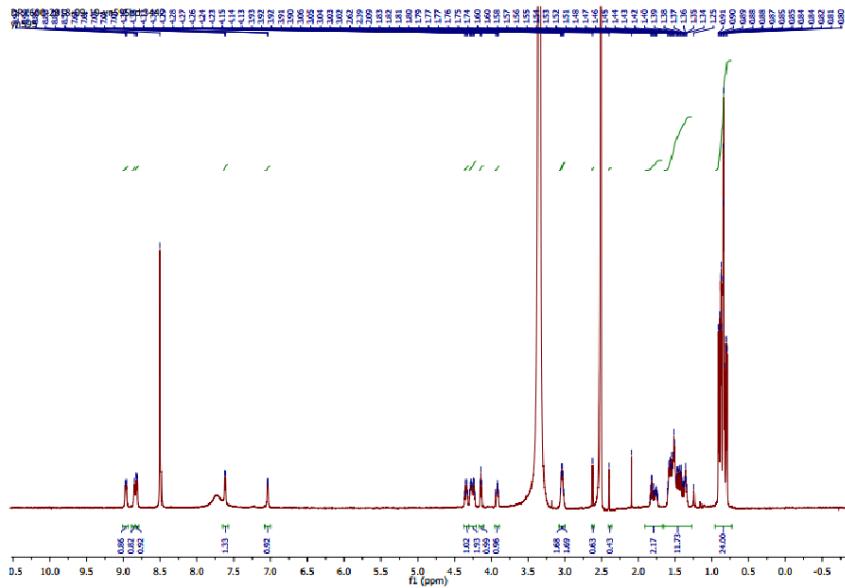
Supplementary Figure 40. ¹H NMR spectrum of compound **42**.

210

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).
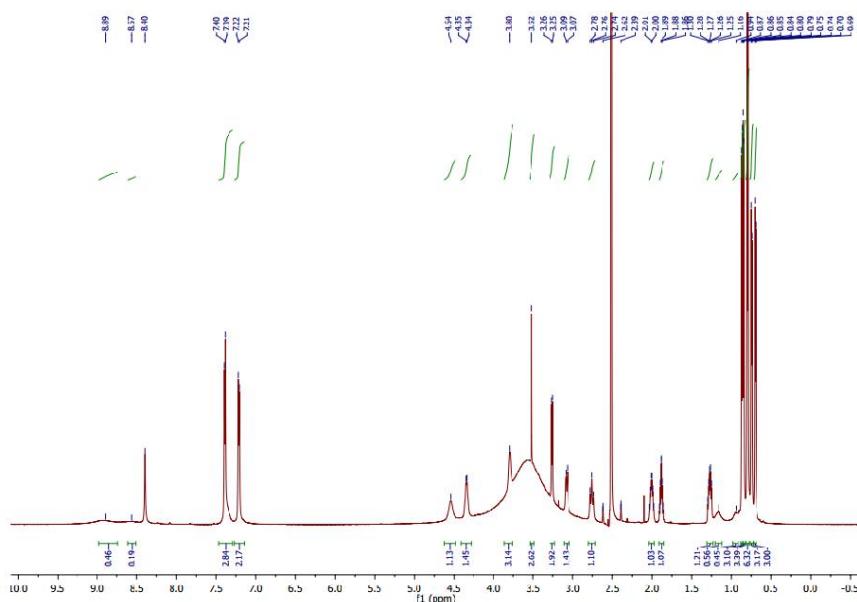
Supplementary Figure 41. ^{1}H NMR spectrum of compound **71**.

72

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**.
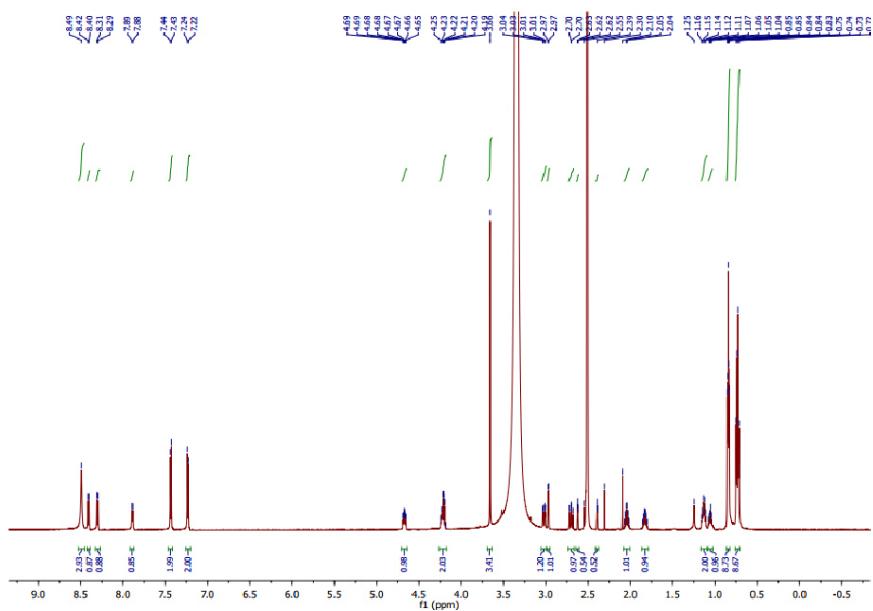
References

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).

75

214

6.3   Non-ribosomal peptides produced by minimal and engineered synthetics with terminal reductase domains

6.3.1 Erklärung zu den Autorenanteilen an der Publikation

Status: accepted

Name der Zeitschrift: *ChemBioChem* 10.1002/cbic.202000176[165]

Autoren: Andreas Tietze (AT), Yan-Ni Shi (YNS), Max Kronenwerth (MK) und Helge B. Bode (HBB)

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

AT (70 %), HBB (30 %)

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Klonierung von Plasmiden und Herstellung von Mutanten, Kultivierung und heterologe Expression, Fütterungsexperimente, HPLC-MS, chemische Synthese, SDS-PAGE: AT (85 %); Isolation von Peptiden: AT (5 %), MK (5 %); NMR Experimente: AT (1 %), YNS (4 %)

(3) zur Erstellung der Datensammlung und Abbildungen

Promoteraustauch in *Xenorhabdus*, NRPS Reprogrammierung, Verifizierung der Proteinlevel, Biosynthesewege: AT (95 %); NMR Daten: YNS (5 %)

(4) zur Analyse und Interpretation der Daten

Promoteraustauch in *Xenorhabdus*, NRPS Reprogrammierung, Verifizierung der Proteinlevel, Biosynthesewege: AT (90 %); NMR Daten: YNS (8 %), MK (2 %)

(5) zum Verfassen des Manuskriptes

AT (80 %), HBB (20 %)

Ort/Datum   Unterschrift des Promovierenden

Ort/Datum   Unterschrift des Betreuers

Attachments

6.3.2 Publication

Nonribosomal Peptides Produced by Minimal and Engineered Synthetases with Terminal Reductase Domains

Andreas Tietze,[a] Yan-Ni Shi,[a] Max Kronenwerth,[a] and Helge B. Bode*[a, b, c]

Nonribosomal peptide synthetases (NRPSs) use terminal reductase domains for 2-electron reduction of the enzyme-bound thioester releasing the generated peptides as C-terminal aldehydes. Herein, we reveal the biosynthesis of a pyrazine that originates from an aldehyde-generating minimal NRPS termed ATRed in entomopathogenic *Xenorhabdus indica*. Reductase domains were also investigated in terms of NRPS engineering and, although no general applicable approach was deduced, we show that they can indeed be used for the production of similar natural and unnatural pyrazinones.

Introduction

In the early 1960s, peptides were discovered that originate from a mechanism different from that of protein synthesis.[1] These nonribosomal peptides (NRPs) show high structural diversity leading to many different biological activities exemplified by the clinically used antibiotic bacitracin, the anticancer agent bleomycin or the immunosuppressant cyclosporine.[2] Their biosynthetic machinery can be found across all three domains of life,[3] and today major insights into the underlying biochemistry and structural basis have been gained.[4,5]

The assembly line-fashioned biosynthesis of NRPs is carried out by large multifunctional nonribosomal peptide synthetases (NRPSs) which harbour a modular architecture.[2,4] Within one module, the adenylation (A) domain recognises and activates a specific amino acid (AA) under ATP consumption, which is then transferred to the 4'-phosphopantetheinyl moiety of a post-translationally modified peptidyl carrier protein also called thiolation (T) domain. The condensation (C) domain forms the peptide bond between two adjacent T domain-bound AAs donating the nascent peptide chain to the following module, where it can be further elongated. Beside this multimodular system, also monomodular,[6] NRPS-like or minimal NRPSs lacking a C domain[7] or even stand-alone domains[8] are known and commonly found in bacteria.[3] Additionally, optional domains, for example, for fatty acid attachment, methylation, cyclization or epimerization of AAs and no restriction to the 20 proteinogenic AAs leads to aforementioned structural diversity.[2] Instead of the most prevalent terminal thioesterase (TE) domains, which release the peptide chain from the NRPS, reductase (R) domains can be an alternative route for peptide release.[9] They catalyse an NAD(P)H dependent two-electron reduction of the thioester to an aldehyde which can be further reduced to an alcohol.[10] Due to their electrophilic properties, aldehydes can contribute as intermediates, for example, for imine formation and subsequent modification as in tilivalline biosynthesis[11] or are often associated with protease inhibitors, for example, by reversible binding of the active site's threonine of the *Mycobacterium tuberculosis* proteasome by fellutamide B.[12]

To get access to more NRPs that either can be modified to improve biological properties, circumvent bacterial resistances or are completely *de novo* peptides, engineering of NRPSs is a powerful tool.[13] Since 1995,[14] this has been the focus of many groups but no general applicable guidelines for NRPS engineering have been established.[15] We recently introduced the concept of exchange units (XU), defining three rules for reproducible NRPS engineering: 1) the tridomain A–T–C is used as XU, 2) the C domain's acceptor site specificity has to be considered and 3) the conserved WNATE sequence depicts the fusion point within the flexible linker connecting the C and A domain.[16] An improved technique (XUC) circumvents the limitation of the C domain specificity by using a fusion point within the linker connecting both subdomains of the C domain.[17] Although the use of TE and even C domains have been investigated as termination domains, the final step within NRP biosynthesis remains a challenging factor in NRPS engineering. Furthermore, aldehyde-generating R domains would provide an alternative route for peptide release and would increase structural diversity. Here, we describe the discovery of an R domain-containing minimal NRPS and show examples of R domains in engineered NRPSs.

[a] A. Tietze, Dr. Y.-N. Shi, Dr. M. Kronenwerth, Prof. Dr. H. B. Bode
Fachbereich Biowissenschaften, Molekulare Biotechnologie
Goethe-Universität Frankfurt
60438 Frankfurt am Main (Germany)
E-mail: h.bode@bio.uni-frankfurt.de
[b] Prof. Dr. H. B. Bode
Buchmann Institute for Molecular Life Sciences (BMLS), Goethe-Universität Frankfurt
60438 Frankfurt am Main (Germany)
[c] Prof. Dr. H. B. Bode
Senckenberg Gesellschaft für Naturforschung
60325 Frankfurt (Germany)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202000176

216

Results and Discussion

AntiSMASH analysis[18] identified a biosynthetic gene *xind01729* in the genome of the entomopathogenic *Xenorhabdus indica* DSM 17382 encoding a monomodular NRPS with a predicted terminal R domain that was not linked to any natural product (Figure S1 in the Supporting Information). Due to the domain arrangement of an A, T and R domain, this minimal NRPS was termed ATRed. Such a three domain architecture has already been described in, say, the biosynthesis of chloramphenicol in *Streptomyces*,[19] virulence factors in *Pseudomonas*,[20] piperazines in *Aspergillus*[21] and for CAR enzymes – a distant relative to the NRPS family - -responsible for the reduction of carboxylic acid substrates to the corresponding aldehydes in bacteria and fungi.[22] An exchange of the promoter upstream of *xind01729* against an arabinose-inducible promoter (P_BAD) showed that compound **1a** is associated with the encoded ATRed in the induced *X. indica* mutant compared to the uninduced mutant (Figures 1A and S2). The production of **1a** was also observed upon heterologous expression of *xind01729* in *Escherichia coli* (Figure S3). Isolation and NMR analysis of **1a** confirmed a structure of a pyrazine that is produced with a titre of 2.1 ± 0.5 mg/L in the wild-type strain (Figures S4–10, Table S4). Based on the domain arrangement and structure, we propose that phenylalanine is activated by the A domain, bound as thioester to the T domain and from there released as aldehyde by the R domain. Two amino aldehydes then form a cyclic Schiff base

which subsequently oxidizes to a pyrazine (Figure 1B). This NRPS-mediated pyrazine biosynthesis is also known from other R domain-containing NRPSs.[20,21,23] Furthermore, pyrazine derivatives with tryptophan (**1b**) or tyrosine (**1c**) instead of one phenylalanine residue were detected in small amounts suggesting a slightly relaxed A domain specificity (Figures S2 and S3) as well as a pyrazinone side product (**1d**) made of two phenylalanines in *E. coli* (Figure S3).

Next, our aim was to analyse the potential application of R domains as release mechanism in engineered NRPS systems. Therefore the identified ATRed_{xind01729}_R domain was fused with the initiation module of the GameXPeptide-producing NRPS (GxpS) in *Photorhabdus laumondii* subsp. *laumondii* TTO1[24] (Figure S11) to keep the overall protein architecture (NRPS-1, Figure 2a). This construct was also elongated by one GxpS module to a bimodular NRPS (NRPS-2) similar to the NRPS involved in the biosynthesis of aureusimine in *Staphylococcus aureus*.[25,26] The engineered proteins were heterologously expressed in *E. coli* but, despite the presence of the expected proteins (Figure S12), no production of peptides was observed after LC–MS analysis (Figure 2A). We also tested the R domain from the tilivalline-producing NRPS (XtvB) in *Xenorhabdus eapokensis* DL20[11] instead of ATRed_{xind01729}_R with the initiation module as well as the first two modules from GxpS (NRPS-3 and -4). In contrast to the monomodular NRPS-3, bimodular NRPS-4 produced compounds **2a** and **2b** with yields up to 24.1 mg/L (Figure S13). NMR analysis of the purified compound **2a** (Fig-
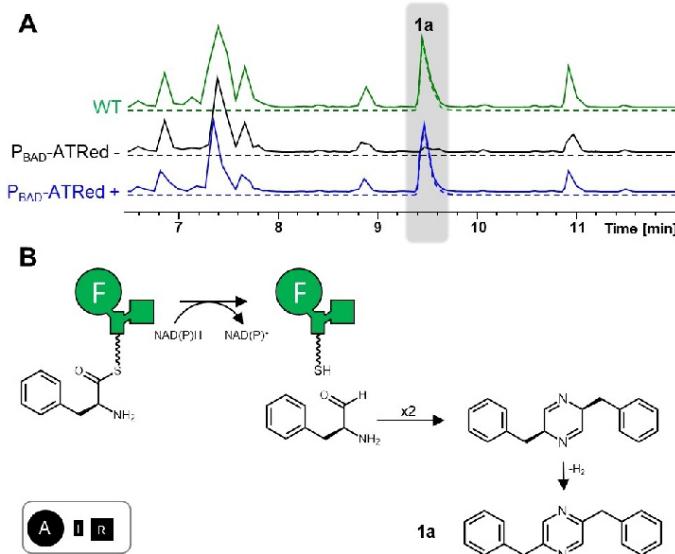
Figure 1. The ATRed NRPS in *X. indica*. A) High-resolution LC–MS analysis of *X. indica* WT (green), uninduced promoter exchange mutant (black) and induced promoter exchange mutant (blue). The base peak chromatogram (BPC) is indicated by continuous lines, and the extracted ion chromatogram (EIC; **1a**; m/z [M + H⁺]⁺ = 261.13) by dashed lines. B) Proposed biosynthesis and structure of **1a**. The ATRed consists of an A (large circle with activated AA substrate indicated by one-letter code; here F), a T (rectangle) and an R (small square) domain.

These are not the final page numbers! ↗↗

217

Attachments
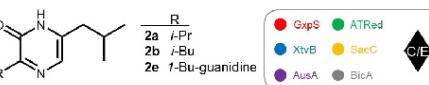
Figure 2. R domains for peptide release in engineered NRP biosynthesis. A) Schematic representation of engineered NRPSs with different R domains and peptide production as determined in triplicate. B) Structures of **2a**, **2b** and **2e**. See Figure 1 for assignment of the domain symbols; further symbol: dual condensation/epimerization (C/E; diamond) domain. The colour code at the bottom identifies NRPSs used as building blocks (Figure S9).

ure S14-20, Table S5) confirmed the structure of a 3-isopropyl-6-isobutyl-pryrazin-2(1H)-one (Figure 2B), and the appearance of two derivatives with valine and leucine as first amino acid is in line with the substrate promiscuity of the GxpS_A1 domain for both AAs.[24] Due to the NRPS architecture and NRP structure, we assume an aureusimine-like biosynthesis via a T-domain-bound dipeptide thioester that is reduced by XtvB_R, thus enabling intramolecular condensation of the generated aldehyde **2c** with its amino group to a cyclic imine and subsequent oxidation to **2a** and **2b** (Figure S21).[25] The aldehyde intermediate **2c** was confirmed by using O-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine (PFBHA;[27] Figure S22).

Recent work showed that retaining the natural T−R interface in bacterial hybrid CAR enzymes leads to higher k_{cat} values.[28] However, a version of NRPS-4 maintaining the T−R didomain from XtvB (NRPS-5) results in an approximately 6.5-fold lower production of both derivatives. Preservation of the natural A−T didomain has also been reported previously in engineered NRPS systems with A−T−TE architecture.[29] Furthermore, the fusion point C-terminal to the last helix of the T domain used in this work (Figure S23) was shown in our development of the XU concept to be applicable for introducing terminal C domains for

peptide release.[16] Beyond ATRed$_{xind01729}$_R and XtvB_R, two more R domains from the postulated safracin-producing NRPS (SacC)[30] in *Xenorhabdus* sp. TS4 as well as the aureusimine-producing NRPS (AusA) in *Staphylococcus lugdunensis* IVK28 (NRPS-6 and −7) were tested with an analogous domain architecture to NRPS-4 (Figure 2A). SacC processes 3-hydroxy-5-methyl-O-methyltyrosine whereas AusA_R has been reported to accept a wide variety of substrates (Figure S11).[31] Unfortunately, no production was observed. This suggests that domain-domain interaction or the R domain's substrate specificity might be crucial for NRPS engineering with R domains as addressed in a molecular docking analysis of a T−R didomain[32] and shown as a common limiting factor for engineering approaches.[15]

Due to the fact that XtvB_R does not exhibit strict substrate specificity (the domain reduces 3-hydroxy anthranilic acid-proline as part of the tilivalline biosynthesis and valine/leucine-leucine in NRPS-4) and the unnatural interaction with GxpS_T2 lead to good production titre, we modified NRPS-4 at its N-terminal position. Exchange of the first D-valine-/leucine-specific XU against the D-arginine specific XU from the bicornutin-producing NRPS (BicA)[33] in *Xenorhabdus budapestensis* DSM 16342 (NRPS-8) resulted in the expected compound **2e** (Figure S24). This was verified by labelling experiments (Figure S25) and comparison to a synthetic NMR standard (Figures S26–31, Table S6).

Conclusion

Although NRPs with aldehydes are relatively rare, their appearance has often been reported with bioactivity like the cysteine protease and proteasome inhibitor flavopeptin from *Streptomyces*.[34] In *Staphylococcus*, an R domain-derived aldehyde serves as important intermediate in the biosynthesis of lugdunin, a promising novel antibiotic against methicillin-resistant *S. aureus*.[35] In this study, we identified the biosynthetic gene responsible for pyrazine biosynthesis in *X. indica* through an R domain containing minimal NRPS termed ATRed. The function of the NRP has not been addressed; however, compounds with pyrazine structures are shown to fulfil biological functions like cell-to-cell communication,[36] thus qualifying them for further studies in order to elucidate their biological purpose. R domains were subsequently tested in engineered NRPSs and we could show that the R domain from the tilivalline-producing NRPS can be used to introduce an aldehyde group in unnatural NRPs. Along with other NRPS engineering approaches, this allows the NRP to be further modified. Nevertheless, the majority of our engineered NRPSs were nonfunctional, thus suggesting that NRPS engineering with terminal R domains is not (yet) generally applicable and further experiments are needed. The limiting factor is probably due to substrate specificity or domain interactions; an issue which should be investigated more in detail with resolving the structure of a T−R didomain and further enzyme/cultivation optimisation.[23,37]

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA

These are not the final page numbers! ↗↗

218

Experimental Section

Strain cultivation: All *E. coli* and *X. indica* strains (Table S1) were grown in liquid or solid lysogeny broth (LB; pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl). Solid medium contained 1.5% (*w/v*) agar. *Saccharomyces cerevisiae* strain CEN.PK 2-1 C and derivatives were grown in liquid and solid yeast extract peptone dextrose (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, and all other strains were cultivated at 30°C. *E. coli* ST18 cells were supplemented with 50 µg/mL 5-aminolevulinic acid. For production of **1a**, *X. indica* was inoculated from an overnight culture in 10 mL volume and grown for 48 h at 160 rpm with 2% (*v/v*) Amberlite XAD-16. P_{BAD} promoters were induced with 0.02% L-arabinose. For the detection of aldehydes,[27] 0.2 mM PFBHA was added to the LB culture.

Generation of promoter exchange mutants: The first 700 bp of *xind01729* were cloned in the PCR-amplified backbone of pCEP_Kan[38] and *E. coli* ST18 cells were transformed with the plasmid. ST18 and *X. indica* wildtype cells were grown in 10 mL LB from an overnight culture to an OD_{600} of 0.6–0.8, washed twice and mixed on an LB plate without 5-aminolevulinic acid in a ratio of 1:3. After incubation for 24 h at 30°C, the cells were harvested and incubated for another 72 h on selection medium containing kanamycin.

Cloning of plasmids and transformation of cells: Genomic DNA of *Xenorhabdus* and *Photorhabdus* strains was isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. Genomic DNA of *S. lugdunensis IVK28* was provided by B. Krismer (Eberhard Karls University of Tübingen, Germany). PCR was performed with oligonucleotides obtained from Eurofins Genomics (Table S3). Cloning was done by Hot fusion[39] or transformation-associated recombination (TAR),[40] and the fragments were amplified in a two-step PCR program with homology arms (20 or 40–80 bp, respectively). For PCR, S7 Fusion high-fidelity DNA polymerase (Biozym) and Q5 high-fidelity DNA polymerase (New England BioLabs) were used according to the manufacturers' instructions. The vector pFF1 was digested with EcoRI and SgsI. All fragments were digested with DpnI (Thermo Fisher Scientific). DNA purification was performed with MSB® Spin PCRapace (STRATEC Biomedical AG) or from 1% TAE agarose gel using Invisorb® Spin DNA Extraction (STRATEC Biomedical AG). Plasmids (Table S2) were transformed into *E. coli* DH10B::mtaA by electroporation and verified by restriction digest. Plasmid was isolated from *E. coli* by using Invisorb® Spin Plasmid Mini Two (STRATEC Biomedical AG).

Heterologous expression of NRPSs and extract preparation: *E. coli* cells harbouring the constructed plasmids were inoculated from an overnight culture to 10 mL cultures containing 2% (*v/v*) Amberlite XAD-16, kanamycin and arabinose for 48 h at 22°C and 160 rpm.

The XAD-16 beads were harvested by sieving and incubated with one culture volume MeOH for 30 min at 160 rpm. The organic phase was filtered and evaporated to dryness under reduced pressure as described before.[17] Extracts were solved in 1 mL MeOH and diluted 1:10 for LC-MS measurements.

LC–MS analysis: All measurements were carried out by using an Ultimate 3000 LC system (Dionex; gradient of MeCN/0.1% formic acid in H_2O/0.1% formic acid, 5% to 95%, 15 min, flow rate 0.4 mL/min, ACQUITY UPLC BEH C18 column 1.7 µm 2.1 mm×100 mm (Waters)) coupled to an AmaZonX (Bruker) electron spray ionization (ESI) mass spectrometer in positive ionization mode or to an Impact II qTof (Bruker) with internal 10 mM sodium formate calibrant for high-resolution data. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

SDS-PAGE analysis: A 20 mL LB culture was inoculated to an $OD_{600}=0.05$ with an overnight culture of *E. coli* cells with the respective NRPS-expressing plasmid and was grown for 18 h at 160 rpm. Cells with IPTG-inducible plasmids were grown at 37°C until an $OD_{600}=0.7$ for induction and subsequently grown at 16°C; cells with arabinose-inducible plasmids were grown at 26°C and induced upon inoculation. The OD_{600} was normalized with LB, the cell pellet (3200 rpm, 10 min, 4°C) of 20 mL was resuspended in 10 mL lysis buffer (100 mM HEPES pH 7.6, 200 mM NaCl, 0.1% Triton X-100, 1 mM dithiothreitol, 1 mM EDTA, protease inhibitor and lysozyme) and incubated for 20 min on ice. After sonication on ice, the supernatant (13 300 rpm, 15 min) was mixed with 3x loading buffer (100 mM Tris·Cl pH 6.8, 4% (*w/v*) SDS, 0.2% Bromophenol Blue, 200 mM β-mercaptoethanol), incubated at 37°C for 20 min and separated on 8% SDS-PAGE gels.

Labeling experiments: *E. coli* cells with the respective NRPS-expressing plasmid were grown in ISOGRO®-^{13}C or -^{15}N (Sigma–Aldrich) medium.[24] 2 mM unlabelled AA was added to the culture; cultivation and extract preparation were performed as described above.

Peptide synthesis, purification and quantification: Compound **2e** was chemically synthesized as described by Schilling et al. by using H-Leu-H NovaSyn TG resin (15.8 µmol, Sigma-Aldrich) and Fmoc-D-Arg(Pbf)-OH (63 µmol, Iris Biotech) with 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU; 63 µmol, Carbolution), 1-hydroxy-1*H*-benzotriazole (HOBt; catalytic, Sigma-Aldrich) and NMM (126 µmol, Sigma–Aldrich) in ACN for 1 h coupling reaction.[41] After Fmoc deprotection with 20%(*v/v*) piperidin (Iris Biotech) in DMF, cyclization occurred after cleavage from the resin (79.95% ACN/20% water/0.05% TFA (*v/v/v*)) and the Pbf group was finally deprotected with TFA.

Compounds **1a**, and **2a** were purified from 1 L culture by using a 1260 Infinity II LC system and 1260 Semiprep LC system (Eclipse XDB-C18 7 µm 21.2×250 mm) coupled to a G6125B LC/MSD ESI-MS (Agilent). Synthesised **2e** was purified by using a 1260 Infinity II LC system (Agilent).

All peptides were quantified in triplicates using a calibration curve (11 values ranging from 100 µg/mL to 0.02 µg/mL) and HPLC-MS measurements. As standards, purified **1a** (for quantification of **1a**), **2a** (for quantification of **2a** and **2b**) and synthetic **2e** (for quantification of **2e**) were used.

NMR analysis: Structures of **1a**, **2a** and **2e** were elucidated by 1D and 2D NMR experiments. ^{1}H, ^{13}C, COSY, HSQC and HMBC spectra were measured on a Bruker AV500 spectrometer using CD_3OD and [D_6]DMSO as solvent.[17] Coupling constants are expressed in Hz and chemical shifts are given on a ppm scale.

219

Attachments

Conflict of Interest

The authors declare no conflict of interest.

[1] B. Mach, E. Reich, E. L. Tatum, *Proc. Natl. Acad. Sci. USA* **1963**, *50*, 175–181.
[2] R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821; *Angew. Chem.* **2017**, *129*, 3824–3878.
[3] H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9259–9264.
[4] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715–738.
[5] J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, *Curr. Opin. Struct. Biol.* **2018**, *49*, 104–113.
[6] D. Reimer, K. N. Cowles, A. Proschak, F. I. Nollmann, A. J. Dowling, M. Kaiser, R. ffrench-Constant, H. Goodrich-Blair, H. B. Bode, *ChemBioChem* **2013**, *14*, 1991–1997.
[7] W.-W. Sun, C.-J. Guo, C. C. C. Wang, *Fungal Genet. Biol.* **2016**, *89*, 84–88.
[8] S. L. Wenski, D. Kolbert, G. L. C. Grammbitter, H. B. Bode, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 565–572.
[9] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255–278.
[10] A. Chhabra, A. S. Haque, R. K. Pal, A. Goyal, R. Rai, S. Joshi, S. Panjikar, S. Pasha, R. Sankaranarayanan, R. S. Gokhale, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5681–5686.
[11] H. Wolff, H. B. Bode, *PLoS One* **2018**, *13*, e0194297.
[12] G. Lin, D. Li, T. Chidawanyika, C. Nathan, H. Li, *Arch. Biochem. Biophys.* **2010**, *501*, 214–220.
[13] K. A. Bozhüyük, J. Micklefield, B. Wilkinson, *Curr. Opin. Microbiol.* **2019**, *51*, 88–96.
[14] T. Stachelhaus, A. Schneider, M. A. Marahiel, *Science* **1995**, *269*, 69–72.
[15] A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nat. Prod. Rep.* **2018**, *35*, 1210–1228.
[16] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275–281.
[17] 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.
[18] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81-W87.
[19] L. T. Fernández-Martínez, C. Borsetto, J. P. Gomez-Escribano, M. J. Bibb, M. M. Al-Bassam, G. Chandra, M. J. Bibb, *Antimicrob. Agents Chemother.* **2014**, *58*, 7441–7450.
[20] A. M. Kretsch, G. L. Morgan, J. Tyrrell, E. Mevers, I. Vallet-Gély, B. Li, *Org. Lett.* **2018**, *20*, 4791–4795.
[21] R. R. Forseth, S. Amaike, D. Schwenk, K. J. Affeldt, D. Hoffmeister, F. C. Schroeder, N. P. Keller, *Angew. Chem. Int. Ed.* **2013**, *52*, 1590–1594; *Angew. Chem.* **2013**, *125*, 1632–1636.

[22] a) D. Gahloth, M. S. Dunstan, D. Quaglia, E. Klumbys, M. P. Lockhart-Cairns, A. M. Hill, S. R. Derrington, N. S. Scrutton, N. J. Turner, D. Leys, *Nat. Chem. Biol.* **2017**, *13*, 975–981; b) M. Winkler, *Curr. Opin. Chem. Biol.* **2018**, *43*, 23–29; c) D. Gahloth, G. A. Aleku, D. Leys, *J. Biotechnol.* **2020**, *307*, 107–113.
[23] M. A. Wyatt, N. A. Magarvey, *Biochem. Cell Biol.* **2013**, *91*, 203–208.
[24] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün, *Chem. Eur. J.* **2012**, *18*, 2342–2348.
[25] M. Zimmermann, M. A. Fischbach, *Chem. Biol.* **2010**, *17*, 925–930.
[26] M. A. Wyatt, W. Wang, C. M. Roux, F. C. Beasley, D. E. Heinrichs, P. M. Dunman, N. A. Magarvey, *Science* **2010**, *329*, 294–296.
[27] T. Wichard, S. A. Poulet, G. Pohnert, *J. Chromatogr. B Biomed. Sci. Appl.* **2005**, *814*, 155–161.
[28] L. Kramer, X. Le, E. D. Hankore, M. A. Wilson, J. Guo, W. Niu, *J. Biotechnol.* **2019**, *304*, 52–56.
[29] E. Hühner, K. Öqvist, S.-M. Li, *Org. Lett.* **2019**, *21*, 498–502.
[30] A. Velasco, P. Acebo, A. Gomez, C. Schleissner, P. Rodríguez, T. Aparicio, S. Conde, R. Muñoz, F. de La Calle, J. L. Garcia, et al., *Mol. Microbiol.* **2005**, *56*, 144–154.
[31] M. A. Wyatt, M. C. Y. Mok, M. Junop, N. A. Magarvey, *ChemBioChem* **2012**, *13*, 2408–2415.
[32] J. F. Barajas, R. M. Phelan, A. J. Schaub, J. T. Kliewer, P. J. Kelly, D. R. Jackson, R. Luo, J. D. Keasling, S.-C. Tsai, *Chem. Biol.* **2015**, *22*, 1018–1029.
[33] S. W. Fuchs, C. C. Sachs, C. Kegler, F. I. Nollmann, M. Karas, H. B. Bode, *Anal. Chem.* **2012**, *84*, 6948–6955.
[34] Y. Chen, R. A. McClure, Y. Zheng, R. J. Thomson, N. L. Kelleher, *J. Am. Chem. Soc.* **2013**, *135*, 10449–10456.
[35] A. Zipperer, M. C. Konnerth, C. Laux, A. Berscheid, D. Janek, C. Weidenmaier, M. Burian, N. A. Schilling, C. Slavetinsky, M. Marschal, et al., *Nature* **2016**, *535*, 511–516.
[36] Y.-M. Shi, H. B. Bode, *Nat. Chem. Biol.* **2017**, *13*, 453–454.
[37] M. A. Fischbach, J. R. Lai, E. D. Roche, C. T. Walsh, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11951–11956.
[38] E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, C. Dauth, Q. Zhou, M. Kaiser, P. Klemmt, H. B. Bode, *ChemBioChem* **2015**, *16*, 1115–1119.
[39] C. Fu, W. P. Donovan, O. Shikapwashya-Hasser, X. Ye, R. H. Cole, *PLoS One* **2014**, *9*, e115318.
[40] R. D. Gietz, R. H. Schiestl, *Nat. Protoc.* **2007**, *2*, 1–4.
[41] N. A. Schilling, A. Berscheid, J. Schumacher, J. S. Saur, M. C. Konnerth, S. N. Wirtz, J. M. Beltrán-Beleña, A. Zipperer, B. Krismer, A. Peschel, et al., *Angew. Chem. Int. Ed.* **2019**, *58*, 9234–9238.

220

COMMUNICATIONS

Peptide aldehyde production: An aldehyde-releasing reductase (R) domain has been identified as part of a minimal nonribosomal peptide synthetase (NRPS) in *X. indica*. We also show that the R domain from the tilivalline-producing NRPS can be used in engineered synthetases to introduce an aldehyde group and subsequently produce natural and unnatural pyrazinones.

WT NRPS   engineered NRPS

reductive release
as aldehydes
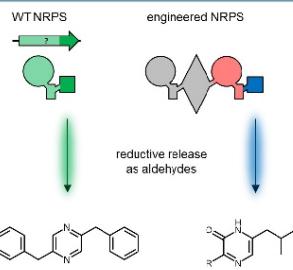
A. Tietze, Dr. Y.-N. Shi, Dr. M. Kronenwerth, Prof. Dr. H. B. Bode*

1 – 6

Nonribosomal Peptides Produced by Minimal and Engineered Synthetases with Terminal Reductase Domains

6.3.3 Supplementary information

Supplementary Table 1. Strains used and generated in this work.

Strain	Genotype	Reference
E. coli BL21 DE3	F– ompT hsdSB(rB- mB-) gal dcm lon λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Invitrogen
E. coli BL21 DE3 pET11a_*xind01729* pCK_mtaA	*E. coli* BL21star DE3 pET11a_*xind01729* pCK_mtaA, Amp[R], Cm[R]	This work
E. coli DH10B	F_ mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lac*ZΔ, M15, Δ*lac*X74 *rec*A1 *end*A1 *ara*D 139 Δ(*ara, leu*)7697 *gal*U *gal*K λ*rpsL* (*Strr*) *nup*G	[1]
E. coli DH10B::mtaA	DH10B with mtaA from pCK_mtaA ∆entD	[2]
E. coli ST18		[3]
E. coli ST18 pCEP-Kan_*xind01729*	*E. coli* ST18 pCEP-Kan_*xind01729,* Kan[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
Xenorhabdus sp. TS4		DSMZ
X. eapokensis DL20		DSMZ
X. budapestensis DSM 16342		DSMZ
X. indica DSM 17382		DSMZ
X. indica DSM 17382::pCEP-Kan_*xind01729*	*X. indica* DSM 17904::pCEP-Kan_*xind01729,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-1*	*E. coli* DH10B::mtaA pAT41_*NRPS-1,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-2*	*E. coli* DH10B::mtaA pAT41_*NRPS-2,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-3*	*E. coli* DH10B::mtaA pAT41_*NRPS-3,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-4*	*E. coli* DH10B::mtaA pAT41_*NRPS-4,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-5*	*E. coli* DH10B::mtaA pAT41_*NRPS-5,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-6*	*E. coli* DH10B::mtaA pAT41_*NRPS-6,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-7*	*E. coli* DH10B::mtaA pAT41_*NRPS-7,* Kan[R]	This work
E. coli DH10B::mtaA pAT41_*NRPS-8*	*E. coli* DH10B::mtaA pAT41_*NRPS-8,* Kan[R]	This work

Supplementary Table 2. Oligonucleotides used in this work.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pCEPKan_xind01729	pCEP_fw_gib	ATGTGCATGCTCGAGCTC	pCEP-Kan
	pCEP_rv_gib	ATGCTAGCCTCCTGTTAGC	
	PF_7	TTTTGGGCTAACAGGAGGCTAGCATATGATAGAAATACCACCCCCTATAAATTTCG	X. indica DSM 17382
	PF_8	ATCTGCAGAGCTCGAGCATGCACATCGTGGCCGTCATAATCAGAC	
pET11a_xind01729	pET11a_for_strep-tag	GTAGGATGGAGCCATCCACAGTTCGAGAAGTAAGGATCCGGCTGCTAAC	pET11a-modified
	pET11a rev	ATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTA	
	PF_15	TTAACTTTAAGAAGGAGATATACATATGATAAATACCACCCCCTATAAATTTCG	X. indica DSM 17382
	PF_16	CGAACTGTGGATGGCTCCATCCTACTGATAAAAAACCTATTTTTTTCCAGTAAGTAAGATAAC	
pAT41_NRPS-1	AT_293	GATACCTATCTGAATAGTGATAAAAAATCAAATAATG	pAT41_NRPS-2
	AT_470	TCAGATTTCGTGATGTTCGTC	
	AT_471	ACGAACATCACGAAATCTGACGCTCAAATCAGTGGTGGC	pAT41_NRPS-4
	AT_483	TCACTATTCAGATAGGTATCCAATGTTTGGGCCAACTCCG	
pAT41_NRPS-2	pAT41_bb+Ypet_fw	GGATCCGCTGGCTCC	pAT41_NRPS-4
	AT_451	GATTTTCTCGGTAAATGTCGCC	
	AT_454	ATTGGCGACATTTACCGAGAAAATCCAACAAAACAAGAACGAGCTCACTG	X. indica DSM 17382
	AT_455	AACCAGCAGCGGAGCCAGCGGATCCCTATGATAAAAAACCTATTTTTTTCCAGTAAGTAAGATAAC	
pAT41_NRPS-3	AT_293	GATACCTATCTGAATAGTGATAAAAAATCAAATAATG	pAT41_NRPS-4
	AT_470	TCAGATTTCGTGATGTTCGTC	
	AT_471	ACGAACATCACGAAATCTGACGCTCAAATCAGTGGTGGC	pAT41_NRPS-4
	AT_483	CGGTGATGTTCTGTCTGGTCTACACTCAGAGTCTGGGCGACAAA	
pAT41_NRPS-4	pAT41_bb+YPet_fw	GGATCCGCTGGCTCC	pAT41
	pAT41_bb_rv	GGAATTCCTCCTGTTAGCCC	
	AL_GxpS-2-1	ACTGTTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	P. luminescens TT01
	AT_328	TTTCATTATTTGATTTTTTATCACTATTCAGATAGGTATCGATTTTCTCGGTAAATGTCGCC	
	AT_308	GATACCTATCTGAATAGTGATAAAAAATCAAATAATGAAATAAAAAATAC	X. eapokensis DL20
	AT_289	TCATGAACTCGCCAGAACCAGCGCGAGCCAGCGGATCCCTTACTTTCAGGTTTATATGACGGTATGCTTG	
pAT41_NRPS-5	pAT41_bb+Ypet_fw	GGATCCGCTGGCTCC	pAT41_NRPS-4
	AT_458	CGCGACATAAATTTGGCGAG	
	AT_460	TTTTGCTCGCCAAATTTATGTCGCGGTTCTGACTTCAACCGAACAAACAG	X. eapokensis DL20
	AT_459	(AACCAGCAGCGGAGCCAGCGGATCCTTACTTACTTTCAGGTTTATATGACGGTATGC	
pAT41_NRPS-6	pAT41_bb+Ypet_fw	GGATCCGCTGGCTCC	pAT41_NRPS-4
	AT_451	GATTTTCTCGGTAAATGTCGCC	
	AT_452	ATTGGCGACATTTACCGAGAAAATCGAAATTTATCGGCGCGAAGG	X. TS4
	AT_453	AACCAGCAGCGGAGCCAGCGGATCCTTATTTCTGTTCCTGTGTCTGGTGTTG	
pAT41_NRPS-7	pAT41_bb+Ypet_fw	GGATCCGCTGGCTCC	pAT41_NRPS-4
	AT_451	GATTTTCTCGGTAAATGTCGCC	
	AT_456	ATTGGCGACATTTACCGAGAAAATCGAACAGCAAAGTGACGAATCGTG	S. lugdunensis
	AT_457	AACCAGCAGCGGAGCCAGCGGATCCTCATCTATGGTATTCTTTACATTCAAATTTTTCATTG	
pAT41_NRPS-8	pAT41_bb+YPet_fw	GGATCCGCTGGCTCC	pAT41_NRPS-4

pAT41_bb_rv	GGAATTCCTCCTGTTAGCCC
JK-P1	CGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTCTCCATACCCGTTTTTTTGGGCTAACAGAGAGG pFF1_NRPS_5*
	AATTCCATGAAAGATAACATTGCTACAGTGG
AT_328	TTTCATTATTGATTTTTATCACTATTCAGATAGGTATCGATTTTCTCGGTAAATGTCGCC
AT_308	GATACCTATCTGAATAGTGATAAAAAATCAAATAATGAAATAAAAAATAC
AT_289	TCATGAACTCGCCAGAACCAGCGGAGCCAGCGGATCCCTTACTTTCAGGTTTATATGACGGTATGCTTG X. eapokensis DL20

Supplementary Table 3. Plasmids used and generated in this work. pFF1_*NRPS_5* from ([4]) is indicated with an * to avoid confusion with NRPSs constructed in this work.

Plasmid	Genotype	Reference
pAT41	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	[4]
pCK_mtaA	Cm^R, ori p15A, *mtaA*	[5]
pET11a-modified	pBR322 ori, P_{T7} promoter, Amp^R, lacI, His$_6$-smt3 tag	[6]
pET11a_*xind01729*	pBR322 ori, P_{T7} promoter, Amp^R, lacI, *xind01729*, strep tag	This work
pFF1	2μ ori, kanMX4, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	[7]
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}$	[4]
pCEP-Kan	R6Kγ ori, oriT, Kan^R, araC, P_{BAD} promoter	[8]
pCEP-Kan_*xind01729*	R6Kγ ori, oriT, Kan^R, araC, P_{BAD} promoter, *xind01729* (bp 1-700)	This work
pAT41_*NRPS-1*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2-*xind01729*_R	This work
pAT41_*NRPS-2*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2T2-*xind01729*_R	This work
pAT41_*NRPS-3*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2-*xtvB*_R	This work
pAT41_*NRPS-4*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2T2-*xtvB*_R	This work
pAT41_*NRPS-5*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2-*xtvB*_T2R	This work
pAT41_*NRPS-6*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2T2-*sacC*_R	This work
pAT41_*NRPS-7*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2T2-*ausA*_R	This work
pAT41_*NRPS-8*	2μ ori, URA3, P_{BAD} promoter, pCOLA ori, Ypet-Flag, Kan^R, *bicA*_A1T2CE2-*gxpS*_A2T2-*xtvB*_R	This work
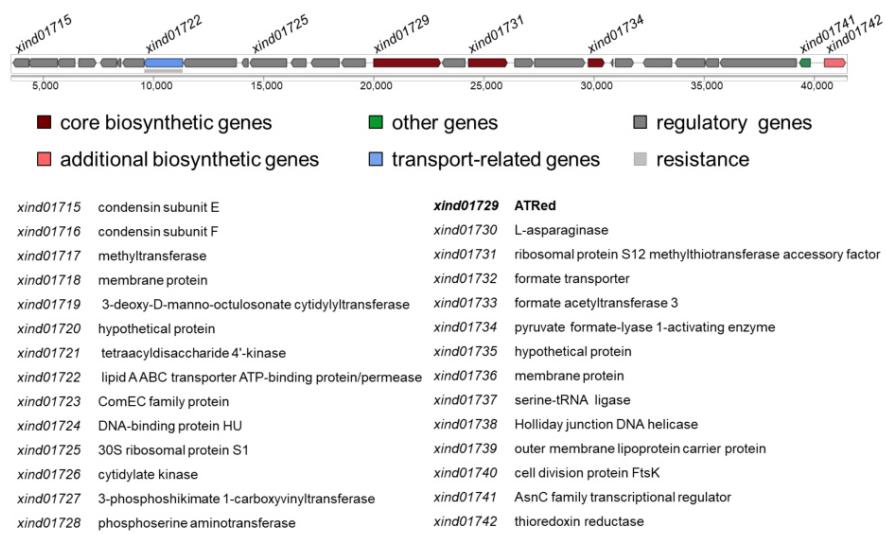
Supplementary Figure 1. Genomic region of *X. indica* DSM 17382 containing the ATRed encoding gene *xind01729*. The gene annotations and details are according to antiSMASH 5.1.2.[9]
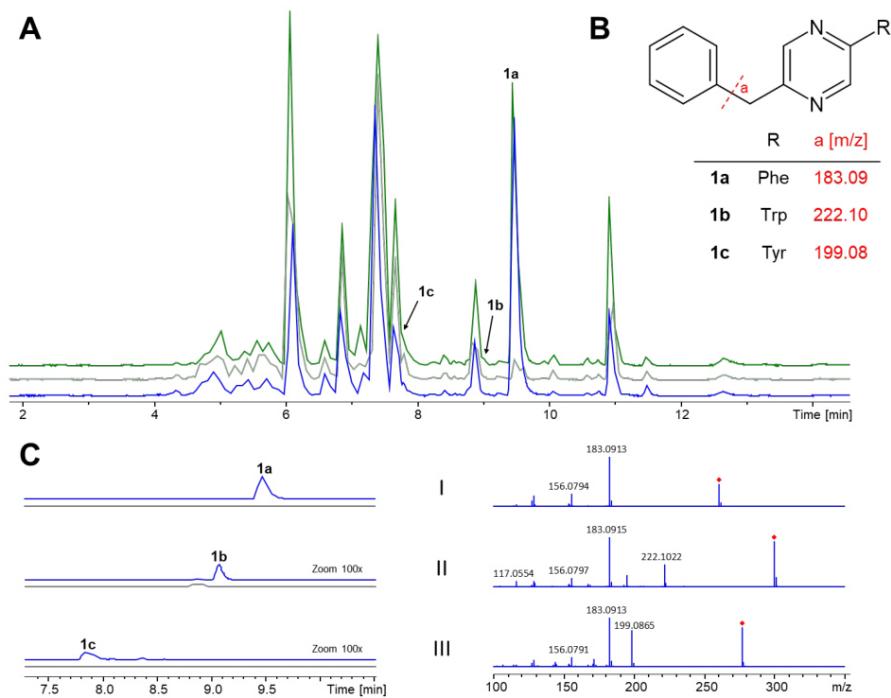
Supplementary Figure 2. HR-HPLC-MS data of **1a**, **1b** and **1c** produced by *X. indica* WT and promoter exchange mutant of *xind01729*. **(A)** Stacked BPC of production from *X. indica* WT (green) and promoter exchange mutant of *xind01729* (grey, non-induced; blue, induced). **(B)** Structure of **1a**, **1b** and **1c** and MS^2 fragments (red). **(C)** Stacked EIC (left) and MS^2 spectra (right) of **1a** (I, rt = 9.5 min, *m/z* [M+H$^+$]$^+$ = 261.138; calculated ion formula $C_{18}H_{17}N_2$; Δppm 1.4), **1b** (II, rt = 9.1 min, *m/z* [M+H$^+$]$^+$ = 300.149; calculated ion formula $C_{20}H_{18}N_3$; Δppm -0.3) and **1c** (III, rt = 7.9 min, *m/z* [M+H$^+$]$^+$ = 277.130; calculated ion formula $C_{18}H_{17}N_2O$; Δppm 0.5)
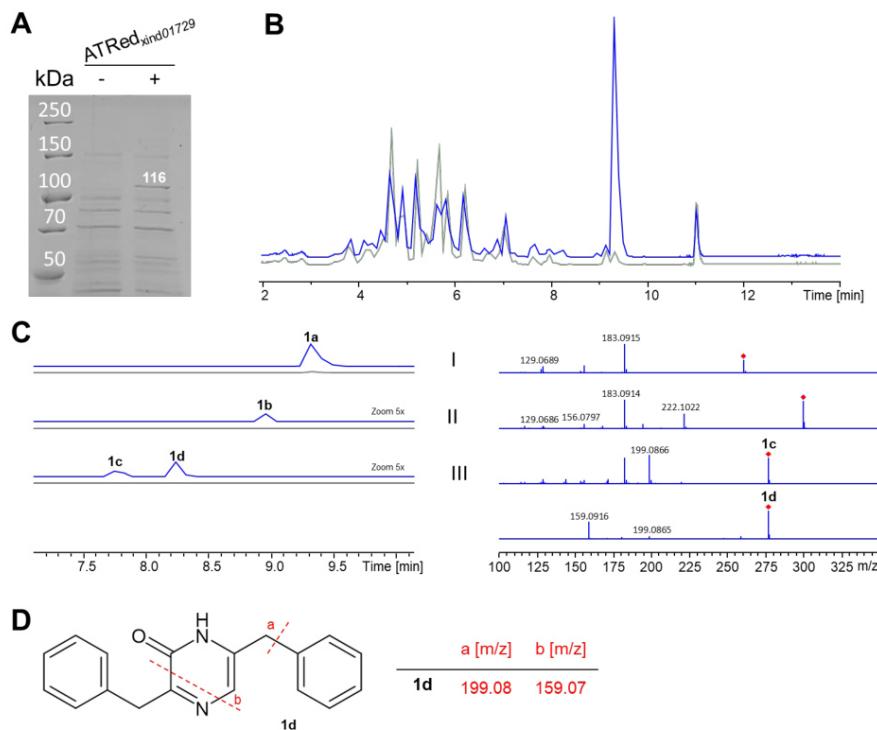
Supplementary Figure 3. HR-HPLC-MS data of **1a**, **1b**, **1c** and **1d** produced by ATRed$_{xind01729}$ after heterologous expression in *E. coli*. **(A)** SDS-PAGE analysis of protein extracts of non-induced (-) and induced (+) sample. The calculated molecular weights of the protein and the size of the marker proteins are indicated. **(B)** Stacked BPC of non-induced (grey) and induced (blue) production from ATRed$_{xind01729}$. **(C)** Stacked EIC (left) and MS^2 spectra (right) of **1a** (I, rt = 9.4 min, m/z [M+H$^+$]$^+$ = 261.138; calculated ion formula C$_{18}$H$_{17}$N$_2$; Δppm 1.0), **1b** (II, rt = 9.0 min, m/z [M+H$^+$]$^+$ = 300.149; calculated ion formula C$_{20}$H$_{18}$N$_3$; Δppm 1.3), **1c** (III, rt = 7.8 min, m/z [M+H$^+$]$^+$ = 277.133; calculated ion formula C$_{18}$H$_{17}$N$_2$O; Δppm 0.0) and **1d** (III, rt = 8.3 min, m/z [M+H$^+$]$^+$ = 277.133; calculated ion formula C$_{18}$H$_{17}$N$_2$O; Δppm 0.8). **(D)** Postulated structure of **1d** and MS^2 fragments (red).
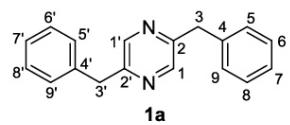
Supplementary Figure 4. Structure of compound **1a.**

Supplementary Table 4. ¹H (500 MHz) and ¹³C (125 MHz) NMR spectroscopic data for compound **1a** in DMSO-d_6 (δ in ppm and J in Hz).

no.	1a	
	δ_C	δ_H (mult., J)
1, 1'	143.7	8.30 (s)
2, 2'	153.7	
3, 3'	41.6	4.05 (s)
4, 4'	138.4	
5, 5', 9, 9'	129.0	7.18 (m)
6, 6', 8, 8'	128.8	7.22 (m)
7, 7'	126.7	7.15 (m)
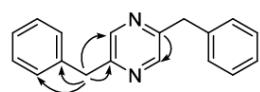
HMBC H→ C

Supplementary Figure 5. Key HMBC correlations of **1**a.
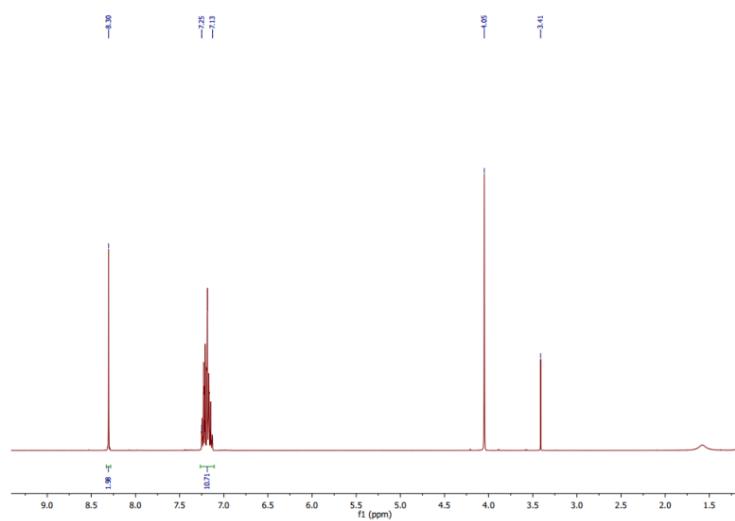
Supplementary Figure 6. ¹H NMR spectrum of compound **1a**.
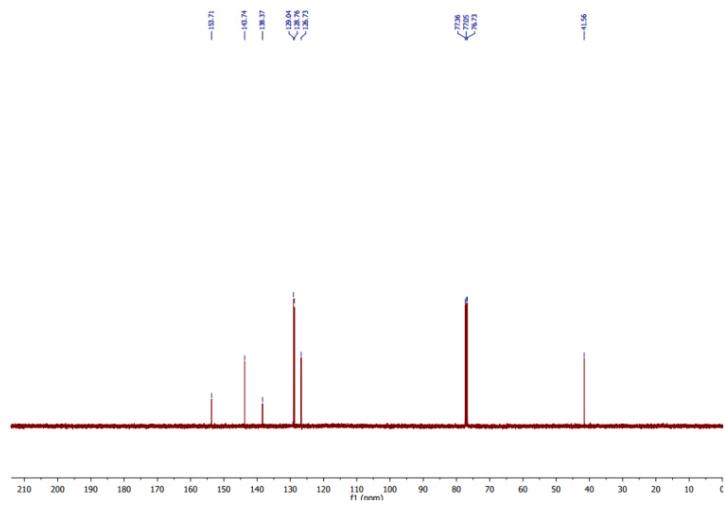
Supplementary Figure 7. ^{13}C NMR spectrum of compound **1a**.
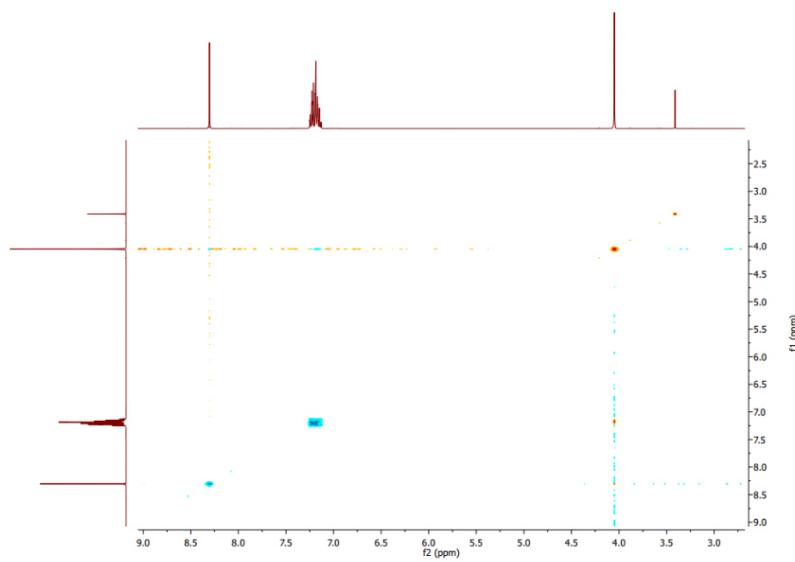
Supplementary Figure 8. COSY spectrum of compound **1a**.
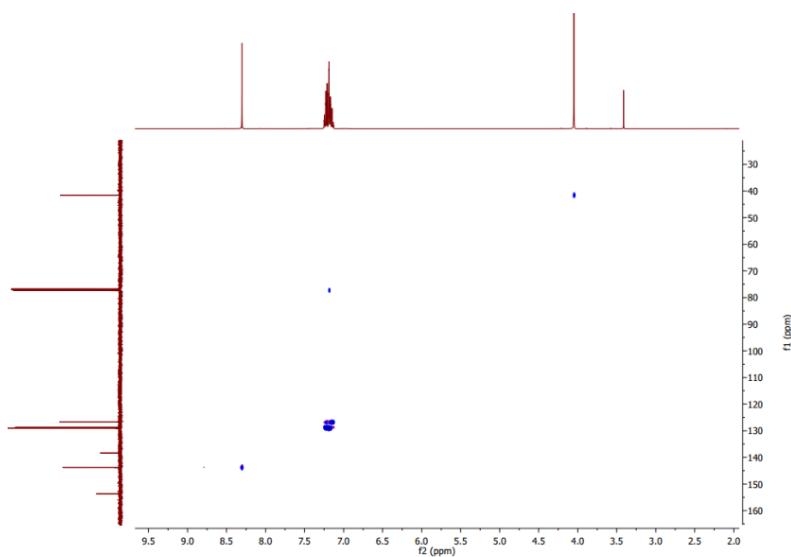
Supplementary Figure 9. HSQC spectrum of compound **1a**.
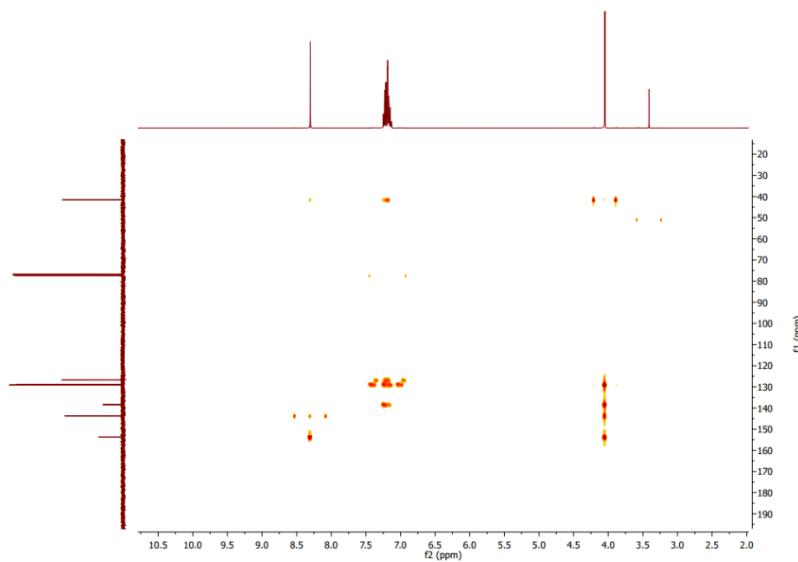
Supplementary Figure 10. HMBC spectrum of compound **1a**.
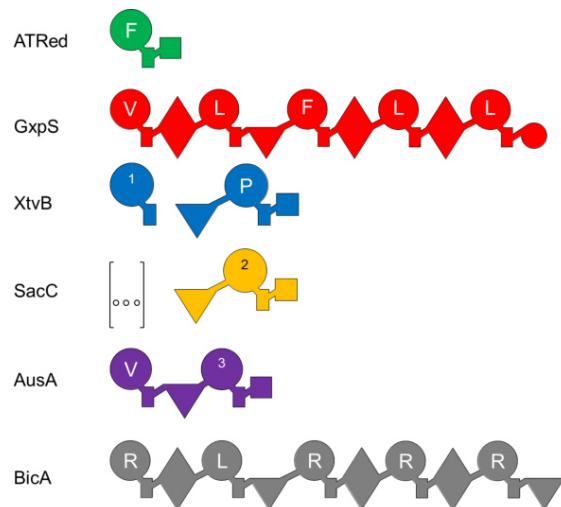
Supplementary Figure 11. Schematic overview of all NRPSs used in this work. ATRed*xind01729* from *X. indica* (WP_047678938), GxpS from *P. laumondii* subsp. *laumondii* TT01[10], XtvB from *X. eapokensis* DL20[11], SacC from *Xenorhabdus* sp. TS4 (PRJNA328577), AusA from *S. lugdunensis* (WP_012990658) and BicA from *X. budapestensis*[12]. Substrate specificities are assigned for all A domains with (1) as 3-hydroxy anthranilic acid, (2) as 3-hydroxy-5-methyl-*O*-methyltyrosine (3) as leucine, tyrosine, phenylalanine, 4-fluoro-phenylalanine, 4-chloro-phenylalanine, 3-chloro-tyrosine and (S)-(+)-a-amino-cyclohexane propionic acid. See Fig. 1 and 2 for assignment of the domain symbols.
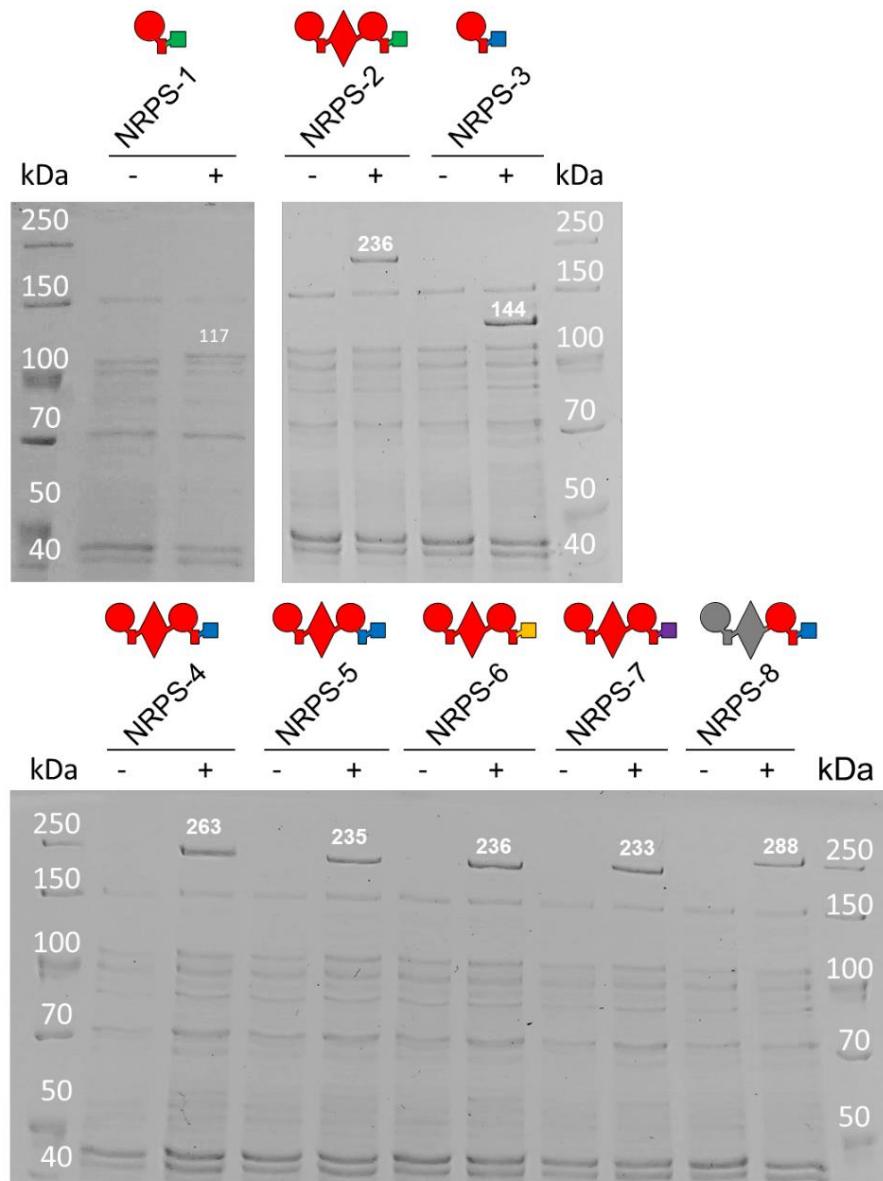
Supplementary Figure 12. SDS-PAGE analysis of engineered proteins. Culture extracts of *E. coli* cells with the respective plasmids after induction with (+) or without arabinose induction (-). The calculated molecular weights of the proteins and the size of the marker proteins are indicated. See Fig. 1 and 2 for assignment of the domain symbols. The colour identifies NRPSs used as building blocks (Supplementary Fig 9).
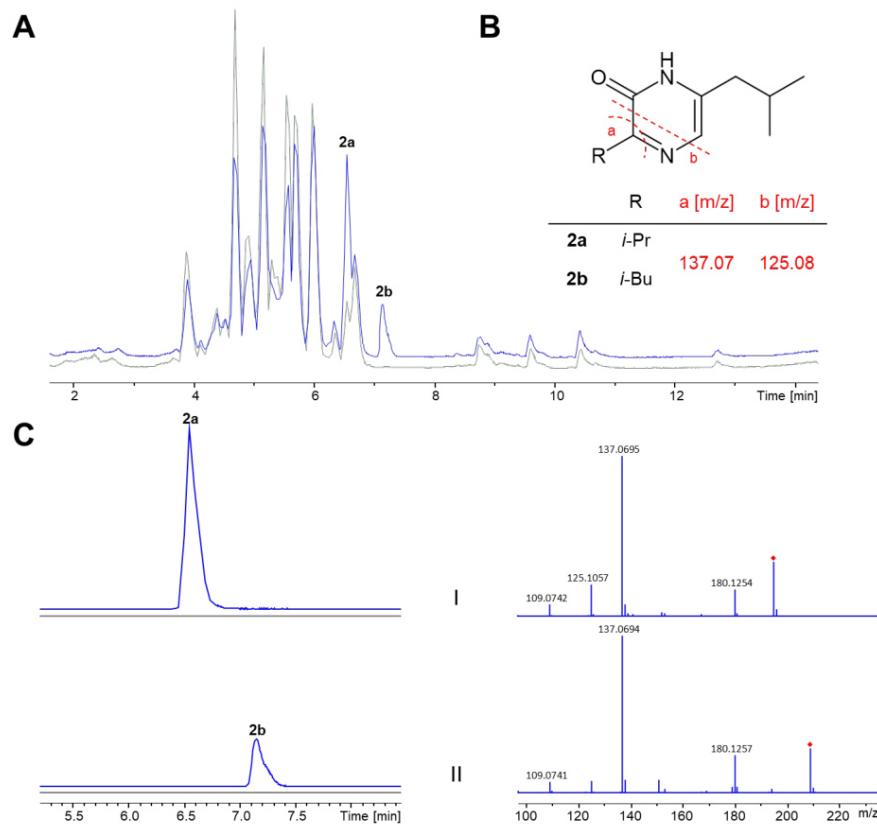
Supplementary Figure 13. HR-HPLC-MS data of **2a** and **2b** produced by NRPS-4 after heterologous expression in *E. coli* DH10B::*mtaA*. **(A)** Stacked BPC of non-induced (grey) and induced (blue) production from NRPS-4. **(B)** Structure of **2a** and **2b** and MS^2 fragments (red). **(C)** Stacked EIC (left) and MS^2 spectra (right) of **2a** (I, rt = 6.6 min, m/z [M+H$^+$]$^+$ = 195.149; calculated ion formula C$_{11}$H$_{19}$N$_2$O; Δppm 1.0) and **2b** (II, rt = 7.2 min, m/z [M+H$^+$]$^+$ = 209.164; calculated ion formula C$_{12}$H$_{21}$N$_2$O; Δppm 1.6).
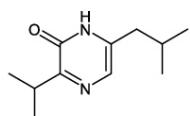
Supplementary Figure 14. Structure of compound **2a**.

Supplementary Table 5. ^{1}H (500 MHz) and ^{13}C (125 MHz) NMR spectroscopic data for compound **2a** in DMSO-d_6 (δ in ppm and J in Hz).

no.	2a	
	δ_C	δ_H (mult., J)
1	160.5	
2	156.2	
3	121.4	7.05 (s)
4	138.4	
5	38.8	2.26 (d, 7.3)
6	28.0	1.92 (m)
7	22.4	1.11 (d, 6.9)
8	22.4	1.11 (d, 6.9)
1'	29.7	3.24 (m)
2'	17.6	0.86 (d, 6.6)
3'	17.6	0.86 (d, 6.6)

HMBC H → C
COSY H → H
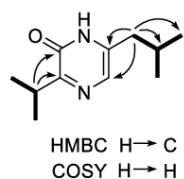
Supplementary Figure 15. Key HMBC and COSY correlations of **2a**.

Attachments
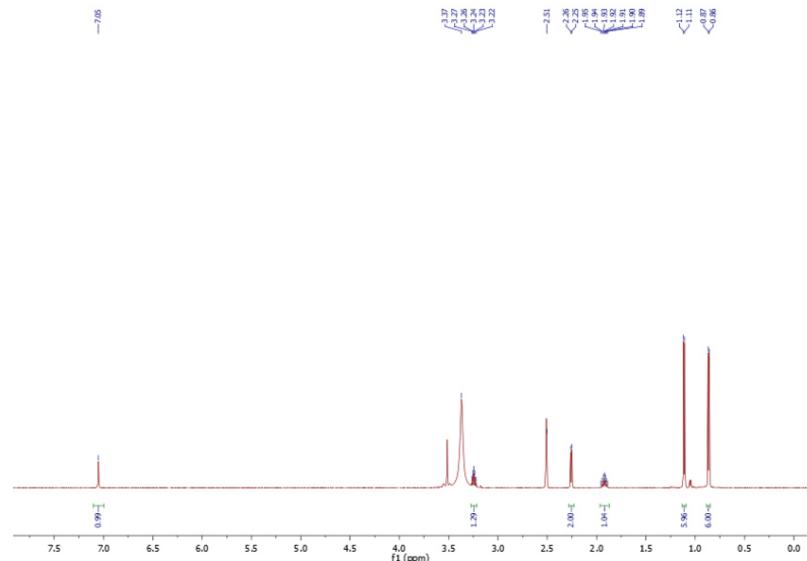
Supplementary Figure 16. ¹H NMR spectrum of compound **2a**.
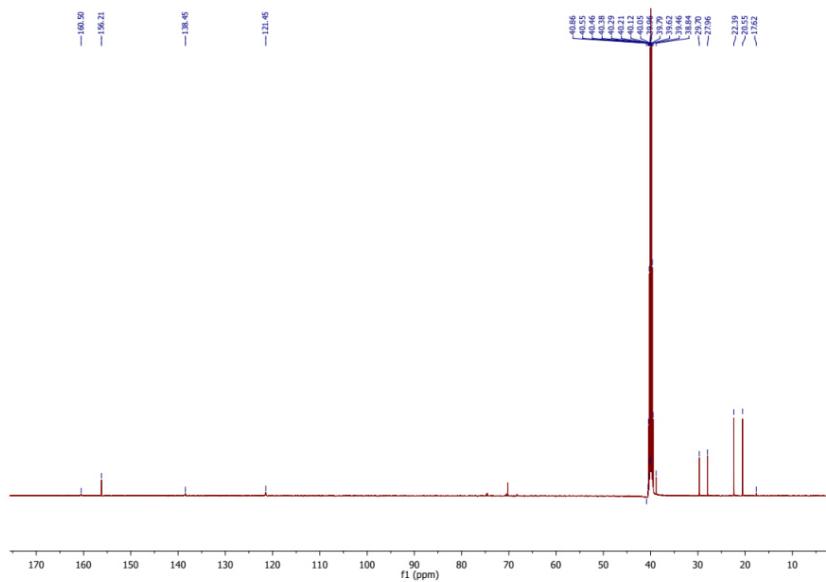
Supplementary Figure 17. ¹³C NMR spectrum of compound **2a**.

236
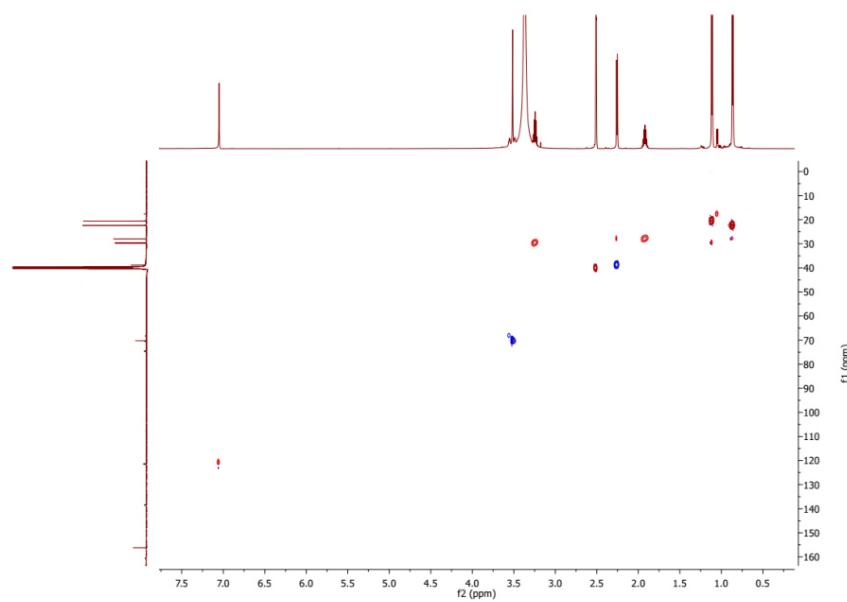
Supplementary Figure 18. HSQC spectrum of compound **2a**.
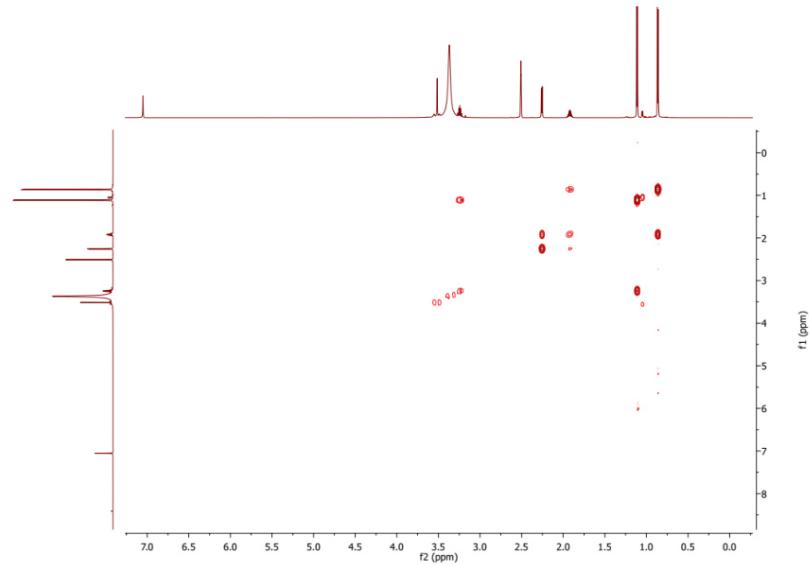
Supplementary Figure 19. COSY spectrum of compound **2a**.
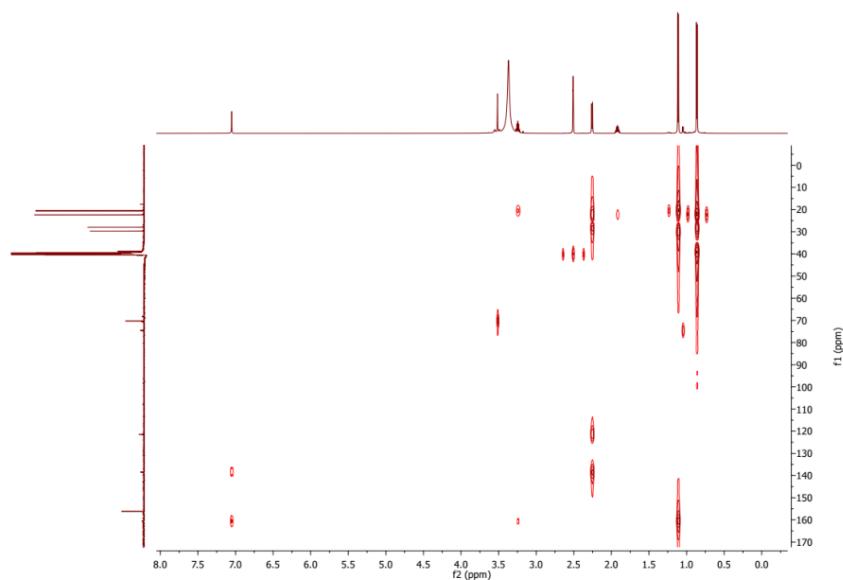
Supplementary Figure 20. HMBC spectrum of compound **2a**.
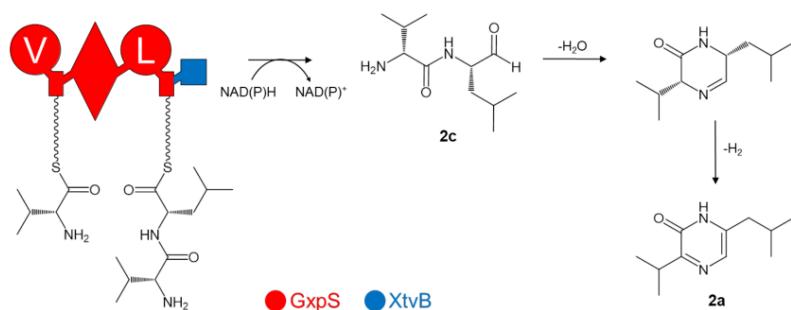
Supplementary Figure 21. Biosynthesis of **2a** by NRPS-4. Standard NRPS biochemistry attaches the nascent D-Val-L-Leu dipeptide on the T2 domain which is released by the R domain via an NAD(P)H-dependent 2-electron reduction of the thioester to produce **2c**. Intramolecular nucleophilic attack of the amino group onto the aldehyde generates a 6-membered Schiff base which oxidizes to yield **2a**. The relaxed substrate specificity of GxpS_A1 can also incorporate Leu beside Val leading to **2b**. See Fig. 1 and 2 for assignment of the domain symbols.The colour code at the bottom identifies NRPSs used as building blocks (Supplementary Fig 9).
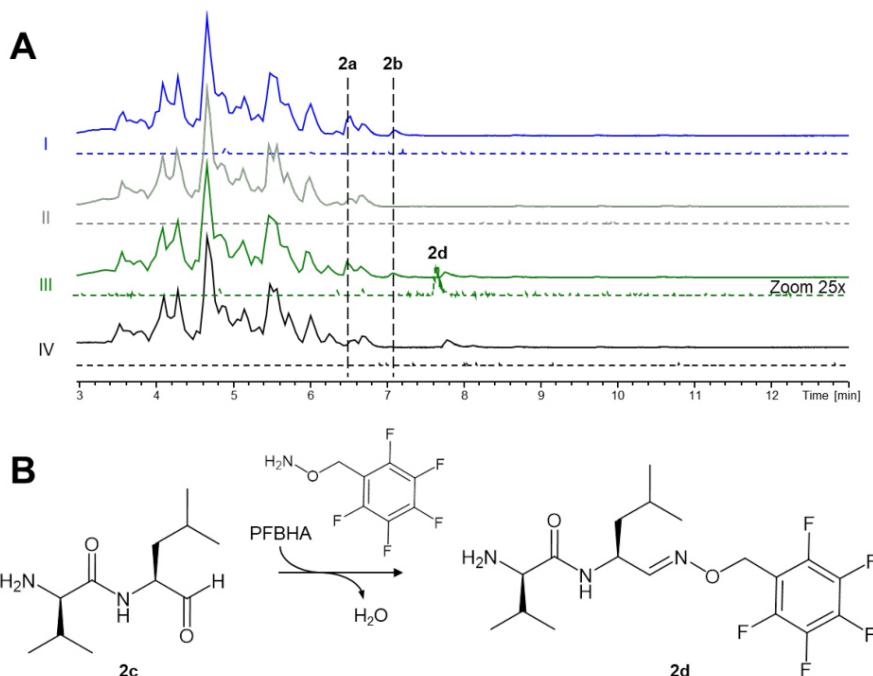
Supplementary Figure 22. (A) HR-HPLC-MS data of **2a** and **2b** as well as derivatization product **2d** of intermediate **2c** produced by NRPS-4 after heterologous production with PFBHA in *E. coli* DH10B::*mtaA*. (I) blue, induced, without PFBHA, (II) grey, non-induced without PFBHA, (III) green, induced with PFBHA and (IV) black, non-induced with PFBHA. The BPC is indicated by continuous lines and the EIC (**2d**; *m/z* [*M*+H]$^+$ = 410.186; rt = 7.6 min; calculated ion formula $C_{18}H_{25}F_5N_3O_2$; Δppm -1.9) by dashed lines. The y-axes of the EICs are increased 25-fold compared to the BPCs. **(B)** Derivatisation of **2c** with PFBHA resulting in **2d**.
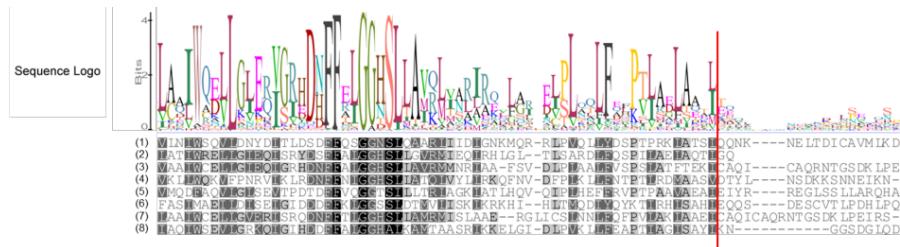
Supplementary Figure 23. Sequence logo and alignment of 86 NRPS T domains. Color code is due to their similarity (black, high similarity; white, low similarity). (1) ATRed_T1 from *X. indica*, (2) GxpS_T1 from *P. luminescens*, (3) GxpS_T2 from *P. luminescens*, (4) XtvB_T2 from *X. eapokensis*, (5) SacC_T3 from *Xenorhabdus* sp. TS4, (6) AusA_T2 from *S. lugdunensis*, (7) GxpS_T5 from pFF1_gxpS_C2int[7] and (8) SrfA-C from *B. subtilis*[13]. All sequences are from *Xenorhabdus* and *Photorhabdus* except (6) and (8). The fusion point of T and R domains is indicated by a red line. The data were analyzed with Geneious 6.1.7.
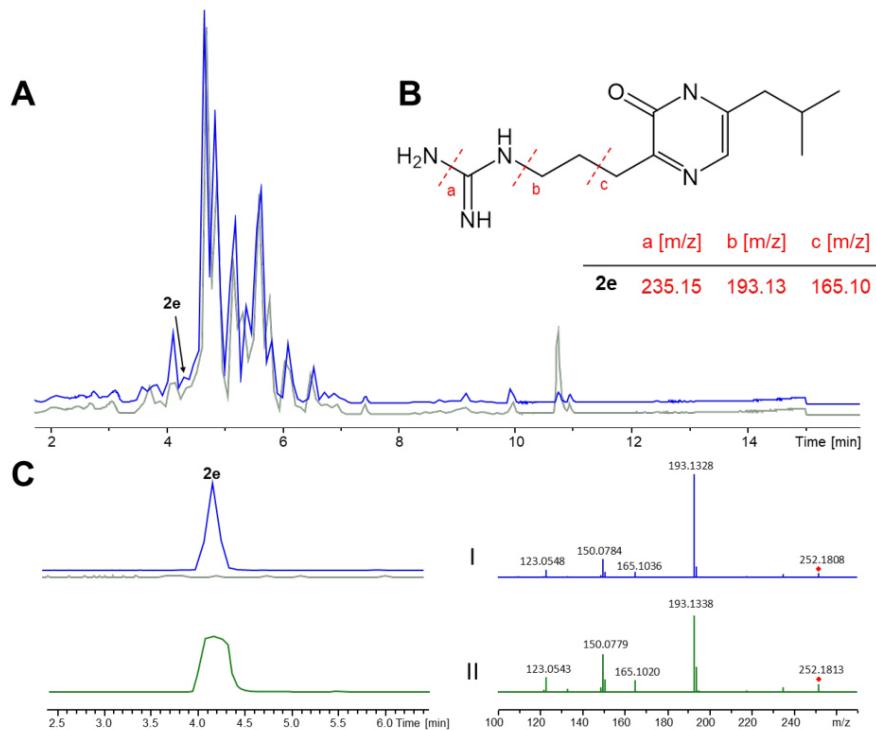
Supplementary Figure 24. HR-HPLC-MS data of compound **2e** produced by NRPS-8 after heterologous expression in *E. coli* DH10B::*mtaA*. **(A)** Stacked BPC of non-induced (grey) and induced (blue) production from NRPS-8. **(B)** Structure of **2e** and MS^2 fragments (red). **(C)** Stacked EIC (left) and MS^2 spectra (right) of **2e** (I, blue, rt = 4.2 min, *m/z* [*M*+H$^+$]$^+$ = 252.181; calculated ion formula C$_{12}$H$_{22}$N$_5$O; Δppm 1.3) and chemically synthesized **2e** (II, green, rt = 4.2 min, *m/z* [*M*+H$^+$]$^+$ = 252.181; calculated ion formula C$_{12}$H$_{22}$N$_5$O; Δppm 1.3)
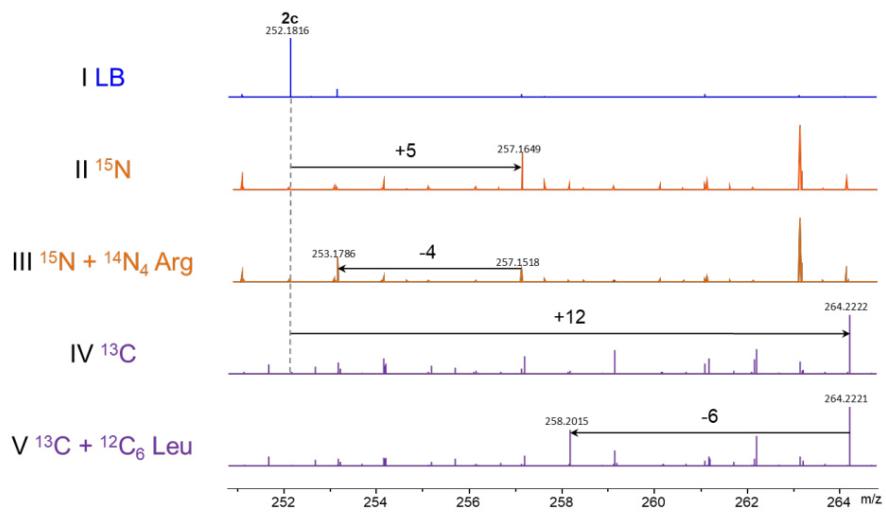
Supplementary Figure 25. Labeling experiments and HR-MS of compounds **2e** produced by NRPS-8 in *E. coli*. MS data of inverse labeling experiments in (I) LB media (blue), (II) ^{15}N media (orange), (III) ^{15}N media supplemented with $^{14}C_4$ Arg (IV) ^{13}C media (purple) and (V) ^{13}C media supplemented with $^{12}C_6$ Leu. The shifts due to incorporation of labelled precursors are indicated by arrows.
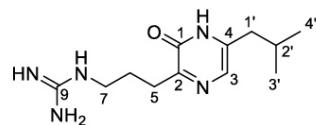
Supplementary Figure 26. Structure of compound **2e.**

Supplementary Table 6. ^{1}H (500 MHz) and ^{13}C (125 MHz) NMR spectroscopic data for compound **2e** in DMSO-d_6 (δ in ppm and J in Hz).

no.	2e	
	δ_C, type	δ_H (mult., J)
1	undetected	
2	156.2	
3	121.2	7.05 (s)
4	138.4	
5	29.3	2.65 (t, 7.4)
6	25.6	1.82 (m)
7	40.8	3.16 (m)
8		7.50 (s)
9	157.0	
1'	38.8	2.28 (d, 7.2)
2'	28.1	1.93 (m)
3'	22.4	0.87 (d, 6.2)
4'	22.4	0.87 (d, 6.2)
-N<u>H</u>CO		12.1 (s)
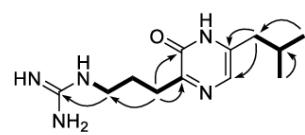
HMBC H→ C
COSY H→ H

Supplementary Figure 27. Key HMBC and COSY correlations of **2e**.
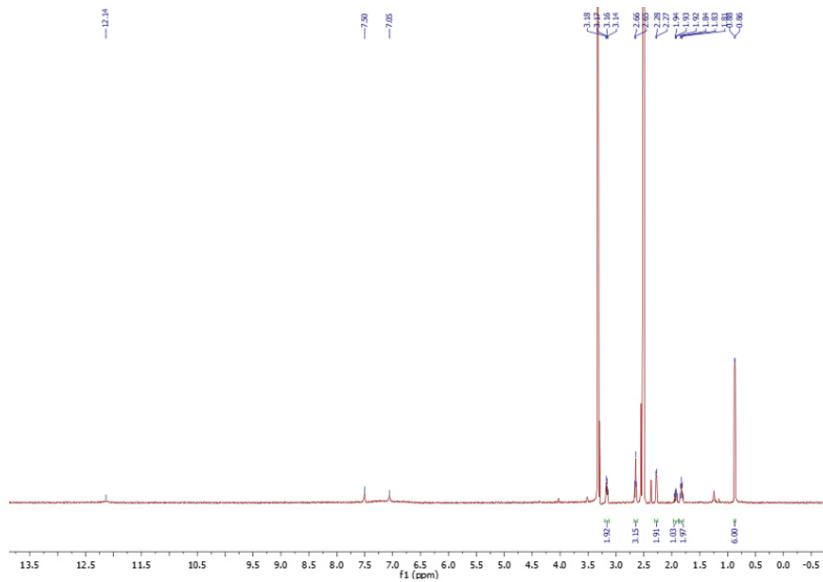
Supplementary Figure 28. ^{1}H NMR spectrum of synthesized **2e.**
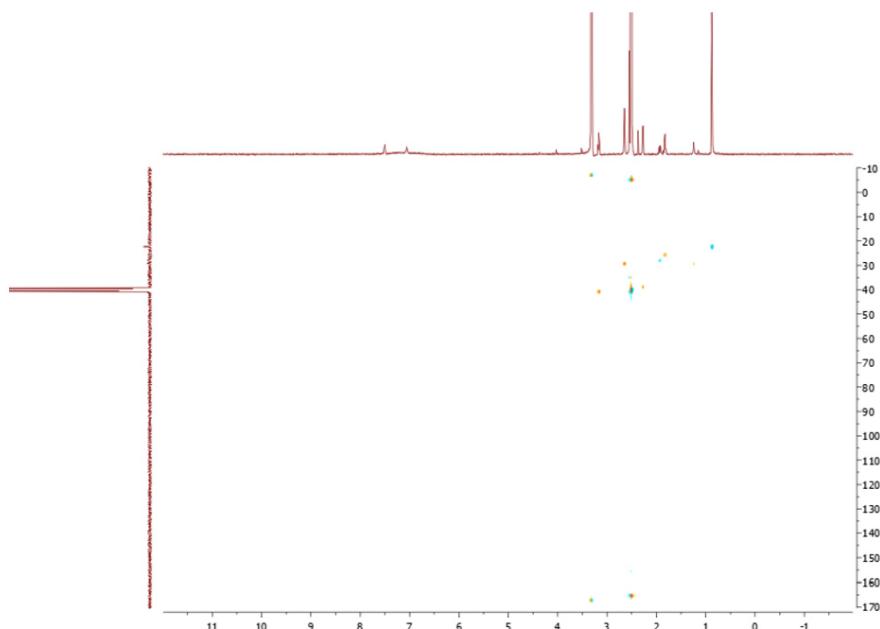
Supplementary Figure 29. HSQC NMR spectrum of synthesized **2e.**
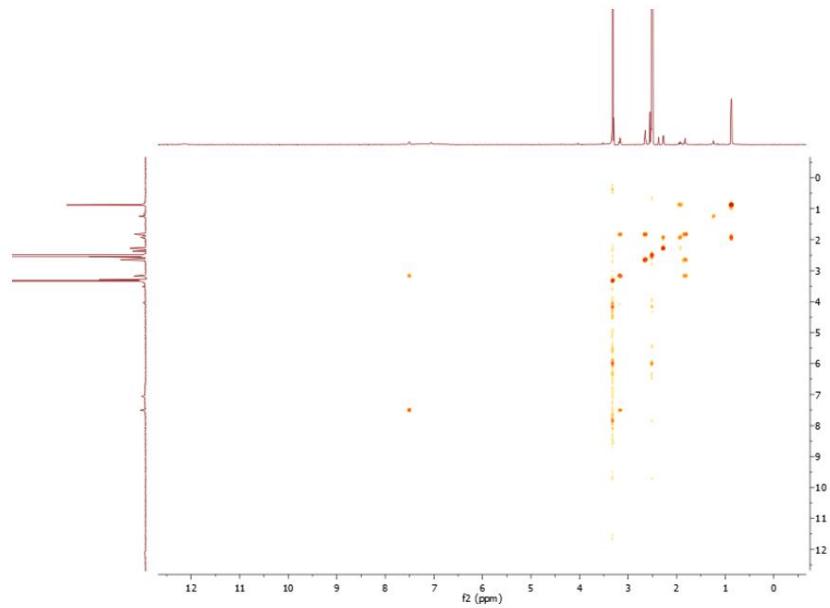
Supplementary Figure 30. COSY NMR spectrum of synthesized **2e.**
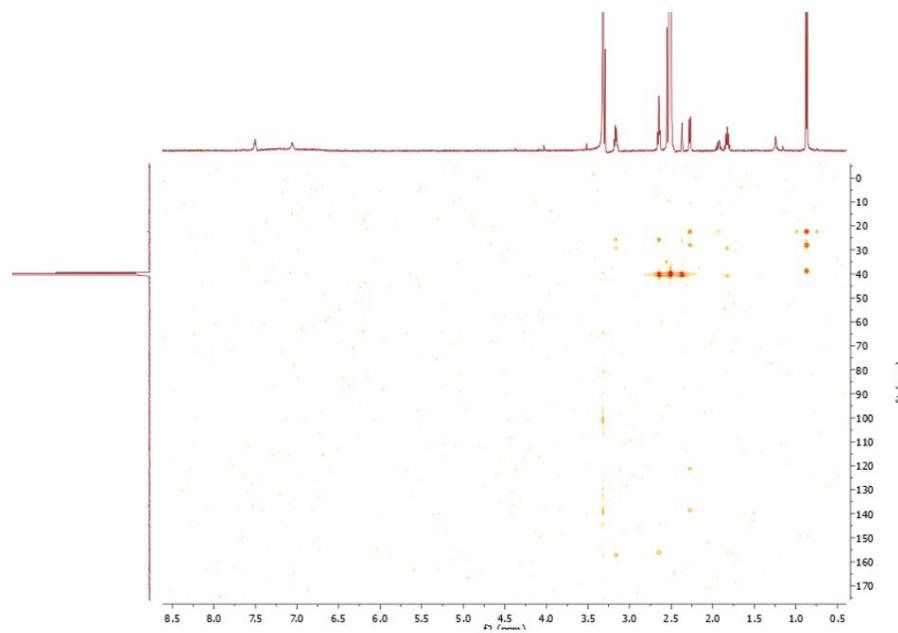
Supplementary Figure 31. HMBC NMR spectrum of synthesized **2e.**

References

[1]   D. Hanahan, *J. Mol. Biol.* **1983**, *166*, 557–580.

[2]   O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290–1294.

[3]   S. Thoma, M. Schobert, *FEMS Microbiol. Lett.* **2009**, *294*, 127–132.

[4]   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.

[5]   C. Kegler, F. I. Nollmann, T. Ahrendt, F. Fleischhacker, E. Bode, H. B. Bode, *ChemBioChem* **2014**, *15*, 826–828.

[6]   C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert, *Nat. Comm.* **2018**, *9*, 4366.

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

[8]   E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, C. Dauth, Q. Zhou, M. Kaiser, P. Klemmt, H. B. Bode, *ChemBioChem* **2015**, *16*, 1115–1119.

[9]   K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81-W87.

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

[11]   H. Wolff, H. B. Bode, *PloS one* **2018**, *13*, e0194297.

[12]   S. W. Fuchs, C. C. Sachs, C. Kegler, F. I. Nollmann, M. Karas, H. B. Bode, *Anal. Chem.* **2012**, *84*, 6948–6955.

[13]   A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science* **2008**, *321*, 659–663.

6.4   Supporting information

6.4.1 R domains in engineered NRPSs

6.4.1.1   Material and methods

Strains

E. coli and *X. eapokensis* cells were grown in liquid or solid (1.5 % (w/v) agar) lysogeny broth (LB) medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH 7.5). *S. cerevisiae* cells were grown in liquid or solid (1.5 % (w/v) agar) yeast extract peptone dextrose (YPD) medium (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) at 30 °C. *E. coli* cells were cultivated at 37 °C, all others at 30 °C. Kanamycin (50 µg/ml) and G418 (200 µg/mL) were used as selection markers. All strains that were used and generated in this work are summarized in Supplementary Tab. 1.1.

Supplementary Table 1.1. Strains used and generated in this work. NRPS-4 from [165] is indicated with an * to avoid confusion with NRPSs constructed in this work.

Strain	Genotype	Reference
E. coli DH10B	F_ mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lac*ZΔ, M15, Δ*lac*X74 *rec*A1 *end*A1 *ara*D 139 Δ(*ara, leu*)7697 *gal*U *gal*K λ*rpsL* (Strr) *nup*G	[211]
E. coli DH10B::*mtaA*	DH10B with *mtaA* from pCK_*mtaA* Δ*entD*	[212]
E. coli DH10B::*mtaA* pXst4_*pxaA*	*E. coli* DH10B::*mtaA* pXst4_*pxaA*, Kan^R	[166]
E. coli DH10B::*mtaA* pFF1_*NRPS-1*	*E. coli* DH10B::*mtaA* pFF1_*NRPS-1*, Kan^R	This work
E. coli DH10B::*mtaA* pFF1_*13A_xabABC_kolS_txlA_gxpS*	*E. coli* DH10B::*mtaA* pFF1_*13A_xabABC_kolS_txlA_gxpS*, Kan^R	[163]
E. coli DH10B::*mtaA* pAT41_*NRPS-2*	*E. coli* DH10B::*mtaA* pAT41_*NRPS-2*, Kan^R	This work
E. coli DH10B::*mtaA* pAT41_*NRPS-4**	*E. coli* DH10B::*mtaA* pAT41_*NRPS-4**, Kan^R	[165]
S. cerevisiae CEN.PK 2-1C	MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2	Euroscarf
X. eapokensis DL20		DSMZ

Isolation and purification of DNA

Genomic DNA was isolated using the Gentra Puragene Yeast/Bact Kit (Qiagen). Plasmids from *E. coli* were isolated using Invisorb Spin Plasmid Mini Two (STRATEC Biomedical AG). DNA from polymerase chain reactions (PCRs) was purified with MSB Spin

Attachments

PCRapace (STRATEC Biomedical AG) or from 1% Tris-acetate-ethylenediaminetetraacetic acid (TAE) agarose gel using Invisorb Spin DNA Extraction (STRATEC Biomedical AG) and additionally digested with *Dpn*I (Thermo Fisher Scientific) if the PCR template was plasmid-based. DNA from yeast was isolated by alkaline lysis.

PCR, cloning of plasmids and transformation of cells

PCR was performed with oligonucleotides (Supplementary Tab. 1.2) obtained from Eurofins Genomics and S7 Fusion High-Fidelity DNA Polymerase (Biozym) or Q5 High-Fidelity DNA polymerase (New England BioLabs) according to the manufacturers' instructions. Homology arms for cloning were introduced via primer design and a two-step PCR. The vector pFF1 was digested with *Eco*RI and *Sgs*I (Thermo Fisher Scientific).

Supplementary Table 1.2. Oligonucleotides used in this work.

Plasmid	Oligonucleotide	Sequence (5'->3')	Template
pFF1_*NRPS-1*	AT_286	TTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAACTTCACAATTAGTACCTCTTACCCAG	pCX2_*bm76III*
	AT_292	TTTCATTATTTGATTTTTTATCACTATTCAGATAGGTATCGATATGTGCAGCTAACTGAGCAACC	
	AT_293	GATACCTATCTGAATAGTGATAAAAAATCAAATAATG	*X. eapokensis* DL20
	AT_289	TCATGAACTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCCTTACTTTCAGGTTTATATGACGGTATGCTTG	
pAT41_*NRPS-2*	AT_226	TGGAACGCGACAGAAACC	pAT41_*NRPS-4**
	pAT41_bb_rv	GGAATTCCTCCTGTTAGCCC	
	AT_491	TTGGGCTAACAGGAGGAATTCCATGCCTATGTCATGCAATGGTATTAAC	pFF1_*13A_xabABC_kolS_txlA_gxpS*
	AT_492	GATAGGGGGTTTCTGTCGCGTTCCAAGTTTCCAATAACAACTTGCGCTC	

Cloning was done by transformation-associated recombination (TAR)[213] in yeast and NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) according to the manufacturers' instructions. *E. coli* cells were transformed with the plasmids by electroporation. Plasmids were verified by restriction digest or sequencing (Eurofins Genomics). All plasmids that were used and generated in this work are summarized in Supplementary Tab. 1.3.

Supplementary Table 1.3. Plasmids used and generated in this work.

Plasmid	Genotype	Reference
pCX2_ pXst4_*pxaA*	2μ ori, kanMX4, T7lac promoter, P$_{BAD}$ promoter, pBR322 ori, Kan^R, G418, *pxaA*_C1A1T1C2A2T2TE	[166]
pFF1	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, MCS	[212]
pFF1_*NRPS-1*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *pxaA*_C1A1T1C2A2T2-*xtvB*_R	This work
pAT41_*NRPS-2*	2μ ori, URA3, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *xabABC*_C1A1T1C2-*kolS*_A2T2C3-*gxpS*_A2T2-*xtvB*_R	This work
pAT41_*NRPS-4**	2μ ori, URA3, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *gxpS*_A1T2CE2A2T2-*xtvB*_R	[165]
pFF1_*13A_xabABC_kol S_txlA_gxpS*	2μ ori, kanMX4, P$_{BAD}$ promoter, pCOLA ori, Ypet-Flag, Kan^R, *xabABC*_C1A1T1C2-*kolS*_A2T2C3-*txlA*_A3T3C4-*gxpS*_A5T5TE	[163]

Heterologous expression and extract preparation

An overnight culture of *E. coli* DH10B::*mtaA* with the plasmid of interest was inoculated (1:100) into 10 mL LB medium with respective selection marker, 0.02 mg/mL L-arabinose for induction and 2 % (v/v) amberlite XAD-16 (Sigma-Aldrich). After 48 h at 22 °C and 160 rpm, the XAD-16 was harvested by decanting the supernatant and incubated for 30 min with one culture volume MeOH with 160 rpm. The organic phase was filtered and evaporated to dryness under reduced pressure as described previously.[164] For LC/MS analysis, the extracts were dissolved in 1 mL MeOH and a 1:10 dilution centrifuged for 20 min at 13.300 rpm.

In vivo aldehyde derivatization

The cultures for heterologous expression were supplemented with 0.5 mM PFBHA.[214]

LC/MS analysis

All measurements were carried out as described previously[165] by using an UltiMate 3000 liquid chromatography (LC) system (Dionex) on a C18 column (ACQUITY UPLC BEH, 1.7 μm, 2.1mm*100 mm (Waters); gradient of acetonitrile (ACN)/0.1 % formic acid in H$_2$O/0.1 % formic acid, 5 % to 95 %, 15 min, flow rate 0.4 mL/min) coupled to an

AmaZonX electrospray ionization (ESI)-ion trap-MS (Bruker). For HR-MS, an Impact II ESI-quadrupole orthogonal time of flight-MS (Bruker) with internal 10 mM sodium formate calibrant was used. The software DataAnalysis 4.3 (Bruker) was used to evaluate the measurements. The software MetaboliteDetect 2.1 (Bruker) was used to calculate differences in chromatograms.

Stable isotope labelling

Stable isotope labelling[53] was carried out in 5 mL ^{13}C or -^{15}N medium (20 g/L ISOGRO-^{13}C or -^{15}N powder (Sigma-Aldrich), 1.8 g/L K_2HPO_4, 1.4 g/L KH_2PO_4, 1 g/L $MgSO_4$, 10 mg/L $CaCl_2$, pH 7.0 in water) under cultivation conditions as mentioned above. The overnight culture was washed in ^{13}C respectively -^{15}N medium before inoculation. Feeding experiments in ^{13}C- or ^{15}N medium were supplemented with 2 mM ^{12}C-L-AAs.

Peptide purification

Peptides were isolated from extracts of 4 L *E. coli* DH10B::*mtaA* cultures using a 1260 Infinity II LC system on a phenyl hexyl column (Kinetex; 5μm Phenyl Hexyl 100 Å, AXIA Packed LC Column 250 x 21.2 mm; gradient of ACN/0.1 % formic acid in H_2O/0.1 % formic acid, 35 % to 45 %, 22 min, flow rate 20 mL/min) coupled to a G6125B LC/mass selective detector ESI-MS (Agilent).

NMR analysis

NMR analysis was performed by Yi-Ming Shi (Goethe-university Frankfurt) on an AVANCE III HD 500 MHz spectrometer (Bruker) using DMSO-d_6 as solvent.

Sequence alignment

Multiple protein sequence alignments were prepared with the Clustal Omega algorithm (European Bioinformatics Institute)[215] and visualized with Geneious 6.1.7 (Biomatters).

Homology modelling

Homology modelling was carried out using the software MOE 2016 (Chemical Computing Group).

Chemical synthesis

Chemical synthesis of the peptide aldehyde was performed in a Syro Wave peptide synthesizer (Biotage) using 100 mg (21 μmol) H-Leu-H NovaSyn TG resin (Sigma-Aldrich) as described by Schilling et al.[216]

For AA coupling, 6 eq. fluorenylmethoxycarbonyl (Fmoc)-L-Ala respectively 6 eq. Fmoc-L-Pro (Iris Biotech, c = 0.2 M) in dimethylformamide (DMF), 6 eq. *O*-(6-chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HCTU, Carl Roth, c = 0.54 M) in DMF and 12 eq. *N,N*-diisopropylethylamine (DIPEA, Iris Biotech, c = 2.4 M) in *N*-methylpyrrolidone (NMP) were used. The Fmoc protection group was cleaved with 40 % piperidine (Iris Biotech) in NMP and 20 % piperidine in NMP. The resin was washed after every coupling and deprotection step with NMP and finally with dichlormethane (DCM).

Fatty acid coupling was performed as described previously with 10 eq. butyric acid, 10 eq. *O*-(7-azabenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HATU, Carbolution Chemicals), 10 eq. 1-hydroxy-7-azabenzotriazole (HOAt, Carbolution Chemicals) and 20 eq. DIPEA in DMF over night at 37 °C.[163]

The peptide was washed extensively with DCM, cleaved from the resin with 2 mL of 79.95 % ACN/20 % water/0.05 % trifluoro acetic acid (v/v/v) overnight as described by Schilling *et al.*[216] and evaporated to dryness under reduced pressure.

6.4.1.2 Supplementary data
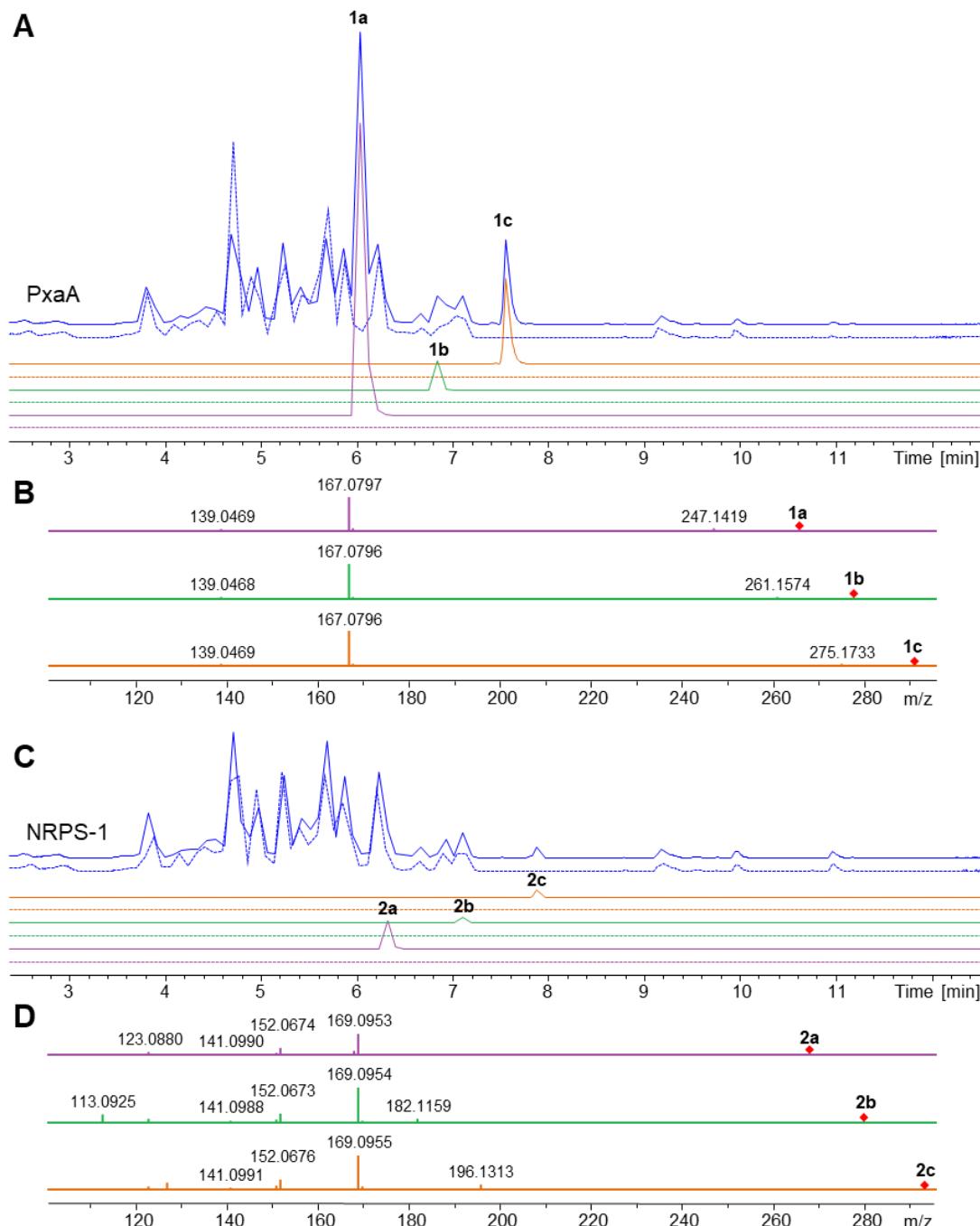
Supplementary Figure 1.1. High-resolution HPLC-MS data of 1a, 1b, 1c, 2a, 2b and 2c produced by PxaA and NRPS-1. A. BPC (blue) and EIC (**1a**, *m/z* [M+H]$^+$ = 265.155; **1b**, *m/z* [M+H]$^+$ = 279.170 and **1c**, *m/z* [M+H]$^+$ = 293.184) after heterologous expression (induced: continuous line; non-induced: dashed line) of PxaA in *E. coli* DH10B::*mtaA*. The colours of the EIC chromatograms are according to the length n of the fatty acid side chain. **B.** MS^2 spectra of **1a**, **1b** and **1c**. The parental ions are depicted by red diamonds. **C.** BPC (blue) and EIC of **2a** (*m/z* [M+H]$^+$ = 267.170; calculated ion formula $C_{14}H_{23}N_2O_3$; Δppm 1.5), **2b** (*m/z* [M+H]$^+$ = 281.186) and **2c** (*m/z* [M+H]$^+$ = 295.202) after heterologous expression (induced: continuous line; non-induced: dashed line) of NRPS-1 in *E. coli* DH10B::*mtaA*. The colours of the EIC chromatograms are according to the length n of the fatty acid side chain. **D.** MS^2 spectra of **2a**, **2b** and **2c**. The parental ions are depicted by red diamonds.

A

C

	a [m/z]	b [m/z]	c [m/z]
2a	169.09 $C_8H_{13}N_2O_2$	152.07 $C_8H_{10}N_1O_2$	100.07 $C_5H_{10}N_1O_1$

B
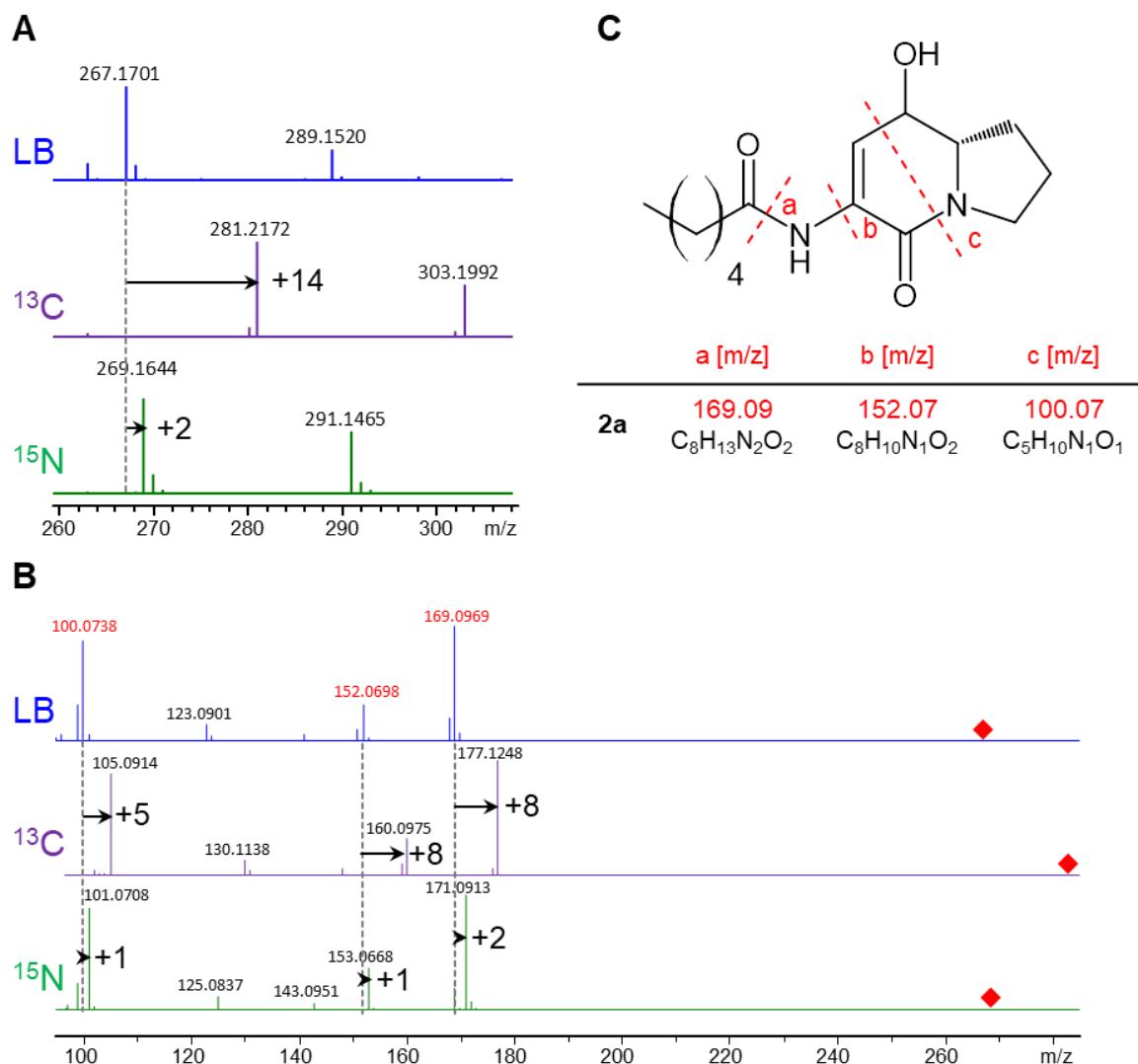
Supplementary Figure 1.2. Stable isotope labelling of 2a. A. High-resolution MS spectra at 6.25 min of HPLC-MS analysis after heterologous production of NRPS-1 in *E. coli* DH10B::*mtaA* in LB (blue), ^{13}C (purple) and ^{15}N (green) media. The shifts due to stable isotope incorporation are indicated by arrows. **B.** MS^2 spectra of *m/z* $[M+H]^+ = 267.17$ (LB, blue), *m/z* $[M+H]^+ = 281.17$ (^{13}C, purple) and *m/z* $[M+H]^+ = 269.16$ (^{15}N, green). The shifts due to stable isotope incorporation are indicated by arrows and the parental ions by red diamonds. **C.** Structure and fragmentation of **2a** (*m/z* $[M+H]^+ = 267.1699$).

Supplementary Figure 1.3. Structure of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).
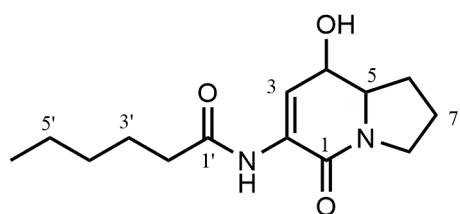
Attachments

Supplementary Table 1.4. [1]H (500 MHz) and ^{13}C (125 MHz) NMR spectroscopic data for compound **2a** in DMSO-d6 *(δ in ppm)*. Table prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).

#	2a	
	δ_H	δ_C
1		158.8
2		134.8
3	5.99 (s)	111.4
4	5.80 (s)	73.4
5	3.66 (m)	60.1
6	1.70 (m)	27.6
	1.94 (m)	
7	1.85 (m)	22.1
	1.70 (overlap)	
8	3.33 (m)	45.7
1'		171.7
2'	2.45 (m)	34.0
	2.56 (m)	
3'	1.49 (m)	24.4
4'	1.22 (m)	30.8
5'	1.22 (m)	22.0
6'	0.82 (t, 7.0)	13.9
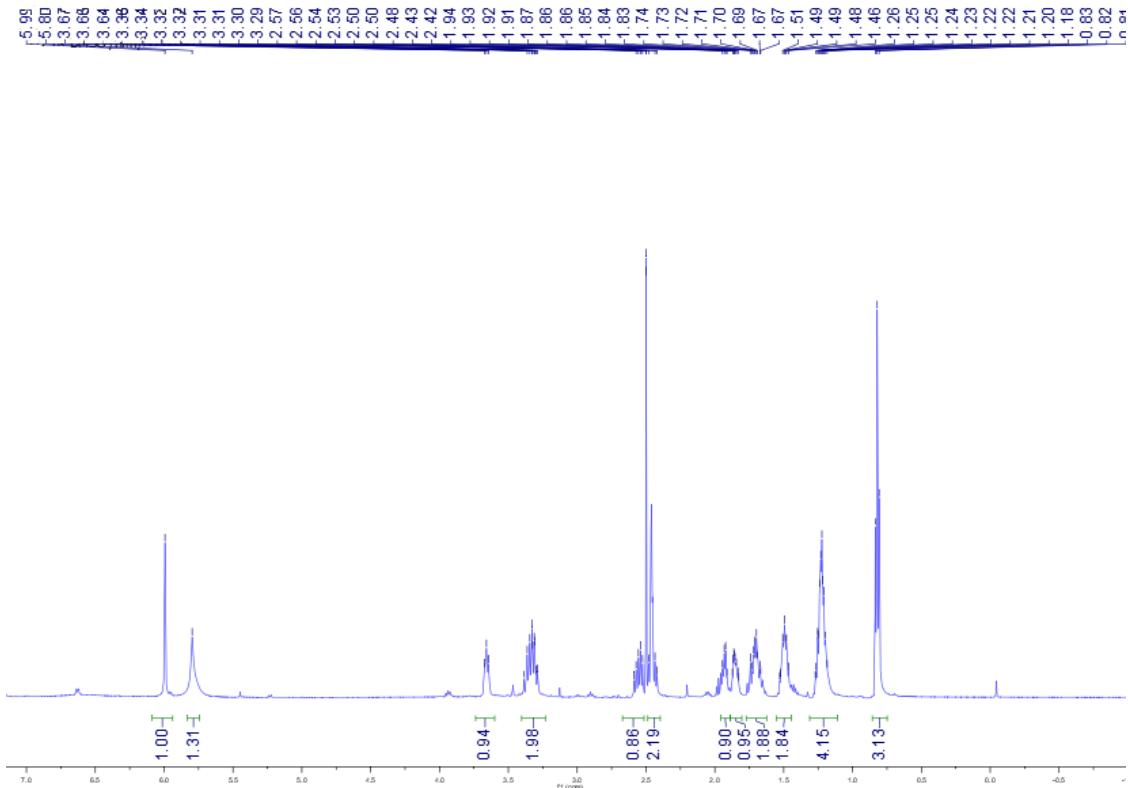
Supplementary Figure 1.4. ^{1}H NMR spectra of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).
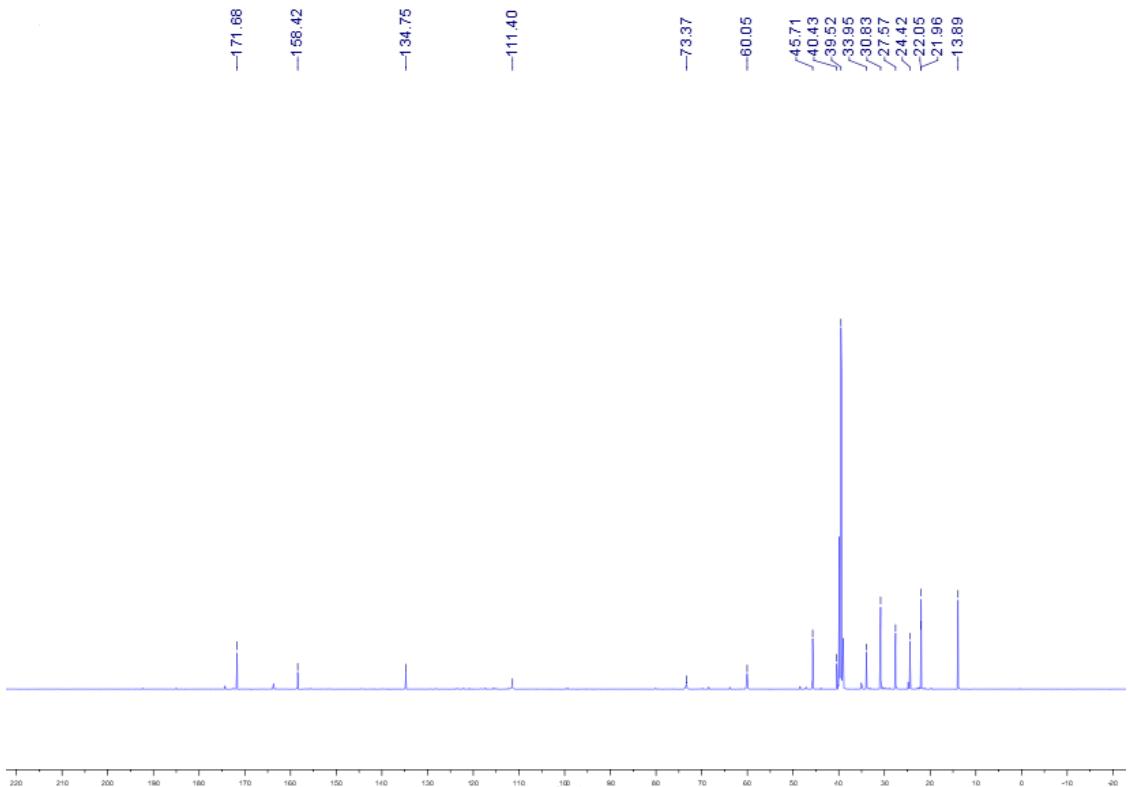
Supplementary Figure 1.5. ^{13}C NMR spectra of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).

Attachments
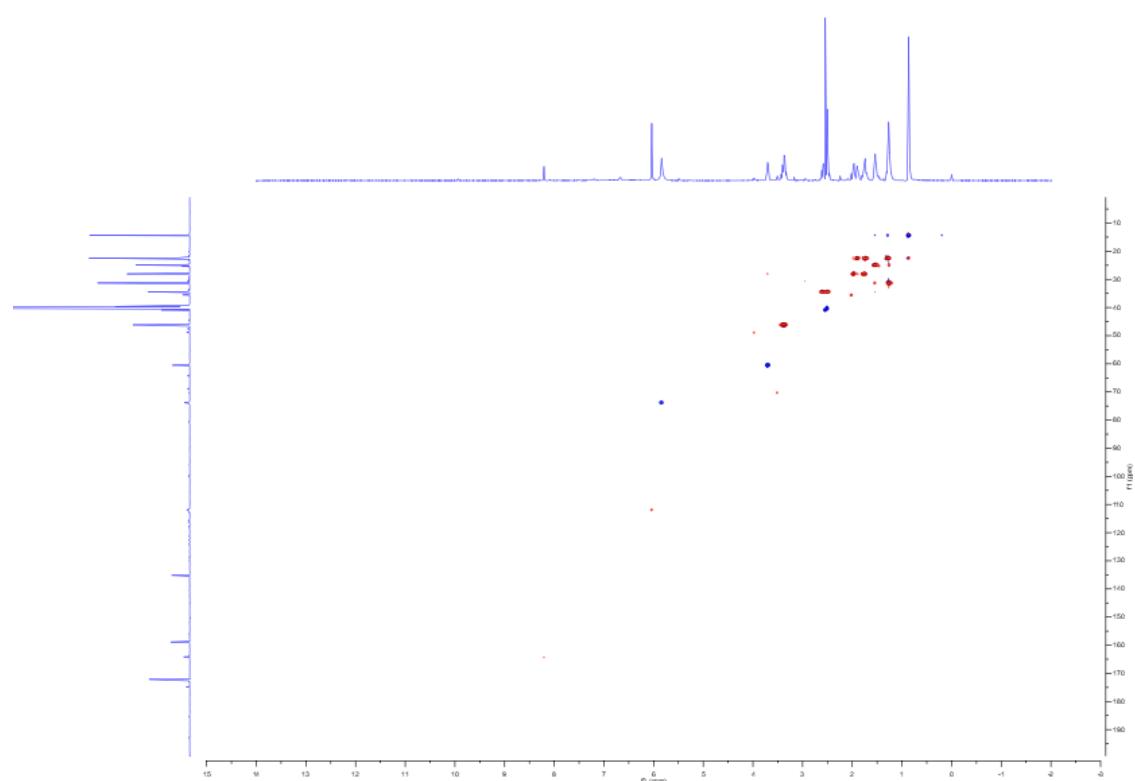
Supplementary Figure 1.6. HSQC spectra of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).
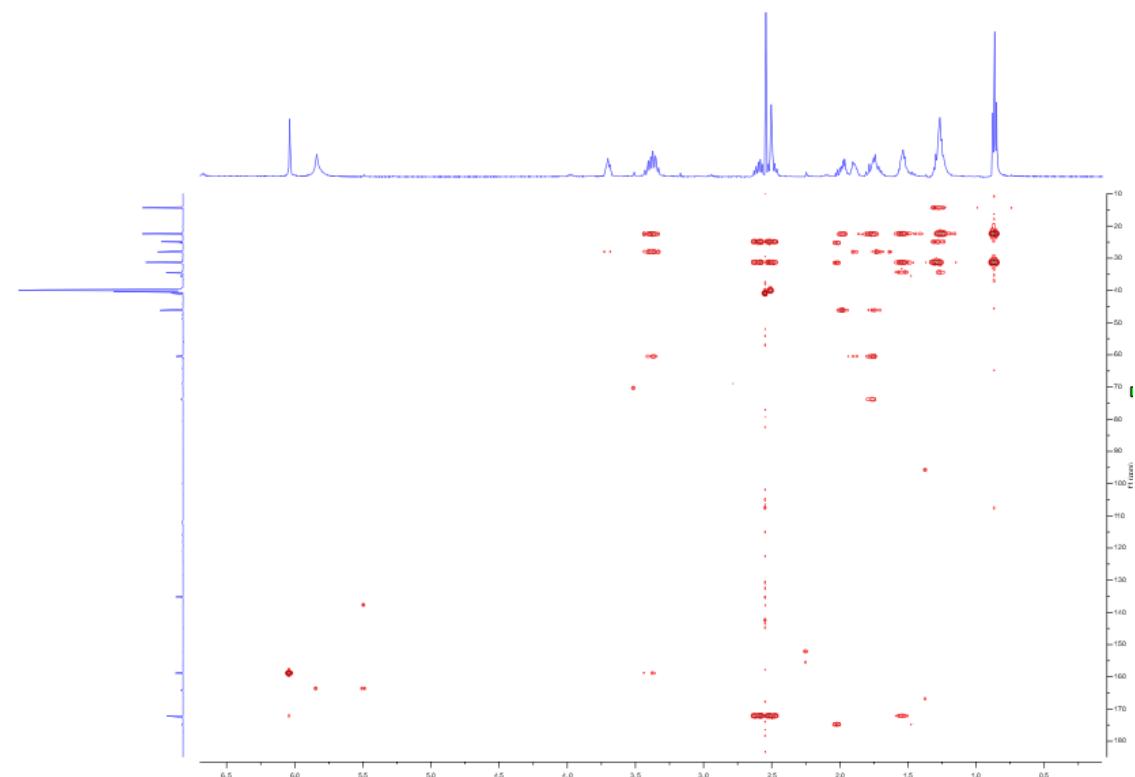
Supplementary Figure 1.7. HMBC spectra of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).
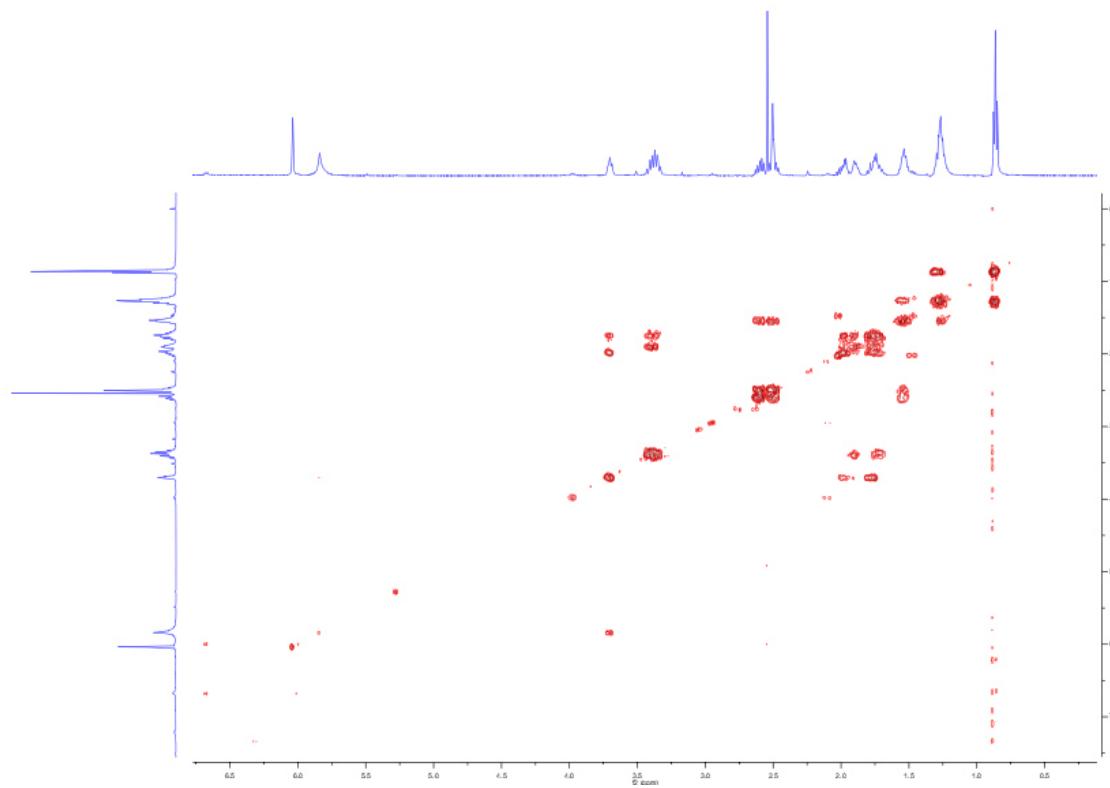
Supplementary Figure 1.8. COSY spectra of 2a. Figure prepared and provided by Yi-Ming Shi (Goethe-university Frankfurt).
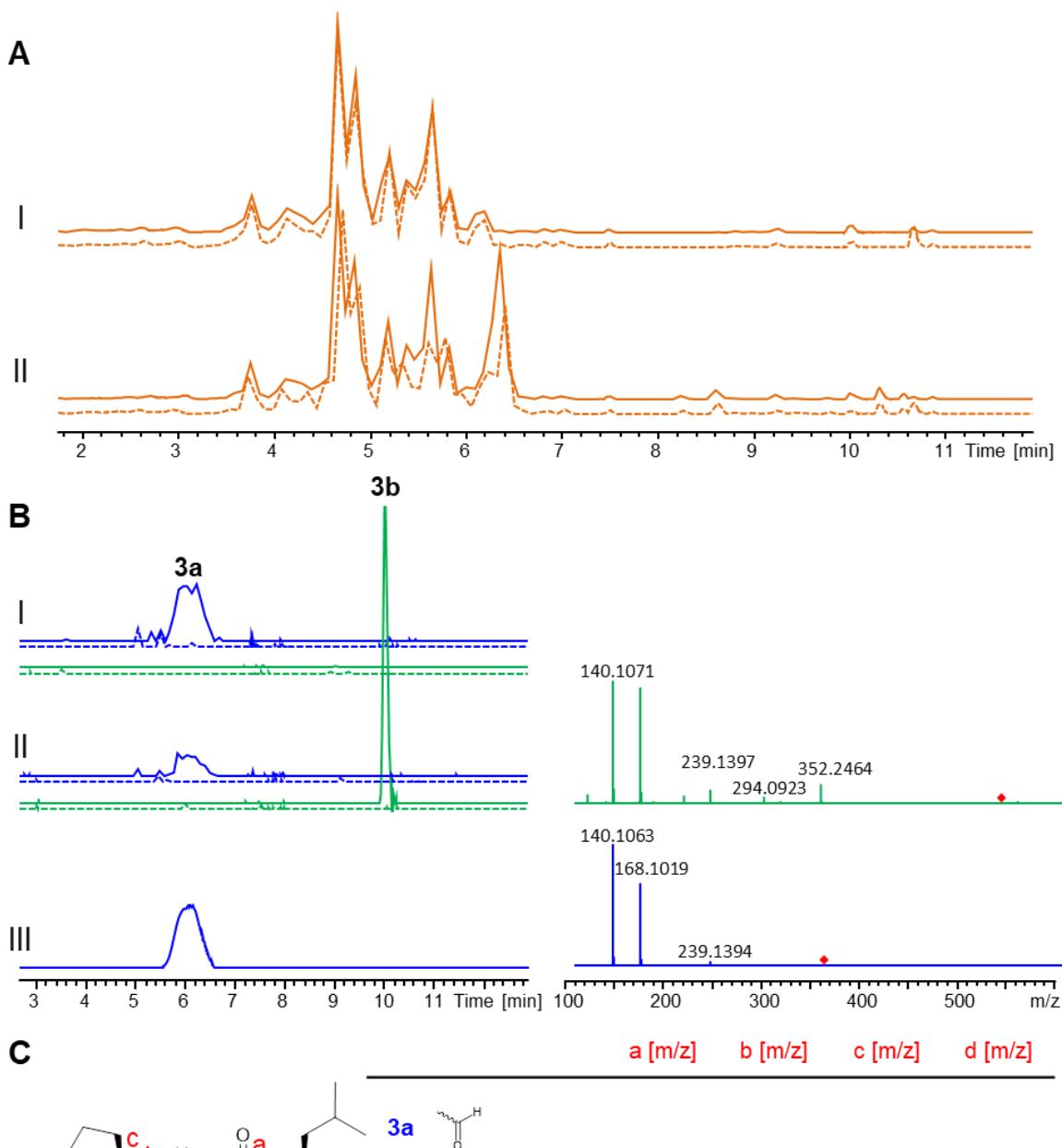
Supplementary Figure 1.9. High-resolution HPLC-MS data of 3a and 3b. A. BPC (orange) after heterologous expression (induced: continuous line; non-induced: dashed line) of NRPS-2 in *E. coli* DH10B::*mtaA* (I) without and (II) with PFBHA. **B.** EIC of **3a** (blue; m/z $[M+H]^+$ = 354.239; calculated ion formula $C_{18}H_{32}N_3O_4$; Δppm 0.5) and **3b** (green; m/z $[M+H]^+$ = 549.248; calculated ion formula $C_{25}H_{34}F_5N_4O_4$; Δppm 1.0) and MS^2 spectra of (I), (II) and (III) chemical standard (calculated ion formula $C_{18}H_{32}N_3O_4$; Δppm 0.4). The parental ions are depicted by red diamonds. **C.** Structure and fragmentation of **3a** and **3b**.

Supplementary Figure 1.10. Stable isotope labelling of 3b. A. High-resolution MS spectra at 10.0 min of HPLC-MS analysis after heterologous production of NRPS-2 in *E. coli* DH10B::*mtaA* in LB (blue), [15]N (green) and [13]C (purple) media with or without addition of ^{12}C AAs. The shifts due to stable isotope incorporation are indicated by arrows. **B.** Structure of **3b**. The PFBHA-derived moiety (red) contains ^{12}C and ^{14}N isotopes and is therefore not considered in the analysis of stable isotope labelling experiments.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1		31.4	18.4	15.2	16.6	16.3	15.8	17.8	18.0	16.5	16.6	21.7	16.1	15.2	16.6	14.2	12.3	11.7
2	31.4		19.7	16.7	19.0	17.6	19.6	18.2	18.3	17.6	18.7	22.8	19.3	18.0	18.7	13.7	12.2	12.6
3	18.4	19.7		19.0	22.7	22.9	22.4	24.0	23.8	23.3	24.2	30.7	23.9	24.7	24.7	13.0	13.7	13.6
4	15.2	16.7	19.0		27.3	27.2	25.2	16.6	17.3	14.7	17.0	20.4	15.4	16.3	16.7	13.5	13.6	12.3
5	16.6	19.0	22.7	27.3		65.1	53.4	20.5	19.1	17.9	20.0	23.5	21.6	23.0	19.2	17.4	17.9	16.2
6	16.3	17.6	22.9	27.2	65.1		54.7	19.4	17.7	16.5	19.5	21.6	20.0	21.4	19.7	15.4	16.3	15.9
7	15.8	19.6	22.4	25.2	53.4	54.7		18.3	18.7	16.8	19.3	21.1	18.8	19.0	18.1	14.9	14.8	14.6
8	17.8	18.2	24.0	16.6	20.5	19.4	18.3		43.0	40.4	81.6	26.0	20.8	22.2	21.1	14.0	14.5	13.0
9	18.0	18.3	23.8	17.3	19.1	17.7	18.7	43.0		49.8	45.2	26.3	17.8	18.4	19.5	15.2	13.8	13.3
10	16.5	17.6	23.3	14.7	17.9	16.5	16.8	40.4	49.8		42.5	24.3	18.1	17.6	18.2	14.0	14.3	14.5
11	16.6	18.7	24.2	17.0	20.0	19.5	19.3	81.6	45.2	42.5		25.1	20.6	21.0	20.1	14.9	12.4	13.0
12	21.7	22.8	30.7	20.4	23.5	21.6	21.1	26.0	26.3	24.3	25.1		28.0	27.3	25.1	17.2	15.7	17.1
13	16.1	19.3	23.9	15.4	21.6	20.0	18.8	20.8	17.8	18.1	20.6	28.0		71.7	36.9	13.3	12.5	14.5
14	15.2	18.0	24.7	16.3	23.0	21.4	19.0	22.2	18.4	17.6	21.0	27.3	71.7		37.2	13.2	13.5	13.5
15	16.6	18.7	24.7	16.7	19.2	19.7	18.1	21.1	19.5	18.2	20.1	25.1	36.9	37.2		11.6	13.2	11.5
16	14.2	13.7	13.0	13.5	17.4	15.4	14.9	14.0	15.2	14.0	14.9	17.2	13.3	13.2	11.6		30.9	26.5
17	12.3	12.2	13.7	13.6	17.9	16.3	14.8	14.5	13.8	14.3	12.4	15.7	12.5	13.5	13.2	30.9		26.1
18	11.7	12.6	13.6	12.3	16.2	15.9	14.6	13.0	13.3	14.5	13.0	17.1	14.5	13.5	11.5	26.5	26.1	

Supplementary Figure 1.11. Similarities of C domains. Heatmap of the Clustal Omega alignment (Supplementary Fig. 1.12) of PxaA_C2 (1) and other C domains which are further specified in Supplementary Tab. 1.5. Shown is the pairwise identity (white, low identity; dark grey, high identity).

Supplementary Table 1.5. Overview of C domains for PxaA sequence alignment. Number and subclass of C domain, respective NRPS, organism and the sequence accession number. Further abbreviations are: XtpS, Xenotetrapeptide-producing synthetase; PAX, PAX peptide-producing synthetase; XfpS, Xefoampeptide-producing synthetase; TxlS, Taxllaid-producing synthetase; XnmS, Xenematid-producing synthetase; BacA, bacitracin-producing synthetase and HMWP, yersiniabactin-producing synthetase.

#	C domain	NRPS	organism	accession number
1	C2	PxaA	*X. stockiae* DSM 17904	WP_099124966
2	C2	BraB	*Pseudomonas* strain SH-C52	WP_084213812
3	C5	NocB	*N. uniformis subsp. tsuyamanensis*	AAT09805
4	$^{L}C_{L}2$	LgrA	*B. brevis* ATCC 8185	PDB-ID 6MFY
5	$^{L}C_{L}2$	PAX	*X. stockiae* DSM 17904	WP_099124752
6	$^{L}C_{L}4$	PAX	*X. stockiae* DSM 17904	WP_099124752
7	$^{L}C_{L}2$	XabS	*X. stockiae* DSM 17904	WP_099124276
8	C/E2	GxpS	*X. stockiae* DSM 17904	WP_099123840
9	C/E2	XtpS	*X. nematophila* ATCC 19061	WP_013184203
10	C/E5	PAX	*X. stockiae* DSM 17904	WP_099124752
11	C/E5	XabS	*X. stockiae* DSM 17904	WP_099124276
12	$^{D}C_{L}5$	LgrB	*B. brevis* ATCC8185	Q70LM6
13	$^{D}C_{L}6$	TxlS	*X. bovienii* SS-2004	WP_080515938
14	$^{D}C_{L}2$	XnmS	*X. mauleonii* DSM 17908	WP_092509235
15	$^{D}C_{L}2$	XfpS	*X. bovienii* SS-2004	WP_041573262
16	Cy1	BacA	*B. licheniformis* ATCC 10716	WP_020452079
17	Cy1	HMWP2	*X. szentirmaii* DSM 16338	WP_038233793
18	Cy3	HMWP1	*X. szentirmaii* DSM 16338	WP_038233794
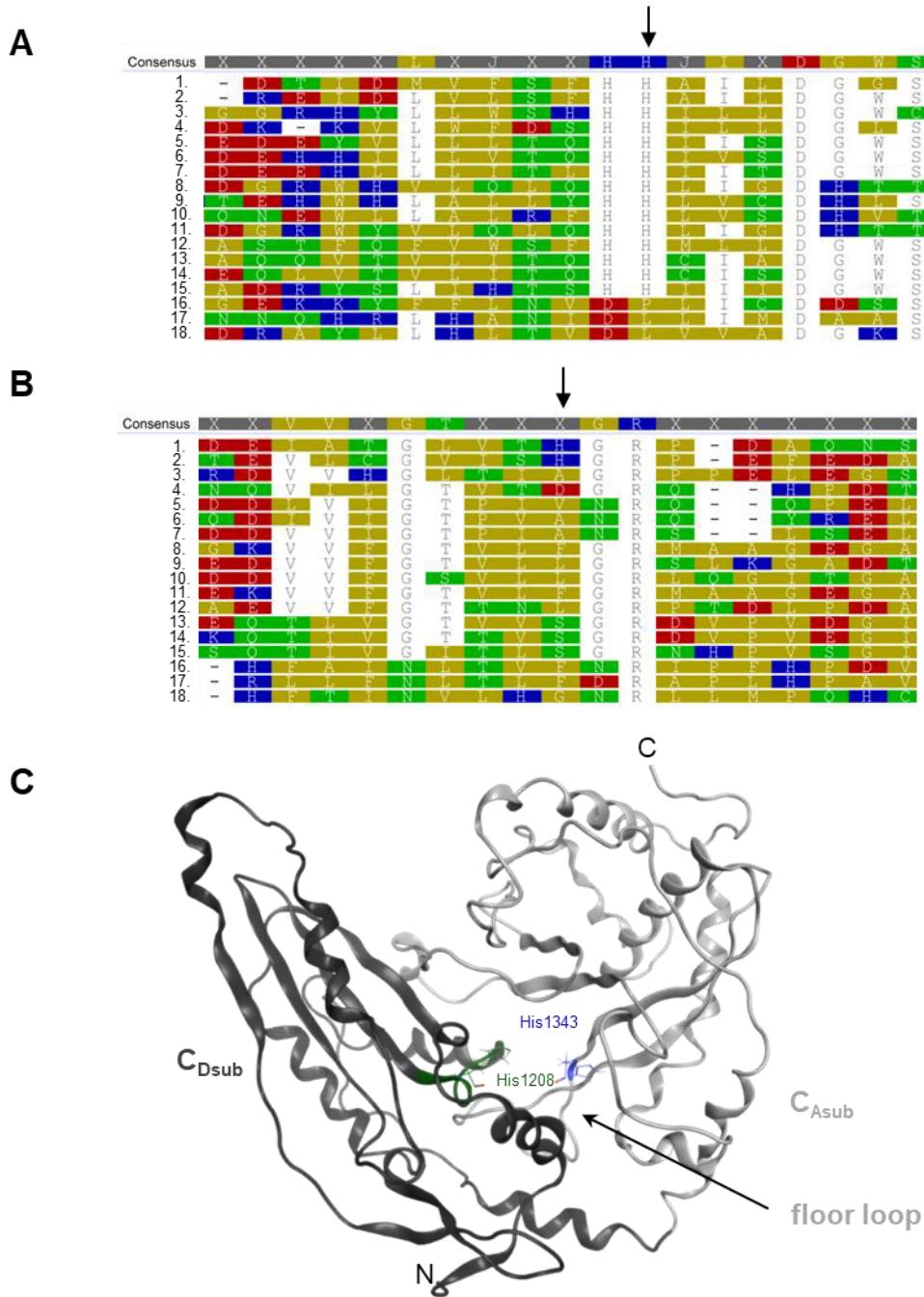
Supplementary Figure 1.12. Sequence alignment and homology model of PxaA_C2. Extracted Clustal Omega alignment of different C domains which are further specified in Supplementary Tab. 1.5. Highlighted are disagreements to the consensus sequence (upper line) in AA polarity colour (red, DE; green, CNQSTY; yellow, AFGILMPVW; blue, HKR). **A.** Region of the catalytic His-motif of C domains. The position of His1208 of PxaA_C2 (1) is highlighted by an arrow. **B.** Region within the *C*-terminal floor loop of C domains. The position of His1343 of PxaA_C2 (1) is highlighted by an arrow. **C.** Ribbon diagram of homology model (RMSD = 13.8 Å) of PxaA_C2 from *X. stockiae* based on LgrA_C2 from *B. brevis* (PDB-ID 6MFY)[85] with *N*-terminal C_Dsub (black) and *C*-terminal C_Asub (grey) subdomains. The catalytic triad with His1208 is highlighted in green and His1343 in blue.

6.4.2 Investigation of ATReds in *Xenorhabdus*

6.4.2.1   Material and methods

Strains

Xenorhabdus strains were grown in liquid or solid LB medium at 30 °C. All strains that were used in this work are summarized in Supplementary Tab. 2.1.

Supplementary Table 2.1. Strains used and generated in this work.

Strain	Genotype	Reference
X. bovienii SS-2004	WT	DSMZ
X. budapestensis DSM 16342		DSMZ
X. cabanillasii DSM 17905		DSMZ
X. cabanillasii JM26		[217]
X. hominickii DSM 17903		DSMZ
X. indica DSM 17382		DSMZ
X. innexi DSM 16336		DSMZ
X. khoisanae DSM 25463		DSMZ
X. kozodoi DSM 17907		DSMZ
X. mauleonii DSM 17908		DSMZ
X. miraniensis DSM 17902		DSMZ
X. nematophila ATCC 19061		ATCC
X. poinarii DSM 4768		DSMZ
X. stockiae DSM 17904		DSMZ
X. szentirmaii DSM 16338		DSMZ
X. szentirmaii US		[217]
X. vietnamensis DSM 22392		DSMZ
Xenorhabdus sp. KJ12.1		[218]
Xenorhabdus sp. KK7.4		[212]
Xenorhabdus sp. PB62.4		[219]

Extract preparation and LC/MS analysis

After 48 h of cultivation in 10 mL LB medium, 0.2 mL of the culture was diluted 1:5 in MeOH and centrifuged for 20 min at 13.300 rpm. For further information, please refer to chapter 6.4.1.1.

Stable isotope labelling

Please refer to chapter 6.4.1.1.

Sequence alignment

The software PRISM 8 (GraphPad Software) was used to prepare the heatmap. For further information, please refer to chapter 6.4.1.1.

Homology modelling

Please refer to chapter 6.4.1.1.

6.4.2.2 Supplementary data

Supplementary Table 2.2. Overview of ATReds from *Xenorhabdus*. Organism, ATRed encoding gene and the sequence accession number when available. The presence of an *mbtH*-like gene within the gene cluster is indicated by a checkmark and the classification.

#	organism		*gene*	accession number	*mbtH*	subtype
1	*M. tuberculosis* H37 Rv	Mtub	*nrp*	PDB-ID 4DQV		
2	*S. aurantiaca* SG15A	Saur	*mxaA*	PDB-ID 4U7W		
3	*X. khoisanae* DSM 25463	Xkho	*03561*	WP_047964749	✗	2
4			*03948*	KMJ43401	✓	1
5	*Xenorhabdus* sp. KK7.4	Xkk	*01108*	WP_099122076	✗	2
6			*02190*	PHM51946	✓	1
7	*X. bovienii* SS-2004	Xbov	*00464*	WP_012987143	✓	1
8	*X. cabanillasii* DSM 17905	XcabDSM	*01493*	WP_115826099	✓	1
9			*03579*	WP_115827345	✗	3
10	*X. poinarii* DSM 4768	Xpoi	*02758*	WP_084717361	✗	1
11	*X. nematophila* ATCC 19061	Xnem	*00646*	WP_010848042	✓	1
12			*01475*	WP_041977370	✗	2
13			*01561*	WP_013183919	✗	3
14	*X. miraniensis* DSM 17902	Xmir	*01976*	PHM48830	✓	1
15	*X. vietnamensis* DSM 22392	Xvie	*00828*	WP_086108296	✓	1
16			*03245*	WP_086110177	✗	3
17	*Xenorhabdus* sp. KJ12.1	Xkj	*01708*	PHM70463	✓	1
18			*02365*	WP_099110217	✗	2
19	*X. stockiae* DSM 17904	Xsto	*02049*	WP_099124989	✗	2
20			*03518*	PHM63926	✓	1
21	*X. cabanillasii* JM26	XcabJM	*01329*	WP_038269276	✗	3
22			*03628*	WP_038260646	✓	1
23	*X. indica* DSM 17382	Xind	*00627*	n/a	✓	1
24			*01729*	WP_047678938	✗	3
25	*Xenorhabdus* sp. PB62.4	Xpb	*01459*	n/a	✓	1
26	*X. innexi* DSM 16336	Xinn	*00707*	WP_086953644	✗	3
27			*02671*	WP_086954558	✓	1
28			*02976*	WP_086953155	✗	2
29	*X. szentirmaii* DSM 16338	XszeDSM	*01262*	WP_038240738	✓	1
30			*03484*	WP_051462298	✗	2
31	*X. mauleonii* DSM 17908	Xmau	*04014*	WP_092511953	✓	1
32			*04297*	WP_092514341	✗	2
33	*X. szentirmaii* US	XszeUS	*00630*	WP_038234872	✓	1
34			*03375*	WP_038240738	✗	2
35	*X. kozodoi* DSM 17907	Xkoz	*00716*	PHM74478	✓	1
36	*X. budapestensis* DSM 16342	Xbud	*02951*	WP_099136729	✗	3
37			*03352*	WP_099137116	✓	1
38	*X. hominickii* DSM 17903	Xhom	*01101*	WP_069317741	✓	1

Supplementary Figure 2.1. Sequence alignment of ATReds from *Xenorhabdus* strains. Extracted Clustal Omega alignment of Nrp_R,[118] MxaA_R[116] and 36 ATRed sequences from 20 *Xenorhabdus* strains which are further specified in Supplementary Tab. 2.2. Highlighted are disagreements to the consensus sequence (upper line) in AA polarity colour (red, DE; green, CNQSTY; yellow, AFGILMPVW; blue, HKR). **A.** The position of His736 in ATRed$_{xind01729}$ (24.) and Leu130 in MxaA_R (1.) is indicated by an arrow. **B.** The position of Gln813 in ATRed$_{xind01729}$ and Lys253 (24.) in MxaA_R (1.) is indicated by an arrow.
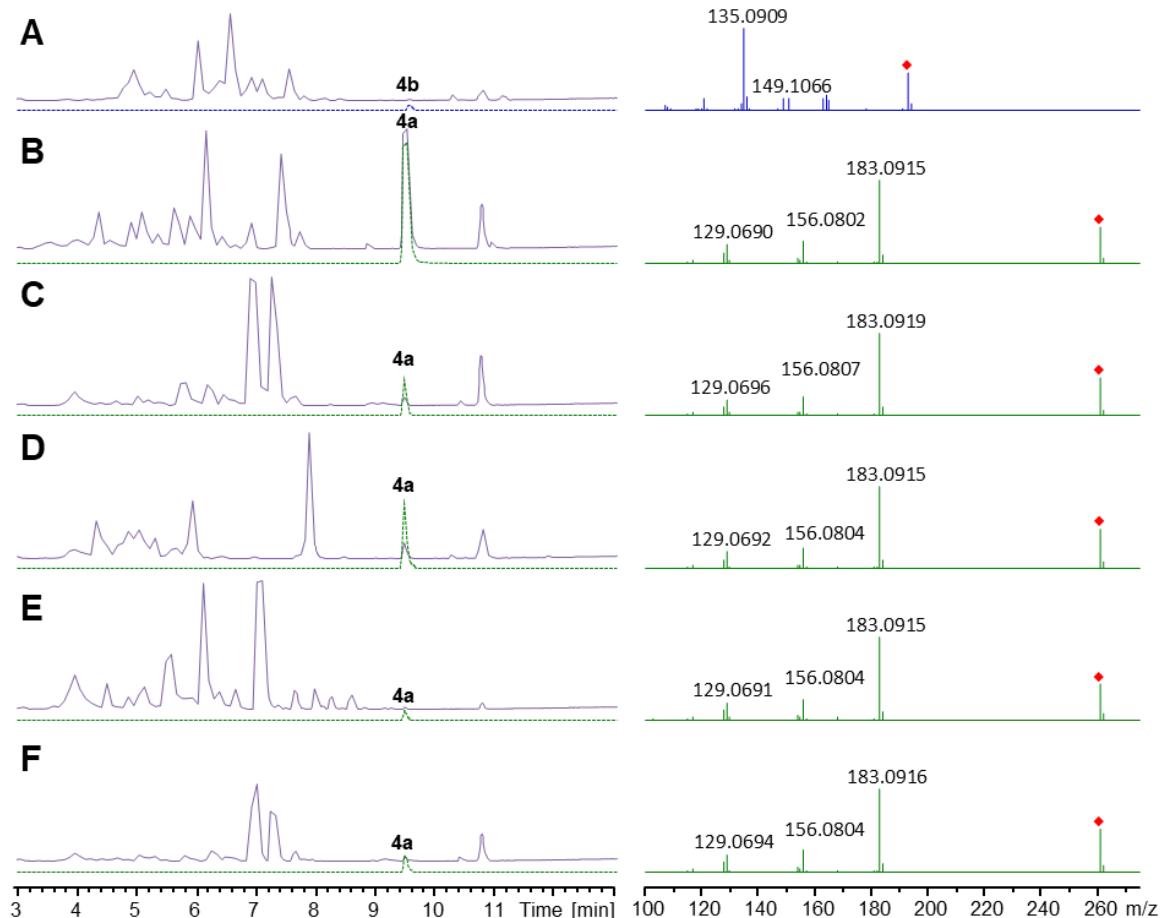
Supplementary Figure 2.2. High-resolution HPLC-MS data of 4a and 4b produced by *Xenorhabdus* WT strains. Left. BPCs (purple) in continuous lines and EICs of **4a** (green, *m/z* $[M+H]^+$ = 261.138) and **4b** (blue, *m/z* $[M+H]^+$ = 193.169) in dashed lines (4-fold increased intensity except for B) after cultivation of **A.** *X. innexi* DSM 16336 (calculated ion formula $C_{12}H_{21}N_2$; Δppm -0.3), **B.** *X. indica* DSM 17382 (calculated ion formula $C_{18}H_{17}N_2$; Δppm -0.1), **C.** *X. cabanillasii* JM26 (calculated ion formula $C_{18}H_{17}N_2$; Δppm -0.5), **D.** *X. vietnamensis* DSM 22392 (calculated ion formula $C_{18}H_{17}N_2$; Δppm 0.8), **E.** *X. nematophila* ATCC 19061 (calculated ion formula $C_{18}H_{17}N_2$; Δppm -0.2) and **F.** *X. cabanillasii* DSM 17905 (calculated ion formula $C_{18}H_{17}N_2$; Δppm -0.1). **Right.** Corresponding MS^2 spectra of **4a** and **4b**, respectively. The parental ions are depicted by red diamonds.

266
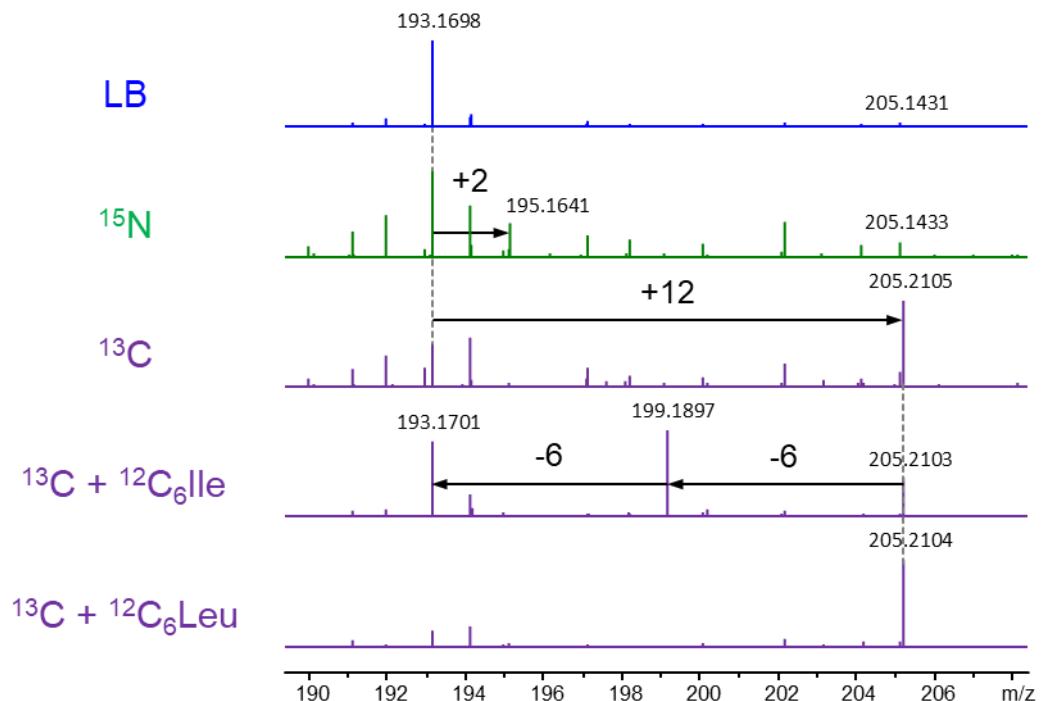
Supplementary Figure 2.3. Stable isotope labelling of 4b. High-resolution MS spectra at 9.5 min of HPLC-MS analysis of *X. innexi* in LB (blue), ^{15}N (green) and ^{13}C (purple) media with or without addition of ^{12}C AAs. The shifts due to stable isotope incorporation are indicated by arrows
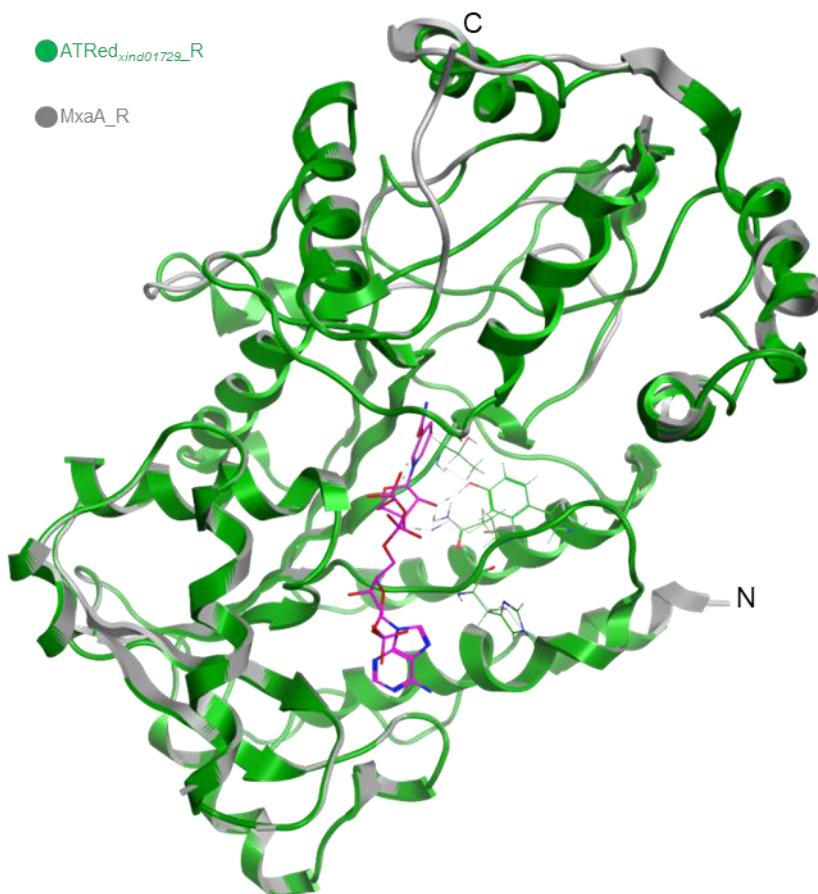
Supplementary Figure 2.4. Homology model of ATRed$_{xind01729}$_R based on MxaA_R. ATRed$_{xind01729}$_R from *X. indica* (green) and MxaA_R from *S. aurantiaca* (grey; PDB-ID 4U7W)[116] share a pairwise identity of 28.2 % and the homology model has an RMSD of 0.8 Å. Important residues and the NADPH cofactor are highlighted as in Fig. 10B.
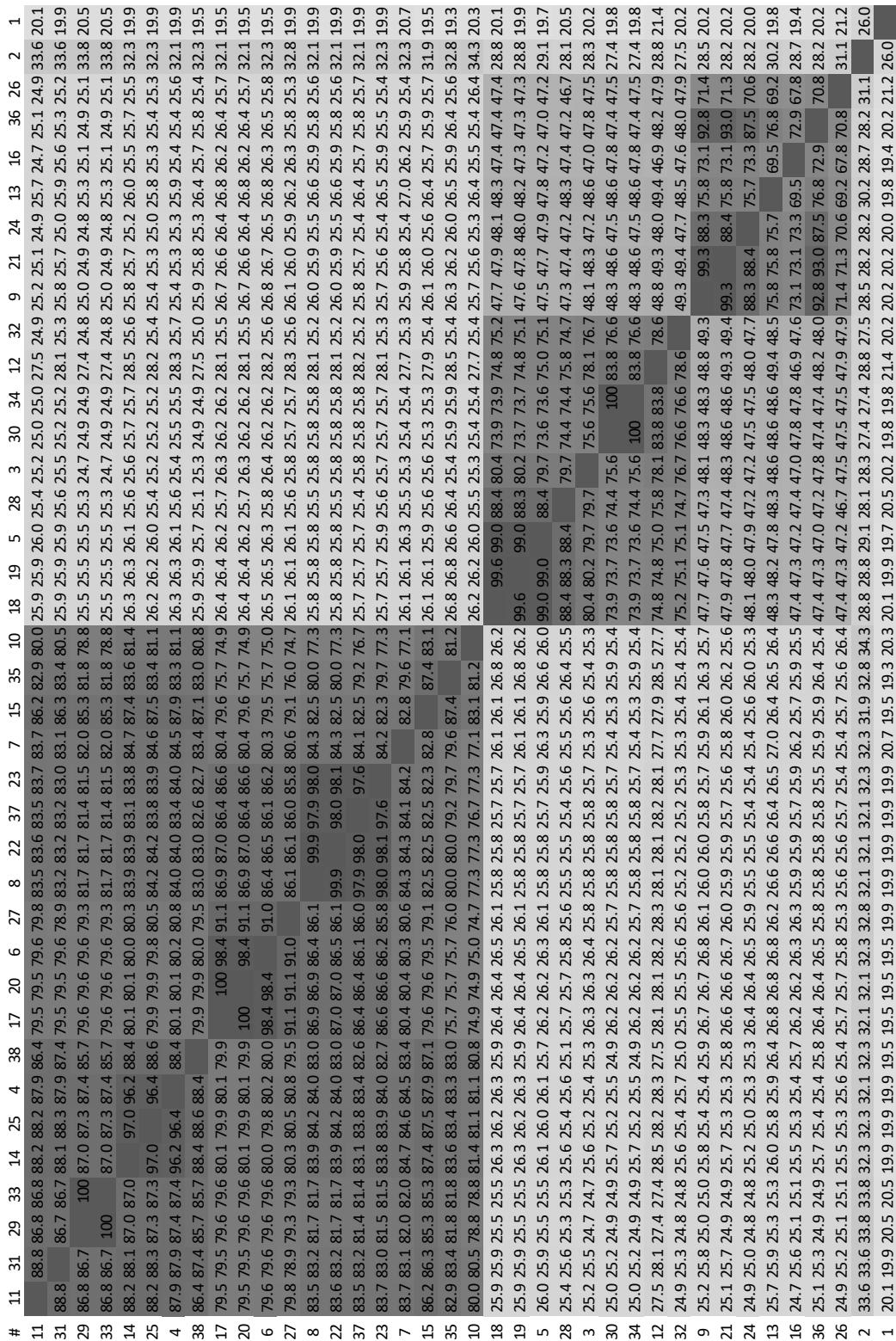
Supplementary Figure 2.5. Similarities of ATRed sequences from *Xenorhabdus*. Heatmap of the Clustal Omega alignment (Supplementary Fig. 2.1) of R_{NRP},[118] MxaA_R[116] and 36 ATReds from *Xenorhabdus* which are further specified in Supplementary Tab. 2.2. Shown is the pairwise identity (white, low identity; dark grey, high identity) in [%].

7   Curriculum vitae

Personal information

Name:	Andreas Tietze
Date and place of birth:	█████████████████████
Nationality:	██████
Contact:	████████████████
	████████████████
	████████████

Education

Goethe University Frankfurt

11.2016 – today   PhD in Biology

10.2014 – 08.2016   M.Sc. in Molecular Biotechnology (overall grade: 1.1 - excellent)

Darmstadt University of Applied Sciences

09.2011 – 09.2014   B.Sc. in Biotechnology (overall grade: 1.5)

Gymnasium Lindenberg i. Allgäu

09.2002 – 05.2011   Allgemeine Hochschulreife (overall grade: 2.1)

Work Experience

Goethe University Frankfurt

11.2016 – 07.2020   Scientific assistant

Research group *Molecular Biotechnology* of Prof. Dr. Helge B. Bode, Institute for Molecular Bio Science

Paul Ehrlich Institute Langen

03.2014 – 09.2014   Internship

Research group *Transposition and Genome Engineering* of Prof. Dr. Zoltán Ivics, Head of the Division of Medical Biotechnology

8   List of publications and record of conferences

Publications

Non-ribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains

Tietze A., Shi Y.-N., Kronenwerth M. and Bode H. B.

ChemBioChem 10.1002/cbic.202000176[165]

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

Bozhüyük K. A. J.*, Linck A.*, Tietze A.*, Kranz J.*, Wesche F., Nowak S., Fleischhacker F., Shi Y.-N., Grün P. and Bode H. B.

Nat. Chem. **11**, 653 – 661 (2019)[164]

*These authors contributed equally.

De novo design and engineering of non-ribosomal peptide synthetases.

Bozhüyük K. A., Fleischhacker F., Linck A., Wesche F., Tietze A., Niesert C. P. and Bode H. B.

Nat. Chem. **10**, 275 – 281 (2018)[163]

List of publications and record of conferences

Protein Engineering Congress EU

> Posterpresentation: Modified peptides from engineering of non-ribosomal peptide synthetases. Frankfurt am Main (Germany), 12.11 – 13.11.2019

International VAAM Workshop on Biology of Microorganisms Producing Natural Products

> Posterpresentation: Modified peptides from engineering of non-ribosomal peptide synthetases. Jena (Germany), 15.09 – 17.09.2019

3rd European Conference on Natural Products

> Posterpresentation: 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

> Posterpresentation: Dividing condensation domains by half – A new strategy for the engineered biosynthesis of non-ribosomal peptides. Frankfurt am Main (Germany), 31.08 – 02.09.2018

> Winner of the poster award

International VAAM Workshop on Biology of Bacteria Producing Natural Products

> Posterpresentation: Production of Novel Natural Product Like Peptides by de novo Design of Non-Ribosomal Peptide Synthetases. Tübingen (Germany), 27.09 – 29.09.2017

9 Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im Mathematisch-Naturwissenschaftlichen Bereich unterzogen habe.

_____   _____
Ort/Datum   Andreas Tietze

10 Versicherung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation mit dem Titel

Engineering and characterisation of non-ribosomal peptide synthetases

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   Andreas Tietze