Non-ribosomal peptide synthetase docking domains: structure, function and engineering strategies

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften (15)
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Jonas Watzel
aus Mainz

Frankfurt am Main 2021
(D 30)

vom Fachbereich Biowissenschaften (15) der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel
Gutachter: Prof. Dr. Helge B. Bode
Zweitgutachter: Prof. Dr. Martin Grininger

Datum der Disputation:

ACKNOWLEDGMENTS

Nach knapp vier Jahren mit unzähligen Stunden im Labor und Büro ist es vollbracht, die Doktorarbeit ist gedruckt. Und so heißt es zu Beginn dieser Arbeit erst einmal all denjenigen Dank zu sagen, ohne welche ein Abschluss dieses Projekts gar nicht erst möglich gewesen wäre.

Zuallererst gilt mein besonderer Dank meinem Doktorvater **Prof. Dr. Helge B. Bode**, der mich in den letzten Jahren als Wissenschaftler ausgebildet und dabei stets unterstützt hat. Ich bedanke mich an dieser Stelle nicht nur dafür, dass er es mir ermöglicht hat an solch einem großartigen Projekt zu forschen, sondern auch für die Rahmenbedingungen, Ressourcen, Ideen und stetige Motivation, die einem die Zeit des Promovierens sehr erleichtert haben. Ein ganz herzlicher Dank gilt auch meinem inoffiziellen Doktorvater **Prof. Dr. Jens Wöhnert**, der mir mit seinen schier grenzenlosen Ideen und seinem Wissen in Bezug auf die Strukturbiologie und die Anwendung von NMR-Methoden immer mit Rat und Tat zur Seite stand.

Für die Übernahme des Zweitgutachtens dieser Arbeit möchte ich **Prof. Dr. Martin Grininger** ganz herzlich danken.

Natürlich gilt mein Dank auch allen **ehemaligen** und **aktuellen Mitgliedern** des **AK Bode** und **AK Wöhnert**, sowie allen **Co-Autoren der Publikationen/ Manuskripte**. Einerseits habt ihr mir mit wissenschaftlichen Ratschlägen stets zur Seite gestanden und andererseits erst das äußerst angenehme Arbeitsumfeld geschaffen. Hervorzuheben sei hier die Protein-Crew des AK Wöhnert alias *insane in the brain* durch deren Expertise ich immer wieder etwas Neues über die Proteinaufreinigung, Proteinkristallisation und NMR-Spektroskopie habe lernen können und ohne die meine Mittagspausen/ Kaffeepausen nur halb so lustig gewesen wären. Auch bedanke ich mich herzlichst bei meinen Bürokollegen für die fachliche Unterstützung und die rege Erörterung sportlicher Höchstleistungen. Besonders hervorheben möchte ich im Folgenden noch einmal einzelne Personen, die für das Gelingen meiner Arbeit unersetzlich waren:

Danke an **Carolin Hacker**, die mich schon während meiner Masterarbeit an die Thematik der Proteinstrukturaufklärung mittels NMR-Spektroskopie herangeführt und

mich in das Aufnehmen von NMR-Spektren, dem Assignment-Prozess und der Strukturrechnung eingewiesen hat.

Elke Duchardt-Ferner gebührt besonderer Dank für ihre NMR-Expertise, mit deren Hilfe beim Ansetzen von NMR-Experimenten und beim Modifizieren von Pulssequenzen erst das Maximum aus meinen NMR-Proben herausgeholt wurde.

Zusätzlicher Dank gebührt meinen Jahrgangskollegen **Sebastian Wenski**, **Nick Neubacher** und **Andreas Tietze (Team JA!)** für die gegenseitige Motivation.

Ein herzliches Dankeschön geht auch an **Peter Grün**, der sich, egal ob präparatives oder analytisches LC-MS-System, immer die Zeit genommen hat sein Wissen und seine Erfahrung mit mir zu teilen und auch an noch so stressigen Tagen immer die Zeit gefunden hat Geräte gemeinsam zu reparieren.

Für das Korrekturlesen dieser Arbeit sage ich herzlichst Danke an **Elke Duchardt-Ferner**, **Carsten Kegler**, **Kenan Bozhüyük** und **Sonja Watzel**. Zusätzlich noch vielen Dank an **Carsten** für dein umfangreiches Wissen auf dem Gebiet von NRPS und dein fachliches Knowhow insbesondere auch bei Klonierungsfragen, die mir so manchen Stolperstein aus dem Weg geräumt haben.

Zum Schluss richte ich ein herzliches Dankeschön an all meine Freunde und meine Familie, die mich in der Zeit außerhalb der Universität immer unterstützt und so zum erfolgreichen Abschluss des Projekts PhD beigetragen haben. Die letzten Sätze meiner Danksagung sind meiner Frau Sonja gewidmet. Vielen Dank für deinen immerwährenden Rückhalt, auch wenn aus den ursprünglich angepeilten 3 Jahren doch ein wenig mehr wurden und dein stets offenes Ohr für meine Erzählungen, was ich denn eigentlich im Labor genau mache, welche für einen Fachfremden nicht immer ganz leicht zu verstehen waren. Danke, dass du mich auf diesem nicht immer einfachen Weg begleitet hast.

TABLE OF CONTENTS

TABLE OF ABBREVIATIONS

4’-PPant	4’-phosphopantetheine
AA	amino acid
A	adenylation
A_{core}	N-terminal adenylation core domain
AMP	adenosine monophosphate
ANN	artificial neural network
A_{sub}	C-terminal adenlyation subdomain
AT	acyltransferase
ATP	adenosine triphosphate
ACP	acyl carrier protein
C	condensation
C_{Asub}	N-terminal acceptor condensation subdomain
CDA	calcium-dependent antibiotic
C_{Dsub}	C-terminal donor condensation subdomain
C/E	dual condensation/epimerization
CSI	chemical shift index
CoA	coenzyme A
COM	communication-mediating
cryo-EM	cryogenic electron microscopy
Cur	curacin-producing synthase
Cy	cyclization
DD	docking domain
DEBS	6-deoxyerythronolide B-producing synthase
DNA	desoxyribonucleic acid
E	epimerization
Ent	enterobactin-producing synthetase
Epo	epothilone-producing hybrid PKS-NRPS
ESI	electrospray ionization
F	formyltransferase
GxpS	GameXPeptide-producing synthetase
HPLC	high performance liquid chromatography
HR	high resolution
HSQC	heteronuclear single quantum correlation spectroscopy
ITC	isothermal titration calorimetry
kDa	kilodalton

KR	ketoreductase
LB	lysogeny broth
Lgr	linear gramicidin-producing synthetase
MT	methyltransferase
Mln	macrolactin-producing synthase
MLP	MbtH-like protein
MS	mass spectrometry
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
PaxS	PAX peptide-producing synthetase
Pik	pikromycin-producing polyketide synthase
PCR	polymerase chain reaction
PCP	peptidyl carrier protein
PDB	Protein Data Bank
PPTase	phosphopantheteine transferase
PKS	polyketide synthase
Pvd	pyoverdine-producing synthetase
RMSD	root-mean-square deviation
RXP	rhabdopeptide/xenortide-like peptide
SAM	*S*-adenosyl methionine
Ser-AVS	serine adenosine vinylsulfonamide
Srf	surfactin-producing synthetase
SLiM	short linear motif
SZ	synthetic zipper
T	thiolation
TE	thioesterase
TOCSY	total correlation spectroscopy
Tub	tubulysin-producing hybrid PKS-NRPS
Tyc	tyrocidine-producing synthetase
VAAM	Vereinigung für Allgemeine und Angewandte Mikrobiologie
VibH	free-standing condensation domain from vibriobactin-producing synthetase
Vir	virginiamycin-producing synthase
XU	exchange unit
XUC	exchange unit condensation domain

SUMMARY

Non-ribosomal peptide synthetases (NRPSs) are known for their capability to produce a wide range of natural compounds and some of them possess interesting bioactivities relevant for clinical application like antibiotics, anticancer, and immunosuppressive drugs. The diverse bioactivity of non-ribosomal peptides (NRPs) originates from their structural diversity, which results not only from the incorporation of non-proteinogenic amino acids into the growing peptide chain, but also the formation of heterocycles or further peptide modifications like methylation, hydroxylation and acetylation.

The biosynthesis of NRPs is achieved via the orchestrated interplay of distinct catalytic domains, which are grouped to modules that are located on one or more polypeptide chains. Each cycle starts with the selection and activation of a specific amino acid by the adenylation (A) domain, which catalyzes the aminoacyl adenylate formation under ATP consumption. This activated amino acid is then bound via a thioester bond to the 4'-phosphopantetheine cofactor (PPant-arm) of the following thiolation (T) domain. Before substrate loading, the PPant-arm is post-translationally added to the T domain by a phosphopantetheinyl transferase (PPTase), which converts the inactive *apo*-T domain in its active *holo*-form. In the last step of the catalytic cycle, two T domain bound peptide building blocks are connected by the condensation (C) domain, resulting in peptide bond formation and transfer of the nascent peptide chain to the following module. Each catalytic cycle is performed by a C-A-T elongation module until the termination module with a C-terminal thioesterase (TE) domain is reached. Here, the peptide product is released by hydrolysis or intramolecular cyclisation.

In comparison to single-protein NRPSs, where all modules are encoded on a single polypeptide chain, multi-protein NRPS systems must also maintain a specific module order during the peptide biosynthesis. Therefore, small C-terminal and N-terminal communication-mediating (COM) domains/docking domains (DD) were identified in the C- and N-terminal regions of multi-protein NRPSs. It was shown that these domains mediate specific and selective non-covalent protein-protein interaction, even though DD interactions are generally characterized by low affinities.

The first publication of this work focuses on the Peptide-Antimicrobial-*Xenorhabdus* peptide-producing NRPS called PaxS, which consists of the three proteins PaxA, PaxB and PaxC. Here, in particular the *trans* DD interface between the C-terminal attached DD of PaxB and N-terminal attached DD of PaxC was structurally investigated and thermodynamically characterized by isothermal titration calorimetry (ITC), yielding a dissociation constant (K_D) of ~25 µM, which is a DD typical affinity known from further characterized DD pairs. The artificial linking of the PaxB/C $^{C/N}$DD pair via a glycine-serine (GS) linker facilitated the structure determination of the DD complex by solution nuclear magnetic resonance (NMR) spectroscopy. In comparison to known docking domain structures, this DD complex assembles in a completely new fold which is characterized by a central α-helix of PaxC ^{N}DD wrapped in two V-shaped α-helices of PaxB ^{C}DD.

The first manuscript of this work focuses on the application of synthetic zippers (SZ) to mimic natural docking domains, enabling the easy assembly of NRPS building blocks encoded on different plasmids in a functional way. Here, the high-affinity interaction of SZs unambiguously defines the order of the synthetases derived from single-protein NRPSs in the engineered NRPS system and allows the recombination in a plug-and-play manner. Notably, the SZ engineering strategy even facilitates the functional assembly of NRPSs derived from Gram-positive and Gram-negative bacteria. Furthermore, the functional incorporation of SZs into NRPS modules is not limited to a specific linker region, so we could introduce them within all native NRPS linker regions (A-T, T-C, C-A).

The second publication and the second manuscript of this thesis again focus on the multi-protein PaxS, in particular on the *trans* interface between the proteins PaxA and PaxB on a molecular level by solution NMR. Therefore, the PaxA ^{C}DD adjacent T domain was included into the structural investigation besides the native interaction partner PaxB ^{N}DD. Before a three-dimensional structure could be obtained from NMR data, the NH groups located in the peptide bonds had to be assigned to the respective amino acids of the proteins (backbone assignment). Based on these backbone assignments, the secondary structure of PaxA T_1-^{C}DD and PaxB ^{N}DD in the absence and presence of the respective interaction partner were predicted.
The structural and functional characterization of the PaxA T_1-^{C}DD:PaxB ^{N}DD complex is summarized in manuscript two. The thermodynamic analysis of this complex by

ITC determined a K_D value of ~250 nM, whereas the discrete DDs did not interact at all. The high-affinity interaction allowed to determine the solution NMR structure of the PaxA T_1-^{C}DD:PaxB ^{N}DD complex without the covalent linkage of the interaction partners and an extended docking domain interface could be determined. This interface comprises on the one hand α-helix 4 of the PaxA T_1 domain together with the α-helical ^{C}DD, and on the other hand the PaxB ^{N}DD, which is composed of two α-helices separated by a sharp bend.

In an additional chapter the PaxB/C *trans* interface of the first publication 1 is analyzed with regard to the question if this DD interface is also extended by regions of the T domain. In contrast to the docking domains between PaxA and PaxB, the PaxB/C DD pair assembles in a stable DD complex even if the adjacent T domain is not present. Nonetheless, ITC measurements determined a 25-fold increased affinity (~1 µM) if the ^{C}DD flanking T domain was included into the analyses. The structural investigation of this PaxB T_4-^{C}DD:PaxC ^{N}DD complex by solution NMR was not successful because it was not possible to assign most of the backbone amide signals to the respective amino acids. However, based on the incomplete chemical shift assignment, the secondary structure could be predicted for the unbound and bound PaxB T_4-^{C}DD. Further NMR titration experiments imply large chemical changes for residues located in the T domain α-helix 4, which gives a first hint to that also the PaxB/C DDs possess an extended interface.

ZUSAMMENFASSUNG

Nicht-ribosomale Peptidsynthetasen (NRPSs) sind für die Biosynthese einer Vielzahl von Naturstoffen verantwortlich, welche teilweise Bioaktivitäten aufweisen, die insbesondere für die klinische Anwendung interessant sind wie bspw. Antibiotika, Medikamente gegen Krebs und Immunsuppressiva. Die strukturelle Vielfalt von nicht-ribosomalen Peptiden (NRPs) ist hierbei nicht nur auf die proteinogenen Aminosäuren beschränkt, sondern es können darüber hinaus auch nicht-proteinogene Aminosäuren eingebaut werden. Außerdem sorgen die Ausbildung von Heterocyclen und die Modifizierung des Peptides durch Methylierungen, Hydroxylierungen und Acetylierungen für eine zusätzliche strukturelle Diversität.

Die Biosynthese von nicht-ribosomalen Peptiden wird hierbei durch das geregelte Zusammenspiel voneinander unabhängiger katalytischer Domänen realisiert, die zu Modulen zusammengefasst und auf einem oder mehreren Proteinen lokalisiert sind. Jeder katalytische Zyklus startet mit der Auswahl und Aktivierung einer spezifischen Aminosäure unter Bildung eines Aminoacyladenylates und dem Verbrauch von ATP durch die Adenylierungsdomäne (A). Die aktivierte Aminosäure wird dann über einer Thioesterbindung mit dem 4′-Phosphopantethein (PPant)-Kofaktor der darauffolgenden Thiolierungsdomäne (T) verknüpft. Bevor die T-Domäne das Substrat jedoch binden kann muss diese zuerst mit dem PPant-Kofaktor posttranslational modifiziert werden, wodurch die Domäne durch die Phosphopantetheinyl-Transferase vom inaktiven *apo*- in den aktiven *holo*-Zustand überführt wird. Der letzte Schritt eines Biosynthesezykluses umfasst die Ausbildung der Peptidbindung durch die Kondensationsdomäne (C) zwischen zwei T-Domänen gebundenen Bausteinen, wodurch gleichzeitig die wachsende Peptidkette auf das darauffolgende Modul übertragen wird. Die katalytischen Zyklen werden in der Regel durch sogenannte C-A-T Elongationsmodule durchgeführt, bis das finale Produkt die C-terminale Thioesterasedomäne (TE) im Terminationsmodul erreicht, von welcher das Peptid durch Hydrolyse abgespalten oder durch einen intramolekularen Ringschluss abgelöst wird.

Im Vergleich zu NRPSs, bei welchen alle Module auf einem Protein lokalisiert sind, müssen auch NRPS aus mehreren einzelnen Proteinen eine definierte Modulabfolge während der Peptidbiosynthese gewährleisten um einen Naturstoff mit einer

spezifischen Aminosäureabfolge zu synthetisieren (Kolinearität). Dafür sind an den C- und N-Termini von multi-protein NRPS sogenannte *communication-mediating* (COM) *domains*/ *docking domains* (DD) ausgebildet, die durch nicht-kovalente Interaktionen für eine spezifische und selektive Proteinassemblierung unumgänglich sind, auch wenn diese häufig nur eine sehr geringe Affinität zueinander besitzen.
Trotz der Tatsache, dass die meisten NRPSs aus mehreren Proteinen aufgebaut sind, ist das strukturelle Wissen über DDs bisher sehr eingeschränkt, weshalb die Durchführung weiterer Studien, die sich mit der strukturellen Aufklärung dieser *trans*-Interaktionsflächen und deren zugrundeliegenden Mechanismen beschäftigen, unerlässlich ist. Des Weiteren war bisher die Anwendung von DDs als Engineering-Werkzeuge durch deren geringe Affinität limitiert, was in Theorie durch die Verwendung synthetischer und hoch-affiner DDs überwunden werden könnte.

Die erste Publikation mit dem Titel „A new docking domain type in the Peptide-Antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*“ befasst sich mit der PAX-Peptid produzierenden Synthetase (PaxS), die sich aus den drei Proteinen PaxA, PaxB und PaxC zusammensetzt. Besonderes Augenmerk wurde in dieser Untersuchung auf die Interaktionsfläche zwischen der C-terminalen DD (^{C}DD) von PaxB und der N-terminalen DD (^{N}DD) von PaxC gelegt, welche strukturell charakterisiert wurden. Durch die thermodynamische Analyse mittels isothermer Titrationskalorimetrie (ITC) konnte darüber hinaus eine Dissoziationskonstante (K_D) von ~25 µM bestimmt werden. Die artifizielle Verknüpfung des PaxB/C $^{N/C}$DD-Paares mittels eines Glycin-Serin-Linkers ermöglichte die Strukturaufklärung des Proteinkomplexes in Lösung durch Kernspinresonanzspektroskopie (NMR). Im Vergleich zu bekannten DD-Strukturen zeichnet sich die gelöste Komplexstruktur durch ein komplett neues Faltungsmuster aus, in dem eine zentrale PaxB ^{N}DD α-Helix von zwei in einem V-angeordneten PaxB ^{C}DD α-Helices umhüllt ist. Neben der Strukturanalyse beider DD im gebundenen Zustand wurden diese ebenfalls im ungebundenen Zustand mittels NMR untersucht, um die aus der Interaktion resultierenden strukturellen Veränderungen nachzuweisen. Die PaxB ^{C}DD wies in der Abwesenheit des Bindungspartners bspw. lediglich eine α-Helix am C-Terminus auf, wohingegen die übrigen Aminosäuren in einem intrinsisch ungeordneten Bereich lokalisiert sind. In diesem Bereich bildet sich während der Proteininteraktion eine zweite α-Helix aus. Die PaxC ^{N}DD setzt sich hingegen unabhängig vom Interaktionsstatus immer aus

einer α-Helix zusammen. Auf molekularer Ebene konnte nachgewiesen werden, dass der DD-Komplex durch die Kombination von hydrophoben Wechselwirkungen, Salzbrücken und Wasserstoffbrückenbindungen stabilisiert ist. Bei dieser NMR-Struktur handelt es sich erst um die zweite aufgeklärte DD-Struktur eines NRPS-Systems.

Das erste Manuskript dieser Arbeit beschäftigt sich mit der Anwendung von synthetischen Zippern, welche natürliche DD imitieren und so die funktionale Assemblierung von NRPS-Bausteinen, die auf verschiedenen Plasmiden kodiert sind, ermöglichen. Durch die hohe Affinität der synthetischen Zipper wird die Reihenfolge der kombinierten Synthetasen eindeutig definiert, wodurch der Austausch einzelner NRPSs per Plug-and-Play möglich ist. Dabei konnten sogar Synthetasen aus Gram-negativen Bakterien mit Synthetasen aus Gram-positiven Bakterien funktional kombiniert werden. Des Weiteren wurde nachgewiesen, dass die Position zum Einfügen der synthetischen Zipper nicht auf einen spezifischen Linker-Bereich beschränkt, sondern in allen Linker-Bereichen (A-T, T-C, C-A) möglich ist. Durch die Verwendung mehrerer heterospezifischer Zipper-Paare konnten sogar mehr als zwei NRPS-Bausteine funktional zusammengefügt werden, wodurch eine strukturell diverse Peptidbibliothek samt einiger neuer Naturstoffe generiert wurde. Die nahezu grenzenlosen Kombinationsmöglichkeiten von NRPS-Bausteinen die auf verschiedenen Plasmiden kodiert sind, birgt darüber hinaus zwei weitere große Vorteile: erstens ist die Klonierung der relativ kleinen kodierenden NRPS-Sequenzen viel schneller realisiert und zweitens werden etwaige Größenlimitierungen bei der Translation von NRPSs im heterologen *Escherichia coli*-Expressionssystem umgangen.

Die zweite Publikation und das zweite Manuskript dieser Arbeit beschäftigen sich erneut mit der zuvor eingeführten PaxS NRPS, dieses Mal jedoch aus dem Bakterium *Xenorhabdus cabanillasii*. Hierbei liegt der Fokus auf der Strukturaufklärung der *trans*-Interaktionsfläche zwischen den Proteinen PaxA und PaxB durch Flüssig-NMR, wobei die PaxA ^{C}DD benachbarte T-Domäne neben der PaxB ^{N}DD in die Untersuchung miteingeschlossen wurde.
Bevor jedoch die dreidimensionale Struktur aus NMR-Daten berechnet werden konnte, mussten die chemischen Verschiebungen (*chemical shifts*) der Peptidbindungs-NH-Gruppen zunächst den jeweiligen Aminosäuren zugeordnet

werden. Dieser sogenannte *Backbone-Assignment*-Prozess ist in Publikation zwei für die ungebundene und gebundene PaxA T_1-^{C}DD (12 Kilodalton (kDa)) und PaxB ^{N}DD (3,5 kDa) beschrieben. 98,0 % aller Amid-Peptidrückgrat-NMR-Signale konnten für die ungebundene und gebundene PaxA T_1-^{C}DD zugeordnet werden und 96,4 % dieser für die beiden Bindungszustände von der PaxB ^{N}DD. Basierend auf den *Backbone-Assignments* konnte die Sekundärstruktur beider Proteine (PaxA T_1-^{C}DD/ PaxB ^{N}DD) in der An- und Abwesenheit des jeweiligen Bindungspartners bestimmt werden. Die PaxA T_1-^{C}DD ist aus vier α-Helices im ungebundenen Zustand aufgebaut, welche typisch für T-Domänenstrukturen sind und die ^{C}DD-Region ist unstrukturiert. Die ungebundene PaxB ^{N}DD weist lediglich eine α-Helix auf. Im gebundenen Zustand kommt es auf beiden Seiten des Proteinkomplexes zu strukturellen Veränderungen: erstens bildet sich eine weitere α-Helix aus Aminosäuren der ^{C}DD aus und zweitens ordnen sich Aminosäuren des N-Terminus der PaxB ^{N}DD ebenfalls in einer weiteren α-Helix an. Neben dem *Backbone-Assignment* konnten auch nahezu alle Seitenkettenatome der Aminosäuren den jeweiligen chemischen Shifts zugeordnet werden.

Die strukturelle und funktionelle Charakterisierung des PaxA T_1-^{C}DD:PaxB ^{N}DD-Komplexes ist in Manuskript zwei zusammengefasst. Die thermodynamische Analyse dieser Interaktion mittels ITC ergab einen K_D-Wert von ~250 nM, wohingegen keine Bindung zwischen der PaxA ^{C}DD und der PaxB ^{N}DD nachgewiesen werden konnte. Dies widersprach allen bisherigen Erkenntnissen über DDs, welche besagen, dass DDs unabhängig von ihren benachbarten Domänen miteinander interagieren und dabei maximal eine Affinität im niedrigen mikromolaren Bereich kennzeichnet. Nichtsdestotrotz konnte in weiteren NMR-Titrationsexperimenten erstmals bestätigt werden, dass α-Helix 4 der T-Domäne an der Interaktion zwischen der PaxA T_1-^{C}DD und der PaxB ^{N}DD beteiligt ist. Die hohe Affinität erlaubte es die PaxA T_1-^{C}DD:PaxB ^{N}DD-Komplexstruktur mittels Flüssig-NMR zu lösen ohne die Bindungspartner mittels eines künstlichen Linkers, wie bspw. in Publikation 1 beschrieben, kovalent zu verknüpfen. Dabei wurde zum ersten Mal eine erweiterte DD-Interaktionsfläche nachgewiesen. Diese setzt sich einerseits aus α-Helix 4 der T-Domäne sowie der ^{C}DD α-Helix und andererseits aus den zwei PaxB ^{N}DD α-Helices, die durch einen scharfen Knick voneinander getrennt sind, zusammen. Im Vergleich zu bekannten T-Domänenstrukturen hebt sich die PaxA T_1-Domäne durch eine signifikant längere α-Helix 4 hervor, wodurch erst diese Interaktionsplattform für die PaxB ^{N}DD geschaffen

wird. Eine Vielzahl von Aminosäureaustauschen in der PaxB ^{N}DD und deren thermodynamische Analyse mittels ITC bestätigte die Struktur noch einmal auf molekularer Ebene. Dabei konnte nachgewiesen werden, dass das erweiterte DD-Interface durch das Zusammenspiel von Salz-Brücken, hydrophoben Wechselwirkungen und Wasserstoffbrückenbindungen stabilisiert ist.

In einem weiteren Kapitel wird beleuchtet, inwieweit die PaxB/C *trans*-Interaktionsfläche der PaxS von *Xenorhabdus bovienii* SS-2004, deren DD-Komplexstruktur in der ersten Publikation gelöst wurde, ebenfalls durch die ^{C}DD flankierende T-Domäne erweitert ist. Im Gegensatz zu den DDs von PaxA und PaxB ordnet sich das PaxB/C $^{N/C}$DD-Paar in einem stabilen DD-Komplex an, auch wenn die T-Domäne nicht anwesend ist. Interessanterweise führte das Hinzufügen der C-terminalen T-Domäne von PaxB (PaxB T$_4$-^{C}DD-Konstrukt) in ITC-Experimenten zu einer 25-fach höheren Affinität (~1 µM) im Vergleich zur Affinität der reinen DDs. Leider war jedoch die Strukturaufklärung des PaxB T$_4$-^{C}DD:PaxC ^{N}DD-Komplexes mittels Flüssig-NMR nicht erfolgreich, da keine ausreichenden Anzahl an NH-chemischen Shifts der Peptidbindungen den jeweiligen Aminosäuren im *Backbone-Assignment*-Prozess zugeordnet werden konnte. Dennoch konnte basierend auf den unvollständigen *Backbone-Assignments* der ungebundenen und gebundenen PaxB T$_4$-^{C}DD deren Sekundärstrukturen bestimmt werden. Die ungebundene PaxB T$_4$-^{C}DD ist aus vier T-Domänen-typischen α-Helices und einer zusätzlichen α-Helix am C-Terminus (PaxB ^{C}DD-Region) aufgebaut. Während der Interaktion mit der PaxC ^{N}DD bleibt die generelle Sekundärstruktur der T-Domäne unangetastet und lediglich eine weitere α-Helix wird in der PaxB ^{C}DD ausgebildet. Diese strukturellen Veränderungen in den DD-Bereichen gleichen denen, die bereits in Publikation 1 veröffentlicht wurden. In zusätzlichen NMR-Titrationsexperimenten konnten jedoch nicht nur große chemische Shiftveränderungen für Aminosäuren der PaxB ^{C}DD, sondern auch für Aminosäuren die in der α-Helix 4 der T-Domäne beheimatet sind nachgewiesen werden. Diese vorläufigen Ergebnisse implizieren, dass auch das PaxB/C DD-Interface durch die T-Domäne erweitert ist.

I. INTRODUCTION

1. NRPSs produce diverse natural products

Non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), which possess a similar modular architecture, produce a wide range of natural products with a large structural and functional diversity. The biosynthesis is realized by enormous multi-enzyme machineries (200–2,000 kDa) harboring repeated functional units - denoted as modules. Each module is responsible for the incorporation of one specific building block in the growing peptide chain. Therefore, each module is equipped with a set of distinct domains with different catalytic functions necessary for a single chain elongation cycle until in the final step the peptide product is released from the NRPS assembly line, leading to the release of non-ribosomal peptides (NRPs). These peptides are synthesized by almost all bacteria, archaea and fungi and for some of them a bioactivity was determined, which resulted in their application as potent pharmaceuticals (antibiotics, anticancer, immunosuppressive), agrochemicals and veterinary agents (Fig. 1).[1–3]

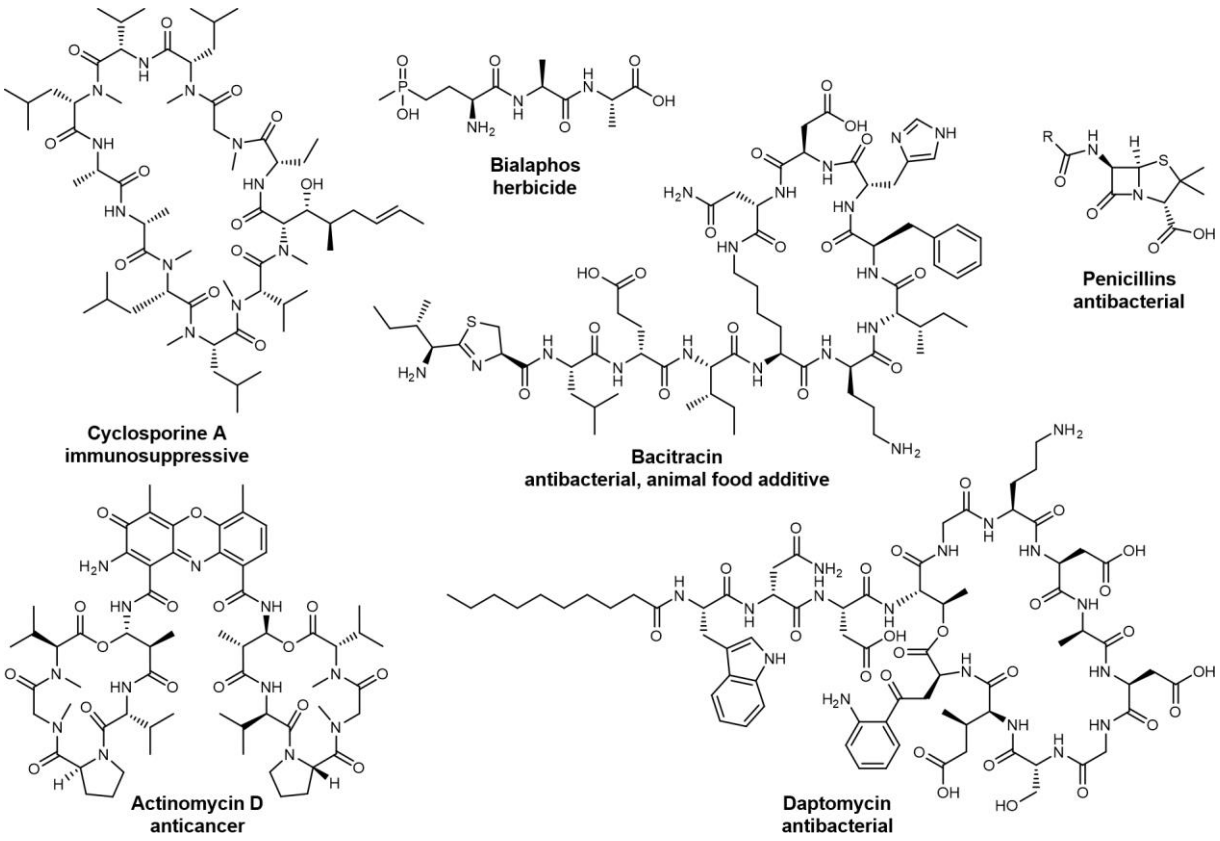


Fig. 1. Selection of marketed NRP drugs.

The structural complexity of these linear, cyclic or branched cyclic natural compounds often makes it impossible to produce them completely by chemical synthesis and therefore their production relies on bacterial fermentation to obtain the final product or a precursor molecule, which is further modified chemically. Especially the simple incorporation of *D*-amino acids, heterocyclic rings, N-methylated residues, sugar and fatty acid moieties, catalyzed by module incorporated domains or tailoring enzymes, expands the chemical diversity of natural products (NPs) and increases their biological activity.[1,2]

Gram-negative entomopathogenic bacteria of the genus *Xenorhabdus* and *Photorhabdus* are known as a rich source for NRPs.[4] The bioactivity testing of cell-free culture supernatants of promoter-induced biosynthetic gene clusters in *P. laumondii* TTO1-Δ*hfq* and *X. szentirmaii*-Δ*hfq* mutants confirmed several known bioactivities[5] like the quorum quenching activity of the phenylethylamides and tryptamides against C*hromobacterium violaceum*[6] and the apoptosis-inducing activity of the proteasome inhibitor glidobactin A.[7,8] The bioactivity of further NRPs produced by bacteria of the genus *Xenorhabdus* and *Photorhabdus* is currently not known, but the NRPS clusters in these bacteria have turned out to be an excellent model to establish re-engineering strategies[9,10] in order to gain further information in the NRPS assembly line arrangement.[11] Especially the afore mentioned property of NRPSs to synthesize structurally complex peptides has driven the efforts to re-engineer these megaenzyme machineries and the structural insights already obtained into the NRPS single domain and overall domain architecture (chapter 2.2), increased our understanding of how NRPs are produced on molecular level. Finally, this knowledge could be used to develop different strategies (chapter 3) that can specifically modify NRPSs in a functional way.

2. The protein architecture of NRPSs

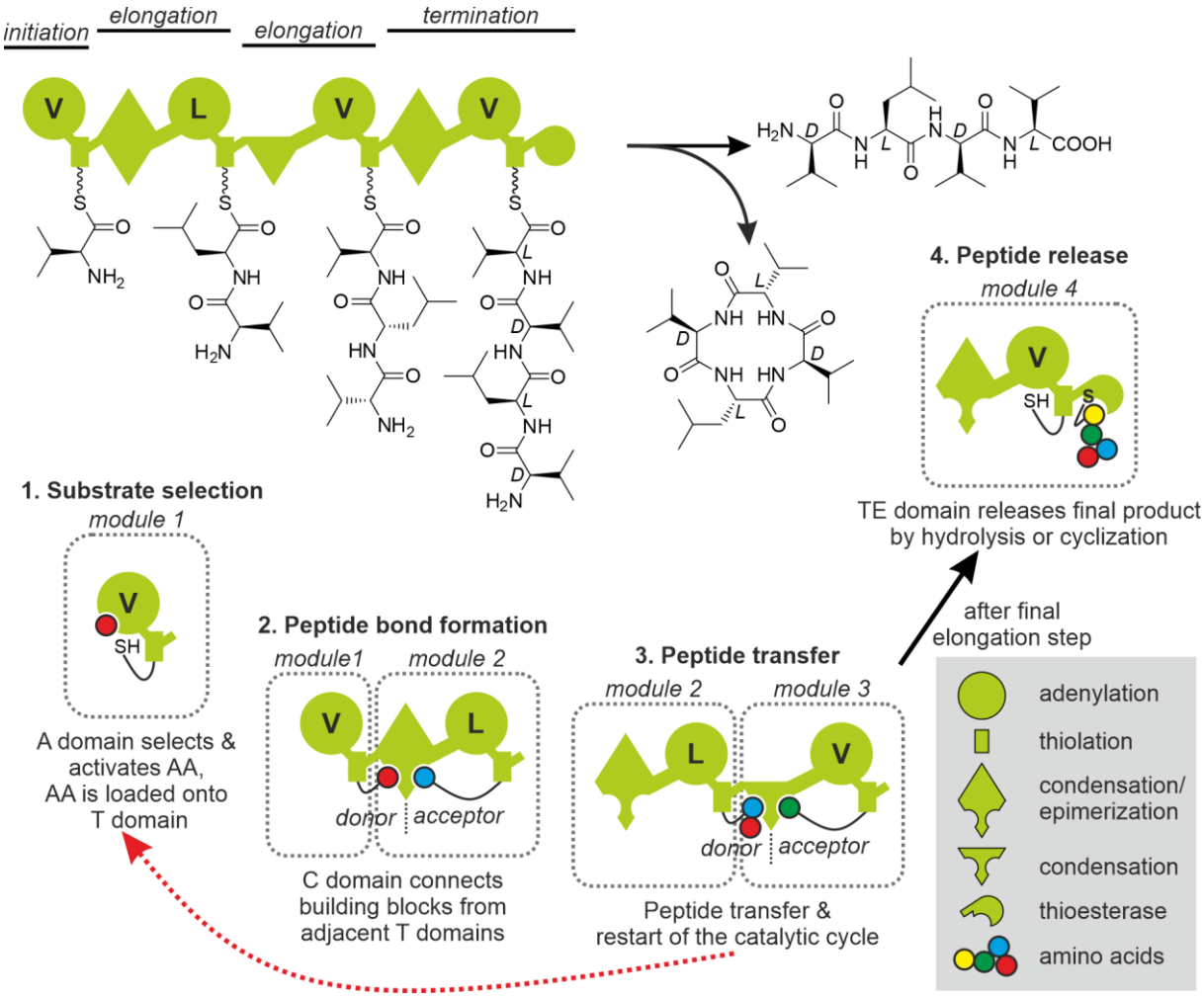


Fig. 2. Example of a non-ribosomal peptide synthetase, producing the linear and cyclic xenotetrapeptide.[12] The different steps of a typical NRP synthesis are displayed below and the respective roles of the domains during each step are explained (adapted from [13]). Furthermore, for each adenylation domain the amino acid specificity is given as a one letter code.

The biosynthesis of natural products by NRPSs is achieved by distinct steps in a catalytic cycle performed by distinct catalytic domains. These domains are in general grouped into modules and each of them is responsible for the incorporation of one amino acid building block in the nascent peptide chain. Therefore, an elongation module is basically equipped with a set of three domains: (I) an adenylation (A) domain, (II) a thiolation (T) domain also called peptidyl carrier protein (PCP) and (III) a condensation domain. Furthermore, the biosynthesis generally follows the collinearity rule, so the number of building blocks correlates with the number of NRPS modules. Additionally, the NRPSs' sequential order defines the primary sequence of the peptide product.

The first step of a catalytic chain initiation or elongation cycle is mediated by the A domain, which is responsible for the selection and activation of an amino acid (AA) by the hydrolysis of adenosine triphosphate (ATP), forming an energy-rich aminoacyl-adenylate intermediate.[14] This activated AA is then tethered to the 4′-phosphopantetheine (PPant)-arm of the downstream T domain,[15] which transfers the building block to the C domain. The C domain catalyzes the peptide-bond formation between the T domain-linked activated AAs and transfers in parallel the donor substrate to the acceptor side.[16] The final module, defined as termination module, often harbors a C-terminally located thioesterase (TE) domain. This domain is responsible for the release[17] of linear, cyclic or branched cyclic peptides, after the transfer of the peptide chain to a serine residue positioned within a conserved Ser-His-Asp catalytic triad has taken place.[18]

To exploit the full potential of NRPSs, further tailoring enzymes can be installed within the NRPS machinery (acting *in cis*) or acting *in trans* on the way to the mature NRPS product. One example is the cyclization (Cy) domain, which replaces the core C domain and catalyzes the formation of 5-membered heterocyclic rings,[19] namely oxazolines derived from serine/threonine residues and thiazolines derived from cysteine residues. In addition, the integration of epimerization (E) domains, located at the C-terminal end of modules, catalyzes the epimerization of *L*-amino acids to *D*-amino acids which are located at the C-terminus of T domain-bound NRPs.[20] Three further examples of tailoring enzymes are the methyltransferases (MT), responsible for the N- or C-methylation of amino acid residues by the transfer of the methyl group of its cofactor *S*-adenosyl methionine (SAM),[21] oxidation domains, which catalyze the oxidation of oxazoline/thiazoline to thiazole/oxazole[22], and the N-terminally located formyltransferase (F) domains in initiation modules, which incorporate a formyl group at the N-terminus of the natural product.[23]

2.1. Types of NRPSs

The potential for a diverse library of NRPs is based on the structural architecture of NRPSs, where distinct catalytic domains interact in an assembly line fashion. Mootz and colleagues[24] initially classified NRPS systems into three groups with respect to the modular arrangement: (I) linear NRPSs (type A), (II) iterative NRPSs (type B), or (III) nonlinear NRPSs (type C).

The type A NRPSs are described as megaenzyme machineries in which the primary sequence of the final peptide product is unambiguously defined by the number and order of modules (initiation module-(elongation module(s))$_x$-termination module) as previously seen for the xenotetraptide[12] NRPS (Fig. 2). Further examples of this type A NRPSs are the tyrocidine,[25] surfactin,[26] bacitracin[27] and cyclosporine[21] synthesizing NRPSs. An additional example of a type A NRPS is the peptide-antimicrobial-*Xenorhabdus* (PAX) peptide-producing NRPS from *Xenorhabdus nematophila* HGB081,[28] which will be described in detail later.

In iterative type B NRPS modules or domains are used more than once during the peptide biosynthesis. Therefore, the number of modules in these NRPS systems is lower, compared to the number of building blocks in the final product. In the biosynthesis of the iron-chelating siderophore enterobactin[29] and gramicidin S[30] the growing peptide chain, for example, is stalled on the active site serine residue of the C-terminal TE domain, thereby allowing preceding modules to assemble further building blocks until the full-length product is formed and released from the TE domain.[31]

Nonlinear type C NRPSs on the one hand often possess modules with an atypical architecture in comparison to the classical domain arrangement (C-A-T) of elongation modules in linear type A NRPSs and on the other hand conflict with the collinearity rule. Two examples of type C NRPSs are the syringomycin[32] and capreomycin[33] systems, in which *in trans* interacting di-domains (SyrB1: A-T; CmnI: T-C) deliver substrates (*L*-Thr) to modules with an unusual order [(C-A-T)$_8$-(C-T)$_9$] and therefore compensate for a missing A domain - or are loaded by a preceding A domain to elongate the growing peptide chain by a further building block.

Notably, especially in the context of linear type A NRPSs, all domains can be located on a single polypeptide chain (all domains interact *in cis*) or divided on different polypeptide chains as seen for example in the tyrocidine, surfactin and PAX NRPSs. Each of these NRPS systems is composed of three polypeptide chains, which have to interact *in trans* during the peptide biosynthesis. Hahn and Stachelhaus[34] discovered short elements, denoted as “communication-mediating (COM) domains”, located at the C- and N-terminal ends between the tyrocidine NRPSs TycA, TycB and TycC, and confirmed their important role in the correct and functional assembly of multi-protein systems by the formation of non-covalent protein-protein interactions.

The accurate order of subunits is crucial also in related megaenzymes of polyketide synthases and it was shown that C- and N-terminal regions, called “docking domains (DDs)”[35], are responsible for the specific protein-protein interaction. To decrypt the specificity achieving properties of these domains on a molecular level, structural analyses are indispensable.

2.2. NRPS domains

2.2.1. Thiolation domain/peptidyl carrier protein

The peptidyl carrier protein or thiolation domain is the central domain of the three core domains (C, A, T) and has the important role of shuffling the substrates and peptide intermediates between different catalytic domains.[13] This task is carried out by this 70-90 amino acids long domain, which for this purpose has to be activated from its inactive *apo*- to its active *holo*-form by the posttranslational modification with a PPant prosthetic group derived from coenzyme A (CoA).[36] The PPant group is attached to the sidechain hydroxy group of a conserved serine residue (GGX**S**) within the T domain by a PPant-transferase (PPTase)[36,37] and facilitates the T domain to covalently bind peptide building blocks via a thioester linkage. The T domain structurally benefits from this posttranslational modification, which expands the binding range of this domain by ~18 Å[13] and allows to reach the catalytic centers of flanking A and C domains.

The first structure of a T domain was solved by NMR, focusing on the third T domain of tyrocidine synthetase 3 (TycC). This study uncovered the general fold of T domains which are composed of four α-helices 1−4 arranged in a four-helical bundle. Compared to α-helices 3 and 4 the first two α-helices are longer and the conserved serine residue, to which the PPant-arm is post-translationally added, is located at the N-terminus of α-helix 2. All α-helices are connected by short loops except for the connection between α-helix 1 and α-helix 2.[38] Here a long loop is installed which is diverse in sequence and structure as seen in the comparison of known PCP structures.[39]

During the NRP biosynthesis, the T domain has to form transient interactions with adjacent A and C domains. Koglin *et al*.[40] gave one explanation how these transient interactions are guided by identifying three structurally different states (*apo* (A), *apo*/*holo* (A/H), *holo* (H)) of a PCP domain form the tyrocidine TycC3 T domain. Therefore, it was assumed that this conformational switch regulates the T domain

interactions, but this observation was challenged by Lohman *et al*.[39] comparing known T domain structures and almost entirely identifying the A/H state.

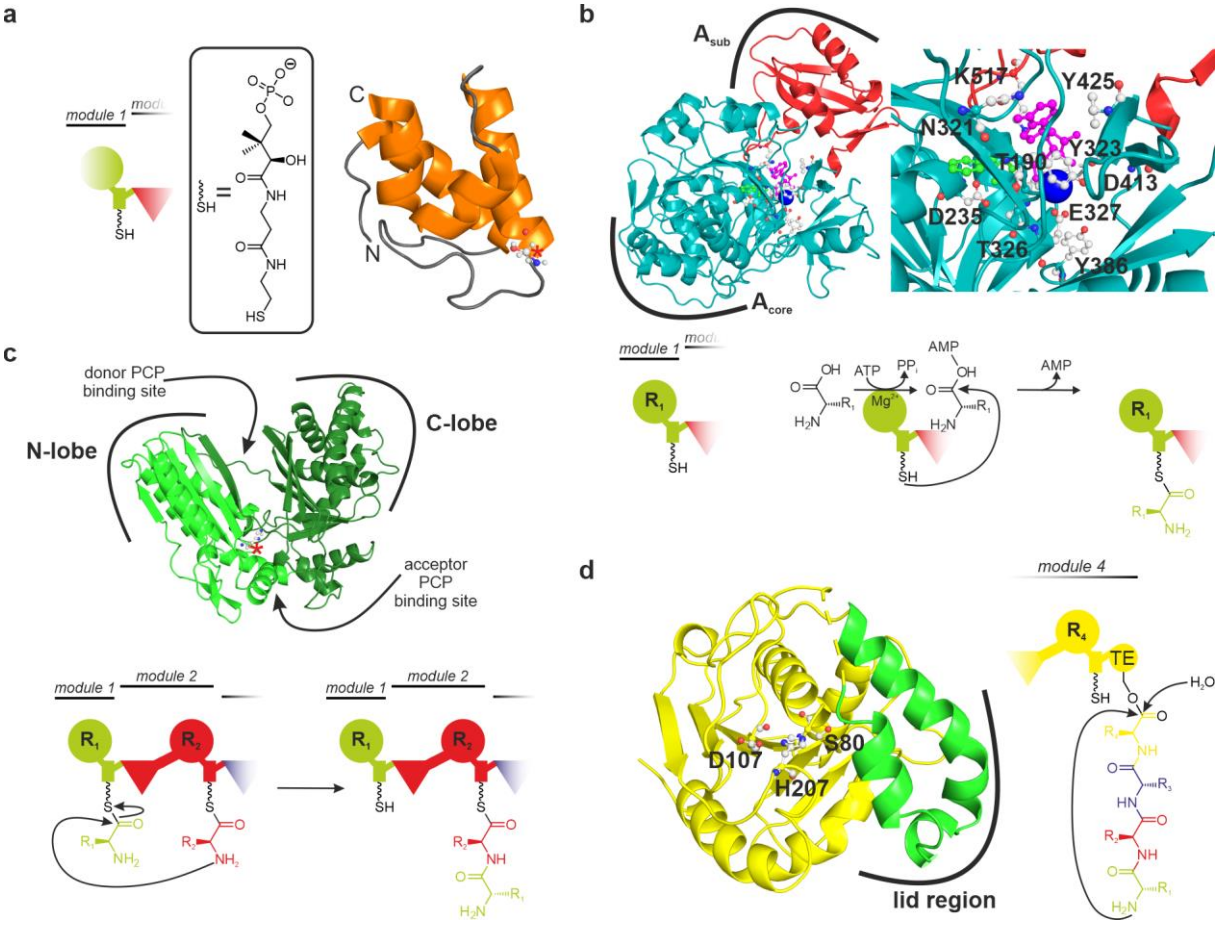


Fig. 3. Structures of NRPS thiolation, adenylation, condensation, and thioesterase domains and their catalytic activity. (a) NMR solution structure of an NRPS extracted *apo*-PCP domain from the NRPS TycC which possesses the typical four-helical bundle arrangement of the A/H state (PDB: 1DNY). The conserved serine residue is labeled with an asterisk and the chemical structure of the PPant-arm is given, which is attached to the conserved serine residue in the active *holo*-state.[38] (b) X-ray crystal structure of the "closed" A domain from the gramicidin S NRPS system in a complex with adenosine monophosphate (AMP; magenta), *L*-phenylalanine (green) and Mg^{2+} (blue ball). The A domain specificity defining residues located in the binding pocket are labeled additionally.[41] (c) X-ray crystal structure of the free standing Vibrobactin (Vib)H condensation domain (Protein Data Bank (PDB): 1L5A). The pseudo-dimer is composed of a N- and C-terminal domain (N-/C-lobe) and the C domain typical H**H**XXXDG (asterisk: catalytically important **histidine**) motif is located at the interface between the two domains right next to the solvent channel.[42] (d) Cartoon representation of the TE domain from the surfactin A-producing NRPS (PDB: 1JMK). The x-ray crystal structure represents the "closed" conformation, in which the peptide binding pocket is surrounded by the three-helical lid region (green; αL1–αL3) and the catalytic triad residues (S80, H207 and D107) are highlighted as ball and stick models. An unordered loop region preceding the lid region is shown as a dotted line.[43]

Furthermore, structural analysis of the *apo*- and *holo*-form of the terminal T domain ($PCP7_{tei}$[44]) from the teicoplanin NRPS by NMR determined highly similar structures (Cα root-mean-square deviation (RMSD) of 0.8 Å) and suggests that the T domain interactions are not driven by conformational changes during the catalytic cycle, but by fine structural rearrangements. The sequence alignment form PCP7 domains from

related glycopeptide biosynthetic NRPS systems further revealed that the C-terminal half following the conserved serine residue is the most diverse region of the compared sequences, whereas the N-terminal sequence half is highly conserved, leading to the assumption that especially this region composed of α-helices 2−4 are responsible for interdomain contacts.[44] This hypothesis could be confirmed in the x-ray crystal structure of the *holo*-AB3403 (C-A-T-TE) protein of an uncharacterized biosynthetic pathway, in which the PCP interface with the condensation domain is defined by α-helix 2 and the loops that precede and follow α-helix 2, comprising specific hydrophobic interactions and hydrogen bonds.[45] The T-TE di-domain structure of the protein EntF from the enterobactin-producing synthetase, locked in an active conformation with a phosphopantetheine analog, shows as well numerous hydrophobic interactions between the TE and T domain, involving mainly α-helix 2 and 3.[46]

In summary, the T domain is catalytically inactive, but the described transient interactions with further NRPS core domains make this domain one of the key players of NRP biosynthesis.

2.2.2. Adenylation domain

The adenylation domain is responsible for the selection and activation of the correct substrate(s). In a canonical module (C-A-T), the A-domain is located directly downstream of the adjacent T domain[47] and was for the first time structurally characterized by Conti and colleagues[41], focusing on the phenylalanine-activating subunit of gramicidin synthetase 1 (PheA). Here, the A domain typical large N-terminal core domain (A_{core}: ~450 AAs) and the much smaller C-terminal subdomain (A_{sub}: ~100 AAs) were determined, which are connected by a four residue “hinge”.[48] Co-crystallization of *L*-phenylalanine and AMP revealed that the N- and C-terminal domain are involved in the substrate recognition via a network of specific hydrogen bonds.[41] Based on these structural insights, especially into the A domain binding pocket, Stachelhaus *et al*.[49] were able to define a “specificity-conferring code” nowadays also known as “Stachelhaus code”. It was found out that substrate recognition of PheA involved ~10 amino acids residues. Stachelhaus *et al*. realized that the discovered signature sequences (A domain “code”) of cognate A domains could be leveraged to *in silico* predict A domain substrate specificity from the primary sequence.[49] Today, this knowledge is still useful especially in the context of uncharacterized NRPSs, where a prediction of the final product composition can be

made by simply analyzing a given protein sequence derived from the gene sequence.[50]

A domains catalyze two half-reactions: (I) ATP-dependent activation of a specific amino acid building block by forming an aminoacyl adenylate and (II) transfer of the activated-amino acid to the respective *holo*-T domain and covalent linkage of the building block via a thioester bond to the PPant-arm.[51] Thus, structures of A domains in different functional states could be solved and it was confirmed that this two-step mechanism goes along with structural alterations.[48] At the beginning of the A domain catalytic cycle an “open” conformation is assumed for NRPS A domains as seen for the firefly luciferase, which also belongs to the superfamily of adenylate-forming enzymes.[52] In this conformation, the C-terminal A_{sub} domain, previously termed as “lid” of the binding pocket, is separated by a large cleft from the N-terminal A_{core} domain[53] by a 94° angle and the substrates (ATP, amino acid, Mg^{2+}) can enter the active site. Then the adenylation reaction takes places after closing of the A_{sub} “lid”, which further provides residues involved in the catalytic center.[41,53] This closed conformation additionally protects the produced aminoacyl-adenylate from bulk solvent. Finally, the A_{sub} domain rotates ~140°[54,55] and allows the T domain to bind and the thioester bond between adenylate and the PPant-arm is formed.

Notably, some bacterial A domains rely on the presence of MbtH-like proteins (MLPs), usually encoded in the gene cluster.[56] These proteins bind to MLP-dependent A domains and are necessary for their adenylation activity,[57] as well as the efficient production of NRPs by non-ribosomal peptide synthetases.

2.2.3. Condensation domain

As mentioned above, C domains are responsible for peptide-bond formation, originally demonstrated by Stachelhaus and colleagues.[16] In general, C domains are composed of ~450 AAs and show only low sequence homology to each other, except for their conserved HHXXXDG motif, which was previously shown as typical for the superfamily of acyltransferases like the chloramphenicol and dihydrolipoamide acetyltransferases.[58]

The first crystal structure of a C domain identified a pseudo-dimeric configuration, in which a N-terminal and C-terminal subdomain, also referred to as N- and C-lobe,[59] adopt a V-shaped conformation.[60] Each subdomain consists of an αβα sandwich

connected via a 56-residue linker composed of two α-helices. The HHXXXDG motif is located in a solvent channel that runs through the molecule in the interface between these two subdomains, which allows the access of two loaded PCP PPant-arms to the second histidine of the active site motif during the peptide bond formation.[42] Hereby, the ~15 Å distance of each of the C domain surfaces (N-/C-face) to the second histidine (H**H**XXXDG) ideally fits the range spanned by a PPant-arm.[59]

The biochemical function of some of the amino acids, involved in the conserved HHXXXDG motif, was analyzed in different mutational analyses. Bergendahl *et al*.[61] for example tackled the catalytically important residues in a C domain of the tyrocidine NRPS system and could show that mutations of the second histidine (H147A) and aspartate (D151N) completely abolished condensation activity. With the first crystal structure of a free-standing NRPS C domain (VibH) in hand, Keating *et al*.[42] also mutated the HHxxxDG residues H125L, H126A/E and G131L, which showed no (H125L) or only a minor effect (H126A/E, G131L) on product formation. Especially the observation made in the context of the VibH H126A/E mutants challenges the proposed critical role of the second histidine in the H**H**xxxDG motif to function as the general base catalyst for peptide bond formation.[61] So far, all condensation domain structures are lacking native ligands in the active site which would help to reveal this mechanism. With respect to the peptide bond formation between two *L*-amino acids, C domains can be further classified as $^{L}C_{L}$ domains or $^{D}C_{L}$ domains if they are stereoselective for *D*-substrates as upstream donor and *L*-substrates as downstream acceptor.[62] Typically, $^{D}C_{L}$ domains are selective for the correct enantiomer and are found downstream of E domains.[63] In addition, C domains are classified as dual condensation/epimerization (C/E) domains, which are $^{D}C_{L}$-catalysts that possess an intrinsic epimerase activity. Therefore, *L*-amino acids incorporated by the upstream module into the growing peptide chain are epimerized into *D*-configuration and connected with the *L*-residue activated by the module containing the C/E domain.[64]

Interestingly, the C domain is not only responsible for the condensing of two amino acids via a peptide bond, but also contains an editing function especially on its acceptor site, which enables the selection of amino acids with the correct stereochemistry and side chain size, whereas the donor side is even more tolerant in its substrate spectrum.[65,66] The fact that this C domain gatekeeper function is not

limited to the C domain itself, but also affects the specificity of the adjacent A domain was shown in further analyses. Meyer *et al*. compared the adenylation specificity of di- (A-PCP) and tri-domain (C-A-PCP) constructs extracted from modules of the microcystin NRPS system *in vitro* and showed that the substrate specificity of the adenylation reaction is reduced if the C domain was excluded from the analyses, whereas only the native amino acid substrates were activated (McyB: Leu; MycC: Arg) in the presence of the cognate C domain.[67] This substrate specificity dependency on the C-A linkage can be structurally explained by the crystal structure of the *Bacillus subtilis* surfactin termination (SrfA-C) module. Here, Tanovic and colleagues[68] determined an overall surface area of 1621 $Å^2$, which is shielded by the A and C domains from solvent access and numerous interactions could be identified between residues of the A_{core} and C domain. Therefore, it was assumed that the C-A domain interface is unaffected during the catalytic cycle and a rigid catalytic C-A platform was proposed.[68]

2.2.4. Thioesterase domain

The non-ribosomal peptide synthesis is (in most cases, but not in fungal NRPS systems) terminated by a thioesterase domain, located at the C-terminal end of the termination module.[17] Therefore, the ~280 AAs long enzyme hydrolyzes the PCP-bound peptide chain in a two-step mechanism.[69] Firstly, a serine residue, positioned in the active site of the TE domain, attacks the PCP-bound peptide thioester as nucleophile and generates an acyl-*O*-TE intermediate. Secondly, the TE-bound peptide chain is released either directly by a water molecule, functioning as an intermolecular nucleophile leading to a linear peptide, or by the attack of an intramolecular nucleophile, like the N-terminal amino group (-NH_2) or the hydroxyl-group (-OH) of a serine or threonine sidechain, producing a macrocyclic product,[31] e.g. as observed for surfactin A and tyrocidine A. The latter mechanism is most likely favored, highlighted by an increased proteolytic degradation resistance.[63]

Mechanistic insights into the TE release mechanism eventually was obtained from the crystal structure of SrfTE published by Bruner *et al*.[43] The SrfTE structure possesses a distinct bowl-shaped active site cavity, of which an “open” and “closed” conformation could be deduced. As a member of the α/β hydrolase enzyme family, a typical catalytic triad composed of S80, H207 and D107 could be identified in the active site. The active residue S80, where the acyl-*O*-TE intermediate is attached,

can easily be identified by the signature sequence GX**S**XG. This intermediate is stabilized in the oxyanion hole formed by the backbone amide group of A81 and amino acid V27. Also noteworthy is the detection of a three-helical lid region (αL1–αL3), which can be folded back into the “open” conformation, allowing unrestricted access to the active site, whereas this cavity is only accessible by small molecules in the “closed” conformation.[43]

2.2.5. Interdomain and intermodular linkers

In this NRPS domain architecture overview it is also necessary to mention the interdomain and intermodular linker sequences as an essential module component. One example, showing the importance of maintaining the native interdomain linker in re-engineering approaches was shown by the translocation of the thioesterase domain of the ten-modular plipastatin synthetase to the end of internal T domains to generate truncated assembly lines. Only when instead of T-C linkers the T-TE linker was used to connect the TE domain to internal T domains, truncated branched cyclic or linear hepta–nonapeptides were produced. This suggests that this interdomain region is important for the catalytic activity of the TE domain.[70]

Another study has focused on the T domain substitution effect in the single-module NRPS IndC from the indigoidine biosynthetic pathway of *Photorhabdus luminescens*. The exchange of the endogenous IndC T domain for natural and synthetic ones generated mutants that partially produced indigoidine, with higher yields even than the unmodified IndC, but surprisingly, the T domain exchange mutant of the closely related indigoidine synthetase BpsA was non-functional. It was supposed that an incompatible A-T or T-TE interdomain linker disrupts the domain-domain crosstalk and therefore different genetic constructs which include interdomain linker sequences of varying origin (IndC or BpsA) were analyzed. Interestingly, only if the bioinformatically suggested, complete A-T interdomain linker of the incoming BpsA T domain was included, a functional NRPS assembly line was built.[71]

A recent study bioinformatically analyzed 39,804 intermodular linkers of well annotated and putative NRPS biosynthetic gene clusters from 39,232 bacterial genomes and identified a relationship between the A domain activated amino acid substrate and the intermodular linker between pairs of modules. Therefore, each module was represented by their activated amino acid substrates (A_n), defining an “A_1-linker-A_2” workspace. Hence, it could be determined that module pairs activating

the same amino acids tend to be linked by the same intermodular linkers, regardless of the bacterial species, which was explained by horizontal gene transfer.[72]

This theoretical observation was proven *in vivo* in the work of Bozhüyük *et al.*[10], in which A-T-C(/E) exchange units were substituted at a conserved WNATE sequence in the C-A linker, which maintained an intact domain connection and led to functional re-engineered NRPSs.

In summary, it can be assumed that interferences in the linker regions can hinder module-module contacts and lead to decreased product yields or even non-functional NRPSs.

2.2.6. NRPS overall domain arrangement

During a peptide elongation cycle (Fig. 4a) the previously mentioned distinct catalytic domains have to interact in an orchestrated manner and only structural investigations of whole modules can enlighten what really happens on a molecular level during this process.

The structures of a C-A-T-TE termination module for the first time showed the NRPS domain arrangement in a more holistic context.[45,68,73] The first x-ray crystal structure of such a multi-domain complex of the 144 kilodalton *Bacillus subtilis* termination module SrfA-C (Fig. 4b) was solved at 2.6 Å resolution in the Essen lab.[68] This structure captured an intermediary step of SrfA-C, characterized by a 63 Å distance between the C domain catalytic residue H147 and the A domain-bound substrate leucine, which suggested the necessity of large conformational domain rearrangements in the catalytic cycle. The overall structure is dominated by a huge C-A interface (1621 $Å^2$) and includes a large number of residues that stabilize this interface. Therefore, a fixed C-A “catalytic platform” was proposed, which remains unaffected during the peptide elongation process. Both domains are connected via a 32 AAs long L-shaped C-A linker, whose 11 residue long helical segment shows specific interactions with the A_{Core} domain. The A_{sub} domain is neither positioned in the postulated open, nor in the closed state,[41] which is explained by the missing co-substrate adenosine triphosphate and binding of which may alter the A_{sub} orientation. The T domain is stalled in its *apo*-form due to the S1003A mutation, to which normally the PPant-arm is post-translationally added. Nonetheless, this *apo*-T domain structure partly covers the C domain acceptor site (downstream) and the

attachment of a PPant-arm with an attached aminoacyl group would allow to span the distance to the active site located at H147.[68]

Two further termination module crystal structures of AB3403 and EntF with the same domain architecture were determined by Drake *et al*.[45] from the Gulick lab. In the AB3403 structure (Fig. 4c), the unloaded PPant-arm is pointing into the hydrophobic tunnel of the C domain formed in between the C domain lobes and ends next to the C domain active site. PCP and C domain interactions were determined for residues located in the PCP helix α2 and the loops that precede and follow this helix. The adenylation domain is observed in the adenylate-forming conformation and the substrates AMP and glycine are trapped in the "closed" A domain state. In contrast to the observed adenylation state of AB3403, the EntF structure was solved in its thiolation state (thioester formation). For this purpose, a serine adenosine vinylsulfonamide (Ser-AVS) inhibitor was used which mimics the aminoacyl adenylate and is forced to react via its Michael acceptor with the pantetheine thiol of the PCP domain. That is why in the EntF structure the thioester-forming conformation of the A domain could be analyzed, but unlikely no electron density was observed for the thioesterase domain. It was suggested that the TE domain adopts different conformations in the crystal lattice, explained by the absence of a significant number of contacts to other catalytic domains. By comparison of the AB3403 and EntF structures to the SrfA-C structure, not only the TE domain positioning differs, but also the C-A domain interface varies due to an observed ~25° rotation of the C domain relative to the A domain. This observation leaves considerable room for doubt that the proposed C-A catalytic platform may not be as rigid as previously thought. Moreover, for the first time both structures provide structural evidence of conformational changes of the A domain, which guides the PCP domain between active sites of a NRPS module.[74] Especially the A_{sub} domain rotation from the adenylation to the thiolation state positions the PCP right next to the A domain ideally suited for substrate loading.[45]

In a further study, the complete crystal structure of the termination module EntF, including the C-terminal TE domain in complex with the native MLP YbdZ and a MLP homolog PA2412 was solved (Fig. 4d). Again, these complex structures were trapped in the thioester-forming conformation by the application of a Ser-AVS inhibitor. The interface between the MLP and the termination module is dominated by the A_{core} domain and only minor contacts were determined between the MLP and C domain,

whereas no contacts could be observed neither between MLP and the A_{sub} domain, nor the TE domain. Interestingly, as indicated by a low RMSD (Cα) of 0.3 Å comparing the MLP-free and MLP-bound A domains, MLP binding does not affect the A domain structure, suggesting that an activation of the A domain is not a result of structural alterations. Focusing on the overall module architecture, a linear domain arrangement is seen, in which the C domain is installed alongside the A domain followed by the TE domain. This TE domain orientation, stabilized by a 565 $Å^2$ large interface with the A_{core} domain, is absolutely new in comparison to previously published module structures, again highlighting the dynamic character of TE domains in termination modules.[73]

In the outstanding work of Reimer *et al*., all steps of the catalytic cycle of the formyltransferase containing F-A-PCP initiation module of the linear gramicidin synthetase subunit A (LgrA) could be structurally characterized, comprising x-ray crystal structures for the adenylation (open/closed conformation), thiolation and formylation state. In comparison to the formyltransferase domain, whose orientation stays nearly unaffected along all catalytic steps, especially the A_{sub} and PCP domains undergo different conformational changes driven by numerous transient interactions. Interestingly, two A_{sub} domain guided relocations of the PCP domain could be observed in this study: (I) the known 140° rotation of the A_{sub} domain,[75] which transfers the PCP next to the A domain and so the activated Val-AMP can be tethered to the PPant-arm and (II) a 180° rotation and 21 Å movement of the A_{sub} domain that flips the PCP over (75° rotation, 61 Å movement) to the F domain, where the formyl group from the donor cofactor formyltetrahydrofolate is attached to the amino group of the amino acid. Again, this work underlines the important role of the small (~100 AAs) A_{sub} domain, which not only provides catalytic residues for the adenylation reaction,[41] but also proves to be the driving force in the PCP relocation, enabling the PCP domain to meet different catalytic centers.[76]

So far only structures of initiation and elongation modules have been solved, but recently also the x-ray crystal structure of a MLP-bound cross-module NRPS DhbF construct (MLP-$A_1T_1C_2$) was solved (Fig. 4e), which for the first time contains domains from more than one module.[77] Again, the complex was stalled in the thiolation state by using a Gly-AVS inhibitor and a resolution of 3.1 Å was obtained. Here, the PCP domain is sandwiched between the preceding A_1 domain and the

following C_2 domain, allocating on one side the A domain surface and reaching with its PPant-arm into the A domain active site, whereas the opposite PCP domain surface is packed against the C domain, docking on the C_2 PCP donor binding site. No interactions could be detected between the A_1 and C_2 domains and most of the contacts between the A_1 and PCP_1 are formed between residues located in PCP helix α2 following the PPant-arm linked serine and A_{1core} domain residues. However, also minor hydrophobic interactions are developed in the A_{1sub}-PCP_1 domain interface. The opposite PCP_1-C_2 interface, composed of the PCP back side and the C domain donor site, buries a 664 Å^2 area and is dominated by several hydrogen bonds, but only if the PCP_1 rotates by 180°, a catalytically relevant conformation can be obtained, which is necessary for the condensation reaction. For this necessary PCP_1 rotation there is no space in the observed structure, which is why it is assumed that the A_1 and C_2 domains are separated farther, enabled through the flexible A_1-PCP_1 and PCP_1-C_2 linkers. In addition, three-dimensional information for a multi-modular NRPS ($DhbF_{\Delta Te}$-MLP) was obtained by cryogenic electron microscopy (cryo-EM) with a resolution of ~26–29 Å, identifying a considerable flexibility in the module-to-module conformation.[77] It seems that the higher-order architecture of NRPSs is, other than previously suggested, rather dynamic than rigid,[78] as NRPS domains are bouncing back and forth and functionally interact in a defined order during the NRPS catalytic cycle by forming transient interfaces.

Most recently, the first di-modular NRPS structures were published focusing again on the linear gramicidin synthetase subunit A comprising the domain order $F_1A_1T_1C_2A_2T_2E^o_2$.[79] In all obtained crystal structures both modules are characterized by a rather rigid di-domain "catalytic platform"[68] (F_1A_{1core}/C_2A_{2core}), showing only minor changes in the F_1:A_1 and C_2:A_2 orientation. The crystal structure of the di-module $F_1A_1T_1C_2A_2T_2$ construct (Fig. 4f) for the first time provides insight into the overall condensation conformation. Interestingly, in this structure the A_{1sub} and A_{1core} domains are arranged in the adenylation conformation and the T_1 domain is positioned in parallel in the condensation conformation, which allows restarting the catalytic cycle, even if the peptide chain is not yet transferred to the downstream T domain. This double functional conformation results in a "*skipping*" of the T domain

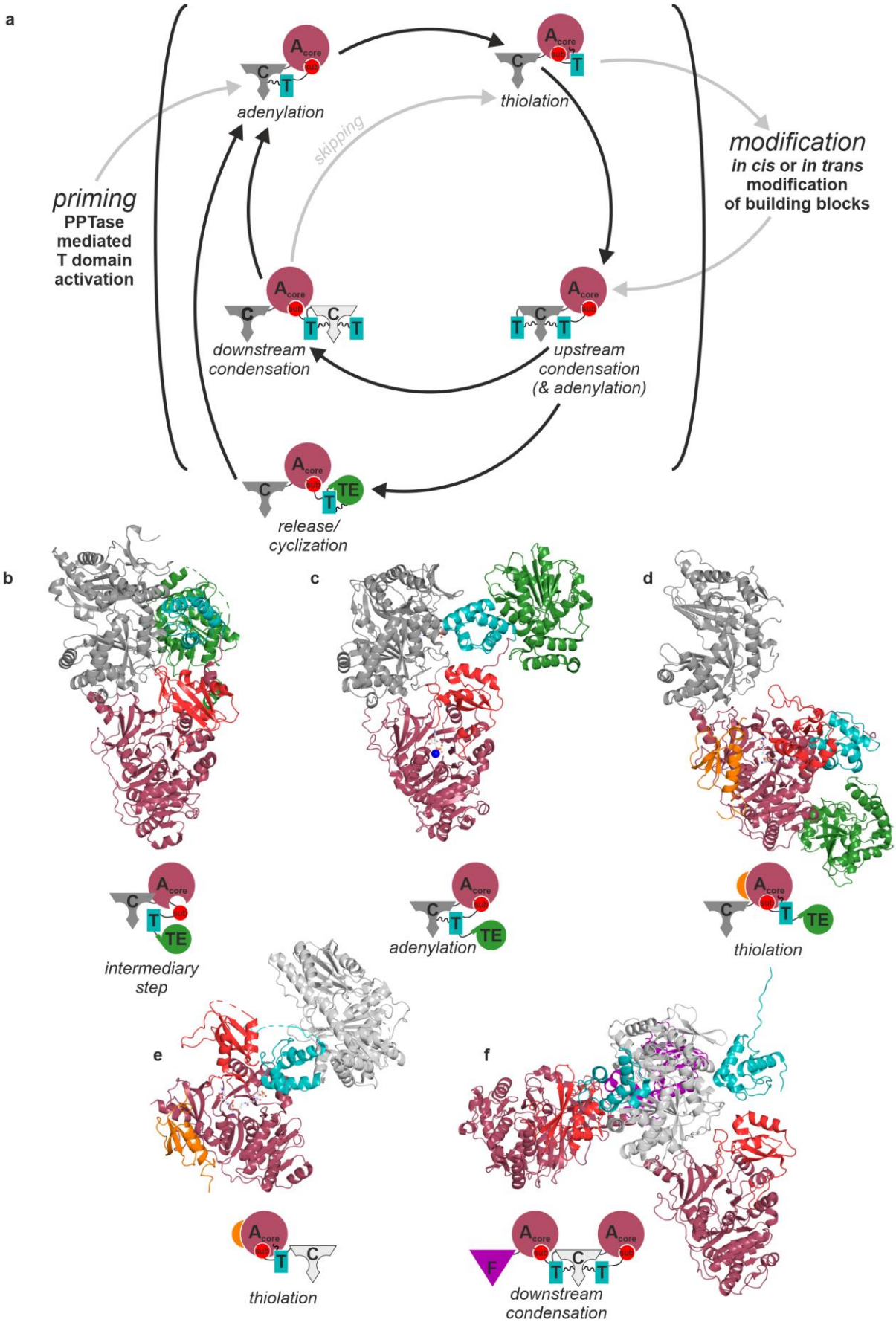


Fig. 4. (a) Model of a NRPS peptide elongation cycle adapted from [2]. Schematic representations of the domain-domain arrangements are given for each step of the catalytic cycle and the module

comprising domains are highlighted by different colors (condensation domain (C), grey shades; adenylation (A) domain, raspberry; thiolation (T) domain, cyan; thioesterase (TE) domain, green; formyltransferase (F) domain, magenta; MbtH-like protein, orange). Before the catalytic cycle can start, the T domain is activated ("*priming*") and transformed into its *holo*-form by a PPTase, which attaches the PPant-arm (wavy line) to a T domain internal conserved serine residue. After "*adenylation*" of the AA substrate by the A domain, the T domain is relocated next to the A domain, guided by an 140° A_{sub} domain rotation, and the thioester-formation can take place, in which the activated AA substrate is tethered to the T domain PPant-arm "*thiolation*". The PPant-arm linked substrate can undergo different modifications by *in cis* or *in trans* interacting tailoring enzymes. The substrate-loaded T domains then migrate to the acceptor site of the C domain and the peptide bond is formed ("*upstream condensation*") and *vice versa* if the substrate-loaded T domain is bound to the donor side of the C domain ("*downstream condensation*"). During the condensation reaction, adenylation can already take place again and "*skipping*" directly to the thiolation step is possible. The C-terminally located TE domain of a termination module "*releases*" the peptide chain. (b) Crystal structure capturing the domain arrangement of an intermediary catalytic step of the SrfA-C termination module (PDB: 2VSQ).[68] (c) Crystal structure of the termination module AB3403 (*holo*-form) with the A domain in the "closed" adenylation conformation, bound to AMP, glycine, and Mg^{+} (blue sphere), while the T domain is positioned for simultaneous upstream condensation (PDB: 4ZXI).[45] (d) X-ray crystal structure of the MLP-bound termination module EntF possesing a linear domain arrangement (PDB: 5JA1)[73] (e) Cross-module fragment ($A_1PCP_1C_2$) crystal structure of the di-modular NRPS DhbF trapped in the thioester formation via a mechanism-based inhibitor and bound to MLP (PDB: 5U89)[77] (f) Di-modular crystal structure of the initiation module LgrA, captured in the peptide bond forming conformation (PDB:6MFZ).[79]

from the condensation directly to the thiolation step. It has to be mentioned that the crystal structure of a shortened $F_1A_1T_1C_2A_2$ construct, which possesses an entirely different overall domain conformation, was also solved in the peptide bond forming conformation, thereby suggesting that different three-dimensional domain architectures can fulfill similar catalytic reactions. Nevertheless, in none of the determined structures domain positions were identified, which would allow module 1 to take over catalytic steps of module 2, and only minimal interactions between the di-domains could be identified at all. This observation should enable an easy module swapping, where only a compatible interface between the T_n:C_{n+1} domains has to be considered.[79]

All single module and di-module structures described above highlight the overall flexibility in regard to domain-to-domain and module-to-module conformations and support the hypothesis that the biosynthesis in NRPSs works as a dynamic machinery and not as a rigid assembly line.

2.3. PAX peptide-producing NRPS – an example of a multi-protein NRPS

The peptide-antimicrobial-*Xenorhabdus* (PAX) peptide-producing NRPS from *Xenorhabdus nematophila* HGB081 analyzed by Fuchs *et al*.[28] is composed of three proteins (PaxA/PaxB/PaxC) that harbor seven modules and interact in an unidirectional way which is characteristic for a type A NRPS. By a combination of labelling experiments and MALDI-Orbitrap MS/MS experiments, the amino acid

composition of the PAX peptides was elucidated and the presence of numerous lysine residues was identified (Fig. 5). In addition, because of the N-terminally located starter C domain, a library of PAX peptides could be detected, resulting from the incorporation of different fatty acid moieties, which differ in length and double bond configuration.

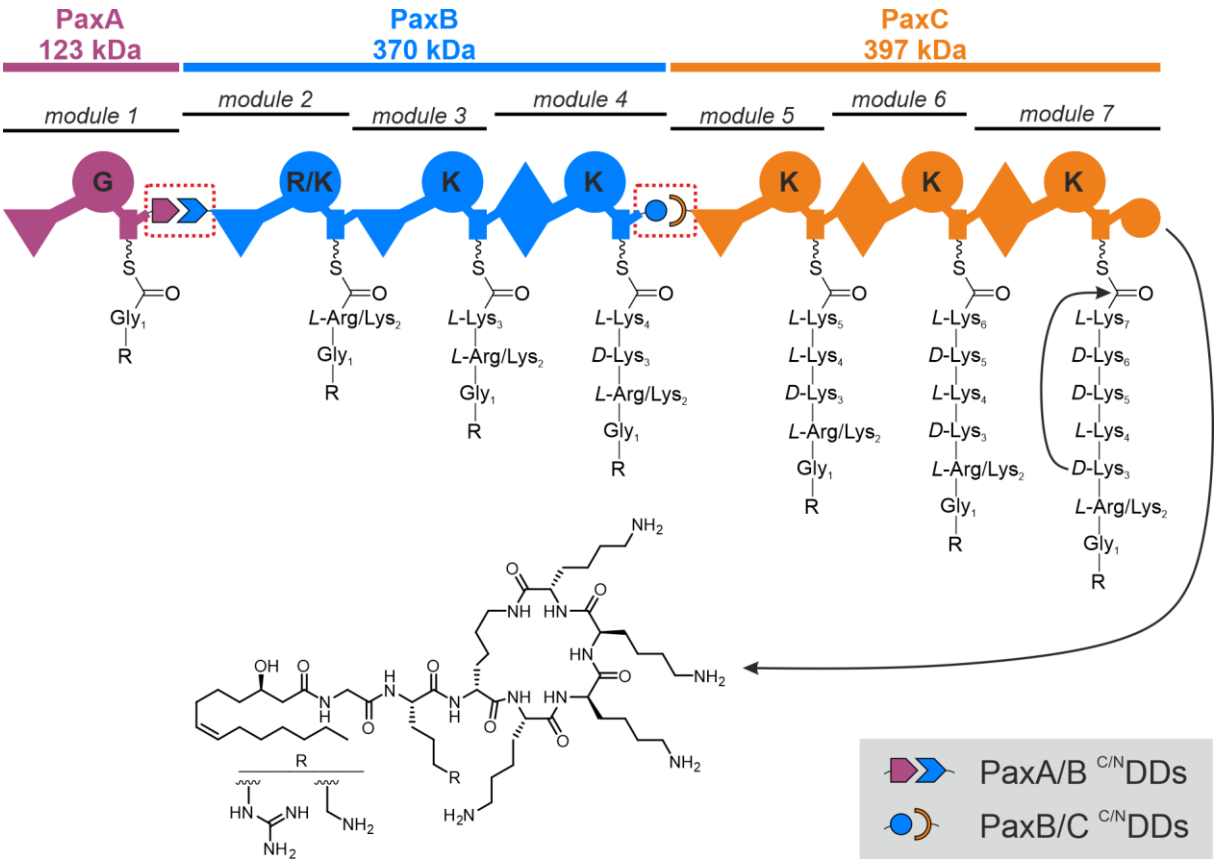


Fig. 5. Biosynthesis of the peptide-antimicrobial-*Xenorhabdus* (PAX) peptide-producing NRPS from *Xenorhabdus nematophila* HGB081. Boxed red are docking domain pairs located between the Proteins PaxA/B and PaxB/C. See Fig. 2 for assignment of the domain symbols.

Although the biosynthetic gene cluster of the PAX peptide-producing NRPS is widely distributed in different *Xenorhabdus* strains, PAX peptide production was detected only in a small number of liquid culture extracts of these strains.[4] Furthermore, the function of the PAX peptides has not been fully understood yet.

Because of the amphiphilic properties (lipophilic head – fatty acid moiety; hydrophilic tail – lysine-rich amino acid composition) it was recently shown that PAX peptides are anchored in the bacterial outer membrane of the Gram-negative producer strains of the genus *Xenorhabdus* via their fatty acid moiety and present the positively charged lysine residues on the bacterial membrane surface. This accumulation of positively charged residues alters the membrane surface potential and functions as a protective

barrier against cytotoxic and positively charged compounds, which target the outer membrane of bacteria.[80]

A further question of the PAX peptide biosynthesis is, how the unidirectional protein assembly (PaxA>PaxB>PaxC) is achieved. As known for further linear type A NRPS systems, N- and C-terminally located docking domains are responsible for the correct protein order, but to date, structural insights into this protein-protein interface, especially of pure NRPS systems, are limited.

2.4. Structural aspects of natural docking elements

In the last decades, the discovery that docking domains play an important role in the assembly of multi-protein NRPS and PKS systems increased the efforts to solve high resolution structures of these elements.[11,35,81–88] This allowed for the identification of structurally diverse docking domain types.[87] Furthermore, by employing bioinformatic methods, docking domains from PKS systems could be grouped into three non-interacting phylogenetic classes, preserved via the interaction of key residues that are responsible for the specificity and selectivity.[89] Besides the evolutionary relationship of docking domains, the following summary of known docking domain structures primarily takes structural aspects of classification into account.

The first docking domains of class 1 (Fig. 6a/b) were identified in the modular 6-deoxyerythronolide B synthase (DEBS) and the pikromycin (Pik)-producing multi-protein polyketide synthase.[35,81] The DEBS PKS system consists of three proteins DEBS1, DEBS2 and DEBS3,[90] each of them being homodimers.[91,92] At the very N- and C-terminal ends of these polypeptide chains, *in trans* interacting linker pairs were identified.[93] Broadhurst *et al*.[35] elucidated the structures of these C-/N-terminal elements attached to the last acyl carrier protein (ACP; related in structure and mechanism to T domains from NRPSs[39]) of DEBS2 and the first KS domain of DEBS3. These elements were termed as “docking domains (DDs)”, which functionally connect multi-protein PKSs and in the example of the erythromycin A-synthesizing PKS system as well in a unidirectional way (DEBS1>DEBS2>DEBS3).[94] With regard to the localization of DDs at the C- or N-terminus of a polypeptide chain, these domains are labeled as ^{C}DD or ^{N}DD and even if the term “docking domain” originally referred only to PKS systems[35] whereas the term “communication-mediating (COM) domain” was used for protein-protein interacting regions in NRPSs,[34] in the following the expression “docking domain” is used for both systems.

By the covalent fusion of the DEBS2 ^{C}DD (~90 amino acids) to the DEBS3 ^{N}DD (~35 amino acids), the proteolytic degradation during the protein purification could be minimized and allowed the investigation of this DD complex structure by NMR (Fig. 6a). The NMR solution structure confirmed the homodimeric appearance of this DEBS2/3 $^{C/N}$DD pair, which comprises three α-helices in the DEBS2 ^{C}DDs and one α-helix in the DEBS ^{N}DDs. The overall structure is dominated by two four-helical bundles: the first is formed between α-helices 1 and 2 of the ^{C}DDs, functioning as a dimerization element, and the second arises from the interaction between the α-helices 3 of the DEBS2 ^{C}DDs and α-helices 1 of the DEBS3 ^{N}DDs, arranged in a coiled-coli motif burying a surface area of ~1100 Å^2. Both four-helical bundles are connected via a 14 amino acid long stretch of highly mobile residues. Focusing on the dimerization and docking interface, mainly hydrophobic residues were identified, suggesting that the interaction within each four-helical bundle is dominated by hydrophobic interactions, which is further supported by a few salt bridges formed by charged residues.[35]

Buchholz and colleagues[81] focused on the DEBS related pikromycin multi-protein PKS system, which is composed of four proteins (PikAI–PikAIV), and were able to solve the crystal structure of a further class 1 DD pair (ACPdd/KSdd) connecting PikAIII and PikAIV by fusing them covalently (Fig. 6b). This DD complex structure lacks the previously introduced dimerization motif consisting of the first four-helical bundle and therefore just comprises the second one, composed of ^{C}DDs α-helices 3 and ^{N}DDs α-helices 1, involved in the docking interface. Nonetheless, the coiled-coil packing of the DD pair and the mainly hydrophobic nature of interactions in the docking interface except for of a single salt bridge (D1545 of PikAIII ^{C}DD <> K17 of PikAIV ^{N}DD) was confirmed. Notably, because of the short distance between the fused DDs, the docking interaction in the PikA structure is formed by neighboring dimers whereas in the NMR solution structure the observed docking interaction is intramolecular. Even if the structural architecture looks highly similar, when comparing the docking domain structures of PikAIII/PikAIV and DEBS2/DEBS3, some aspects differ: (I) ^{C}DD helix 3 is significantly shorter (PikAIII (9-residues)/ DEBS2 (15-residues)) and comprises two smaller α-helices (PikA) compared to only one α-helix in the DEBS system and (II) the positioning of matching hydrophobic and polar key residues in the docking interface differs. The latter observation leads to the

assumption that especially mismatching charge-charge interactions prohibit the mixed assembly of PikAII/III and DEBS2/3 polypeptide chains. Besides structural aspects, binding affinities of discrete docking domains, including all ^{C}DDs and ^{N}DDs of the DEBS and PikA PKS systems, were determined via surface plasmon resonance. For all natively matching pairs K_Ds between 70-130 µM were measured, but no binding was detected if non-natively interacting DDs were analyzed.[81]

Class 2 docking domains (Fig. 6c) firstly were identified in a cyanobacterial multi-polypeptide chain PKS system producing curacin A by focusing on the interface between the proteins CurG/CurH and CurK/CurL. Both ^{C}DDs and ^{N}DDs are of similar sequence length (~40 AA), which is too short for the presence of a dimerization region in the ^{C}DD regions as previously described for class 1 DDs. Interestingly, each crystal structure of CurG/CurH $^{C/N}$DD and CurK/CurL $^{C/N}$DD comprises two α-helices connected by a sharp bend, which separates the two interaction regions of the ^{N}DD. On the one hand, ^{N}DD monomers build a parallel coiled-coil dimer via helices 2 and on the other hand, the polar ^{N}DD helices 1 point away from the dimer interface. On the opposite side, ^{C}DD helices 2 bind in parallel orientation to the coil-coil dimer and helix 1 interacts predominantly via hydrophobic interactions, with both helices of one ^{N}DD monomer all in all forming an eight α-helical bundle. This DD arrangement positions the ^{C}DD preceding ACP domain towards the ^{N}DD followed by the KS domain in proximity. In contrast, class 1 DDs arrange the DD attached ACP and KS domains on opposite sides of the DD interaction platform. Nevertheless, also the docking interface in class 2 is specified by hydrophobic as well as electrostatic interactions, burying a surface area of 1834 Å^2. Furthermore, binding affinities for the ACP-^{C}DD <> ^{N}DD-KS-AT complex formation were determined, yielding K_D values of 4-23 µM.[83]

A third class of docking domains (class 3; Fig. 6d–g) was originally identified in a mixed PKS-NRPS system of the myxobacterial strain *Angiococcus disciformis* An d48 which synthesizes tubulysin, by Richter *et al.*[84], focusing on the docking elements connecting the NRPS proteins TubB (^{C}DD) and TubC (^{N}DD). Here, the TubB ^{C}DD comprises 25 and the TubC ^{N}DD 73 residues, but only the TubC ^{N}DD could be solved in their NMR structure elucidation, which forms a homodimer identified by sedimentation velocity analysis in an analytical ultracentrifuge. Interestingly, the TubC ^{N}DD is characterized by an overall αββαα topology, in which a

central β-hairpin, consisting of β-sheet 1 and β-sheet 2, is flanked by three α-helices (α-helix 1–3; Fig. 6d). Additionally, the dimer interface is formed between β-sheet 1 and 2 plus the extended linker and α-helix 2 forming a hydrophobic core with a surface area of ~750 Å^2. In addition, multiple salt bridges within each promoter and between the polypeptide chains were observed, stabilizing the β-hairpin and α-helix 2 *in cis* and the ^{N}DD dimer *in trans*. In contrast to the well-folded TubC ^{N}DD, the unbound TubB ^{C}DD is unfolded - identified by circular dichroism spectroscopy and NMR analysis. Nevertheless, the quantitative analysis of the TubB ^{C}DD binding by ITC determined a K_D of 47 ± 10 µM. Despite the unknown structure of the TubB ^{C}DD binding partner, the $^{C/N}$DD binding interface could be tracked to residues located in the surface-exposed hydrophilic face of the β-hairpins. These residues possessed significant chemical shift changes in NMR titration experiments. The presence of a high number of charged amino acids in this region indicated an “electrostatic code” assuring the binding of the correct docking partner via salt bridges, which was confirmed by the production of an extensive number of arginine-to-alanine TubC ^{N}DD mutants.[84]

A similar αββαα ^{N}DD topology was identified in a pure multi-protein NRPS system producing rhabdopeptide/xenortide-like peptides (RXPs; Fig. 6e)[11] and the enacyloxin-producing mixed PKS-NRPS system (Fig. 6f).[85]

Hacker *et* al.[11] elucidated highly similar ^{N}DD structures by NMR in their three-protein (KJ12A/KJ12B/KJ12C) NRPS system, characterized by specific charged residues localized in β-sheet 2. The covalently linked $^{C/N}$DD complex structure of the KJ12B/KJ12C protein-protein interface clarified the role of these charged residues, which define the “recognition rules” that have to be fulfilled by oppositely charged residues localized in the ^{C}DD binding partner, composed of a single β-sheet, to allow for a functional protein assembly. Remarkably, this salt-bridge guided DD recognition defines the overall arrangement of NRPSs in this system as a unidirectional or repetitive assembly line, leading to a diverse library of peptides. The determination of binding affinities of all possible DD combinations by ITC yielded dissociation constants of 8–100 µM.[11]

In the enacyloxin synthesizing PKS-NRPS system, the ^{C}DD is described as a “short linear motif (SLiM)”, localized within an intrinsically disordered region at the C-terminus of the protein Bamb_5917 in the absence of the Bamb_5915 ^{N}DD (Fig. 6f). In the DD complex, the ^{C}DD interacts as a β-strand with the β-hairpin

formed by the ^{N}DD.[85] Unfortunately, it was not possible to solve the complex structure of Bamb_5917 (PCP-^{C}DD) and Bamb_5915 (^{N}DD-C domain), but NMR titration and carbene footprinting experiments revealed interactions between the DD connected PCP and C domain besides contacts in the docking interface.[85]

A further example of a class 3 docking domain was uncovered in the epothilone PKS-NRPS system (Fig. 6g). Dowling and colleagues[86] analyzed the link between the proteins EpoA and EpoB, where the final EpoA ACP is C-terminally extended by a ^{C}DD and on the opposite side a ^{N}DD is put in front of a cyclization (Cy) domain. The crystal structure of EpoB ^{N}DD-Cy shows that the ^{N}DD is characterized by the class 3 typical αββαα fold and no contacts between the docking domain and the cyclization domain could be identified.[87] Furthermore, it was shown how widespread this type of DD interface is in NRPS and PKS systems, connecting a wide range of catalytic domains besides ACP/PCP and C/KS domains like cyclization domains, condensation/epimerization domains, thioesterases, thioester reductases, halogenases, methyltransferases and other proteins of unknown function.[85]

The fourth class of docking domain structures was detected in the *trans*-acyltransferase (AT) PKS of virginiamycin,[95] consisting of the proteins VirA, VirFG and VirH, by analyzing the docking domain interaction between VirA and VirFG (Fig. 6h). The VirA ^{C}DD was predicted to contain 50 residues after an ACP domain and the VirFG ^{N}DD 70 residues N-terminally attached to a conserved KS domain, forming a heterodimeric complex during interaction with a K_D of 5.8 ± 0.2 µM. In the unbound state, the VirA ^{C}DD tends to form nonspecific dimers and the CD spectrum showed a high content (73 %) of α-helical secondary structure, whereas the VirFG ^{N}DD seemed to be unstructured and was described as intrinsically disordered protein. The covalent artificial linkage of this DD pair (VirA ^{C}DD-VirFG ^{N}DD) enabled elucidating the NMR complex structure. Thereby, a four-helical bundle with a surface area of ~730 Å^2 was observed to which each DD contributes two α-helices (α-helix 1 and 2). These findings identify a structural transition of the VirFG ^{N}DD if its partner VirA ^{C}DD is present. The docking interface is dominated by hydrophobic interactions and only three further electrostatic interactions were determined, whose crucial role in the assembly of the VirA ^{C}DD/VirFG ^{N}DD complex was confirmed by DD mutants that exchanged charged residues for alanine or reversed the charge.[87]

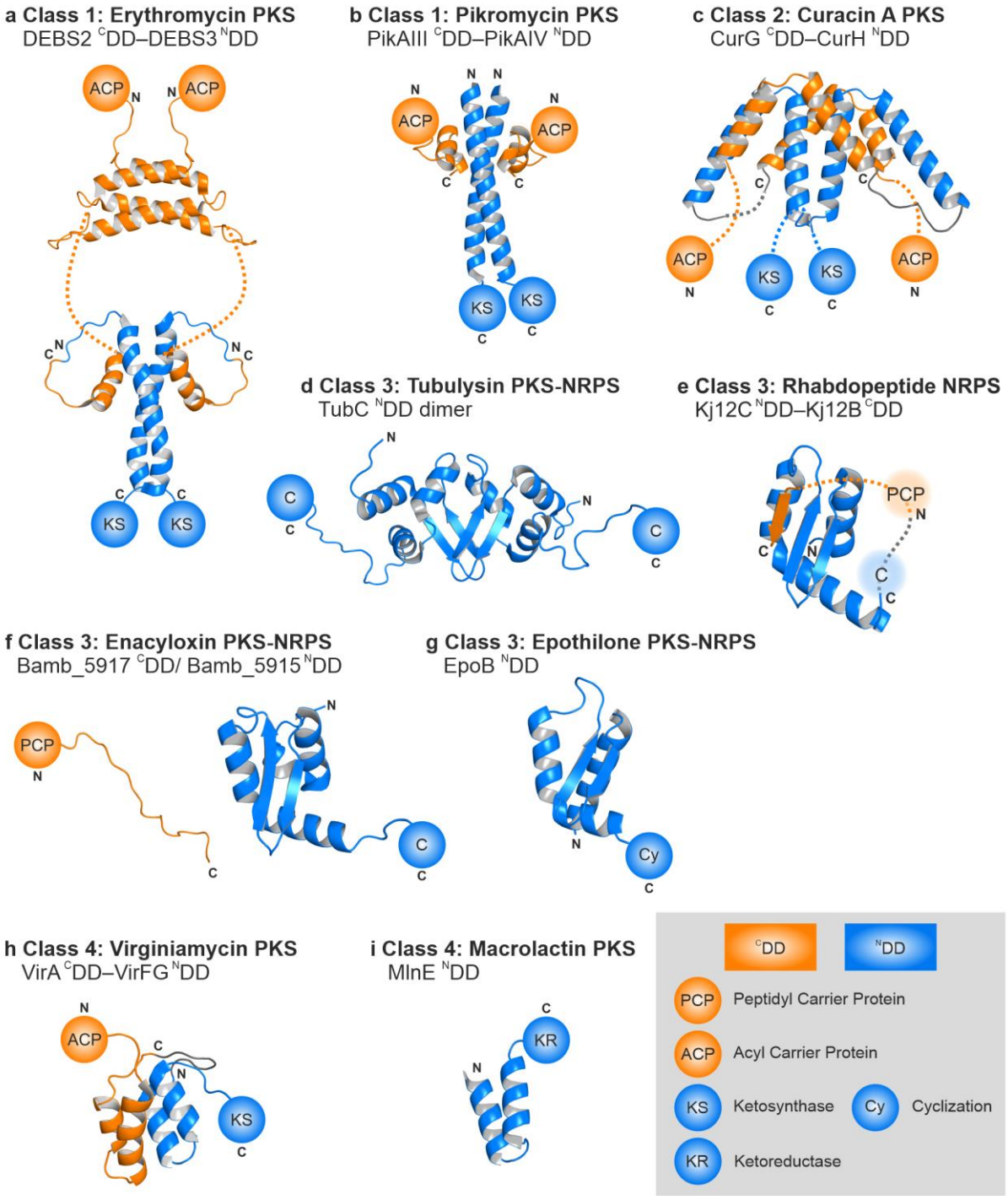


Fig. 6. Structural classes of docking domains identified in modular NRPS and PKS systems. All DDs are shown in the presence of typical up- and downstream flanking domains. The CDDs are colored orange and the NDDs blue. (a–c) Class 1 (PDB: 1PZQ/1PZR; 3F5H)[35,81] and class 2 (PDB: 4MYY)[83] DDs are characterized exclusively for *cis*-AT PKSs, whereas (d–g) Class 3 DDs are found in hybrid PKS-NRPS systems (PDB: 2JUG; 5MTI/6CGO; 5T81)[84–86] (homodimeric NDD) and pure NRPSs (PDB: 6EWV)[11] (monomeric NDD). (h/i) Class 4 DD (PDB: 2N5D; 5D2E)[87,88] structures are described only of *trans*-AT PKSs, in which two α-helices of each C/NDD contribute to the four-helical bundle DD complex.

A structurally similar NDD was solved from the seven polypeptide macrolactin *trans*-AT PKS, analyzing the linkage between the proteins MlnD/MlnE (Fig. 6i). Here, a crystal structure of the N-terminally located MlnE NDD-ketoreductase (KR) domain

was solved, again highlighting the application of similar DD folds in different systems and connecting not only ACP/KS domains but also KS/DH, KS/KR, DH/KR, KR/MT domains.[88]

3. NRPS engineering

3.1. NRPS engineering strategies

The re-engineering of NRPSs is challenging because of different control mechanism established in megasynthetases, for example the gatekeeper function of A domains, the editing function of C domains, and especially the formation of many transient domain-domain interactions during the catalytic cycle. Nonetheless, in the last decades, four major strategies to overcome these barriers and allow the functional re-engineering of NRPSs (Fig. 7) have been developed: (a) A domain exchange, with or without its adjacent T domain, (b) modification of the A domain substrate specificity by site-directed mutagenesis or A subdomain swap, (c) substitution of domain building-blocks comprising C-A, C-A-T, A-T-C, C_{Asub}-A-T-C_{Dsub} domains and (d) docking domain mediated rearrangement of the NRPS assembly line.

The first engineering attempt (Fig. 7a) focuses on the substitution of single A domains or A-T di-domains to activate an alternative substrate. The general success of an A domain exchange was shown in module 2 of the *Pseudomonas aeruginosa* NRPS Pyoverdine (Pvd)D, but only in the case of conservative exchanges of A domains, which activate the native amino acid substrate *L*-Thr, a significant pyoverdine production was observed.[96] In contrast, the A-T di-domain exchange in the surfactin-producing NRPS allowed the incorporation of valine, phenylalanine or cysteine instead of leucine via module 7 in front of the C-terminal TE domain.[97] Furthermore, a most recent study of Calcott *et al*.[98] challenges the dogma that the C-A interface has to be maintained in re-engineering attempts, in order to not disrupt the C domain acceptor site specificity. Here, the successful *in vivo* A domain substitution in the pyoverdine-producing NRPS and the *in vitro* characterization of an A domain exchanged di-domain model system showed good production titers of the desired peptide, leading to the assumption that selecting the correct A domain boundaries in the substitution strategy are more important than maintaining the C-A domain interface.[98]

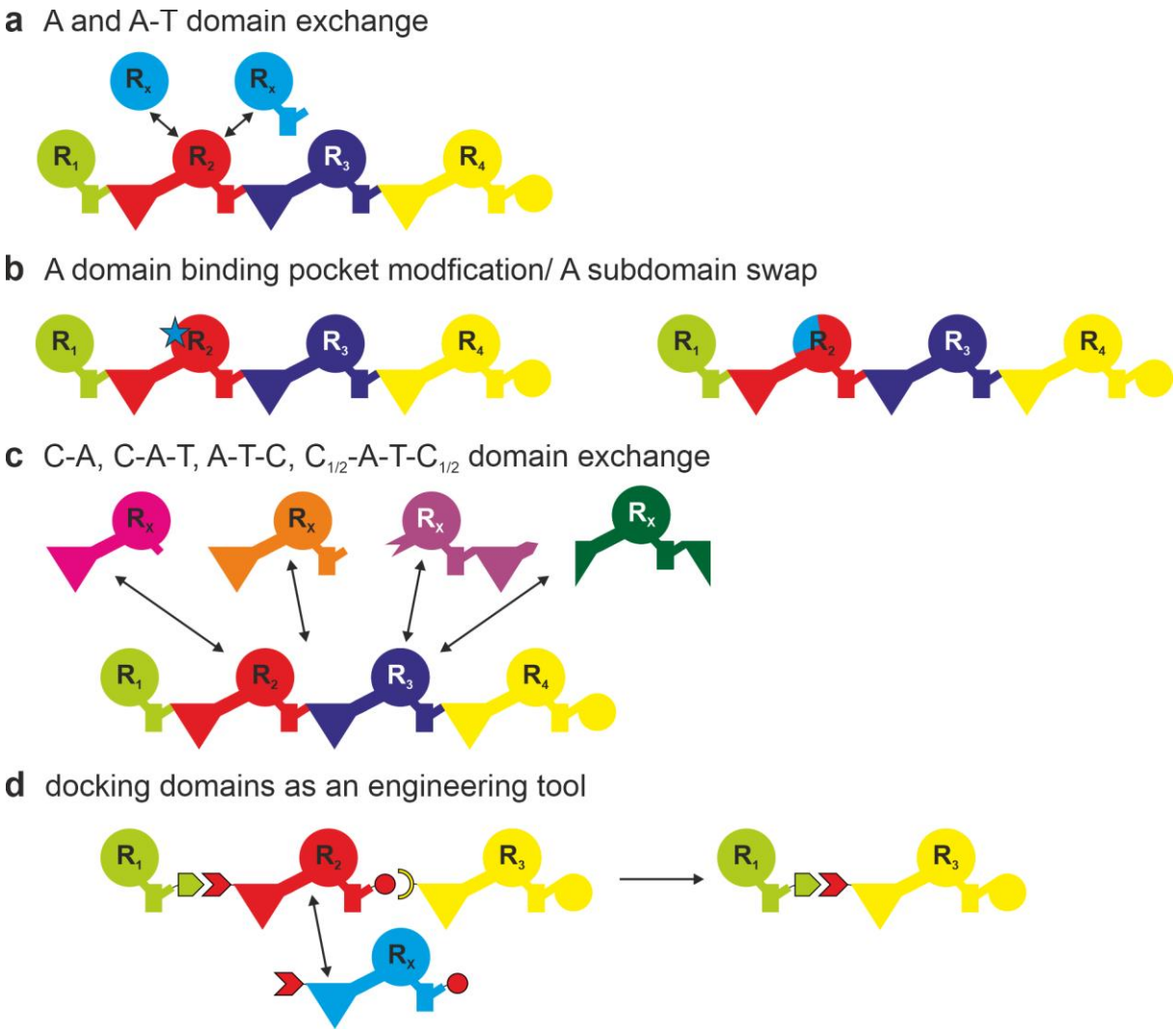


Fig. 7. Toolbox to re-engineer NRPS templates at a genetic level. (a) Exchange of the substrate-specifying A domain, with or without its adjacent T domain. (b) Modification of the A domain substrate specificity by site-directed mutagenesis or A subdomain swap. (c) Strategies that use the substitution of domain building-blocks comprising C-A, C-A-T, A-T-C, C_{Asub}-A-T-C_{Dsub} domains to change the final product composition. (d) Docking domain mediated rearrangement of the NRPS assembly line. See Fig. 2 and Fig. 5 for assignment of the domain symbols.

The second engineering approach (Fig. 7b) is based on the adenylation domain specificity code identified by Stachelhaus *et al*.[49] and has the benefit that amino acid exchanges within the protein structure will most likely not affect the overall structure and, even more important, not influence the NRPS domain-domain interfaces. In separate studies, the A domain binding pocket modification had three main goals: changing the specificity for (I) an alternative substrate[99] (II) an alternative non-natural amino acid and (III) reducing the promiscuous A domain activity. (II) In the calcium-dependent antibiotics (CDA)-producing NRPS, a single mutation of K278Q in the A domain of module 10 alters the A domain specificity from glutamate or 3-methylglutamate to the amino acid glutamine and non-proteinogenic amino acid 3-

methylglutamine.[100] (III) A reduced promiscuous A domain activity was shown for the A domain located in the third module of Plu3262/GxpS, producing luminmides A and B[101]/GameXPeptides (GXP) A and C[102]. Remarkably, the single or double mutants C278M, A301G and I299F/A301G, selected by the comparison of the specificity conferring code of Plu3262 with different A domains, which specifically incorporate Leu, Phe and Val, showed an increased production of luminmide B/GXP C, whereas the Phe-derivative (luminmide A/GXP A) was only produced in trace amounts.[103] A further approach targeting the A domain specificity was established by Crüsemann *et al*.[104], who with bioinformatic methods identified evolutionary recombination sites in A domains of the hormaomycin NRPS, which were further used to create functional chimeric A domains with altered substrate specificity analyzed *in vitro*.[104] This result was confirmed in a proof-of-principle demonstration by Kries and colleagues, who termed this strategy as “subdomain swap”, which yields active and selective A domain chimeras.[105]

The third strategy (Fig. 7c) of successfully exchanging di-domains (C-A)[106] or even whole modules (C-A-T)[107] as an engineering possibility was already shown two decades ago, but two recent publications of Bozhüyük *et al*.[9,10] have paved the way for an unforeseen NRPS recombination potential. Firstly, they proclaimed the e**X**change **U**nit (XU) concept, which uses A-T-C domain units from 15 NRPSs to re-build functional assembly lines of known peptides and even generate chimeric NRPSs that produce new-to-nature peptides in good yield, only limited by the rule that the substrate specificity of the C domain acceptor site has to be met.[10] Secondly, a fusion point in between the two lobes (N-terminal donor/C-terminal acceptor subdomains (C_{Asub}/C_{Dsub})) connecting linker region of the pseudo-dimeric condensation domains was identified, leading to the development of the e**X**change **U**nit **C**ondensation domain (XUC) strategy. This uses C_{Asub}-A-T-C_{Dsub} units to *de novo* design NRPSs with high production yields, but building blocks from different genera (e.g. *Bacillus* and *Photorhabdus*/*Xenorhabdus*) could not be efficiently recombined and the interfaces of C_{Asub}/C/E_{Dsub} units and vice versa were not compatible.

A fourth strategy (Fig. 7d) to re-engineer NRPSs is based on the studies of Hahn *et al*., who identified the previously introduced *in trans* interacting COM domains in multi-protein NRPS systems that select the correct interaction partner via specific key

residues.[34,108] Therefore, similar to DDs that are found in modular PKSs, COM domains also maintain a specific order of synthetases in modular NRPSs. The first attempt to use these COM domain properties *in vivo* focused on the three protein NRPS system of surfactin (SrfA-A/SrfA-B/SrfA-C), where the COM domain pair between the proteins SrfA-A/SrfA-B ($COM^{Donor(D)}_{SrfA\text{-}A}$/$COM^{Acceptor(A)}_{SrfA\text{-}B}$) was targeted. Remarkably, either the exchange of this pair with a matching COM domain pair of the heterologous tyrocidine biosynthetic system (COM^{D}_{TycB}/COM^{A}_{TycC}) or with a cognate COM domain pair $COM^{D}_{SrfA\text{-}B}$/$COM^{A}_{SrfA\text{-}C}$ enabled the functional assembly of this NRPS system. In contrast, the exchange with a heterologous, mismatching pair COM^{D}_{TycA}/COM^{A}_{TycC} led to the skipping of the trimodular SrfA-B resulting in the biosynthesis of a shortened lipopeptide.[109] A further *in vitro* study confirmed that COM domains are suited to form a functional tri-modular NRPS complex (A/B/C) composed of modules derived from the tyrocidine, bacitracin and surfactin NRPSs, which in nature normally do not functionally interact.[108] The *in vitro* installation of compatible linker recognition (docking domain) sequences even allowed the integration of noncognate proteins into the PKS-NRPS epothilone assembly line and an efficient intermodular chain transfer from noncognate proteins could be observed.[110] Furthermore, a more recent study has shown that docking domains are an ideal tool to artificially split the natively single-protein xefoampeptide NRPS system *in vivo* in two or even three independent translated proteins by the insertion of matching DD pairs in between E/C and T/C domains, producing smaller enzymatically active NRPS proteins, which are better suited for the re-engineering of NRPSs. Unfortunately, especially the NRPS split into two proteins by the insertion of matching DDs into the T-C linker region led to significantly reduced peptide production rates.[111]

Additionally, the portability of DDs was shown by Cai *et al.*[112], who transferred DDs from linear type A NRPSs to an iterative type B NRPS to produce a unidirectional NRPS assembly line. This DD exchange alters the normally diverse product spectrum of the rhabdopeptide/xenortide-like peptides (RXPs)-producing NRPSs from a peptide library to a single biosynthesis product. Nonetheless, the replacement of the wild type DDs with naturally compatible DD pairs in some cases tremendously reduced the peptide production level, which suggests that further protein-protein interactions between DD adjacent domains play an important role in a functional NRPS assembly.[112]

Further structural insights into the docking domains installed within the proteins Kj12A/Kj12B/Kj12C identified “recognition rules” of specific key residues in the docking interface, which could be used to reprogram the RXP-synthesis *in vivo* and alter the protein affinity *in vitro*.[11] Especially in PKS systems, native and synthetic docking domains were established as a tool to engineer new biosynthetic machineries.[113–115]

3.2. Characteristics of a potential synthetic docking element

Even if docking domains are structurally diverse in nature and have the indisputable property to select the correct interaction partner in the context of NRPS and PKS systems, consisting of more than one polypeptide chain, they have one major drawback in common: overall low affinities (4–20 µM[11,81,83]). In the cellular context, these weak-affinity, non-covalent interactions between DDs cannot explain how a stable and enzymatically active protein complex is formed between modules, which are distributed across different polypeptide chains. Thus, protein-protein interactions between DD flanking domains most likely contribute to the stabilization of the inter-protein interface in multi-protein NRPSs. This hypothesis would also explain, why the application of native docking domains as a tool to form *de novo* multi-protein NRPS assembly lines, for example by just adding them at the *N*- and *C*-termini of NRPS subunits, is not always highly efficient.[111,112]

Therefore, instead of using natural docking domains in combination with an optimized inter-protein interface, a further idea to recombine NRPS building blocks is to use small (~5 kDa), high-affinity docking elements, which produce a covalent-like connection between modular synthetases. This covalent-like coupling of NRPSs will ideally force them to interact in a functional way.

With regard to structural similarities, α-helical coiled-coils,[116] which are formed by approximately 3–5 % of all amino acids in proteins (Fig. 8a),[117] seem to be an interesting target for the replacement of native docking domains. Typically, coiled-coils are dimers or higher oligomers composed of α-helices wrapped around each other in a left-handed helix and build a supercoil. Luckily, the *de novo* design of such α-helical coiled-coil motifs with high affinities can be accomplished relatively easy using computer-based prediction schemes. The reason for the predictability of primary sequences forming coiled-coil helical bundles is their characteristic interaction termed as ‘knobs-into-holes’ packing, postulated almost 70 years ago,[118]

in which a residue from one helix (knob) packs into a space surrounded by four sidechains of the other helix (hole). Here, in sequences of parallel left-handed coiled-coil proteins, each helix is characterized by a repeating seven amino acid pattern (heptad repeat) with anywhere from two in designed coiled-coils (Fig. 8b/c).[119]

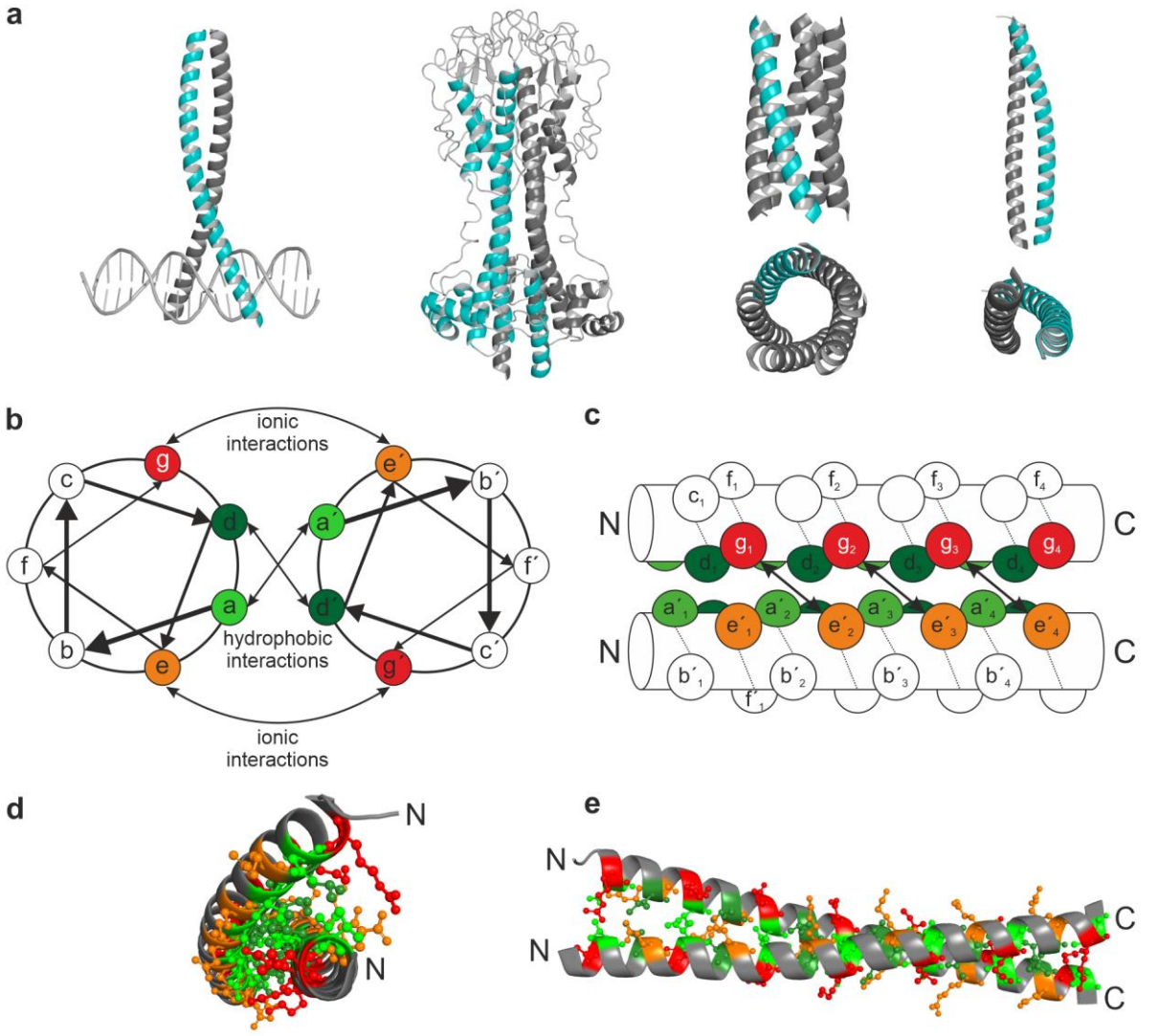


Fig. 8. Structural and functional insights into coiled-coil proteins. (a) Examples of coiled-coil proteins: basic-region leucine zipper (bZIP; (PDB: 1FOS)),[120] *Trypanosoma brucei* surface glycoprotein (PDB: 1VSG),[121] synthetic 5-helical α-barrel (PDB: 4PN8)[122] and synthetic 2-helical heterodimer (PDB: 3HE5).[123] Coiled-coil regions are highlighted in cyan and homo-oligomeric chains are colored dark grey. (b) Front view representation as helical wheel diagram focusing on specific interactions in an exemplarily parallel dimeric coiled-coil looking down the axis from the N-terminus to the C-terminus and (c) as side view representation. The residues are labeled a-g in one helix and a´-g´ in the other while the color coding highlights hydrophobic interactions (green shades; a/a´, d/d´) and ionic interactions (red and orange; e/e´, g/g´) at specific positions in the heptad repeats. (d, e) Ribbon diagram of the x-ray structure of a synthetic leucine zipper (PDB: 3HE5)[123] as front and side view with highlighted residues colored as in b (generated with PyMOL).

The heptad repeat is usually denoted as $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$ in one helix and $(a´\text{-}b´\text{-}c´\text{-}d´\text{-}e´\text{-}f´\text{-}g´)_n$ in the other, both harboring mostly nonpolar residues at the positions a/a´

and d/d´, localized at the interface of the two helices responsible for the formation and stabilization of the coiled-coil, whereas solvent-exposed polar residues are positioned at e/e´ and g/g´.[124] Especially the latter positions e/e´ and g/g´ ensure pairing specificity between helices due to electrostatic interactions. Additionally, the arrangement of α-helices in left-handed coiled-coils leads to distortion induced structural alterations in the helices itself, reducing the normally observed 3.6 residues per turn in α-helices to around 3.5, so heptad repeats occur every two turns of the helix.[116]

Keating *et al*. originally designed anti-bZIP (Fig. 8d/e) peptides by computational techniques that bind specifically to different human basic-region leucine zippers (bZIPs) transcription factors and form parallel coiled-coils ("leucine zipper").[125] The interaction analysis of these synthetic anti-bZIP peptides (SYNZIPs) and human bZIPs comprised 48 synthetic and 7 human bZIP coiled-coils and identified 27 heterodimeric coiled-coil pairs, which mainly interact in parallel (26/27) and generally possess dissociation constants of less than 10 nM. Furthermore, these 27 pairs can be combined in networks of varying topologies with 2–6 nodes to assemble 3 to 6 proteins.[126] Especially orthogonally interacting SYNZIP pairs are ideally suited to connect proteins in a unique way and were already successfully used in the molecular engineering of megasynthases,[115] but so far have not been used in re-engineering strategies of NRPSs.

4. Structure determination of proteins with NMR

4.1. Chemical shift assignment

To obtain structural insights into NRPS docking domains at atomic resolution, nuclear magnetic resonance (NMR) spectroscopy is an ideal method. Because of their small size of below 5 kilodalton, docking domain proteins are ideally suited for NMR whereas they are too small for cryo-EM. Furthermore, DDs often possess flexible protein regions that impair the protein crystallization, which is essential for x-ray crystallography. The focus in this work will be on practical aspects of liquid NMR spectroscopy, specifically, on the application of NMR techniques as a tool to solve molecular structures.

The general procedure to solve at best a molecular structure at atomic resolution by NMR starts with the protein expression, in general in a heterologous host (e.g. *Escherichia coli*[127]), to obtain a uniformly ^{15}N- and ^{15}N, ^{13}C-labeled protein sample.

The presence of nuclei with odd mass numbers is essential for NMR purposes because only these nuclei possess an NMR-necessary nuclear spin angular momentum yielding nuclear magnetism. A further benefit of ^{15}N, ^{13}C-labeled samples is the resolution of ^{1}H resonance overlaps by heteronuclear ^{1}H-^{15}N and ^{1}H-^{13}C correlations.

With these labeled proteins, different NMR experiments are performed to obtain chemical shift values (resonances) of specific atoms in the polypeptide chain. Furthermore, to translate these chemical shifts into structural information, each chemical shift must be unambiguously assigned to the respective atom in the protein, which is described as “assignment process”. This protein NMR resonance assignment process in general consists of two steps: (I) assignment of protein backbone atoms to their respective Hα, Cα, H^N, N and CO resonances and identification of sequentially linked spin systems and (II) assignment of resonances belonging to atoms localized in the aliphatic or aromatic sidechains of the different amino acids. With the assignment of the ^{1}H resonances at hand, ^{1}H,^{1}H distance information can be extracted from nuclear Overhauser effect (NOE) intensities and used for structure calculation by computational methods. Even if these steps look rather simple in theory, in the following the essential but complex NMR experiments for each step are outlined.

The assignment of resonances to the respective amino acids in a protein sequence and finally the assignment of all protons of a protein are essential to finally obtain high-resolution structures. Therefore, triple resonance experiments (Fig. 9) were developed to link resonances intra-residually or to sequentially linked spin systems.[128] All triple resonance experiments are named according to the nuclei they detect and correlate with. For protein backbone assignment, typically a set of NMR experiments is performed, starting here with the most sensitive one: the 3D HNCO.[129] This experiment correlates the H^N and N chemical shifts of one amino acid (*i*) with the carbonyl C´ (CO) of the preceding amino acid (*i*-1), but no sequential information can be identified in this spectrum. Even if the pulse sequences of 3D HNCO and 3D HNCA[129] experiments are identical, except for Cα and C´ specific pulses, only the 3D HNCA experiment correlates H^N(*i*), N(*i*), Cα(*i*) and H^N(*i*), N(*i*), Cα(*i*-1) due to too similar intra- and inter-residual scalar coupling constants (^{1}JCα(*i*),N(*i*) = 11 Hz, ^{2}JCα(*i*-1),N(*i*) = 7 Hz) and allows to extract sequential assignment of backbone

resonances based on Cα chemical shifts. Due to the fact that Cα chemical shift dispersion is often very low, especially in proteins dominated by alpha-helical secondary structures, and because sometimes the differences in signal intensities of intra-residual and sequential cross peaks are rather small, an additional 3D HN(CO)CA[129] experiment can be used to unambiguously assign H^N(*i*), N(*i*), Cα(*i*-1) correlations by abolishing the presence of H^N(*i*), N(*i*), Cα(*i*) cross peaks by correlation of the amide resonances to the preceding Cα and C´(*i*-1). Nonetheless, further three-dimensional experiments are necessary to confirm the backbone assignment. One idea is the sequential linking via carbonyl carbons in a so called 3D HN(CA)CO[130] experiment. The 3D HN(CA)CO experiment correlates H^N(*i*), N(*i*), C`(*i*) and $H^N$(*i*), N(*i*), C`(*i*-1) in analogy to the HNCA experiment but focusses on carbonyl C´ chemical shifts. Also, the pulse sequences of a 3D HNCA and 3D HN(CA)CO are similar, but in an additional step the magnetization is transferred between the Cα and C´ spins. This results in the major drawback of 3D HN(CA)CO experiments, which are lower in sensitivity than the HNCA due to loss in magnetization during a longer magnetization transfer pathway.

Last but not least, 3D HNCACB[131] experiment can be applied, which involves cross peaks for H^N(*i*), N(*i*), Cα(*i*); H^N(*i*), N(*i*), Cα(*i*-1); H^N(*i*), N(*i*), Cβ(*i*) and H^N(*i*), N(*i*), Cβ(*i*-1). The cross peaks representing Cα and Cβ atoms are of opposite sign and enables easily distinguishing the respective resonances. Additionally, sequential cross peaks (H^N(*i*), N(*i*), Cα(*i*-1); H^N(*i*), N(*i*), Cβ(*i*-1)) are significantly weaker. Due to characteristic chemical shift regions for specific Cβ resonances such as the alanine, threonine, and serine Cβ, this experiment allows for an amino acid type specific assignment. Besides the fact that 3D HNCACB experiments are useful in the context of backbone assignment, this experiment already provides resonances of nuclei (Cβ) located in the amino acid sidechain.

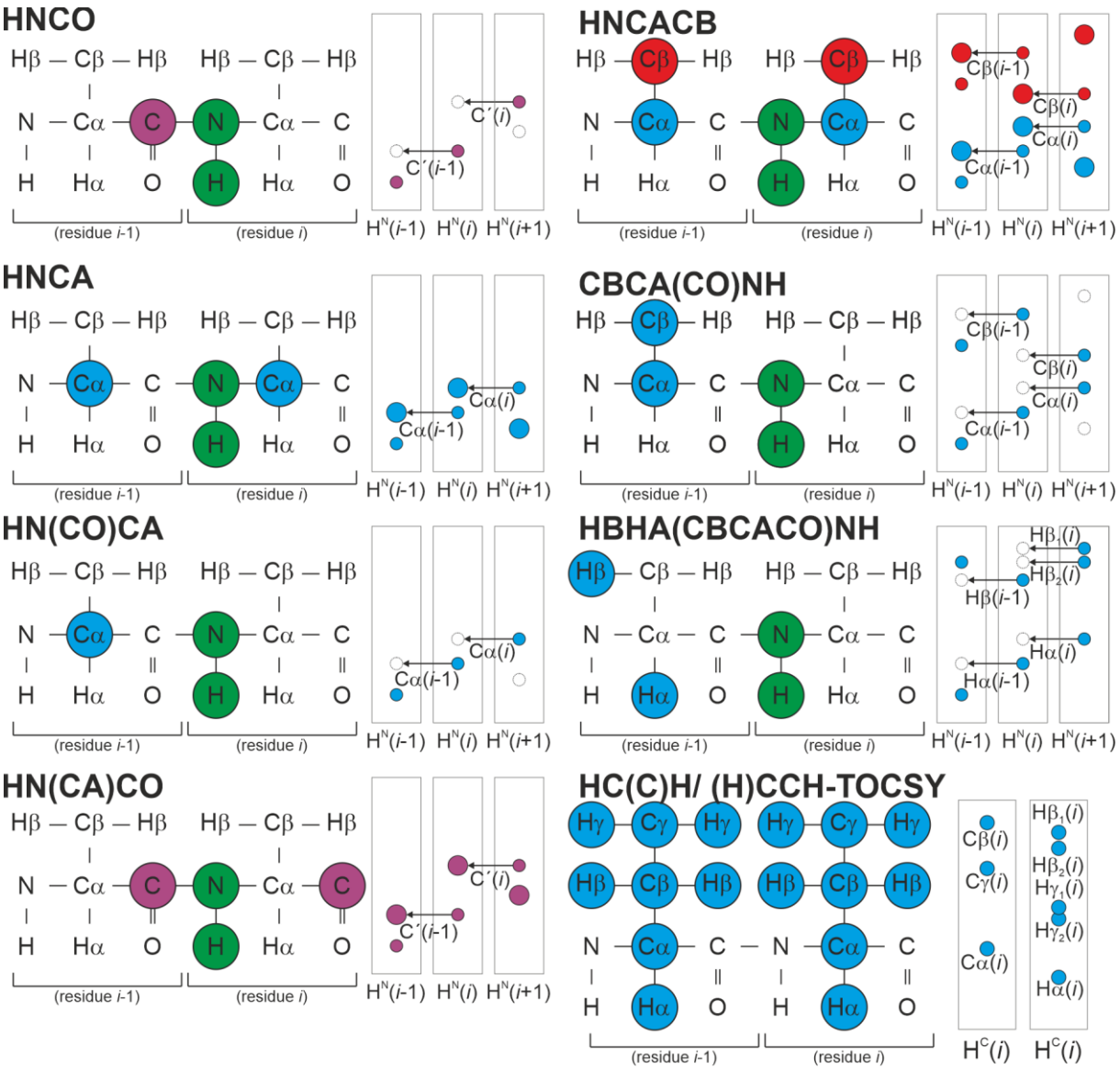


Fig. 9. Overview of triple resonance experiments essential for the backbone and sidechain assignment process. The obtained resonances for each experiment are highlighted and the information content regarding intra-residue and sequential chemical shifts is mentioned in the strip representation. The size of the filled circles in the H^N strips represents the relative NMR signal intensity, which is generally observed in the respective spectra. The blank circles indicate chemical shifts, which are detected in other NMR experiments to reconfirm the respective chemical shifts. The arrows highlight the assignment of sequential chemical shifts.

In summary, the combination of these 3D NMR experiments normally allows the assignment of intra-residue ^{13}C chemical shifts and sequential chemical shift values, which are linked to specific NH groups. The comparison of these ^{13}C cross peaks, which are associated to different NH groups, allows the identification of sequentially linked spin systems and the assignment to the primary protein sequence.

In a second step, the assignment of resonances has to be expanded from backbone to sidechain atoms, starting with nuclei located in aliphatic amino acid side chains. To

this end, 3D CBCA(CO)NH[132] and 3D HBHA(CBCACO)NH[133] experiments are applied, which correlate $H^N(i)$, N(*i*), Cα(*i*-1), Cβ(*i*-1) or $H^N(i)$, N(*i*), Hα(*i*-1), Hβ(*i*-1) chemical shifts, respectively. The latter triple resonance experiment is obtained from 3D CBCA(CO)NH experiment, but the chemical shift evolution of carbon atoms is replaced by that of aliphatic protons. Due to the fact that the magnetization transfer in triple resonance experiments from sidechain carbon atoms, especially beyond Cβ nuclei, to backbone amide protons become insensitive, especially with regard to an increasing molecular weight, it is often difficult to record such spectra efficiently. Therefore, HC(C)H-/(H)CCH-total correlation spectroscopy (TOCSY) experiments are preferred, which connect side chain chemical shifts of protons or carbons to the respective spin systems on basis of sequentially assigned Hα/β and Cα/β chemical shifts, which were previously isolated from 3D CBCA(CO)NH, 3D HNCACB and 3D HBHA(CBCACO)NH experiments. Further proton assignments, for example of asparagine Hδ-Nδ and glutamine Hε-Nε sidechain amino group resonances or protons connected to carbons in aromatic ring systems of histidine, tryptophan or phenylalanine, can often be assigned via $^{13}C_{aromatic}$- or ^{15}N-NOESY-heteronuclear single quantum correlation (HSQC) spectra.

The assignment of proton resonances to specific sidechain atoms via NOESY-experiments is based on the dipole-dipole cross-relaxation to correlate ^1H spins that are close in space. More importantly, ^{15}N-/^{13}C-NOESY-HSQC experiments are essential for the third step in structure elucidation: the isolation of ^1H, ^1H distance restraints derived from nuclear spin polarization from one nuclear spin population to another via dipole-dipole cross relaxation,[134] the so-called nuclear Overhauser effect. Hereby, two protons that are separated by less than 5 Å generate a NOE and the intensity of this peak is dependent on the ^1H-^1H distance with stronger intensities for shorter distances. Many of such NOE peaks, each of them representing one structural restraint of the molecule, are necessary to calculate a three-dimensional structure. How a structure is calculated from this pool of distance information by computational methods is described in the following.

4.2. Calculation of structures from NMR restraints

The raw data of a structure calculation consist of a given chemical shift assignment and most likely unassigned lists of NOESY peaks defined by their positions and intensities. I focus here on the automated NMR structure calculation with the program

CYANA,[135] which combines the methods of working in torsion angle space,[136] relying on the knowledge that the protein conformation is uniquely defined by the sum of all torsion angles values, and using simulated annealing[137] by molecular dynamics simulation allowing the system to escape from local minima. Additionally, the automated analysis of NOESY spectra with regard to the assignment of NOE peaks is realized in CYANA with help of the CANDID algorithm.[138] For this purpose, the assignment of the backbone and side-chain chemical shifts has to be as complete as possible.[139] The iterative structure calculation procedure of CYANA comprises seven cycles of automated NOE assignment and structure calculation, in the end leading to a determined structure with unambiguously assigned distance restraints. Every round of this iterative process optimizes the NOE assignment of the previous step, resulting in a structure represented by a bundle of conformers. This bundle of structures is characterized by two criteria: first, the RMSD describing the similarity of the atomic coordinates in all structures expressed in Å and second, the target function, representing how accurately the structural conformation is fulfilling the NOE defined distance restraints. Both values are ideally small, generating a homogenous bundle of reliable structures in good agreement with the distance restraints.

4.3. Chemical shift based secondary structure determination

Worth mentioning in the context of structural aspects is the knowledge that specific chemical shifts characterize regions of defined secondary structures in proteins,[140] which can be extracted of chemical shift data in general. This observation enabled the correlation of chemical shifts and local geometry, especially the backbone torsion angles Φ and Ψ by for example the programs TALOS[141], TALOS+[142] and TALOS-N[143]. The first uses a database of assigned proteins with known structure and just compares tripeptides with similar chemical shifts and sequences for its prediction, whereas the latter ones are optimized with an artificial neural network (ANN) component, leading to more accurate predictions. In addition to the older TALOS versions, TALOS-N is also able to predict sidechain X^1 angle information with the help of the ANN component, knowing that X^1 is connected particularly to backbone N^H and Cα chemical shifts.[144] In summary, TALOS-N is able to predict backbone Φ/Ψ torsion and sidechain X^1 torsion angles and secondary structures using the chemical shifts assigned to H^N, Hα, Cα, Cβ, CO, N atoms of a given protein primary sequence as input. The graphical representation of a secondary structure prediction is defined

as chemical shift index (CSI),[145] in which the secondary structure probability is plotted against the protein sequence.

5. Overview and aim of this thesis

The structural knowledge of domains and modules that fully interact *in cis* has constantly been enlarged in the last decade, but the structural insights into *trans* interfaces of multi-protein systems is limited so far. Therefore, the aim of this thesis was to reveal further aspects of the underlying mechanism in the contact surfaces between proteins of multi-protein systems by focusing especially on docking domains and their flanking domains. Additionally, the NRPS re-engineering potential of synthetic "docking domains" in form of synthetic zippers was evaluated.

In chapter I the mechanism and structural knowledge of NRPSs is summarized, especially highlighting the structural aspects of already published DD structures. Furthermore, a brief overview is given about actual NRPS engineering strategies and the process, how a NMR solution structure or the protein secondary structure are obtained from NMR raw data.

In chapter II all publications and manuscripts that contribute to this thesis, focusing on three major aspects of NRPS docking domains are listed. First, the structural knowledge of stand-alone DD complexes is expanded by solution NMR ("A new docking domain type in the Peptide-Antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*"), second, the use of synthetic zippers as a tool to functionally recombine synthetases of different NRPS systems is established ("Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases") and third, an extended NRPS docking interface reported for the first time is structurally characterized by solution NMR ("NMR resonance assignments for a docking domain pair with an attached thiolation domain from the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii*" and "Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS".

In chapter III, some additional preliminary structural NMR results are described, which deal with the question if the DD interface between the proteins PaxB and PaxC of the three protein PaxS NRPS from *Xenorhabdus bovienii* SS-2004 highlighted in publication 1 is also extended by the ^{C}DD flanking T domain.

Chapter IV discusses the classification of the solved PaxB ^{C}DD:PaxC ^{N}DD complex structure with regard to known DD structures and compares the newly developed synthetic zipper based NRPS engineering tool to other favorable strategies that allow the exchange of modules to generate novel chemical derivatives. Furthermore, the extended docking interface is considered in context of the overall NRPS domain and module architecture by generating a model structure that unveils a possible ideal fusion point to use this extended interface as another NRPS engineering possibility. In this context also the results of the possibly extended PaxB/PaxC docking interface are discussed.

The PhD thesis closes with the list of references and an attachment which includes all publications/manuscripts and supporting information, a curriculum vitae of the PhD student, the list of publications and manuscripts, a record of conferences, and finally the declaration and affidavit.

II. PUBLICATIONS AND MANUSCRIPTS

1. A new docking domain type in the Peptide-Antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*

Jonas Watzel[1], Carolin Hacker[2], Elke Duchardt-Ferner[2], Helge B. Bode*[3,4,5], Jens Wöhnert*[6]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

2 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

3 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. h.bode@bio.uni-frankfurt.de.

4 Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

5 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

6 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. woehnert@bio.uni-frankfurt.de.

* corresponding authors

Status: **published**

Published in: *ACS Chem. Biol.* **15,** 982–989 (2020)

Digital object identifier: 10.1021/acschembio.9b01022

Attachments: Declaration of the author contributions and the publication including the supporting information.

2. Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases

Kenan A. J. Bozhueyuek[1], Jonas Watzel[1], Nadya Abbood[1], Helge B. Bode*[1,2,3]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. h.bode@bio.uni-frankfurt.de.

2 Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

3 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

* corresponding author

Status: **in revision**

Digital object identifier: 10.1101/2020.05.06.080655

Attachments: Declaration of the author contributions and the manuscript including the supporting information.

3. NMR resonance assignments for a docking domain pair with an attached thiolation domain from the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii*

Jonas Watzel[1], Sepas Sarawi[1,2], Elke Duchardt-Ferner[2], Helge B. Bode[1,3,4,5], Jens Wöhnert*[2]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

2 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. woehnert@bio.uni-frankfurt.de.

3 Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

4 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany.

5 Max-Planck-Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043, Marburg, Germany.

* corresponding author

Status:	**published**
Published in:	*Biomol. NMR Assignments* (2021)
Digital object identifier:	10.1007/s12104-021-10010-1
Attachments:	Declaration of the author contributions and the publication including the supporting information.

4. Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS

Jonas Watzel[1], Elke Duchardt-Ferner[2], Sepas Sarawi[1,2], Helge B. Bode*[1,3,4], Jens Wöhnert*[2]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

2 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. woehnert@bio.uni-frankfurt.de.

3 Max-Planck-Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043, Marburg, Germany. helge.bode@mpi-marburg.mpg.de

4 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

* corresponding authors

Status: **submitted**

Attachments: Declaration of the author contributions and the manuscript including the supporting information.

III. ADDITIONAL RESULTS

1. Structural investigation of the extended PaxB/C docking domain interface in the PAX peptide-producing NRPS from *Xenorhabdus bovienii*

The additional results of this thesis focus again on the PAX peptide-producing NRPS from *Xenorhabdus bovienii* SS-2004. In extension to publication 1, this work includes the ^{C}DD adjacent T domain into the analysis and evaluates if the docking domain interface is also extended between the proteins PaxB and PaxC (Fig. 10a).

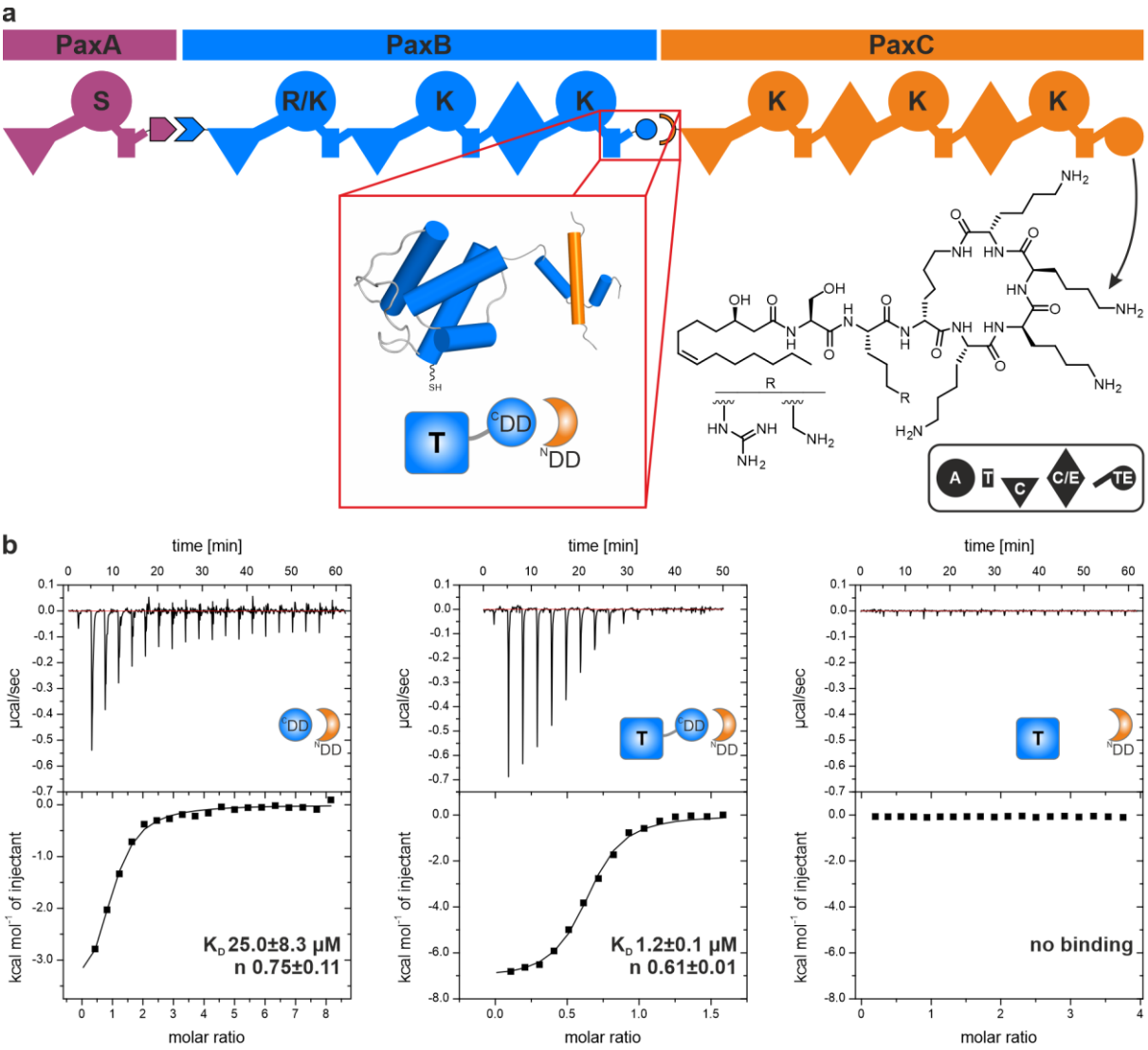


Fig. 10. (a) Schematic representation of the peptide-antimicrobial-*Xenorhabdus* (PAX) peptide-producing NRPS from bacteria of the genus *Xenorhabdus* with focus on the *trans* docking interface between the proteins PaxB and PaxC. (b) ITC thermogram and the derived binding curve for titration between PaxB ^{C}DD:PaxC ^{N}DD, PaxB T$_4$-^{C}DD:PaxC ^{N}DD and PaxB T$_4$:PaxC ^{N}DD are shown (n=3).

In a first step, the interaction between the discrete docking domains PaxB ^{C}DD and PaxC ^{N}DD, the PaxB T$_4$-^{C}DD and PaxC ^{N}DD and the PaxB T$_4$ domain without the ^{C}DD and PaxC ^{N}DD (Supporting Fig. 1a/b) were analyzed by ITC (Fig. 10b,

Supporting Fig. 2). For excised docking domain interaction, a K_D value of 25.0±8.3 µM was determined, but addition of the C-terminal T domain of PaxB leads to a ~25-fold increased affinity (1.2±0.1 µM), whereas no binding was observed between the PaxB T_4 domain and the PaxC ^{N}DD.

This observation should be structurally evaluated by solution NMR and therefore, a backbone assignment of the unbound PaxB T_4-^{C}DD was performed (Fig. 11).

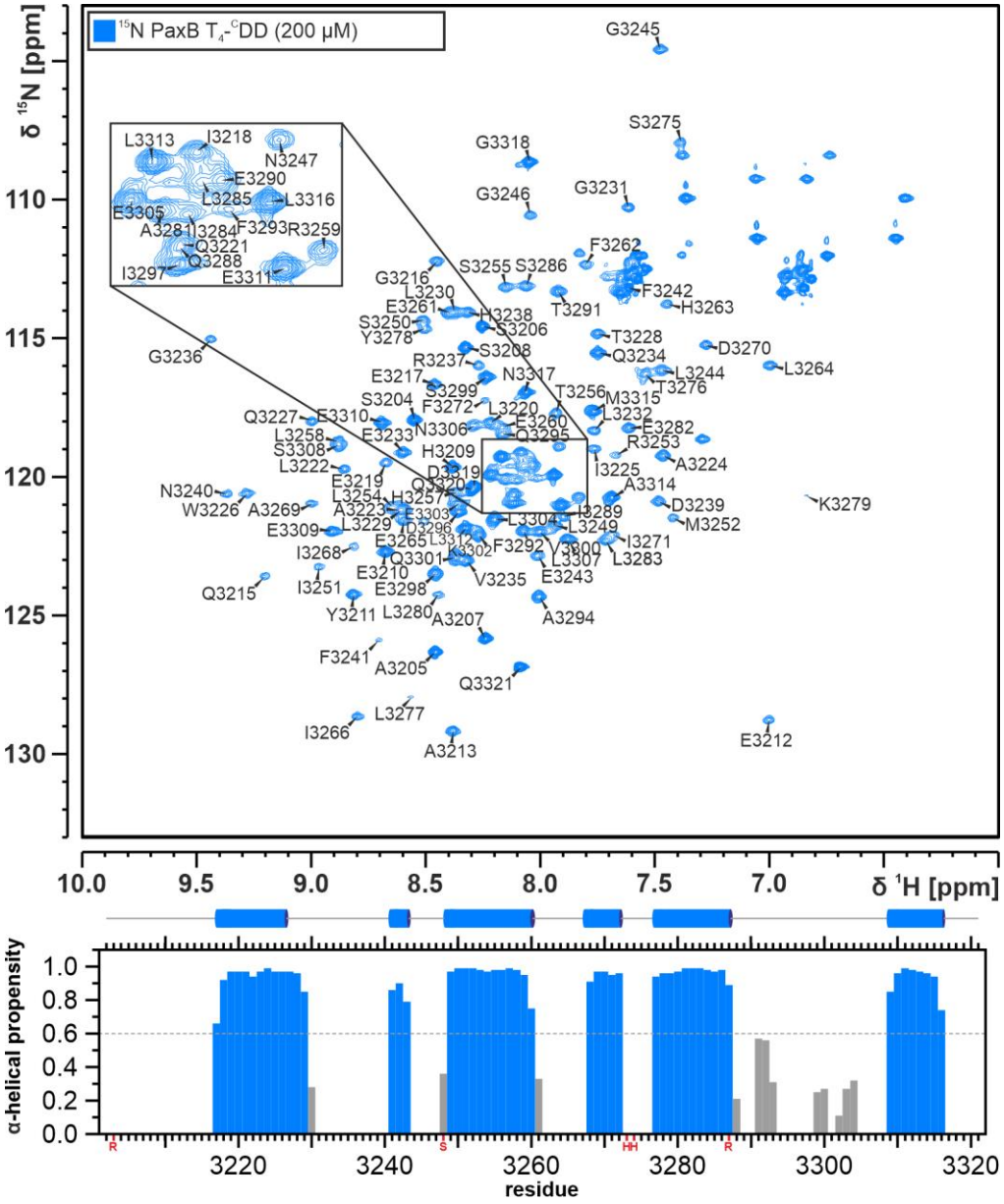


Fig. 11. (a) Assigned ^{1}H,^{15}N-HSQC spectrum (top) of the PaxB T_4-^{C}DD from *Xenorhabdus bovienii* SS-2004 in its unbound state. Below, the TALOS-N-derived chemical shift index is plotted onto the sequence and the predicted secondary structure (confidence value ≥ 0.6) elements according to TALOS-N are depicted on top. Red-labeled residues could not be assigned.

The protein PaxB T_4-^{C}DD with a molecular weight of 13.7 kDa is composed of 120 amino acid residues and in its ^{1}H,^{15}N-HSQC spectrum most of the backbone amide

signals are well dispersed, but some of them partially overlap. 117 peptide bond NH-signals are expected in the $^{1}H,^{15}N$-HSQC spectrum, excluding the two proline residues (P3214, P3267) and the N-terminal aspartate (D3202) residue, but only 112 backbone amide signals could be assigned (112/117), 95.7 %, Fig. 11 (top)) to the respective residues. Furthermore, 97.5 % of all Cα (117/120) and 94.2 % of all CO (113/120) chemical shifts were assigned. The backbone assignment was confirmed by amino acid selective labelling (^{15}N) of phenylalanine and leucine (Supporting Fig. 3). To get insights into the secondary structure of the PaxB T_4-^{C}DD in the absence of the PaxC ^{N}DD, a prediction was performed using TALOS-N (Fig. 11 (bottom) and six amino acid stretches with high (≥ 0.6) α-helical propensity were identified. Here, five out of six α-helices are positioned within the T domain, whereas the sixth α-helix lies at the C-terminus in the ^{C}DD region.

Also, the backbone of the PaxB T_4-^{C}DD bound to the PaxC ^{N}DD was assigned (Fig. 12) with a completeness of 88.9 % for all H^{N}/N (104/117), 94.2 % for all Cα (113/120) and 88.3 % for all CO (106/120) chemical shifts. Based on these chemical shift assignments, the secondary structure composition of the bound PaxB T_4-^{C}DD was determined by TALOS-N. The interaction with the PaxC ^{N}DD induces the formation of an additional α-helix in the ^{C}DD area, whereas the other secondary structure elements, defined for the non-interacting state of the PaxB T_4-^{C}DD as well, remain unaffected. In summary, residues of the PaxB T_4 domain comprise five and of the PaxB ^{C}DD two regions with high α-helical propensity (≥ 0.6).

To further investigate the PaxB T_4-^{C}DD:PaxC ^{N}DD interaction, NMR titration experiments were carried out. Therefore, the unlabeled (^{14}N) PaxC ^{N}DD was added to the ^{15}N-labeled PaxB T_4-^{C}DD in a stepwise manner and evaluated in $^{1}H,^{15}N$-HSQC experiments (Fig. 12a). In all six titration experiments, gradual chemical shift changes and/or peak broadening during the stepwise addition of the ^{N}DD peptide were detected. These findings confirm on the one hand the interaction between both proteins in general and on the other hand the observed K_D value (1.2±0.1 µM), which is in good agreement with the fast to intermediate exchange regime observed in the NMR titrations.

The quantified chemical shift changes of each residue were plotted onto the PaxB T_4-^{C}DD sequence, identifying two regions that are predominantly affected by the complex formation with the PaxC ^{N}DD (Fig. 12b).

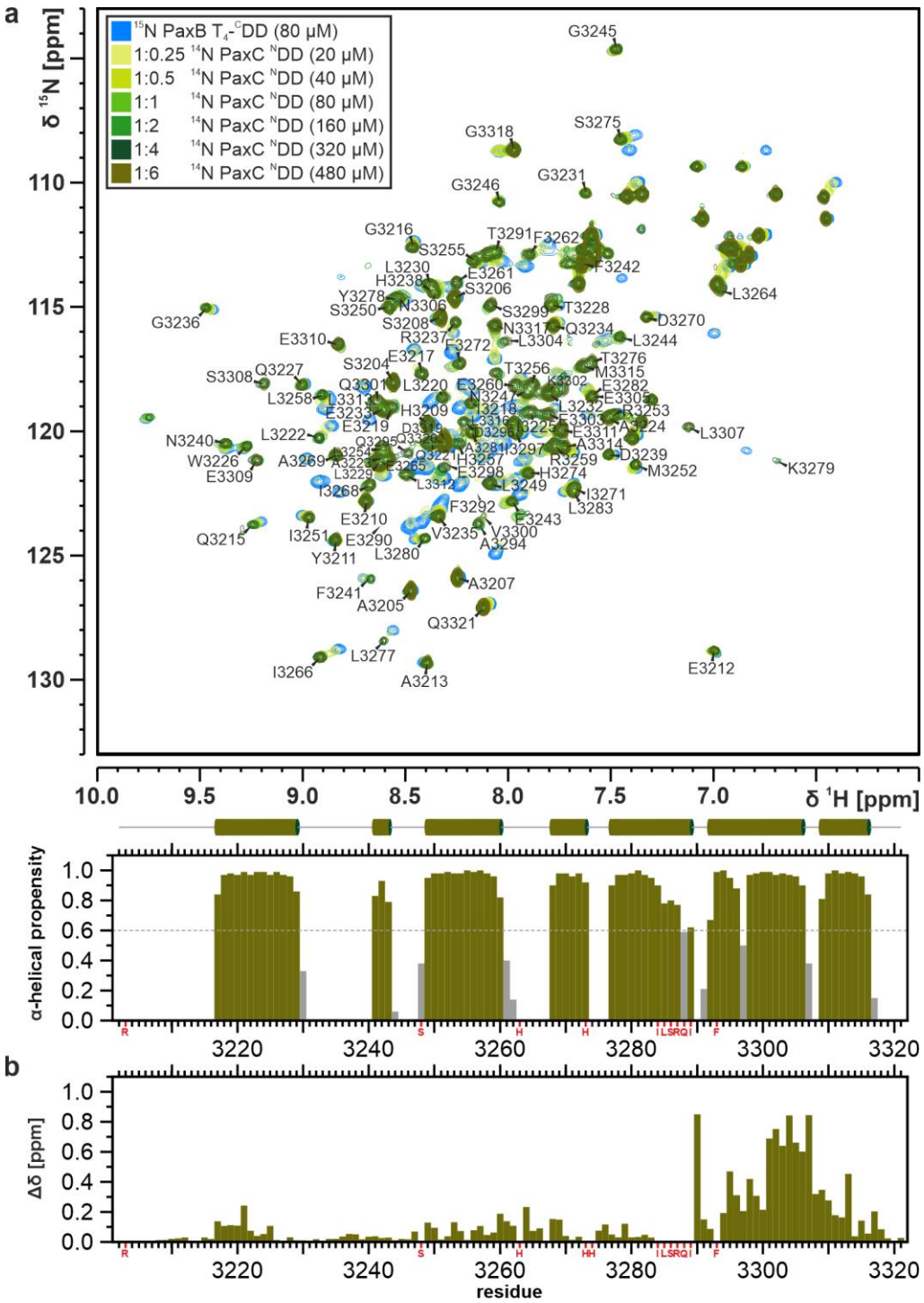


Fig. 12. (a) Overlay of $^{1}H,^{15}N$-HSQC spectra (top) of PaxB T_4-^{C}DD from *Xenorhabdus bovienii* SS-2004 in the absence (blue) and presence (different green shades) of increasing amounts of unlabeled PaxC ^{N}DD. The molar ratios of the two interaction partners are 1:0.25, 1:0.5, 1:1, 1:2, 1:4, 1:6. The assignment is given for the PaxB T_4-^{C}DD, bound in a stable PaxB T_4-^{C}DD:PaxC ^{N}DD complex. Below, the TALOS-N-derived chemical shift index is plotted onto the sequence and the predicted secondary structure (confidence value ≥ 0.6) elements according to TALOS-N are depicted on top. (b) Chemical shift changes observed upon complex formation of unambiguously assigned residues plotted onto the sequence. Red-labeled residues could not be assigned either in the unbound or bound state.

The first region, which possesses one of the largest chemical shift changes is the ^{C}DD, in which an additional α-helix is formed upon binding to the ^{N}DD and the second region is the C-terminal α-helix of the T_4 domain. Even if the backbone assignment of

these residues is incomplete, their obvious chemical shift alterations regarding the observed and assigned chemical shifts in the unbound state, indicate an interaction induced effect.

To get insights into the tertiary structure of the PaxB T_4-^{C}DD bound to the PaxC ^{N}DD, a chemical-shift-ROSETTA (CS-ROSETTA)[146] run was performed, which relies on the experimentally determined NMR chemical shifts to calculate the three-dimensional protein structure. Even if the CS-ROSETTA run did not converge (Fig. 13b), meaning that the selected structure is not necessarily representative of the ensemble of structures formed, it gives an idea of the α-helices three-dimensional arrangement.

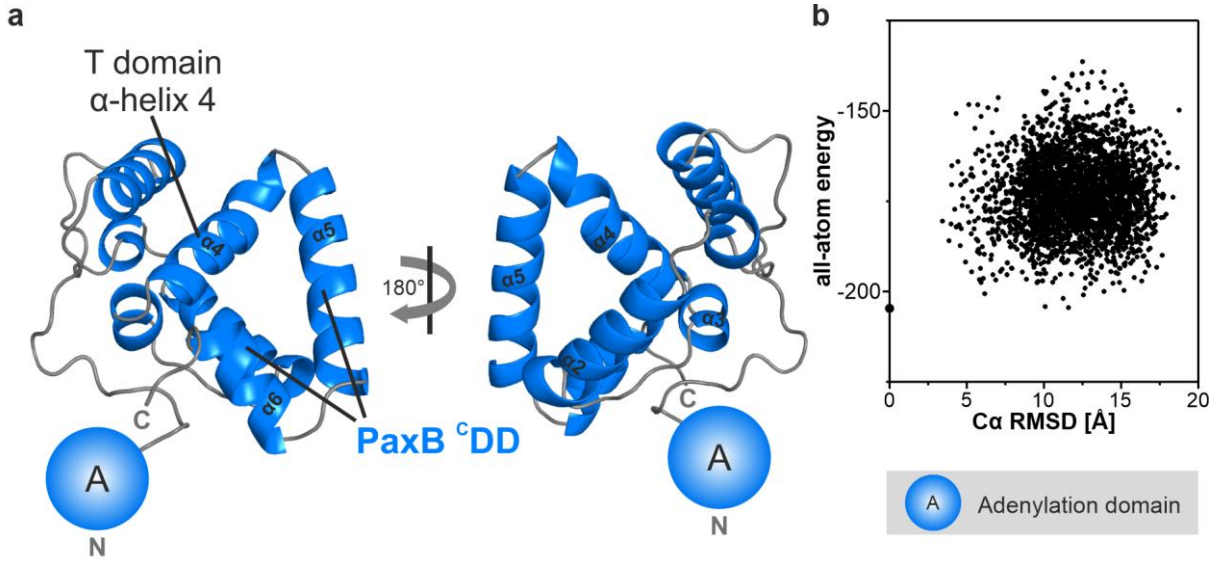


Fig. 13. (a) Lowest-energy CS-ROSETTA model (Rosetta version 3.8, CS-Rosetta Toolkit version 3.3) of bound PaxB T_4-^{C}DD. (b) Plot of ROSETTA all-atom energy versus Cα RMSD relative to the lowest-energy model (bold dot on vertical axis).

In addition, the chemical shifts and TALOS-N secondary structure predictions of the PaxB ^{C}DD residues in the PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD[82] and the PaxB T_4-^{C}DD (bound state) constructs were compared (Fig. 14a/b).

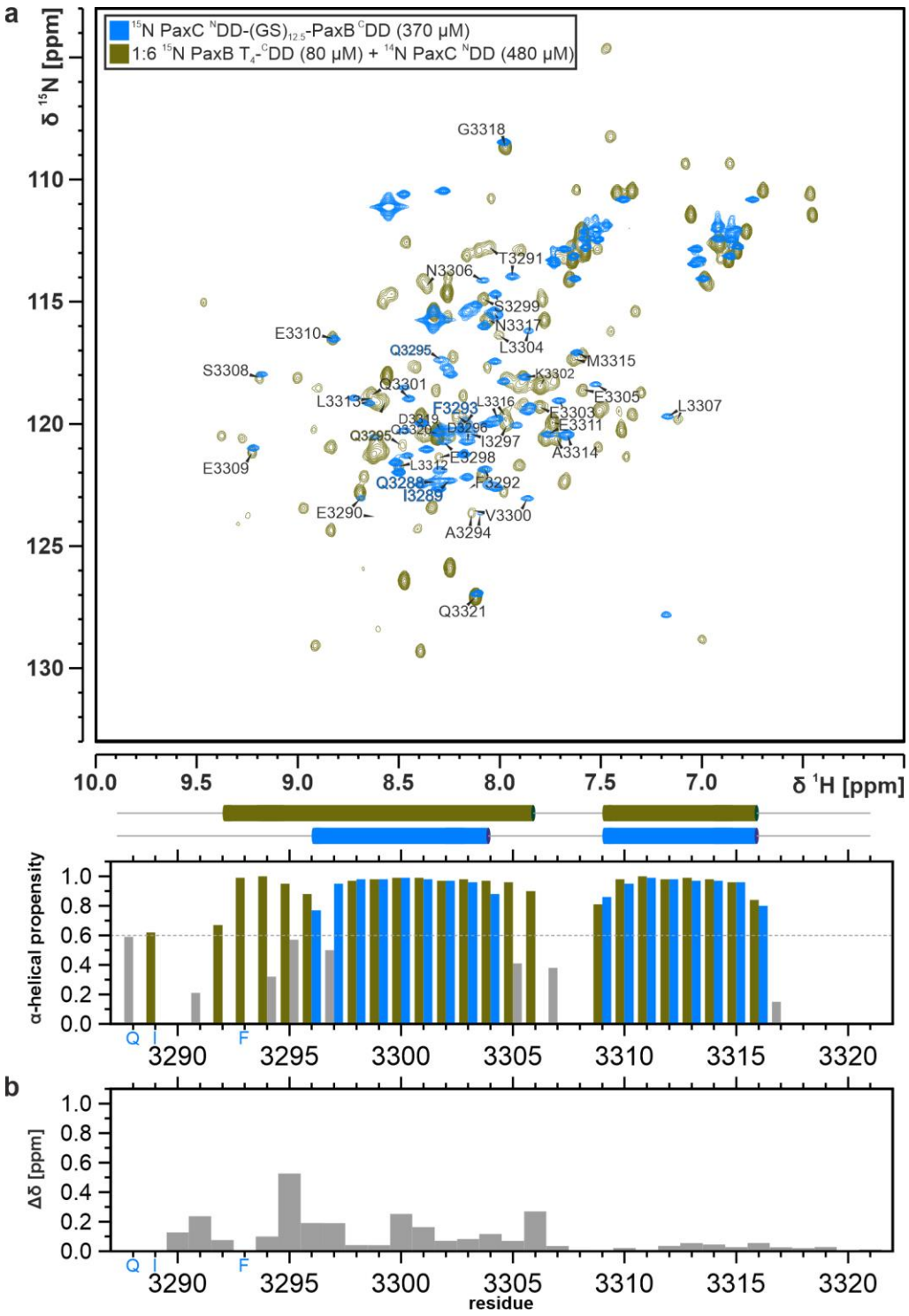


Fig. 14. (a) Overlay of ^{1}H,^{15}N-HSQC spectra (top) of the PaxB T_4-^{C}DD bound to the PaxC ^{N}DD (green) and the covalently linked PaxB/C $^{C/N}$DD complex (blue). Below, the TALOS-N-derived chemical shift index is plotted onto the ^{C}DD sequence and the predicted secondary structure (confidence value ≥ 0.6) of both constructs according to TALOS-N is depicted on top. (b) Chemical shift differences determined for ^{C}DD residues of the proteins PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD and PaxB T_4-^{C}DD (bound state) plotted onto the sequence. Blue-labeled residues could only be assigned for the artificially linked DD complex.

Only minor chemical shift differences of ^{C}DD residues could be defined between the PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD and bound PaxB T_4-^{C}DD constructs, but most likely,

the first PaxB ^{C}DD α-helix is extended at its C- and N-terminus if the preceding T_4 domain is present during the DD interaction.

In a final set of experiments, the effect of the T_4 domain C-terminal α-helix 4 (Supporting Fig. 1a) and the ^{N}DD flanking C_5 domain on the affinity of the overall docking interaction was evaluated in ITC experiments (Fig. 15, Supporting Fig. 4). The attachment of the T_4 domain α-helix 4 to the ^{C}DD increased the binding affinity by factor of ~2 (12.5±1.8 µM), relative to that of the wild type DD interaction (25.0±8.3 µM), whereas the binding affinity between the proteins PaxB T_4-^{C}DD and PaxC ^{N}DD-C_5 (1.4±0.1 µM) is identical to the K_D value of the PaxB T_4-^{C}DD:PaxC ^{N}DD interaction (1.2±0.1 µM).

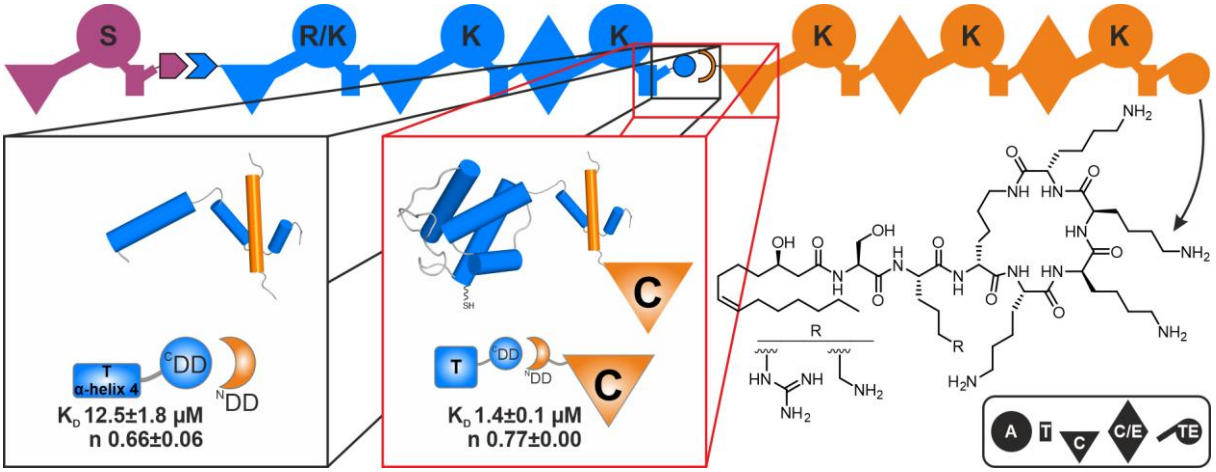


Fig. 15. ITC analysis (n=3) to prove if the T_4 domain α-helix 4 or the ^{N}DD adjacent C_5 domain contribute to the *trans* docking interface between the proteins PaxB and PaxC of PaxS from *Xenorhabdus bovienii* SS-2004.

IV. DISCUSSION

1. Protein-protein interactions in multi-protein NRPSs

As NRPSs, that consist of a single polypeptide chain and in which all modules are covalently connected via linker regions, also multi-protein NRPSs must maintain a defined order during the biosynthesis of natural products. In these systems, multiple modules are distributed across several gigantic polypeptides and only if the correct order of these proteins is achieved, the biosynthesis can take place. The correct order is achieved via high-specificity interactions of short elements originally termed as DDs[35] in PKS and mixed PKS-NRPS systems or as COM domains in NRPSs,[34] which are located at the extreme termini of the different polypeptide chains. With respect to their absolute position at the C- or N-terminus of the protein, they are further described as ^{C}DD or ^{N}DD, respectively.

These DDs prohibit the unnatural protein ordering due to two mechanisms that maintain selectivity and specificity and guarantee that DDs only interact with their respective native partner. Firstly, only structurally compatible DDs can interact and numerous studies have identified multiple types of DDs that significantly differ in their overall fold and secondary structure composition.[11,35,81,83,84] Secondly, strategically positioned key residues in the interaction surface of DDs allow them to select the correct binding partner via specific interactions.[11,89,108]

In the last decades, efforts in several laboratories have uncovered numerous structures of DDs from NRPS, PKS and mixed PKS-NRPS systems, which were further classified with respect to their overall structural appearance (see introduction chapter 2.4). Furthermore, some studies have evaluated the underlying mechanism in all protein interfaces of multi-protein systems with respect to a structure- or key residue interaction-based DD selectivity.[11,83,147]

In the following, the DD structures of PaxS are compared to known DD classes and an explanation is given for how the communication across the different protein interfaces of PaxS (PaxA/B; PaxB/C) is maintained.

1.1. Structural and mechanistical classification of PaxS DDs

The PAX peptide-producing NRPSs from bacteria of the genus *Xenorhabdus* consist of the three proteins PaxA, PaxB and PaxC, which interact in a unidirectional way

(PaxA>PaxB>PaxC) and comprise two *in trans* interacting docking domain pairs (PaxA ^{C}DD:PaxB ^{N}DD; PaxB ^{C}DD:PaxC ^{N}DD).

Initially, putative DDs with partially different sequence lengths were identified by bioinformatic methods at the extreme N-termini of PaxB (~30 AAs) and PaxC (~33 AAs) and C-termini of PaxA (~19 AAs) and PaxB (~34 AAs) of the *Xenorhabdus bovienii* SS-2004 PaxS. Both DD pairs share the accumulation of negatively charged residues in the ^{C}DDs and positively charged ones in the ^{N}DDs and even their positions within the ^{C}DD and ^{N}DD sequences are highly similar (Fig. 16).[82] Thus, no significant differences in the ΔpIs among both DD pairs (ΔpI = theoretical pI ^{N}DD-theoretical pI ^{C}DD:[148,149] PaxB ^{N}DD:PaxA ^{C}DD = 5.7 and PaxC ^{N}DD:PaxB ^{C}DD = 6.1) could be identified.

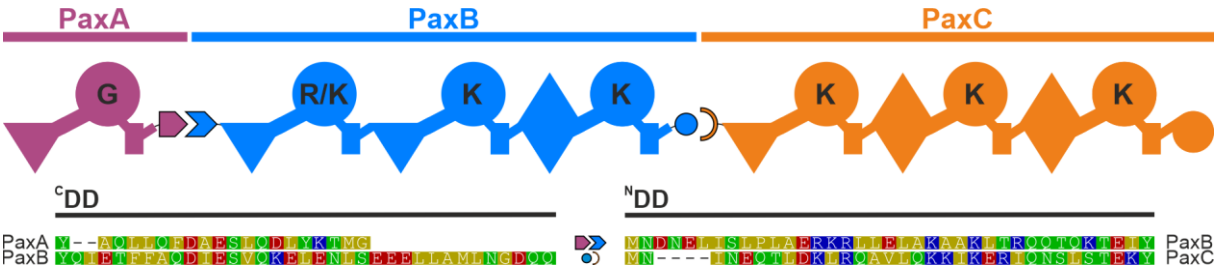


Fig. 16. Schematic representation of the domain and module arrangement in the PaxS. Below, the PaxA/B ^{C}DD and PaxB/C ^{N}DD sequences are given. The residues are colored according to their polarity.

Nonetheless, the interactions of both DD pairs are orthogonal to each other as shown in ITC measurements of the discrete $^{C/N}$DDs of PaxA/C, in which no binding was detected.[82] These findings highlight that divergent electrostatics are most likely not the major driving force in the correct DD partner selection of PaxS, as previously suggested for similar DD junctions in the enacyloxin IIa polyketide synthase.[147] Therefore, the structure of both PaxS DD pairs should be elucidated by solution NMR to identify the underlying mechanisms that maintain docking specificity.

At first, the solution NMR structure of the artificially linked PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD construct was solved. This DD complex structure consists of a three-helical bundle, in which the single α-helix of the PaxC ^{N}DD is surrounded by two α-helices of the PaxB ^{C}DD arranged in a V-shape, and is stabilized mostly by hydrophobic interactions and a few salt bridges between charged amino acids e.g. ^{N}DD R23:^{C}DD D3296 and ^{N}DD K19:^{C}DD E3303.[82]

Furthermore, the discrete DDs were analyzed thermodynamically by ITC and a K_D of 25.0±8.3 µM was measured, which is in good agreement with previously published

DD affinities (~20 µM).[35,81,83] This confirms the previously defined features of DDs, which comprise generally weak affinity interactions and interact with matching partners independent of the presence of their flanking domains and thus form a stable and stand-alone complex.[11,81,83,87]

As DDs of class 1, class 2 and class 4, as well as the PaxB/C $^{C/N}$DDs purely consist of an α-helical secondary structure,[35,81,83,87,88] but in contrast to class 1 and class 2 DDs, none of the PaxB/C $^{C/N}$DD α-helices functions as a dimerization element.[35,83] The reason for this is the generally different oligomerization state of PKSs as homodimers and NRPSs as monomers.[150] Nonetheless, class 1 DDs, class 2 DDs, class 4 DDs, and PaxB/C $^{C/N}$DDs have a similar docking interface, dominated by hydrophobic interactions supported by electrostatic interactions (salt bridges) of strategically positioned residues that are responsible for the docking specificity.[35,82,83,87] Nonetheless, no DD specificity code could be deduced from these structures in contrast to the class 3 DDs of the rhabdopeptide-producing NRPS, where the docking interface is limited to a β-strand of the ^{C}DD that interacts with another β-strand stabilized in a β-hairpin motif of the ^{N}DD.[11]
A further aspect of the unbound PaxB ^{C}DD is its intrinsically disordered region (IDR) - besides a stable C-terminal α-helix in the absence of the PaxC ^{N}DD binding partner. During DD complex formation, a structural transition of the IDR to an additional α-helix is seen, whereas the PaxC ^{N}DD is always composed of a single α-helix independent of the binding state.[82] Also in the virginiamycin *trans*-AT PKS the VirFG ^{N}DD is described as an IDR in the absence of the VirA ^{C}DD, whereas the bound VirFG ^{N}DD contributes to the class 4 DD typical four-helical bundle with two α-helices.[87] Additionally, all ^{C}DDs in the *cis*-AT/*trans*-AT enacyloxin PKS possess strong characteristics of IDRs.[147] It seems that besides obvious structural variations the DD binding induced folding mechanism of IDRs is a further gatekeeper function to maintain DD specificity.

It is shown that the PaxB/C $^{C/N}$DD pair has a lot in common with previously defined DD classes like the similar type of interactions in the docking interface, the pure α-helical secondary structure composition, and the low affinity consistent among all DD classes. However, none of the previously elucidated DD complex structures matches the three-helical bundle topology of the DD complex located in the PaxB/C

inter-protein interface. Therefore, this PaxB/C $^{C/N}$DDs define a structurally new type of DDs.

In publication 2 and manuscript 2 the inter-protein interface between the proteins PaxA and PaxB was analyzed. At first, focusing only on the PaxA/B $^{C/N}$DD regions, the obtained CD data of both isolated DDs identified a completely random coil conformation of the ^{C}DD and a large α-helical content of the ^{N}DD. Thus, the PaxA ^{C}DD possesses characteristics of an IDR, like the previously described PaxB ^{C}DD. Furthermore, the PaxA/B $^{C/N}$DD complex formation was thermodynamically analyzed by ITC, but no interaction between both mutually compatible partners could be seen. This observation leads to the assumption that other regions of the DD flanking domains contribute to the inter-protein interface between PaxA and PaxB; so larger constructs that especially include the ^{C}DD preceding T domain were analyzed in the following.

1.2. Extended docking domain interface between PaxA/B of PaxS

The inclusion of the PaxA ^{C}DD flanking T_1 domain into the docking interface analysis enables the detection of a PaxA T_1-^{C}DD:PaxB ^{N}DD complex. Surprisingly, for this protein-protein interaction a high affinity (K_D of 248±18 nM) could be determined by ITC, which is significantly higher than previously described affinities of pure docking domain interactions (~20 µM).[35,81] The structural evaluation of this protein complex by NMR (see publication 2 & manuscript 2) uncovers the structural alterations in both parts of the complex during the complex formation and identifies the role of the T domain in this inter-protein interface.

In the unbound state, the PaxA ^{C}DD region of the T_1-^{C}DD construct is characterized as an IDR, whereas the PaxB ^{N}DD is composed of a single α-helix, which is separated from an additional IDR at its N-terminus by a proline-induced bend. Again, the disordered conformation of $^{C/N}$DDs as IDRs seems to be a widespread motif in inter-protein interfaces as a tool to achieve specificity.[87,147] In the presence of the interaction partners, the ^{C}DD residues undergo a disorder-to-order transition and form a single α-helix and also at the N-terminus of the ^{N}DD a second α-helix is built. So far, it was still unclear how the T_1 domain participates in this docking interface, but NMR titration experiments indicated that especially α-helix 4 of the T domain takes part in the interaction, which was then proven in the 3D structure of the PaxA T_1-^{C}DD:PaxB ^{N}DD complex.

For the first time, an extended docking domain interface could be proclaimed in which the cooperation of parts of the T domain with the connected ^{C}DD is necessary for the ^{N}DD recognition. Also Hacker *et al.* analyzed the effect of the ^{C}DD preceding T domain on the ^{N}DD affinity by ITC, but no difference in the K_D values could be seen (Kj12B ^{C}DD:Kj12C ^{N}DD: 8 ± 6 µM; Kj12B T-^{C}DD:Kj12C ^{N}DD: 19±6 µM) in this class 3 DD interface of a pure NRPS system. A similar class 3 DD interaction of a hybrid PKS-NRPS system was analyzed by Kosol and colleagues[85] in a more holistic context, where the protein-protein interactions between T-^{C}DD (Bamb_5917) and ^{N}DD-C domain (Bamb_5915) constructs were evaluated. Based on NMR titration and carbene footprinting experiments, three regions were proclaimed that directly interact with Bamb_5915: α-helices 1 and 2 of the T domain plus the ^{C}DD. Additionally, neither a defined interaction surface between the T domain and the ^{N}DD could be detected, nor a complex structure of these *in trans* interacting proteins could be elucidated.[85] Furthermore, in ITC measurements it was observed that the affinity of the DD interaction in *trans*-AT PKSs is not altered by the incorporation of the ^{C}DD flanking ACP domain (VirA ^{C}DD/VirFG ^{N}DD pair: K_D 5.8±0.2 µM; VirA ACP_{5b}-^{C}DD/VirFG ^{N}DD pair: K_D 4 µM).

Contrary results relating to the role of DD adjacent ACPs were obtained most recently in the analysis of a *cis*-AT PKS inter-protein interface, in which a three-fold increased affinity to the ^{N}DD was determined if the ^{C}DD flanking ACP was present.[147] This observation indicates that also in some inter-protein interfaces of PKSs the ACP besides the ^{C}DD takes part in the ^{N}DD recognition.

All mentioned examples lead to the assumption that two different principles have evolved to maintain specificity in inter-protein interfaces of NRPS and PKS systems by docking domains, which either interact independent of adjacent domains or can only interact if the additional interaction surface of the T domain is compatible.

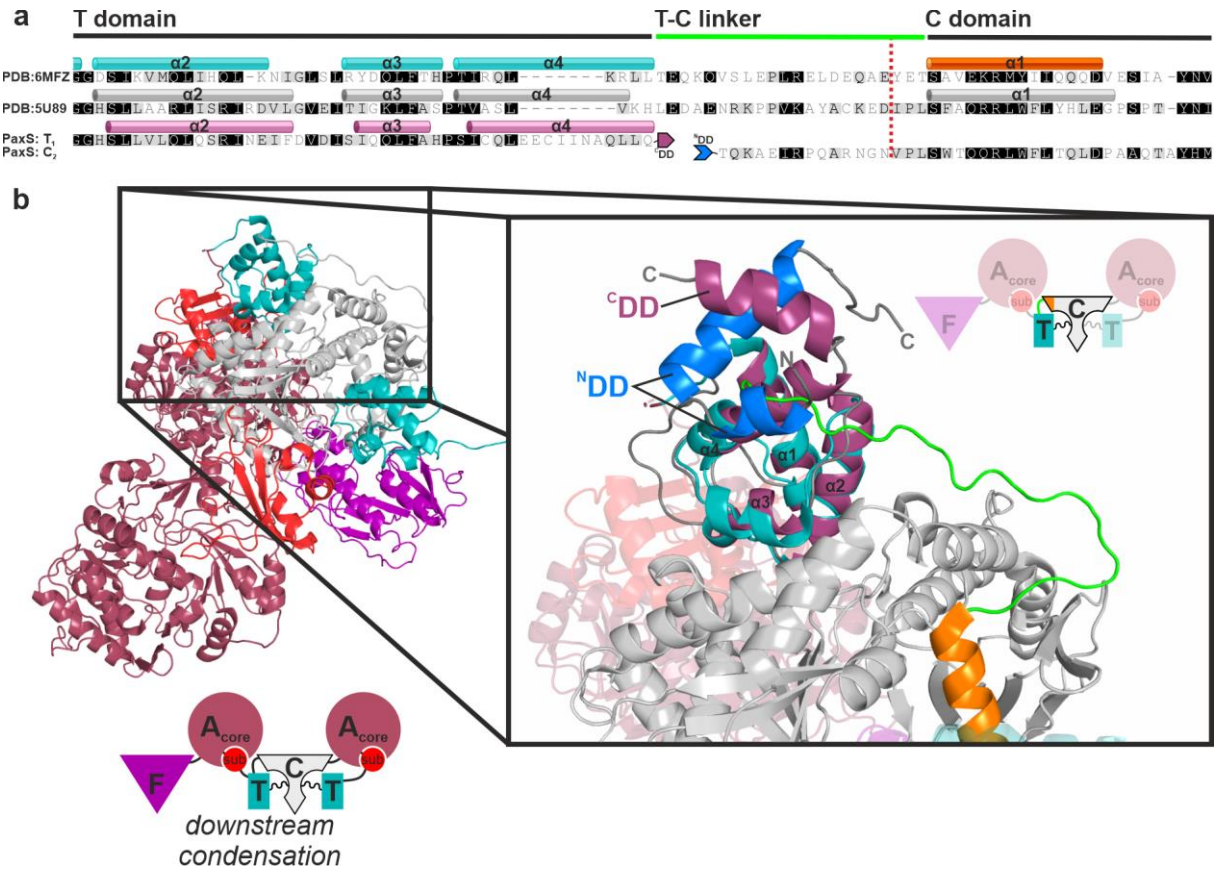


Fig. 17. (a) Sequence alignment of *in trans* interacting T_1 and C_2 domains (DDs are excluded) of PaxS and T_n-C_{n+1} di-domain sequences of published cross-module structures (PDB: 6MFZ, 5U89). The amino acids are colored according to their similarity and the secondary structures are depicted above each sequence. The dotted line represents a possible fusion point to incorporate this type of extended docking domains into single-protein NRPSs, ideally without affecting the structures of the flanking domains. (b) PaxA T_1-^{C}DD:PaxB ^{N}DD complex structure superimposed on the di-modular NRPS structure of LgrA with the command “align” in PyMOL. In the enlarged view the T-C linker is green-colored, and the first α-helix of the C domain is orange-colored.

To get a structural impression of how the extended docking interface interferes with the overall domain arrangement in a multi-modular context, the PaxA T_1-^{C}DD:PaxB ^{N}DD complex structure was superimposed onto the cross-module structure of LgrA (PDB: 6MFZ; Fig. 17).[79] This di-modular structure captures the domain arrangement in the condensation conformation, in which the nascent peptide chain is elongated and transferred from the T_n domain to the downstream T_{n+1} domain. Thereby, it links subsequent modules in a catalytic reaction step. The extended docking interface composed of T domain α-helix 4, ^{C}DD and ^{N}DD is positioned outwards of the T_n:C_{n+1} interaction surface and should not disrupt or hinder any domain interactions between the T and C domains. Therefore, this type of high-affinity extended docking domains should be ideally suited to split single-protein NRPSs into two parts, without negatively affecting the catalytic properties of NRPSs.

In manuscript 1 a functional insertion point for synthetic zippers within the T-C linker region was already shown, which should as well facilitate the insertion of natural docking domains. On the other side of the inter-protein interface it would most likely be beneficial to exchange the whole T domain, even if only α-helix 4 of the T domain contributes to the extended docking domain interface. The substitution of T domain α-helix 4 during the ^{C}DD insertion would presumably impair the formation of intramolecular interactions within the such generated chimeric four-helical bundle of the T domain.

1.3. Extended docking domain interface between PaxB/C of PaxS

In contrast to the docking interface between the proteins PaxA and PaxB in which the pure DDs do not interact, the discrete $^{N/C}$DDs of PaxB and PaxC form a stable DD complex characterized by a binding affinity of ~25 µM,[82] which is well within the range of already published DD affinities.[11,81,83,87] In some of these publications, also the embedding of the ^{C}DD preceding carrier proteins by the expression of ACP-^{C}DD (PKS)[87] and T-^{C}DD (NRPS)[11] constructs was analyzed and showed no impact on the ITC measured binding affinities to the respective ^{N}DD partner. However, the PaxB T_4-^{C}DD construct showed a 25-fold increased affinity to the PaxC ^{N}DD (~1 µM; Fig. 10).

The NMR-based secondary structure prediction and the application of NMR titration experiments should further enlighten the protein regions that contribute to the docking interface. As known from already published T domain structures, also the determined four α-helices of the unbound PaxB T_4-^{C}DD most likely assemble in a T domain characteristic four-helical bundle.[38] An additional single-turn helix within the loop region connecting the α-helices 1 and 2 is not found in all carrier protein structures,[39] but was previously described as fifth loop α-helix (αl).[151] The sixth α-helix positioned at the C-terminus of the ^{C}DD is connected to the T domain via a stretch of intrinsically disordered residues (Fig. 11). This region of undefined secondary structure at the N-terminus of the ^{C}DD was as well seen in the unbound sate of the discrete ^{C}DD.[82] During the PaxB T_4-^{C}DD:PaxC ^{N}DD complex formation, the overall secondary structure of the T domain is not altered, but a second α-helix appears in the previously intrinsically disordered ^{C}DD region (Fig. 12a). This observation alone could not prove any influence of the T domain on a possible extended docking interface, but again confirms the secondary structure appearance within both DDs upon binding as seen in the artificially linked PaxB/C $^{C/N}$DD complex

structure.[82] Also the comparison of NH group chemical shifts of ^{C}DD residues located in the bound ^{C}DD and T-^{C}DD constructs show only minor differences, which again supports the stand-alone character of the ^{C}DD except for ^{C}DD α-helix 1, which presumably benefits from the T domain in terms of further contacts that induce a single-turn extension at its N-terminus (Fig. 14).

In an additional NMR titration experiment, the PaxC ^{N}DD-binding site could be specified. Not only residues of the ^{C}DD but also of the T domain (especially α-helix 4) show large chemical shift changes during the complex formation (Fig. 12b), thus both regions contribute to the docking interface.

To get structural insights into this extended docking interface of the PaxB T_4-^{C}DD bound to the PaxC ^{N}DD, the lowest-energy CS-ROSETTA model of the bound PaxB T_4-^{C}DD was superimposed on the already solved PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD NMR solution structure (Fig. 18).[82] Even if CS-ROSETTA was not capable of calculating a converged low energy fold, maybe because Cβ and Hα chemical shifts were lacking, the assembled complex structure possibly captures the true binding event in which the PaxC ^{N}DD α-helix is wrapped in three α-helices: T_4 domain α-helix 4 and the two α-helices 5 and 6 of the ^{C}DD.

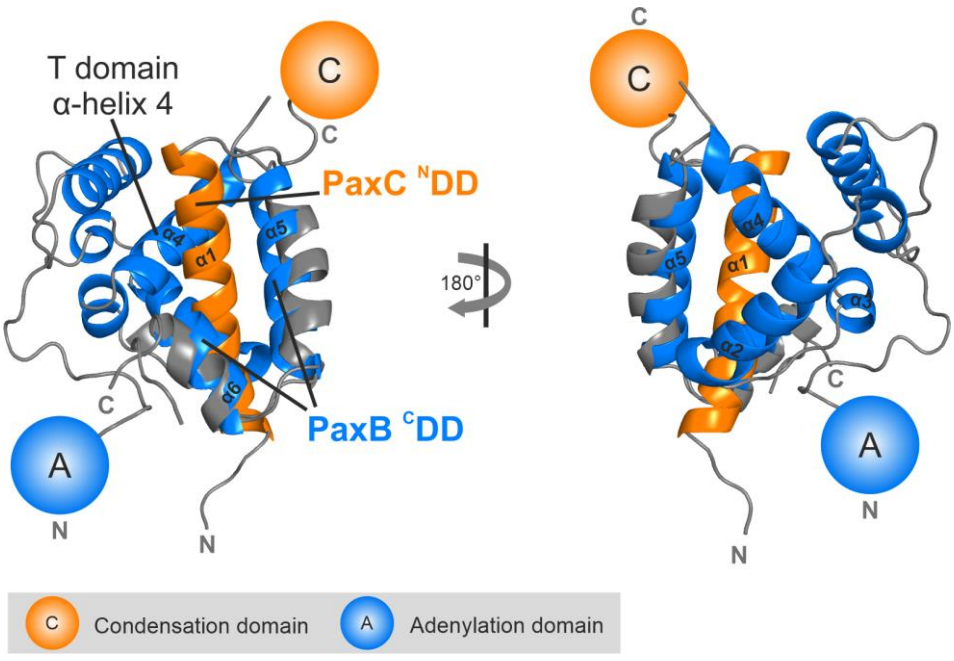


Fig. 18. Lowest-energy CS-ROSETTA model of PaxB T_4-^{C}DD (blue) in its bound state superimposed on the PaxC ^{N}DD (orange)-$(GS)_{12.5}$-PaxB ^{C}DD (grey) NMR solution structure. For reasons of clarity, the 25 amino acid long GS linker is not shown.

The domain arrangement within modules differs significantly in the different steps of the catalytic cycle of NRPSs[45,68,77] and in the publication of Reimer *et al.* the first

crystal structures of di-modular LgrA constructs of these different domain conformations were solved.[79] Here it was proclaimed that only intramodular linkers and the T_n:C_{n+1} domain interaction during the condensation reaction strictly connect adjacent modules of fully *in cis* interacting domains, whereas the overall positions of neighboring modules to each other are not unambiguously defined.[79] Therefore, it was assumed that the binding affinity is also significantly altered if the ^{N}DD subsequent C domain of *in trans* interacting T_n/C_{n+1} domains is present, but no effect on the K_D value was observed. Because no structure of *in trans* interacting domains is available, it can only be speculated that the loss of further A_{n+1} and T_{n+1} domains of the module probably destroys the structural integrity of the C domain by the exposition of hydrophobic patches. Furthermore, the absence of the PPant-arm in the expressed *apo*-form of the T_n domain besides the missing T_{n+1} domain prohibits formation of the overall condensation conformation. The protein structure of the PaxB T_4 α-helix 4-^{C}DD construct is most likely also not fully intact by reason of the isolation of the C-terminal T domain α-helix 4 from the four-helical bundle. Additionally, the CD spectra of the discrete PaxB ^{C}DD and the PaxB T_4 α-helix 4-^{C}DD are highly similar (Supporting Fig. 1c), indicating a similar α-helical composition supporting the hypothesis of an unfolded PaxB T_4 α-helix 4 even though the affinity is slightly increased.

In conclusion, it can now be explained how the unidirectional protein order is maintained in the PaxS, if neither key-residue interactions[11] nor architectural variations of DDs assigned to the same class[83] or topology differences of DDs of different classes[147] are responsible for the NRPS order. It is shown that especially the $^{C/N}$DD occurrence as IDRs is a selectivity feature in the PaxS, which is further supported by the evolution of extended inter-protein interfaces that include the ^{C}DD preceding T domains. This uncovers a further selectivity mechanism, in which the gatekeeper function, so far proclaimed only for DDs, is expanded to C-terminal T domains of multi-protein NRPSs, and explains how interaction specificity and selectivity is achieved even if DDs are highly similar within one system.

2. Synthetic zipper approach vs. standard strategies of NRPS engineering

In the past years different strategies have been developed, which allow to re-engineer NRPSs by module and multi-module exchanges and to produce novel natural products (Fig. 19). Two of them, namely the exchange unit (XU)[10] and

exchange unit condensation domain (XUC)[9] concepts are the most favorable ones nowadays.[152]

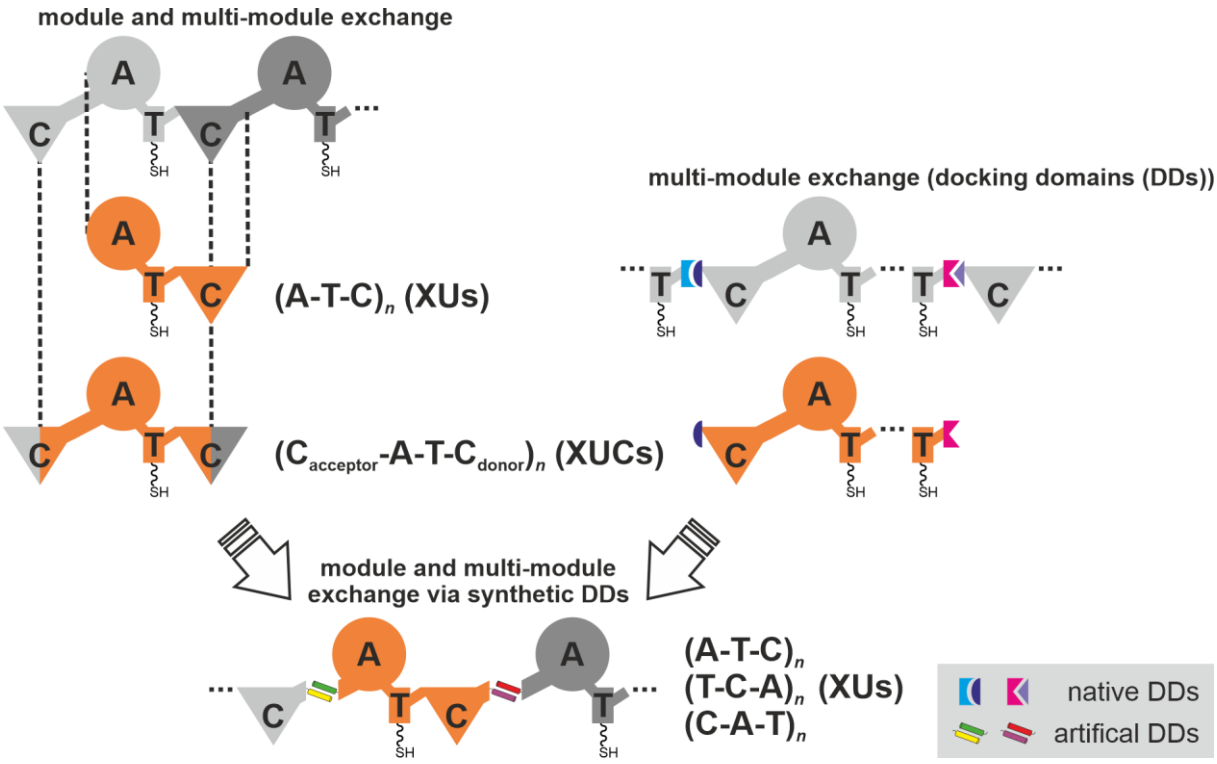


Fig. 19. Engineering strategies of NRPSs. Grey and orange shades indicate the original and modified domains, respectively. Additionally, the exchangeable units are given at the right of the domains for the XU and XUC concepts.

The XU concept is based on exchangeable A-T-C or A-T-C/E units, which can be covalently fused in a conserved WNATE sequence of the C-A linker, but only in a way that the specificity of the downstream C domain is kept. In detail this means that the $(XU)_n$ following $(XU)_{n+1}$ must incorporate the identical amino acid into the nascent peptide chain as the $(XU)_n$ would expect in the wild type NRPS. Nonetheless, this method was capable of generating numerous NRPSs with good production titers, but a correlation between a reduced peptide production and the number of used XUs was seen, most likely explained by the increasing number of non-natural C-A interfaces.[10]

This C domain specificity limitation was tackled in the follow up XUC strategy of Bozhüyük *et al*.[9] This concept is based on the assumption that the C domain structure is composed of two lobes[59] (subdomains) connected via a four-AA-long flexible linker, of which especially the C-terminal lobe (≙ C domains' acceptor site (C_{Asub})) defines the C domain specificity. Therefore, the covalent fusion of C_{Asub}-A-T-C_{Dsub} building blocks within the short subdomain-linking loop region should preserve the C domain specificity. This enabled the emulation of natural NRPSs with XUCs,

but also the *de novo* design of NRPSs, both with good production titers. Nonetheless, this approach also has the following two major drawbacks: C/E and C domains cannot be combined with each other, explained by structural and sequential differences[62] and XUCs from bacteria of different genera (*Xenorhabdus*/*Photorhabdus* vs. *Bacillus*) were incompatible, most likely by reason of the non-natural C subdomains interface.[9]
Other engineering approaches take advantage of low-affinity DDs (COM domains),[34] which connect multi-protein NRPSs in nature by ensuring specificity and selectivity. The use of DDs allowed successful multi-module exchanges, but in both cases the re-assembled NRPSs rely on (multi-)module building blocks, which natively interact *in trans* via DDs.[108,109]

From these different engineering strategies, the idea arises to combine the benefits of the XU concept with the flexibility of DD-mediated module exchanges. This new concept generates a new synthetic type of NRPSs (type S), in which high-affinity synthetic zippers (SYNZIPSs)[125,126] are used to connect XUs generated from the splitting of single-protein NRPSs. SYNZIPs are composed of single α-helices (~5 kDa) that intertwine with dissociation constants less than 10 nM in most cases, to form a coiled-coil motif.[126] This high affinity should facilitate a covalent-like binding state of XUs encoded on different plasmids and should not disrupt any essential domain-domain interactions by their small size.

The possibility to artificially split single-protein NRPSs was already shown in the work of Kegler *et al*.[111], but in this example only natural DDs were inserted within the linker regions connecting E-C and T-C domains. Between these domains (E/C; T/C), DDs are naturally present.[108,109] Nonetheless, the artificial splitting of natively, fully *in cis* coupled domains does not seem to disrupt the non-covalent protein-protein interactions of adjacent domains.[111] The newly developed tool to produce type S NRPSs goes one step further and incorporates *in trans* interacting synthetic zippers in all kinds of linker regions (C-A, A-T, T-C) to connect NRPS building blocks in a functional way. Thus, we were able to produce all kinds of possible XUs ($[C\text{-}A\text{-}T]_n$, $[A\text{-}T\text{-}C]_n$, $[T\text{-}C\text{-}A]_n$) and efficiently recombine them functionally. As with the original XU strategy[10], the recombination of XUs from different bacteria genera (Gram-negative/Gram-positive) was possible. The structural diversity of peptides produced by type S NRPSs does not significantly differ from the peptides described in the XU

and XUC publications,[9,10] as the artificially zipped NRPS building blocks were derived from the identical set of NRPSs found in bacteria of the genus *Xenorhabdus* and *Photorhabdus*. Nonetheless, the unique feature of the use of synthetic zippers as an NRPS engineering tool is the versatile and fast recombination capability, which is further expanded by the introduction of a second heterospecific SYNZIP pair to produce a linear network of three XUs that interact *in trans*.

The answer to why this new strategy is so successful in the context of NRPSs, even if ripping normally *in cis* interacting modules out of their protein environment is a tremendous intervention, is given in recent structural studies of multi-modular NRPS constructs. In 2016, a higher-order architecture of NRPSs was proposed, in which all domains are regularly arranged within a helical filament structure.[78] However, the electron microscopy of the di-modular NRPS DhbF and the x-ray crystallography of the di-modular NRPS LgrA revealed diverse module arrangements without any conserved inter-modular interaction surface, confirming that no rigid architecture underlies in NRPSs.[77,79] In addition, not always is a single conformation of a multi-modular NRPS linked to a distinct step of the NRPS catalytic cycle.[79] However, the domain-domain contacts within a module are not arbitrary and guided by transient interactions, as for example the rather rigid C-A catalytic platform,[68] in which the relative position of both domains to each other remains fairly unaffected during the catalytic cycle.[79] Further examples of defined domain-domain orientations and interactions are for example the ~140° A_{sub} domain rotation during the adenylation reaction and the overall T domain movement in the synthetic cycle.[76,153]

Taking all this structural information into account, it becomes obvious that NRPSs possess no super-modular architecture. Therefore, it is possible to remove individual XUs from single-protein NRPSs and re-assemble new type S NRPSs via synthetic zippers. The replacement of native linker sequences with these synthetic zippers also challenges the role of linkers as the essential regions that stabilize domain-domain contacts.[70,71] In a next step, it is essential to obtain structural insights into the synthetic zipper-coupled interface of type S NRPS to elucidate if and how the artificial *in trans*-linkage really interferes with domain-domain interactions and to allow for the structure-based optimization of this engineering strategy.

V. REFERENCES

1. Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29,** 1074–1098 (2012).
2. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem., Int. Ed.* **56,** 3770–3821 (2017).
3. Payne, J. A. E., Schoppet, M., Hansen, M. H. & Cryle, M. J. Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. *Mol. BioSyst.* **13,** 9–22 (2016).
4. Tobias, N. J. *et al.* Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.* **2,** 1676–1685 (2017).
5. Bode, E. *et al.* Promoter activation in Δ*hfq* mutants as an efficient tool for specialized metabolite production enabling direct bioactivity testing. *Angew. Chem., Int. Ed.* **58,** 18957–18963 (2019).
6. Bode, E. *et al.* Biosynthesis and function of simple amides in *Xenorhabdus doucetiae*. *Environ. Microbiol.* **19,** 4564–4575 (2017).
7. Dudnik, A., Bigler, L. & Dudler, R. Heterologous expression of a *Photorhabdus luminescens* syrbactin-like gene cluster results in production of the potent proteasome inhibitor glidobactin A. *Microbiol. Res.* **168,** 73–76 (2013).
8. Stein, M. L. *et al.* One-shot NMR analysis of microbial secretions identifies highly potent proteasome inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 18367–18371 (2012).
9. 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).
10. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10,** 275–281 (2018).
11. Hacker, C. *et al.* Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nat. Commun.* **9,** 4366 (2018).
12. Kegler, C. *et al.* Rapid determination of the amino acid configuration of xenotetrapeptide. *ChemBioChem* **15,** 826–828 (2014).
13. Kittilä, T., Mollo, A., Charkoudian, L. K. & Cryle, M. J. New structural data reveal the motion of carrier proteins in nonribosomal peptide synthesis. *Angew. Chem., Int. Ed.* **55,** 9834–9840 (2016).

14. Dieckmann, R., Lee, Y.-O., van Liempt, H., Döhren, H. von & Kleinkauf, H. Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Lett.* **357,** 212–216 (1995).

15. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C. & Walsh, C. T. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 2509–2514 (2000).

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

17. Schneider, A. & Marahiel, M. A. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.* **169,** 404–410 (1998).

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

19. Marshall, C. G., Burkart, M. D., Keating, T. A. & Walsh, C. T. Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate. *Biochemistry* **40,** 10655–10663 (2001).

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

21. Weber, G., Schörgendorfer, K., Schneider-Scherzer, E. & Leitner, E. The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.* **26,** 120–125 (1994).

22. 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).

23. Schoenafinger, G., Schracke, N., Linne, U. & Marahiel, M. A. Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J. Am. Chem. Soc.* **128,** 7406–7407 (2006).

24. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3,** 490–504 (2002).

25. 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).

26. Cosmina, P. *et al.* Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **8,** 821–831 (1993).

27. 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).

28. 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).

29. Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37,** 2648–2659 (1998).

30. 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).

31. Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **6,** 385–400 (1999).

32. Guenzi, E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273,** 32857–32863 (1998).

33. 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).

34. Hahn, M. & Stachelhaus, T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 15585–15590 (2004).

35. Broadhurst, R., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10,** 723–731 (2003).

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

37. 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).

38. 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).

39. Lohman, J. R. *et al.* The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. *Proteins* **82,** 1210–1218 (2014).

40. Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312,** 273–276 (2006).

41. 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).

42. 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).

43. 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).

44. Haslinger, K., Redfield, C. & Cryle, M. J. Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins* **83,** 711–721 (2015).

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

46. 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).

47. Stachelhaus, T. & Marahiel, M. A. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270,** 6163–6169 (1995).

48. 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).

49. 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).

50. 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).

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

52. Conti, E., Franks, N. P. & Brick, P. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4,** 287–298 (1996).

53. May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 12120–12125 (2002).

54. Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* **47,** 8016–8025 (2008).

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

56. Baltz, R. H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38,** 1747–1760 (2011).

57. 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).

58. 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., Ser. III* **318,** 927–936 (1995).

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

60. 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).

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

62. 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).

63. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58,** 453–488 (2004).

64. 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).

65. 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).

66. 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).

67. 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).

68. 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).

69. Marahiel, M. A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* **15,** 799–807 (2009).

70. Gao, L. *et al.* Translocation of the thioesterase domain for the redesign of plipastatin synthetase. *Sci. Rep.* **6,** 38467 (2016).

71. Beer, R. *et al.* Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol. BioSyst.* **10,** 1709–1718 (2014).

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

73. 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).

74. 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).

75. 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).

76. 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).

77. 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).

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

79. Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366,** 706 (2019).

80. Vo, T. D., Spahn, C., Heilemann, M. & Bode, H. B. Microbial Cationic Peptides as a Natural Defense Mechanism against Insect Antimicrobial Peptides. *ACS Chem. Biol.* (2021).

81. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4,** 41–52 (2009).

82. Watzel, J., Hacker, C., Duchardt-Ferner, E., Bode, H. B. & Wöhnert, J. A new docking domain type in the peptide-antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*. *ACS Chem. Biol.* **15,** 982–989 (2020).

83. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20,** 1340–1351 (2013).

84. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4,** 75–81 (2008).

85. Kosol, S. *et al.* Structural basis for chain release from the enacyloxin polyketide synthase. *Nat. Chem.* **11,** 913–923 (2019).

86. Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 12432–12437 (2016).

87. Dorival, J. *et al.* Characterization of intersubunit communication in the virginiamycin trans-acyl transferase polyketide synthase. *J. Am. Chem. Soc.* **138,** 4155–4167 (2016).

88. Zeng, J. *et al.* Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases. *ACS Chem. Biol.* **11,** 2466–2474 (2016).

89. Thattai, M., Burak, Y. & Shraiman, B. I. The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3,** 1827–1835 (2007).

90. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252,** 675–679 (1991).

91. Staunton, J. *et al.* Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3,** 188–192 (1996).

92. J F Aparicio, P Caffrey, A F Marsden, J Staunton & P F Leadlay. Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *J. Biol. Chem.* **269,** 8524–8528 (1994).

93. Gokhale, R. Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol.* **4,** 22–27 (2000).

94. Tsuji, S. Y., Cane, D. E. & Khosla, C. Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40,** 2326–2331 (2001).

95. Pulsawat, N., Kitani, S. & Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393,** 31–42 (2007).

96. 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).

97. 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).

98. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11,** 4554 (2020).

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

100. 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).

101. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30,** 440–446 (2012).

102. 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).

103. Bian, X., Plaza, A., Yan, F., Zhang, Y. & Müller, R. Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives in vivo. *Biotechnol. Bioeng.* **112,** 1343–1353 (2015).

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

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

106. 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).

107. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5848–5853 (2000).

108. Hahn, M. & Stachelhaus, T. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 275–280 (2006).

109. 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).

110. O'Connor, S. E., Walsh, C. T. & Liu, F. Biosynthesis of epothilone intermediates with alternate starter units: engineering polyketide-nonribosomal interfaces. *Angew. Chem., Int. Ed.* **42,** 3917–3921 (2003).

111. Kegler, C. & Bode, H. B. Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains. *Angew. Chem., Int. Ed.* **59,** 13463-13467 (2020).

112. Cai, X., Zhao, L. & Bode, H. B. Reprogramming promiscuous nonribosomal peptide synthetases for production of specific peptides. *Org. Lett.* **21,** 2116–2120 (2019).

113. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23,** 1171–1176 (2005).

114. Menzella, H. G., Carney, J. R. & Santi, D. V. Rational design and assembly of synthetic trimodular polyketide synthases. *Chem. Biol.* **14,** 143–151 (2007).

115. Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C. & Grininger, M. Engineering of chimeric polyketide synthases using SYNZIP docking domains. *ACS Chem. Biol.* **14,** 426–433 (2019).

116. Lupas, A. N. & Bassler, J. Coiled coils - a model system for the 21st century. *Trends Biochem. Sci.* **42,** 130–140 (2017).

117. Wolf, E., Kim, P. S. & Berger, B. MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6,** 1179–1189 (1997).

118. Crick, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* **6,** 689–697 (1953).

119. Burkhard, P., Meier, M. & Lustig, A. Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci.* **9,** 2294–2301 (2000).

120. Glover, J. N. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373,** 257–261 (1995).

121. Freymann, D. *et al.* 2.9 Å resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. *J. Mol. Biol.* **216,** 141–160 (1990).

122. Thomson, A. R. *et al.* Computational design of water-soluble α-helical barrels. *Science* **346,** 485–488 (2014).

123. Reinke, A. W., Grant, R. A. & Keating, A. E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132,** 6025–6031 (2010).

124. Lau, S. Y., Taneja, A. K. & Hodges, R. S. Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J. Biol. Chem.* **259,** 13253–13261 (1984).

125. Grigoryan, G., Reinke, A. W. & Keating, A. E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458,** 859–864 (2009).

126. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* **1,** 118–129 (2012).

127. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10,** 411–421 (1999).

128. Sattler, M. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).

129. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson. (1969-1992)* **89,** 496–514 (1990).

130. Clubb, R. T., Thanabal, V. & Wagner, G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N–^{13}C-labelled proteins. *J. Magn. Reson. (1969-1992)* **97,** 213–217 (1992).

131. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson. (1969-1992)* **99,** 201–207 (1992).

132. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114,** 6291–6293 (1992).

133. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J. Biomol. NMR* **3,** 185–204 (1993).

134. Overhauser, A. W. Polarization of nuclei in metals. *Phys. Rev.* **92,** 411–415 (1953).

135. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62,** 453–471 (2015).

136. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298 (1997).

137. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220,** 671–680 (1983).

138. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209–227 (2002).

139. Jee, J. & Güntert, P. Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J. Struct. Funct. Genomics* **4,** 179–189 (2003).

140. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113,** 5490–5492 (1991).

141. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13,** 289–302 (1999).

142. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).

143. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56,** 227–241 (2013).

144. Dios, A. C. de, Pearson, J. G. & Oldfield, E. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science* **260,** 1491–1496 (1993).

145. Wishart, D. in *Encyclopedia of Biophysics,* edited by G. C. K. Roberts (Springer Berlin, Berlin, 2013), pp. 279–280.

146. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4685–4690 (2008).

147. Risser, F. *et al.* Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase. *J. Struct. Biol.* **212,** 107581 (2020).

148. Bjellqvist, B. *et al.* The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14,** 1023–1031 (1993).

149. Bjellqvist, B., Basse, B., Olsen, E. & Celis, J. E. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **15,** 529–539 (1994).

150. Miyanaga, A., Kudo, F. & Eguchi, T. Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* **35,** 1185–1209 (2018).

151. Goodrich, A. C., Harden, B. J. & Frueh, D. P. Solution structure of a nonribosomal peptide synthetase carrier protein loaded with its substrate reveals transient, well-defined contacts. *J. Am. Chem. Soc.* **137,** 12100–12109 (2015).

152. Hwang, S., Lee, N., Cho, S., Palsson, B. & Cho, B.-K. Repurposing modular polyketide synthases and non-ribosomal peptide synthetases for novel chemical biosynthesis. *Front. Mol. Biosci.* **7,** 87 (2020).

153. Gulick, A. M. Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Curr. Opin. Chem. Biol.* **35,** 89–96 (2016).

154. Owen, J. G., Robins, K. J., Parachin, N. S. & Ackerley, D. F. A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environ. Microbiol.* **14,** 1198–1209 (2012).

155. Chaston, J. M. *et al.* The entomopathogenic bacterial endosymbionts *Xenorhabdus* and *Photorhabdus*: convergent lifestyles from divergent genomes. *PLoS One* **6,** e27909 (2011).

156. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. in *Structure and mechanism,* edited by N. J. Oppenheimer (Academic Press, San Diego, Calif., 1989), pp. 44–73.

157. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J. Biomol. NMR* **12,** 1–23 (1998).

158. Favier, A. & Brutscher, B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR* **49,** 9–15 (2011).

159. Solyom, Z. *et al.* BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR* **55,** 311–321 (2013).

160. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187,** 163–169 (2007).

161. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial* (CANTINA Verlag, Goldau, 2004).

162. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696 (2005).

163. Mulder, F. A., Schipper, D., Bott, R. & Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus* subtilisins. *J. Mol. Biol.* **292,** 111–123 (1999).

VI. ATTACHMENTS

1. A new docking domain type in the Peptide-Antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*

1.1. Author contribution statements

Status: published

Journal: *ACS Chem. Biol.* **15,** 982–989 (2020)

Authors: Jonas Watzel (JoWa), Carolin Hacker (CH), Elke Duchardt-Ferner (EDF), Helge B. Bode (HBB) and Jens Wöhnert (JW)

(1) Concept and design

JoWa (60 %), JW (20 %), HBB (20 %)

(2) Conducting tests and experiments

JoWa (60 %): cloning of plasmids, protein expression and purification, NMR measurements, ITC and CD measurements, HPLC-MS analysis; CH (30 %): NMR measurements; EDF (10 %): NMR measurements

(3) Compilation of data sets and figures

JoWa (100 %): compilation of data sets and figures (NMR spectra, protein structures, ITC thermograms, chemical shift index)

(4) Analysis and interpretation of data

JoWa (70 %): structure calculation, analysis and interpretation of NMR data; CH (30 %): analysis and interpretation of NMR data

(5) Drafting of manuscript

JoWa (65 %), JW (25 %), HBB (10 %)

place and date	signature PhD student
place and date	signature supervisor

1.2. Publication

pubs.acs.org/acschemicalbiology Articles

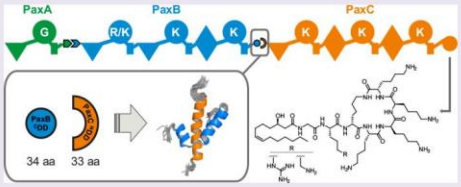
A New Docking Domain Type in the Peptide-Antimicrobial-Xenorhabdus Peptide Producing Nonribosomal Peptide Synthetase from *Xenorhabdus bovienii*

Jonas Watzel, Carolin Hacker, Elke Duchardt-Ferner, Helge B. Bode,* and Jens Wöhnert*

Cite This: *ACS Chem. Biol.* 2020, 15, 982–989 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nonribosomal peptide synthetases (NRPSs) produce a wide variety of different natural products from amino acid precursors. In contrast to single protein NRPS, the NRPS of the bacterium *Xenorhabdus bovienii* producing the peptide-antimicrobial-Xenorhabdus (PAX) peptide consists of three individual proteins (PaxA/B/C), which interact with each other noncovalently in a linear fashion. The specific interactions between the three different proteins in this NRPS system are mediated by short C- and N-terminal docking domains ($^{C/N}$DDs). Here, we investigate the structural basis for the specific interaction between the ^{C}DD from the protein PaxB and the ^{N}DD from PaxC. The isolated DD peptides feature transient α-helical conformations in the absence of the respective DD partner. Isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) titration experiments showed that the two isolated DDs bind to each other and form a structurally well-defined complex with a dissociation constant in the micromolar range as is typical for many DD interactions. Artificial linking of this DD pair via a flexible glycine−serine (GS) linker enabled us to solve the structure of the DD complex by NMR spectroscopy. In the complex, the two DDs interact with each other by forming a three helix bundle arranged in an overall coiled-coil motif. Key interacting residues were identified in mutagenesis experiments. Overall, our structure of the PaxB ^{C}DD/PaxC ^{N}DD complex represents an architecturally new type of DD interaction motif.



INTRODUCTION

Nonribosomal peptide synthetases produce a wide range of different natural products from amino acid precursors. Many of these products have clinical relevance as exemplified by the antibiotic daptomycin or the immunosuppressant cyclosporine A.[1] One reason for the diverse product spectrum is the modular architecture of nonribosomal peptide synthetase (NRPS) in which distinct catalytic domains are linked together into functional modules. Each module is then responsible for the incorporation of one amino acid building block into the final product. Different modules can be covalently linked together on a single long polypeptide chain or can be located on individual proteins that then interact with each other noncovalently. A typical NRPS elongation module consists of an adenylation (A) domain, a thiolation (T)/peptidyl-carrier protein (PCP) domain, and a condensation (C) domain. After the activation of a module specific amino acid (aa) building block as the respective aminoacyl adenylate, the activated aa is loaded onto the PCP domain. Then, an amide bond between amino acids bound to the PCP domains of adjacent modules is formed by the condensation domain.

Many NRPSs consists of one protein (subunit) in which different modules are covalently linked. Thus, the amino acid composition of the produced peptide is defined by the linear arrangement of the modules along the protein chain. This collinearity rule also has to be followed in the context of multiprotein NRPSs in which the individual modules are distributed between more than one protein chain. Hahn and Stachelhaus[2] have shown that short interacting regions at the C- and N-termini of different protein chains maintain the strict sequential order by enabling the linear assembly of non-covalent protein complexes due to specific noncovalent interactions. These short interacting regions are referred to as docking domains (DDs) or COM (communication-mediating) domains. Generally, DDs are small and often comprise only 15−60 amino acids with only low identity scores in sequence alignments. *In vitro* and *in vivo* investigations on the multiprotein NRPS system synthesizing tyrocidine have confirmed the essential role of these DDs in the context of protein−protein interaction between the individual polypeptide chains. Interestingly, native DD pairs

Received: December 20, 2019
Accepted: March 13, 2020
Published: March 13, 2020

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982–989

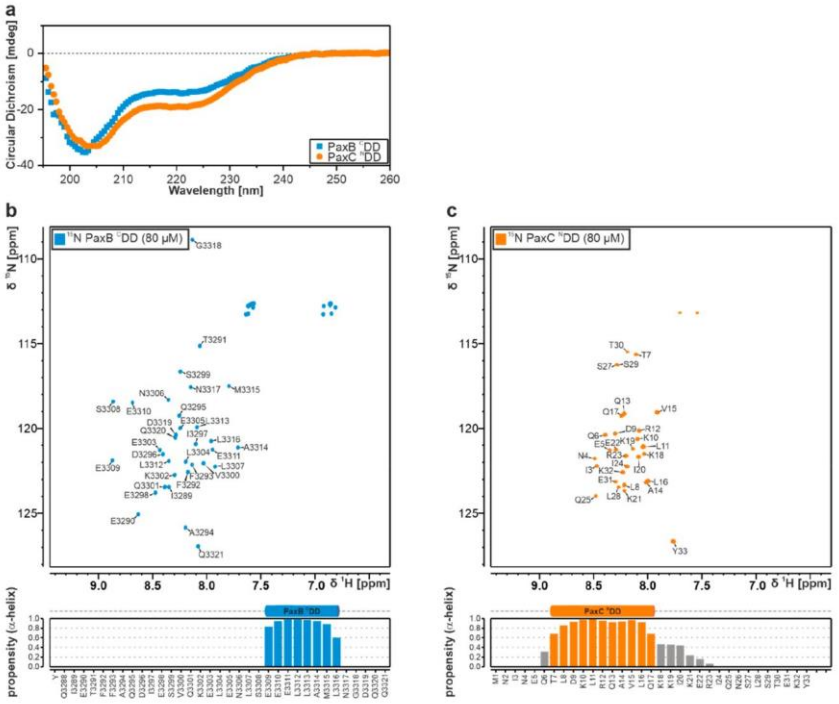


Figure 1. (a) Circular dichroism spectra of 50 μM PaxB/C $^{C/N}$DD samples. (b) Assigned ^{1}H,^{15}N-HSQC spectrum of the PaxB ^{C}DD (blue) in the unbound state (top). Chemical shift index-derived secondary structure of the unbound PaxB ^{C}DD (blue) according to TALOS-N. The defined secondary structure (α-helical propensity $\geq$ 0.6) is given in a schematic representation above each residue (bottom). (c) Assigned ^{1}H,^{15}N-HSQC spectrum of the PaxC ^{N}DD (orange) in the unbound state. Chemical shift index-derived secondary structure of the unbound PaxC ^{N}DD according to TALOS-N. The defined secondary structure (α-helical propensity $\geq$ 0.6) is given in a schematic representation above each residue (bottom).

in the tyrocidine synthesizing NRPS could be replaced with matching COM domain pairs from other NRPS systems.[2] In addition, subunits from the surfactin and tyrocidine multiprotein NRPS systems could be connected with suitable DD pairs into a functional hybrid NRPS system.[3,4] In the context of tyrocidine and rhabdopeptide NRPSs, it was shown that the COM domain or DD specificity is mediated mainly by polar or electrostatic interactions between key residues. The exchange of specific amino acid residues in the docking domain complex interface therefore lead to an altered partner specificity and consequently to an altered product spectrum.[4,5]

Despite the important role of docking/COM domains in the context of multiprotein NRPSs, only little is known about the structure of these interaction platforms. So far, only a single structure of a DD complex from a pure NRPS system has been solved.[5] Additional structures of DD complexes are only described for related polyketide synthase (PKS) systems, which possess a similar modular architecture but are often dimeric. In the PKS field, three structurally different DD classes were identified until now. Furthermore, additional structures for isolated ^{N}DDs without a bound ^{C}DD have been solved for mixed PKS/NRPS systems.[6−11]

The peptide-antimicrobial-Xenorhabdus (PAX) peptide producing NRPS of the entomopathogenic bacterium *Xenorhabdus bovienii* contains seven modules distributed among three individual protein chains called PaxA, PaxB, and PaxC, respectively. The peptide product of this NRPS consists of seven amino acids with a unique amino acid order and contains mainly basic amino acid residues (lysine/arginine) and an N-terminally attached fatty acid moiety incorporated by the C starter domain.[12] Since this NRPS synthesizes a product with a unique amino acid order, the collinearity rule between the product composition and module arrangement requires that the three individual protein chains must interact with each other in a well-defined order. The structural basis for the ordered interaction of the three individual protein chains in this NRPS is not yet known. Our goal here was to define the structural basis for the ordered interaction of PaxA, PaxB, and PaxC with each other. To this end, we were able to identify putative docking domains in this system by bioinformatics methods that help to explain the unilinear specific noncovalent complex formation in this NRPS. These docking domains have no similarities to the NRPS docking domains that have been analyzed structurally so far. Here, we describe the complex structure of the DD pair connecting the proteins PaxB and

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982−989

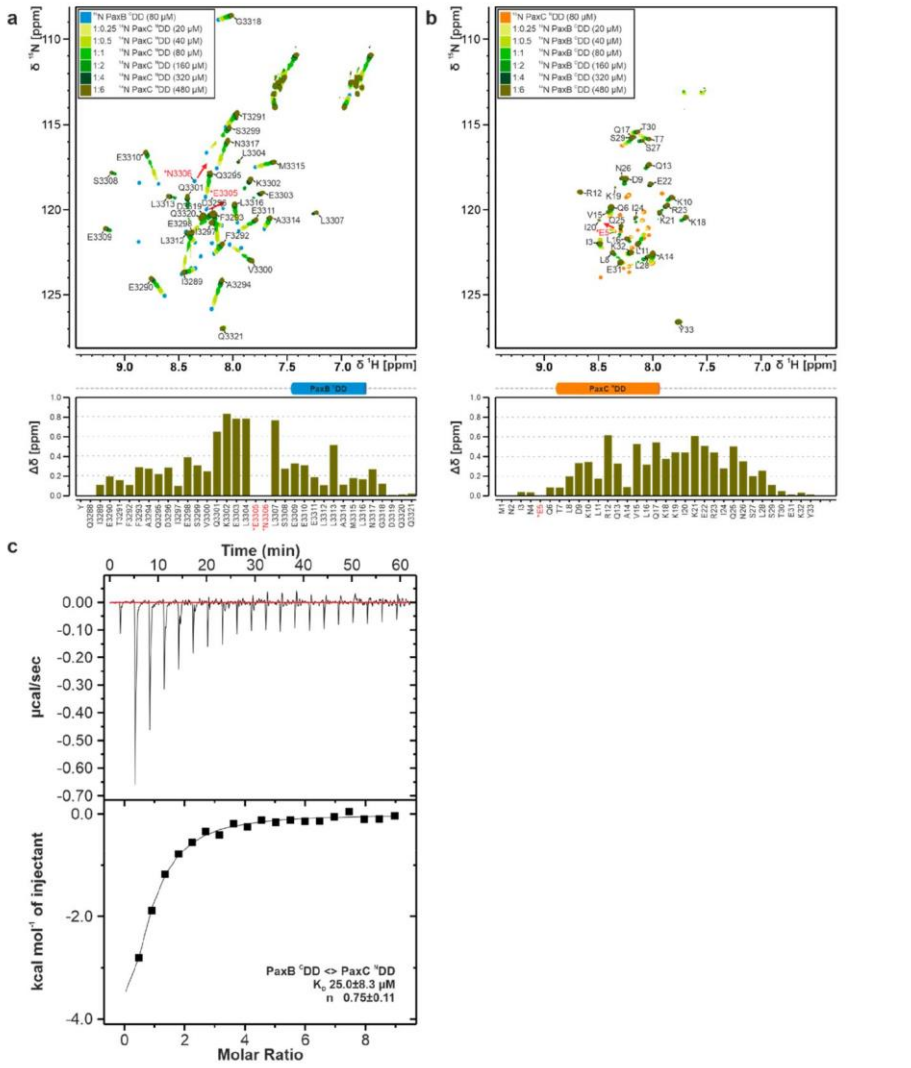


Figure 2. Overlay of $^{1}H,^{15}N$-HSQC spectra (top) of (a) 80 μM ^{15}N-labeled PaxB ^{C}DD in the absence (blue) and presence of increasing amounts of unlabeled PaxC ^{N}DD [green shades: 1:0.25, 1:0.5, 1:1, 1:2, 1:4, and 1:6] and a bar chart of chemical shift changes observed at the titration end point plotted against the residues of PaxB ^{C}DD (bottom). (b) 80 μM ^{15}N-labeled PaxC ^{N}DD in the absence (orange) and presence of increasing amounts of unlabeled PaxB ^{C}DD [green shades: 1:0.25, 1:0.5, 1:1, 1:2, 1:4, and 1:6] and bar chart of chemical shift changes observed at the titration end point plotted against the residues of PaxC ^{N}DD (bottom). The chemical shift derived secondary structure propensities in the unbound state are depicted above each residue for both DDs. For residues colored red and marked with an asterisk (*) no peak could be unambiguously identified at the titration end point [1:6 molar ratio] due to peak overlap. Nonetheless, the peaks of these PaxB/C $^{C/N}$DD residues in the unbound state are marked, and the direction of shift changes as observed in earlier titration points at lower contour levels is indicated by red arrows. (c) ITC thermogram and the derived binding curve for the titration of PaxB ^{C}DD and PaxC ^{N}DD are shown.

PaxC solved by nuclear magnetic resonance (NMR) spectroscopy in solution. The DD complex reveals a structurally new type of DD interaction. In addition to the elucidation of the solution structure of the DD complex, the protein−protein

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982−989

interaction was analyzed by point mutations, NMR titration, and isothermal titration calorimetry (ITC) experiments in order to identify amino acids, which are involved in the binding interface. Sequence alignments revealed that this type of DD interaction is widespread in NRPS and PKS systems.

RESULTS AND DISCUSSION

To gain additional insights into the potential architectures of NRPS derived DD complexes, the three subunit Pax NRPS from *Xenorhabdus bovienii* was analyzed (Figure S1a). This NRPS consists of three multimodular subunits (PaxA/B/C). The chemical structure of the product of this NRPS suggests that the PaxA subunit interacts exclusively with PaxB and PaxC interacts only with the PaxB subunit. Bioinformatic analysis of the three Pax polypeptides identified putative C-terminal docking domains in PaxA and PaxB as C-terminal extension segments of canonical PCP domains. Furthermore, putative N-terminal docking domains are present in PaxB and PaxC since here the canonical condensation domains are apparently N-terminally extended. However, the N-terminal docking domains of PaxB and PaxC differ from each other in length and show only minor sequence similarity. Similarly, the putative C-terminal docking domains of PaxA and PaxB differ from each other with regard to sequence and length (Figure S1b). This would agree well with the observed specificity of the intersubunit interactions in the Pax NRPS and the fact that this cluster synthesizes only a single product with a specific arrangement of the amino acid building blocks. However, putative docking domains with sequence similarity to those found in the Pax NRPS are present in related systems (Figure S2). Here, we focus on the interaction of the ^{C}DD of PaxB and the ^{N}DD of PaxC. Sequence alignments suggested that the PaxB ^{C}DD is ~34 amino acids and the PaxC ^{N}DD is ~33 amino acids long. It has been already experimentally verified that these docking domains are able to function as independent specific communication elements in multisubunit NRPS systems unrelated to the Pax NRPS.[13]

For biophysical interaction studies as well as structural investigations, the two peptides were overexpressed initially as SUMO-fusion proteins and purified to homogeneity by a combination of affinity chromatography and gel filtration as described in the Supporting Information. For concentration measurements, a non-native tyrosine residue was attached to the N-terminal end of the ^{C}DD, but this added amino acid is positioned in a region predicted to be unstructured and has no effect on the overall docking domain structure. The ^{N}DD already contains a tyrosine in its native sequence. The structural integrity of the purified peptides after removal of the SUMO-tag and subsequent purification was checked by HR-HPLC-ESI-MS (Figure S3).

The conformation of the individual ^{N}DD and ^{C}DD peptides was initially analyzed by circular dichroism (CD) spectroscopy. Both peptides yield CD spectra that can be interpreted as indicating a mixture of α-helical and unstructured conformations (Figure 1a). CD spectra for both peptides display the minima typical for α-helical proteins at ~222 and ~208 nm. However, the minimum at 208 nm is much more pronounced compared to the one at ~222 nm in contrast to what is normally observed for α-helical structures.[14] Furthermore, at wavelengths <200 nm, the CD signal is still negative as is typical for unstructured peptides. The α-helical preferences appear to be slightly more pronounced for the isolated ^{N}DD peptide.

To obtain more detailed insights into the conformational properties of the two peptides, we recorded ^{1}H,^{15}N-HSQC (Heteronuclear single quantum coherence) spectra of the unbound DDs (Figure 1b,c top). All observable amide signals in these spectra were assigned to their respective aa residues by applying the standard set of triple-resonance NMR experiments (HNCA, HN(CA)CO, HNCO, HN(CO)CA, HNCACB, and HN(CO)CACB).[15] The complete set of backbone (H_N, N_H, Cα, Cβ, and CO) chemical shift assignments was obtained for residue I3289−Q3321 of the PaxB ^{C}DD and residue I3−Y33 of the PaxC ^{N}DD and allowed the analysis of their secondary structure propensities using the chemical shift index (CSI) as implemented in TALOS-N[16] (Figure 1b,c bottom). For both peptides, regions with high α-helical propensities were identified. In the free PaxB ^{C}DD amino acids, E3309−L3316 show a particularly strong preference for α-helical conformations. In the PaxC ^{N}DD residues T7 to Q17 have a high and continuous α-helical propensity.

The interaction between the PaxB ^{C}DD and the PaxC ^{N}DD peptides was analyzed by NMR titration experiments in which one of the two peptides was ^{15}N-labeled, whereas the other one was unlabeled. In Figure 2a,b top, ^{1}H,^{15}N-HSQC spectra of the titration of ^{15}N-labeled PaxB ^{C}DD with unlabeled PaxC ^{N}DD and ^{15}N-labeled PaxC ^{N}DD with unlabeled PaxB ^{C}DD are shown. Furthermore, the DD interaction was studied thermodynamically in ITC experiments. According to the ITC thermograms, the PaxB ^{C}DD binds the PaxC ^{N}DD with a K_D of 25.0 ± 8.3 μM (Figure 2c). The K_D value was also confirmed by NMR based on the chemical shift changes observed in the titration experiments (Figure S4) yielding K_D's of 14.8 ± 2.0 μM (PaxB ^{C}DD addition to the PaxC ^{N}DD) and 23.9 ± 9.5 μM (PaxC ^{N}DD addition to the PaxB ^{C}DD).

Additional ITC measurements confirmed that the PaxC ^{N}DD is not capable of binding a peptide corresponding to the PaxA ^{C}DD as expected (data not shown).

Stepwise chemical shift changes or a gradual disappearance and reappearance of signals during the titrations can be observed for a large fraction of all observable signals. This shows that a ^{N}DD/^{C}DD complex is formed, which is in fast to intermediate exchange with the free peptides on the NMR time scale as expected on the basis of the K_D measured in the ITC and NMR experiments. A large number of signals for each peptide show chemical shift changes upon the addition of the unlabeled binding partner, suggesting that complex formation is connected to rather large structural changes (Figure 2a,b). The most pronounced chemical shift changes are observed for amino acids Q3301−L3307 of the PaxB ^{C}DD, which are unstructured in the free peptide according to the CSI (Figure 1b, bottom) in agreement with structural rearrangements. The assignments for most of the amide signals in the bound states of both DDs could be obtained in a straightforward manner by following the gradual shift changes during the titration experiments, and therefore, no additional triple resonance backbone assignment experiments were recorded.

The large structural changes in both peptides upon complex formation are also obvious from a CD spectrum recorded for a 1:1 mixture of the two peptides at a concentration of 120 μM, which is significantly higher than the observed K_D (~25 μM, Figure S6). This spectrum shows a much higher α-helical propensity compared to the spectra of the free peptides.

In order to investigate the structure of the DD complex and to get detailed insights into the structural basis for the DD

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982−989

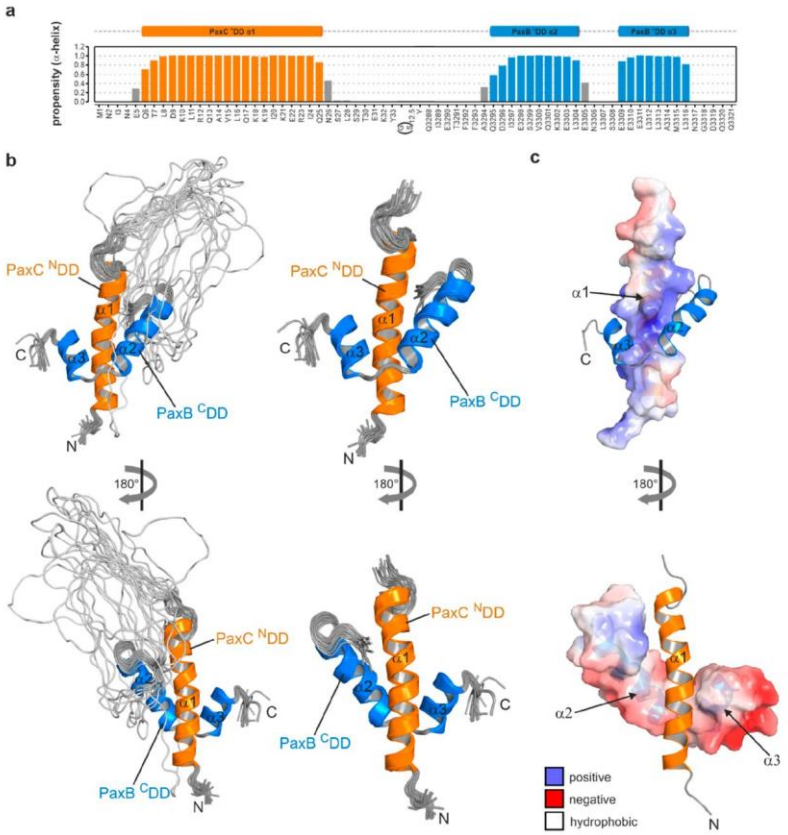


Figure 3. (a) Chemical shift index-derived secondary structure in the linked PaxC ^{N}DD (orange)-$(GS)_{12.5}$-PaxB ^{C}DD (blue) complex according to TALOS-N. The defined secondary structure (α-helical propensity $\geq$ 0.6) is given in a schematic representation above each residue. (b) Solution structure bundle of the 20 energy minimized conformers with the lowest target functions of the PaxC ^{N}DD-$(GS)_{12.5}$ (light gray)-PaxB ^{C}DD (same coloring as in (a)) artificially linked construct in two orientations (bottom/top) (rotated by 180° via the *y*-axis). For clarity reasons, the 25 amino acid long glycine−serine linker is not shown in the right figures. (c) Cartoon representation of the energy minimized structure of PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD (linker is hidden) with the lowest target function and the same color coding and orientation as in (b). The electrostatic surface potentials are mapped on the molecule surfaces, focusing on protein−protein interacting regions (α helices) of the DDs. The orientation is the same as in (a) with negatively charged surface areas colored in red, positively charged areas colored in blue, and white areas corresponding to hydrophobic surfaces.

interaction, both DDs were artificially connected via a glycine−serine linker. Different constructs varying in linker length (10, 25, and 30 residues) and domain order (^{N}DD-linker-^{C}DD or ^{C}DD-linker-^{N}DD) were screened. The ^{N}DD-linker-^{C}DD construct with a 25 amino acid long linker was chosen for structure determination since its ^{1}H,^{15}N-HSQC-spectrum most closely resembled the HSQC spectra of the unlinked peptides at the end points of the titration experiments (Figure S5). The small chemical shift difference between the linked construct and the unlinked complex under saturation conditions also confirms that the glycine−serine (GS)-linked DD pair resembles the native conformation of the DD complex and is not affected by the artificial linkage. A gel filtration experiment confirmed that the linked complex is monomeric as expected (Figure S7). Furthermore, an additional extension of the linker to 30 aa did not lead to additional changes in the NMR spectra. Most signals in the ^{1}H,^{15}N-HSQC-spectrum of the linked DD complex peaks could be assigned simply by comparison with the spectra recorded at the titration end for the noncovalent complexes. However, these assignments were confirmed by recording the standard set of triple resonance backbone assignment experiments, which also yielded the assignments for signals not assigned in the noncovalent complexes at the titration end point (PaxB ^{C}DD: E3305, N3306; PaxC ^{N}DD: E5).

Overall, the backbone assignment was 98% complete; the side chain aliphatic CH group assignment was 92% complete, and the aromatic CH group assignment was 78% complete for

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982−989

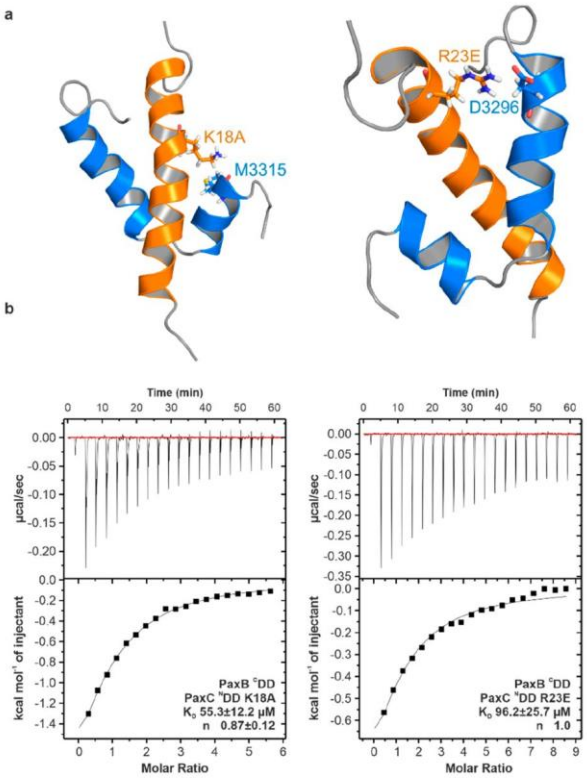


Figure 4. (a) Cartoon representation of PaxC ^{N}DD (orange)-$(GS)_{12.5}$-PaxB ^{C}DD (blue). Amino acids involved in the complex formation (K18, R23) due to the intermolecular interactions are highlighted in stick representation. (b) Raw ITC curves (top) and binding isotherms with a fitted curve (bottom) obtained for the titration of PaxB ^{C}DD with mutant PaxC ^{N}DD ligands. Both PaxC ^{N}DD mutants showed a decreased affinity (K_D of 55.3 μM [K18A], 96.2 μM [R23E]) in comparison to the wildtype PaxB ^{C}DD.

the GS-linked DD complex. The chemical shift derived secondary structure based on TALOS-N demonstrates that in the complex stable α-helices form, which is also in agreement with the CD spectrum of the GS-linked complex (Figure S6). One continuous α-helical region spanning residues Q6 to Q25 is found in the ^{N}DD, whereas two shorter stable α-helical segments (aa Q3295 to L3304 and E3309 to L3316) are present in the ^{C}DD (Figure 3a). 1685 nuclear Overhauser effect (NOE)-derived proton−proton distance restraints corresponding to an average of 25.5 distance restraints per DD residue were obtained from three-dimensional ^{13}C- and ^{15}N-edited NOESY-HSQC (nuclear Overhauser effect spectroscopy-heteronuclear single quantum coherence) experiments, leading to a NMR solution structure for the linked construct with a backbone root-mean-square deviation (RMSD) of 0.40 Å for all ordered residues (aa N4−E31, aa Q60−G90). The full structural statistics according to the recommendations of the NMR-VTF can be found in Table S4.[17] The structural bundle containing the 20 structures with the lowest NMR target function after structure calculation with CYANA (version 3.98)[18] and energy refinement by OPALp[19] has been deposited in the Protein Data Bank (PDB Accession Code: 6TRP, PaxC ^{N}DD-$(GS)_{12.5}$-Y-PaxB ^{C}DD). It represents the solution structure of the DD complex, which is shown in Figure 3b.

The cartoon representation of the complex structure is shown with (left) and without (right) the artificial GS-linker (Figure 3). In the DD complex, the PaxC ^{N}DD forms a long continuous α-helix α1 extending from amino acid N4 to N26. The PaxB ^{C}DD structure features two shorter α-helices, α2 (F3293−N3306) and α3 (E3309−N3317), separated by a short loop (L3307, S3308), which are arranged in a V-shaped structure. The structurally determined helix lengths (N4−N26, F3293−N3306, E3309−N3317) are in good agreement with the CSI determined α-helical propensity (Figure 3a: Q6−Q25, Q3295−L3304, E3309−L3316) which differ only in a range of ±2 residues. The appearance of a second α-helix in the PaxB ^{C}DD due to the intermolecular interaction in the complex also explains the particularly large chemical shift changes observed for the signals of residues Q3301−L3307.

The ^{N}DD helix α1 binds to the center of the V and is oriented antiparallel to the ^{C}DD helix α2 and parallel to ^{C}DD

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982−989

helix α3. The packing of the helices against each other is largely dominated by interactions between hydrophobic side chains (Figure S8). However, a number of polar interactions at the periphery of the interaction surface were also identified (Figure 4a). For instance, the side chain amino group of K18 of the ^{N}DD forms an electrostatic interaction with the backbone carbonyl group of M3315 of the ^{C}DD (N^{ζ}–O distance of 4.2 ± 0.7 Å). Additionally, a salt bridge is formed between R23 of the ^{N}DD and D3296 of the ^{C}DD (N^{η}–O^{δ} distance of 2.9 ± 0.2 Å). An additional salt bridge (N^{ζ}–O^{e} distance of 3.5 ± 0.8 Å) connects the side chains of K19 (^{N}DD) and E3303 (^{C}DD). While the presence of these salt bridges could not be confirmed by the direct observation of NOEs involving the lysine H_N^{ζ} atoms or the arginine H_N^{η} atoms, which were not observable due to the exchange with the solvent, a number of NOEs involving aliphatic side chain atoms to the remainder of the structure define the specific side chain orientations in agreement with the observed salt bridges.

To further evaluate the identified interaction interface in the DD complex structure, different amino acids were exchanged by site-directed mutagenesis and the binding properties of these constructs were analyzed thermodynamically by ITC (Figure 4b).

The exchange of K18A leads to the disruption of an ion–dipole interaction formed by lysine NH_3^+ with the oxygen of the backbone carbonyl atom of M3315. Accordingly, a decreased affinity (K_D of 55.3 ± 12.2 μM) in comparison to the native complex (K_D of 25.0 ± 8.3 μM) was observed. In another PaxC ^{N}DD mutant (R23E), the salt bridge between the anionic carboxylate group of D3296 and the cationic guanidinium group of R23 was eliminated. Again, a decreased affinity (K_D of 96.2 ± 25.7 μM) was determined for the mutant. Nonetheless, neither the disruption of the ion–dipole nor the salt-bridge intermolecular interaction could inhibit the complex formation completely. Thus, a combination of different interaction types including hydrophobic interactions, salt bridges, and hydrogen bonds contribute to maintain a stable intermolecular interaction.

Overall, the DD interaction between the PaxB ^{C}DD and the PaxC ^{N}DD is thermodynamically dominated by hydrophobic interactions (Figures 3c and S8). A few strategically placed salt bridges and polar interactions most likely contribute to the specificity of the interactions.

The structural analysis presented here describes a new type of DD complex architecture based on a three-helical bundle, which has not been previously described for NRPS systems or for PKS complexes. However, sequence alignments of different C- and N-terminal docking domains ($^{C/N}$DDs), identified by BLAST,[20] suggest that this type of DD interaction should be present both in other NRPS and PKS systems (Figure S2). In NRPS systems, the respective ^{N}DDs and ^{C}DDs are most commonly attached to C domains and T domains, respectively, whereas they are connected to acyl carrier proteins (ACPs) and oxidoreductase (OxRed) domains in PKS systems. In comparison to previously described PKS DD[6,8] structures defined as Class 1 DDs,[9] the PaxB/C $^{C/N}$DD complex lacks any dimerization element (Figure S9). This is in agreement with the monomeric character of multiprotein NRPSs. The unique feature of the new DD type presented in this work is its overall arrangement in which a V-shaped C-terminal docking domain consisting of two α-helices connected by a short linker wraps around a single α-helical N-terminal docking domain. It should be mentioned that the V-shaped ^{C}DD interaction platform is only formed if the respective DD partner is present. The overall coiled-coil packing is therefore generated by a combination of the PaxB ^{C}DD and PaxC ^{N}DD, whereas Class 1 DDs are characterized by a homodimer of single α-helical ^{N}DDs forming a stand alone, coiled-coil motif, which is then extended peripherally by the docking of two copies of the ^{C}DD, also consisting of a single α-helix (Figure S9[6]). The Class 1 DD is further elongated by two α-helices that function as a dimerization element. Obviously, there is an architectural divergence of the DD type described here that features a three α-helical bundle in contrast to the four α-helical bundles characteristic of Class 1 DDs. Nonetheless, besides structural differences, the DDs described in this work and the previously described Class 1 DDs share certain properties, for example, that the interaction is largely dominated by hydrophobic interactions supported by charge–charge interactions and a rather low affinity.

Additionally, as previously mentioned, no interaction between the PaxA ^{C}DD and PaxB ^{N}DD was detectable. The PaxA ^{C}DD is rather short with a length of only ~20 aa. This suggests that the connection between the PaxA and PaxB subunit relies on a different type of protein interaction, which is incompatible with the PaxB/PaxC DDs. This supports the hypothesis that the respective DDs are responsible for maintaining the specific sequential order of the three proteins in this NRPS architecture required by the collinearity rule for the observed product composition.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.9b01022.

Supporting tables summarizing the bacterial strains, plasmids, and oligonucleotides used in this work and the structural statistics for the NMR solution structure of the DD complex; Supporting figures describing the PAX peptide producing NRPS, sequence alignments of related DDs identified by BLASTp, HR-HPLC-ESI-MS data, NMR based data (K_D determination, chemical shift comparison), CD spectra of the DDs, the SEC chromatogram of the monomeric DD complex and structural aspects of the new DD type; description of the genetic engineering of plasmids, protein expression and purification conditions, NMR spectroscopy, structure calculation, and ITC, CD, and HR-HPLC-ESI-MS experiments (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Helge B. Bode – *Molecular Biotechnology, Institute of Molecular Biosciences and Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt am Main, Germany;* orcid.org/0000-0001-6048-5909; Email: h.bode@bio.uni-frankfurt.de

Jens Wöhnert – *Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany;* Email: woehnert@bio.uni-frankfurt.de

https://dx.doi.org/10.1021/acschembio.9b01022
ACS Chem. Biol. 2020, 15, 982–989

Authors

Jonas Watzel − *Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany*

Carolin Hacker − *Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany*

Elke Duchardt-Ferner − *Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.9b01022

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the LOEWE program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hesse and was conducted within the framework of the MegaSyn Research Cluster in the laboratories of H.B.B. and J.W. All NMR experiments were carried out at the Center for Biomolecular Magnetic Resonance (BMRZ) at Goethe-University Frankfurt, which is supported by the state of Hesse.

■ REFERENCES

(1) Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., and Thomas, M. G. (2008) Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharmaceutics 5*, 191−211.

(2) Hahn, M., and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. U. S. A. 101*, 15585−15590.

(3) Chiocchini, C., Linne, U., and Stachelhaus, T. (2006) In vivo biocombinatorial synthesis of lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic complex. *Chem. Biol. 13*, 899−908.

(4) Hahn, M., and Stachelhaus, T. (2006) Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl. Acad. Sci. U. S. A. 103*, 275−280.

(5) Hacker, C., Cai, X., Kegler, C., Zhao, L., Weickhmann, A. K., Wurm, J. P., Bode, H. B., and Wöhnert, J. (2018) Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nat. Commun. 9*, 4366.

(6) Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. (2003) The structure of docking domains in modular polyketide synthases. *Chem. Biol. 10*, 723−731.

(7) Dorival, J., Annaval, T., Risser, F., Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B., and Weissman, K. J. (2016) Characterization of Intersubunit communication in the virginiamycin trans-acyl transferase polyketide synthase. *J. Am. Chem. Soc. 138*, 4155−4167.

(8) Buchholz, T. J., Geders, T. W., Bartley, F. E., Reynolds, K. A., Smith, J. L., and Sherman, D. H. (2009) Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol. 4*, 41−52.

(9) Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013) Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol. 20*, 1340−1351.

(10) Richter, C. D., Nietlispach, D., Broadhurst, R. W., and Weissman, K. J. (2008) Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol. 4*, 75−81.

(11) Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T., and Drennan, C. L. (2016) Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U. S. A. 113*, 12432−12437.

(12) Fuchs, S. W., Proschak, A., Jaskolla, T. W., Karas, M., and Bode, H. B. (2011) Structure elucidation and biosynthesis of lysine-rich cyclic peptides in Xenorhabdus nematophila. *Org. Biomol. Chem. 9*, 3130−3132.

(13) Cai, X., Zhao, L., and Bode, H. B. (2019) Reprogramming Promiscuous Nonribosomal Peptide Synthetases for Production of Specific Peptides. *Org. Lett. 21*, 2116−2120.

(14) Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc. 1*, 2876−2890.

(15) Sattler, M. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc. 34*, 93−158.

(16) Shen, Y., and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR 56*, 227−241.

(17) Montelione, G. T., Nilges, M., Bax, A., Güntert, P., Herrmann, T., Richardson, J. S., Schwieters, C., Vranken, W. F., Vuister, G. W., Wishart, D. S., Berman, H. M., Kleywegt, G. J., and Markley, J. L. (2013) Recommendations of the wwPDB NMR Validation Task Force. *Structure 21*, 1563.

(18) Güntert, P. (2009) Automated structure determination from NMR spectra. *Eur. Biophys. J. 38*, 129.

(19) Koradi, R., Billeter, M., and Güntert, P. (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun. 124*, 139−147.

(20) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403−410.

1.3. Supporting information

A New Docking Domain Type in the Peptide-Antimicrobial-Xenorhabdus Peptide Producing nonribosomal Peptide Synthetase from *Xenorhabdus bovienii*

Jonas Watzel[1], Carolin Hacker[2], Elke Duchardt-Ferner[2], Helge B. Bode[1,3,4], Jens Wöhnert[2], Goethe University, Frankfurt/ Germany

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

2 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

3 Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

4 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

TABLE OF CONTENTS

SUPPORTING TABLES

Table S1. Strains used in this work.

Strain	Genotype / NRPS	Reference
E. coli BL21-Gold(DE3)	*E. coli* B F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte / -	(*1*)
Xenorhabdus bovienii SS-2004	wild type / *paxS (2)*	(*3*)

Table S2. Plasmids used in this work.

Plasmids	Description	Reference
pET11a-modified	modified from pET11a, the operon under the control of T7 promoter was modified by introduction of N-terminal His_6-smt3 tag, amp^R	(*4*)
pJW30	vector, the PaxB ^{C}DD was fused C-terminally to *smt3* into pET11a-modified, under control of T7 promoter, amp^R	this study
pJW31	vector, the PaxC ^{N}DD was fused C-terminally to *smt3* into pET11a-modified, under control of T7 promoter, amp^R	this study
pJW35	vector, the PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD was fused C-terminally to smt3 into pET11a-modified, under control of T7 promoter, ampR	this study

Table S3. Oligonucleotides used in this work.

Plasmids	Oligo-nucleotide	Sequence (5´→3´; overlapping ends)	Targeting DNA fragment	Template
pJW30 (PaxB ^{C}DD)	pET11a_FW	TAAGGATCCGGCTGCTAAC	pET11a-modified vector backbone	pet11a-modified
	pET11a_Smt3_RV	ACCACCAATCTGTTCACGA		pet11a-modified
	jw0017_FW	CATCGTGAACAGATTGGTGGTTATCAAATTGAAACT TTTTTCGC	PaxB ^{C}DD insert	*Xenorhabdus bovienii* SS-2004
	ck0045_RV	TTTGTTAGCAGCCGGATCCTTATTGTTGATCTCCATT TAACATGG		*Xenorhabdus bovienii* SS-2004
pJW31 (PaxC ^{N}DD)	pET11a_FW	TAAGGATCCGGCTGCTAAC	pET11a-modified vector backbone	pet11a-modified
	pET11a_Smt3_RV	ACCACCAATCTGTTCACGA		pet11a-modified
	ck0042_FW	CATCGTGAACAGATTGGTGGTATGAACATAAATGAA CAAACTTTGG	PaxC ^{N}DD insert	*Xenorhabdus bovienii* SS-2004
	ck0016_RV	TTTGTTAGCAGCCGGATCCTTAATATTTTTCAGTACT CAGGCTGTTC		*Xenorhabdus bovienii* SS-2004
pJW35 (PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD)	pET11a_FW	TAAGGATCCGGCTGCTAAC	pET11a-modified vector backbone	pet11a-modified
	pET11a_Smt3_RV	ACCACCAATCTGTTCACGA		pet11a-modified
	ck0042_FW	CATCGTGAACAGATTGGTGGTATGAACATAAATGAA CAAACTTTGG	PaxC ^{N}DD-$(GS)_{7.5}$ insert#1	*Xenorhabdus bovienii* SS-2004
	jw0022_RV	ACCTGAACCACTACCCGAACCCGATCCGGAACCTG AACCACTACCATATTTTTCAGTACTCAGGCTGTTC		*Xenorhabdus bovienii* SS-2004
	jw0023_FW	GGTAGTGGTTCAGGTTCCGGATCGGGTTCGGGTAG TGGTTCAGGTTATCAAATTGAAACTTTTTTCGCC	$(GS)_{7.5}$-PaxB ^{C}DD insert#2	*Xenorhabdus bovienii* SS-2004
	ck0045_RV	TTTGTTAGCAGCCGGATCCTTATTGTTGATCTCCATT TAACATGG		*Xenorhabdus bovienii* SS-2004

Table S4. Structural statistics for the NMR solution structures of PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD. For the calculation PSVS 1.5 (*5*) was used.

	PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD.
Conformationally-restricting experimental constraints[a]	
Total NOE distance restraints	1685
intraresidue |i = j|	420
sequential |i - j| = 1	517
medium-range 1 <|i - j| < 5	568
long-range |i - j| ≥ 5	180
NOE constraints per restrained residue[b]	25.5
Dihedral angle restraints (TALOS-N)	94
Total number of restricting constraints[b]	1779
Total number of restricting constraints per restrained residue[b]	27.0
Total number of long-range constraints per restrained residue[b]	2.7
Residual constraint violations[a,c]	
Average number of distance violations per structure	
0.1-0.2 Å	4.9
0.2-0.5 Å	0
>0.5 Å	0
Average number of dihedral angle violations per structure	
1-10	13.15
>10	0
Model quality (ordered residues)	
RMSD backbone atoms (Å)	0.4
RMSD heavy atoms (Å)	0.7
RMSD bond lengths (Å)	0.010
RMSD bond angles (°)	2.2
CYANA target function	2.24±0.35
Ramachandran Plot Statistics from Richardson's lab	
Most favored regions	90.5 %
Allowed regions	8.9 %
Disallowed regions	0.6 %
Global quality scores (raw/Z score)	
Verify3D	0.29/-2.73
Prosall	0.74/0.37
PROCHECK (φ-ψ)	-0.32/-0.94
PROCHECK (all)	-0.63/-3.73
MolProity clash score	7.36/0.26
Model contents	
Ordered residue ranges (user defined)	4-31, 60-90
Total no. of residues	93
Biological Magnetic Resonance Bank (BMRB) accession number	34469
Protein Databank (PDB) ID code	6TRP

[a] analysed for residues 1 to 93, [b] there are 66 residues with conformationally restricting constraints, [c] average distance violations were calculated using the sum over r^{-6}

SUPPORTING FIGURES

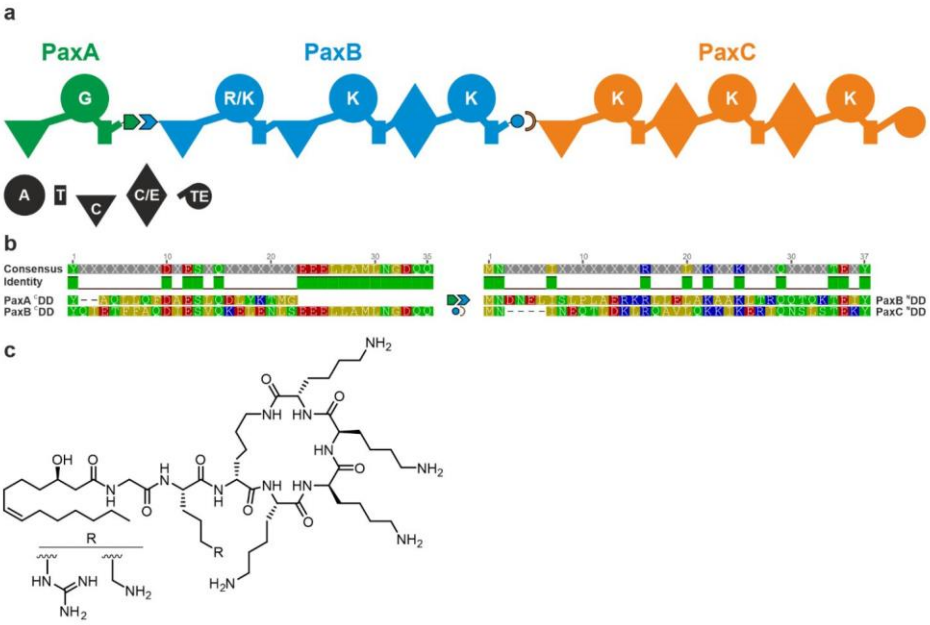


Figure S1. Peptide-antimicrobial-Xenorhabdus (PAX) producing NRPS. (a) Schematic illustration of the peptide-antimicrobial-Xenorhabdus (PAX) producing NRPS of *Xenorhabdus bovienii*. The analysed DD pair (boxed) was artificially linked by an 25 aa long glycine-serine (GS) linker. The NRPS consists of three polypeptides (PaxA/B/C) with an unidirectional interaction order. For domain assignment the following symbols are used: adenylation domain (A, large circle); thiolation domain (T, rectangle); condensation domain (C, triangle); dual condensation/epimerization domain (C/E, diamond); thioesterase domain (TE, C-terminal small circle). (b) Alignment of *Xenorhabdus bovienii* PaxA/B ^{C}DDs and PaxB/C ^{N}DDs. The amino acids are coloured in respect to their polarity. The Alignment was performed using the multiple alignment program MUSCLE (default parameters). (*6, 7*) (c) Structure of two major PAX peptides.

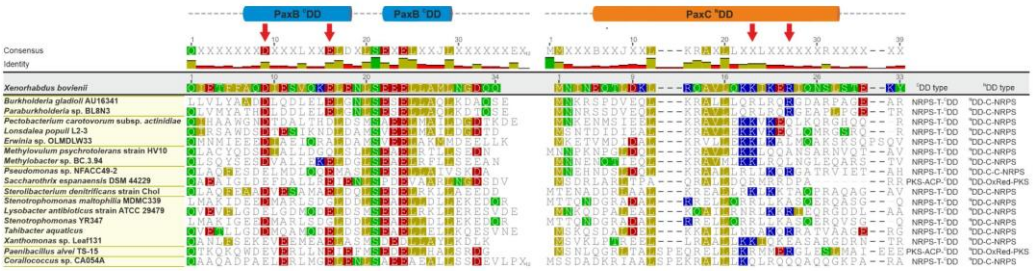


Figure S2. Alignment of selected NRPS/PKS ^{C}DDs and ^{N}DDs identified by BLASTp search using *Xenorhabdus bovienii* PaxB/C ^{C}DD/^{N}DDs with parts of the attached T or C domains as query sequences. Secondary structural elements are depicted above and arrows indicating salt bridge forming residues. Both annotations refer to the solved docking domain complex structure. The bacteria in the alignment are grouped by their order including (from top to bottom): Burkholderiales, Enterobacteriales, Methylococcales, Pseudonocardiales, Rhodocyclales, Xanthomonadales, Bacillales and Myxococcales. This docking domain type is widespread in the field of PKS and NRPS systems, connecting most commonly thiolation (T) and condensation (C) domains [NRPSs] besides acyl carrier proteins (ACPs) and oxidoreductase (OxRed) domains [PKSs]. The consensus alignment is depicted above the alignment sequences with a 50% threshold implemented. Amino acid agreements are highlighted by comparison with the *Xenorhabdus bovienii* reference sequences. Alignments were performed using the multiple sequence alignment program MUSCLE (default parameters). (*6*, *7*)

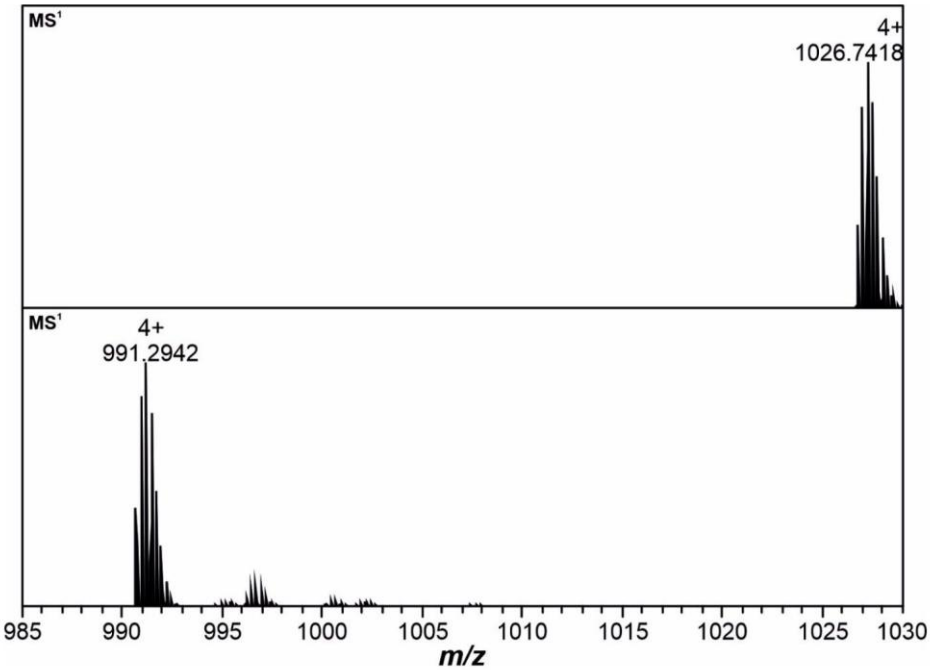


Figure S3. HR-HPLC-ESI-MS analysis of purified proteins. Displayed are the *m/z*-values of the average protein masses (PaxB ^{C}DD (top) and PaxC ^{N}DD (bottom)) of the 4+ charge states (MS^1). *m/z* 1026.7418 corresponds to PaxB ^{C}DD ($m/z_{theoretical}$ 1026.7440; Δppm 2.1) and *m/z* 991.2942 corresponds to PaxC ^{N}DD ($m/z_{theoretical}$ 991.2993; Δppm 5.1) in MS^1.

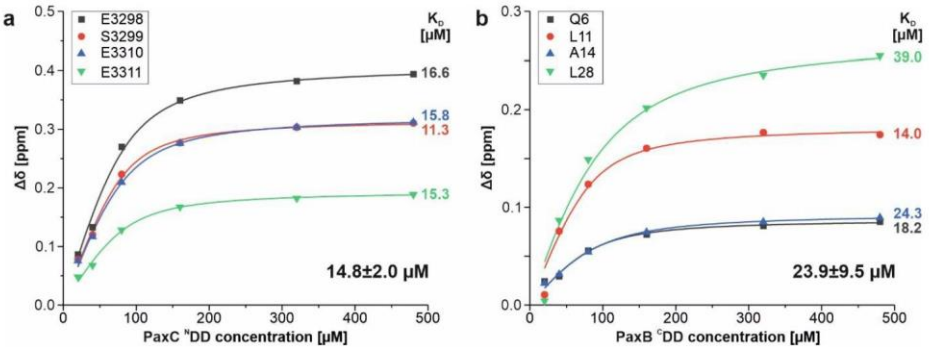


Figure S4. K_D determination based on NMR chemical shift data. The chemical shift changes caused by peptide binding are plotted against the respective peptide concentration for the (a) PaxB ^{C}DD and (b) PaxC ^{N}DD. The mean K_D values and standard deviations are given for residues (PaxB ^{C}DD: E3298, S3299, E3310, E3311; PaxC ^{N}DD: Q6, L11, A14, L28) which are all in the fast exchange regime.

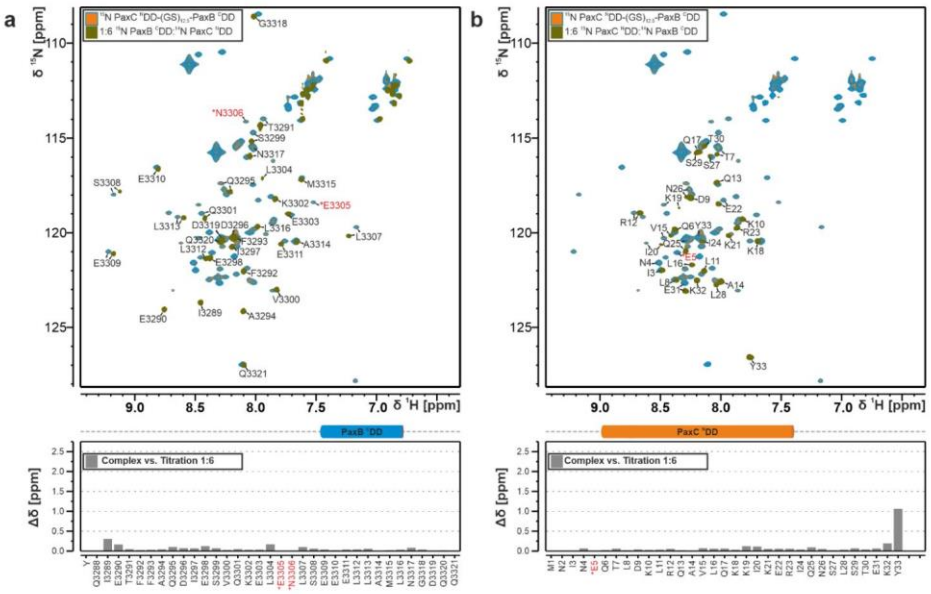


Figure S5. Chemical shift comparison of the unlinked and the GS-linked DD complex. Overlay of $^{1}H,^{15}N$-HSQC spectra (top) of (a) 80 µM ^{15}N-labeled PaxB ^{C}DD upon addition of a six-fold molar excess of unlabelled PaxC ^{N}DD and ^{15}N-labeled PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD; bar chart of chemical shift differences between titration endpoint [1:6 molar ratio] and DD complex plotted against the residues of PaxB ^{C}DD (bottom). (b) 80 µM ^{15}N-labeled PaxC ^{N}DD upon addition of a six-fold molar excess of unlabelled PaxB ^{C}DD and ^{15}N-labeled PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD; bar chart of chemical shift differences between titration endpoint [1:6 molar ratio] and DD complex plotted against the residues of PaxC ^{N}DD (bottom). For residues coloured red and marked with an asterisk (*) no peaks could be clearly identified at the titration endpoint due to peak overlap but these were easily identified in the GS-linked complex.

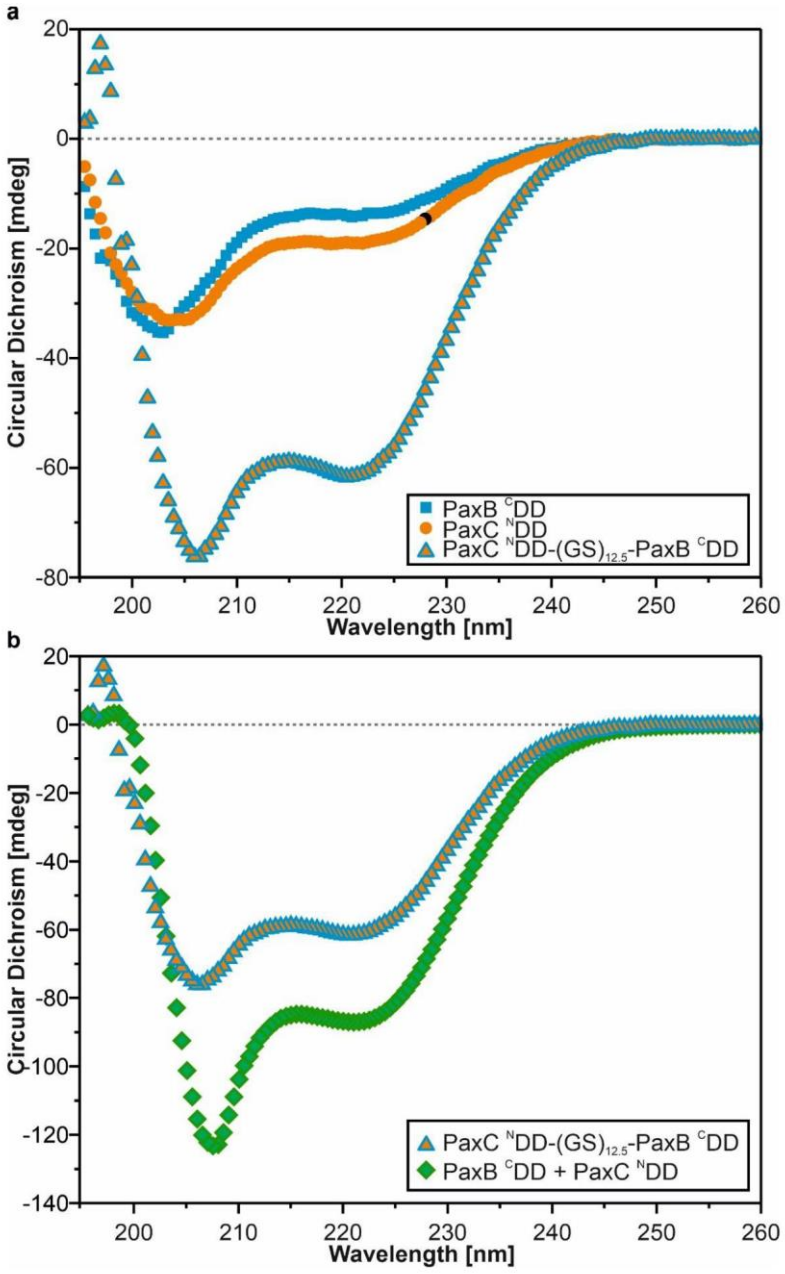


Figure S6. Overlay of circular dichroism spectra. (a) 50 μM PaxB/C $^{C/N}$DD and PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD samples and (b) of 50 μM PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD and unlinked PaxB ^{C}DD:PaxC ^{N}DD (120 μM of each DD) samples.

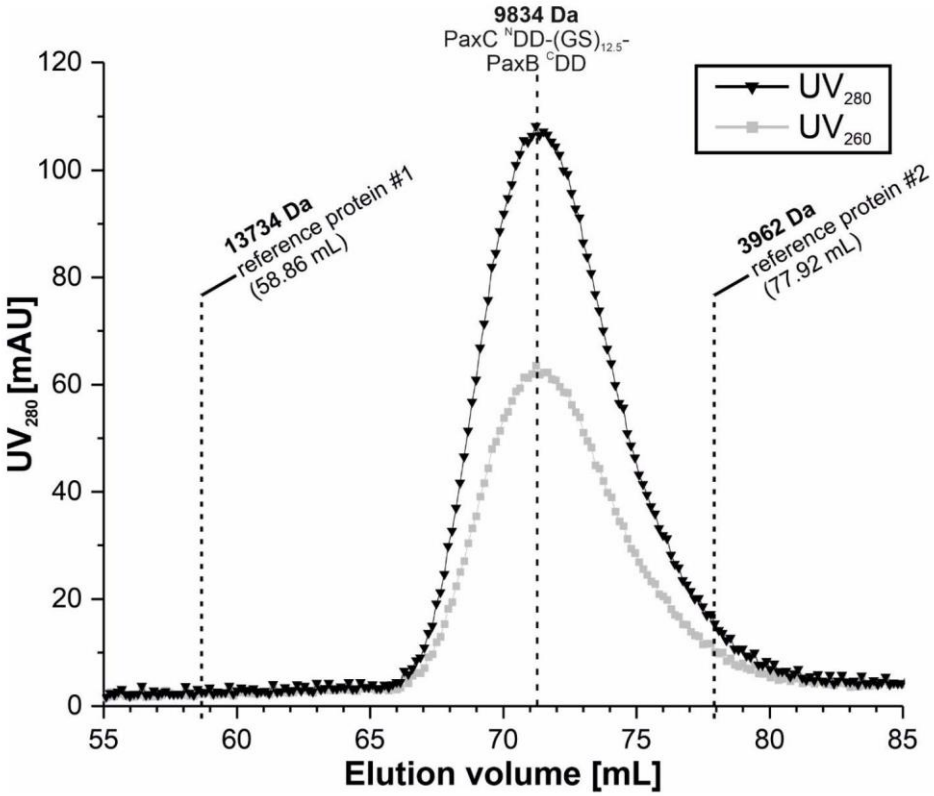


Figure S7. Chromatogram of the size exclusion chromatography (SEC) of the PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD complex. In dashed lines the elution volume of two reference proteins (#1 & #2) and the GS-linked DD complex with the corresponding molecular weights is indicated.

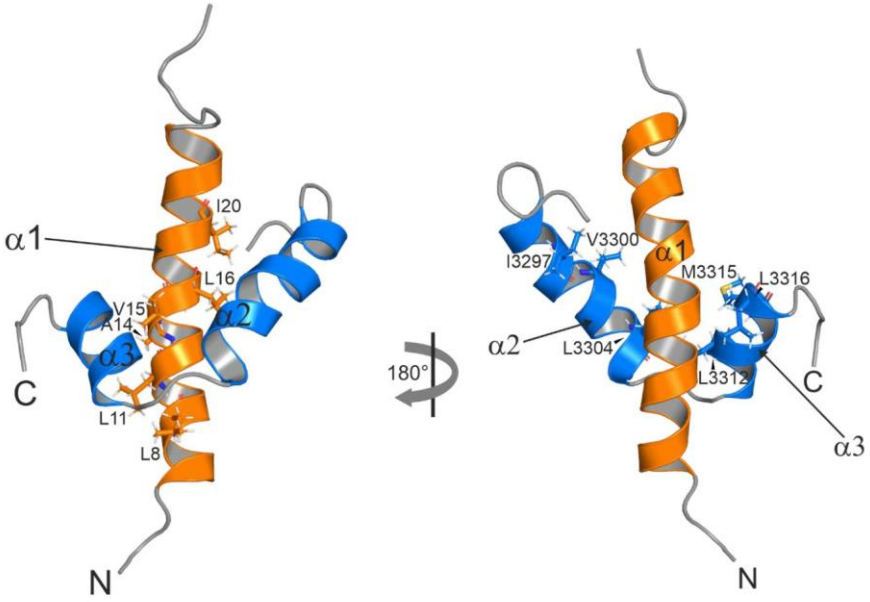


Figure S8. Hydrophobic interface of DD complex. Cartoon representation of the energy minimized structure of PaxC ^{N}DD (orange)-$(GS)_{12.5}$-PaxB ^{C}DD (blue) (linker is hidden) with the lowest target function. The hydrophobic residues which are involved in the DD interaction are highlighted as stick models.

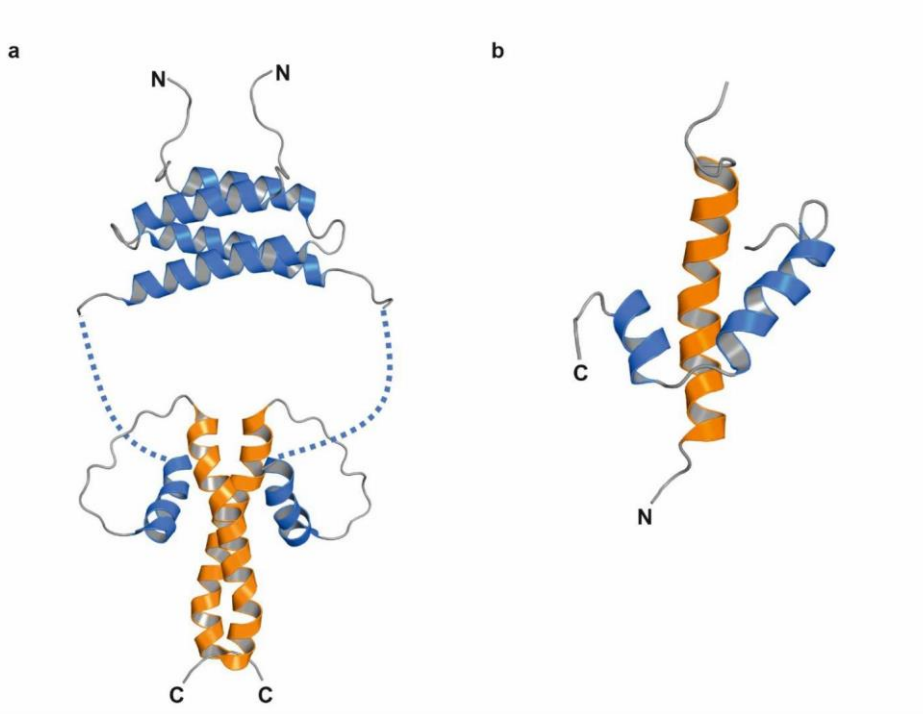


Figure S9. Docking Domain Class comparison. (a) NMR structure of a complex of covalently fused Class 1a docking domains (PDB: 1PZQ, 1PZR) (*8*) from a *cis*-AT PKS. The long, flexible linker connecting the second and third α-helices of the ^{C}DD is represented as a dashed line. (b) NMR structure of the newly identified Docking Domain type from the PAX peptide producing NRPS (GS-linker is hidden). C and N indicate C- and N-termini, respectively. The ^{C}DDs are coloured blue and the ^{N}DDs orange.

MATERIAL AND METHODS

General molecular biology

Molecular biology techniques like plasmid DNA preparation, transformation, restriction digestion and DNA gel electrophoresis, were adapted from standard protocols (*9*). Isolation of genomic DNA was carried out according to the manufacturer's instructions (QIAGEN). Q5High-Fidelity DNA-Polymerase (New England Biolabs) was used for PCR amplifications following the guide of the producer. PCR primers used in this study are listed in Table S3. All the plasmids (Table S2) generated in this study were constructed via Gibson assembly (*10*). The basic cloning was performed directly in the protein expression host *E. coli* BL21-Gold(DE3) (Agilent Technologies).

Expression and purification of DDs and DD complexes

For structure determination, $^{C/N}$DDs from *X. bovienii* (SS-2004) PaxB/C as well as the covalently linked ^{N}DD-^{C}DD complex and ^{N}DD mutants were heterologously expressed in *E. coli* BL21-Gold(DE3) under the control of a T7 promoter. The coding sequences were cloned into a modified pET11a vector containing an N-terminal His_6-SUMO tag, which allows cleaving off the tagged SUMO protein by ULP1 treatment. DNA fragments encoding PaxC ^{N}DD and PaxB ^{C}DD were linked with a 25 aa long GS linker in between. The resulting constructs were grown in uniformly ^{15}N and ^{15}N,^{13}C M9 minimal media containing 1 g L^{-1} $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) or 1 g L^{-1} $^{15}NH_4Cl$ and 2.5 g L^{-1} $^{13}C_6$-d-glucose (Cambridge Isotope Laboratories) and 100 mg mL^{-1} ampicillin. For ITC measurements, proteins were expressed in *E. coli* BL21-Gold (DE3) using LB medium. Protein expression was induced at an OD_{600} of 0.6-0.8 with 1 mM IPTG overnight at 20 °C. After expression, cells were lysed by sonication in lysis buffer containing 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 0,5 µl pure Benzonase (Novagen), and

protease inhibitor (Roche). The lysate was cleared by centrifugation (30 min, 7500 × *g*, 4 °C) and the supernatant was passed through a Bio-Scale Mini Profinity Ni-charged IMAC Cartridge (Biorad) using HisTrap-buffer (50 mM sodium phosphate buffer (pH 8) with 150 mM NaCl and 500 mM imidazole for elution). The His_6-SUMO fusion tag was cleaved off by Ulp1 proteolytic digestion in dialysis buffer (50 mM Tris HCl, pH 8, 50 mM NaCl) and removed by a second purification step with the Ni-charged IMAC Cartridge using HisTrap-buffer. All three proteins (^{N}DD/^{C}DD and the ^{N}DD-^{C}DD complex) were further purified via SEC on a HiPrep 16/60 Sephacryl S-100 High Resolution column (GE Healthcare) in SEC buffer composed of 50 mM sodium phosphate buffer (pH 6.5) with 100 mM NaCl.

Construction of plasmids encoding modified DDs

DDs with amino acid exchanges were generated by site-directed mutagenesis using phosphorylated primers coding for the modified sequences. PCRs based on the wildtype DD expression plasmid were performed and the template was degraded by DpnI (Thermo Scientific) digestion. Finally, the PCR product with the specifically introduced mutation was ligated via T4 DNA ligase.

NMR spectroscopy

For NMR measurements, the DDs (0.5–1 mM) were prepared in 50 mM sodium phosphate buffer pH 6.5, 100 mM NaCl, and 10% D_2O. For NMR titration experiments, a protein concentration of 80 µM was used. NMR spectra were recorded at 20 °C on Bruker AVANCE III 600, 700, 800, and 950 MHz spectrometers equipped with cryogenic triple resonance probes. The proton chemical shifts were internally referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid and the heteronuclear ^{13}C and ^{15}N chemical shifts were indirectly referenced with the appropriate conversion factors (*11*). The standard set of triple resonance experiments

(HNCO, HN(CA)CO, HNCACB) was used for the backbone resonance assignments of the ^{C}DD of PaxB (*12*). For the PaxC ^{N}DD and the ^{N}DD-^{C}DD linker construct, BEST-TROSY versions of the triple resonance spectra were used (*13*). Shaped proton pulses with a bandwidth of 5.0 ppm centered at 8.5 ppm were used. The delay between scans was set to 0.3 s in all experiments. For side chain resonance assignment, 3D HBHA(CO)NH, (H)CCH-TOCSY, and H(C)CH-TOCSY experiments were used. All spectra were recorded and processed using Bruker TopSpinTM 3.5 and analyzed using the programs CARA (*14*) (www.nmr.ch) and CcpNmr Analysis (*15*). For titration experiments with NMR, the concentrations of the ^{C}DD peptide ((PaxB ^{C}DD, YQIETFFAQDIESVQKELENLSEEELLAMLNGDQQ (35 aa)); and ^{N}DD peptide ((PaxC ^{N}DD, MNINEQTLDKLRQAVLQKKIKERIQNSLSTEKY (33 aa)), were determined with UV-Vis spectroscopy. Therefore, a tyrosine was attached to the N-terminus of the ^{C}DD. For titration experiments, ^{1}H,^{15}N HSQCs or ^{1}H,^{15}N BEST-TROSY-HSQCs were recorded after the stepwise addition of lyophilized, unlabeled $^{C/N}$DD (25–480 µM) to a 80 µM ^{15}N-labeled $^{N/C}$DD protein sample. Before lyophylization the respective unlabeled peptide was rebuffered in ddH_2O. To evaluate NMR titration experiments, the chemical shifts were determined using the peak picking function of CcpNmr Analysis (*15*). The chemical shift differences were calculated using the following function (*16*).

$$(1)\ \Delta\delta = \sqrt{\Delta\delta_H^2 + \left(\frac{\Delta\delta_N}{6.5}\right)^2}$$

For the K_D determination based on the NMR titration experiments a quadratic saturation binding equation (*17*) was fitted to the concentration-dependent chemical shift changes of the relevant shifting peaks:

$$(2)\ \Delta\delta_{obs} = \Delta\delta_{max}\frac{[L_0] + [P_0] + K_D - \sqrt{([L_0] + [P_0] + K_D)^2 - 4[L_0][P_0]}}{2[P_0]}$$

Structure calculation

^{15}N-nuclear Overhauser spectroscopy (NOESY)-HSQC, ^{13}C-NOESY-HSQC (aliphatic carbons), and ^{13}C-NOESY- HSQC (aromatic carbons) experiments in H_2O with mixing times of 250 ms were used to obtain distance restraints. The TALOS-N server was used to generate torsion angle restraints (*18*) based on backbone H, N, Cα, Cβ, and CO chemical shifts. Peak picking and NOE assignment was performed with the ATNOS/CANDID module in UNIO (*19*) in combination with CYANA (*20, 21*) using the 3D NOESY spectra. To correct falsely picked artefacts, the peak lists were reviewed manually and corrected. Distance restrains were obtained using the automated NOE assignment and structure calculation protocol available in CYANA (version 3.98) (*21*). An assignment of 87% of the observable NOESY crosspeaks for all NOESY spectra was achieved. Restrained energy refinement with OPALp (*22*) and the AMBER94 force field (*23*) of the 20 structures with the lowest target function was carried out. This set of CYANA generated, energy minimized structures with the lowest target functions were validated with the Protein Structure Validation Software (Table S4) suite1.5 (*5*). Electrostatic surface potential calculations were conducted with the PDB2PQR web server (*24*) using the PARSE force field and visualized with the APBS plug-in (*25*) for PyMOL with a threshold for electrostatic potential shading from -1 kT/e to +1 kT/e (k = Boltzmann's constant, T = absolute temperature, and e = electron charge (The PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC). All figures of structures were prepared with PyMOL.

Isothermal Titration Calorimetry

ITC measurements were performed at 20 °C in 50 mM sodium phosphate buffer, pH 6.5, and 100 mM NaCl using a MicroCal iTC200 (Malvern Instruments) calorimeter. In all experiments, 50 µM of the respective docking partner $^{N/C}$DDs were provided in the reference cell and the interaction partner, in a suitable concentration (PaxB ^{C}DD:

2000 µM, PaxC ^{N}DD R23E: 2100 µM, PaxC ^{N}DD K18A: 1380 µM), was added stepwise. ITC experiments started with an initial delay time of 120 s. The first injection of 0.2 µl was followed by 19 serial injections of 2 µl, separated by an interval of 180 s. For each experiment, the reference power was set to 11 µcal/s, stirring speed to 750 rpm and the high feedback mode was selected. Three independent titrations were performed for each DD pairs. The thermograms were processed using Origin7.0 (OriginLab) assuming a one site binding model. In all ITC measurements, a saturation of the binding partner, depicted by a clear plateau, was observed. This plateau was used for baseline correction. In titrations of PaxC ^{N}DD R23E the n-value had to been set to 1, fulfilling the expected one site binding, to have been able to calculate the K_D value.

Circular Dichroism

The circular dichroism spectra of PaxB/C $^{C/N}$DD and PaxC ^{N}DD-$(GS)_{12.5}$-PaxB ^{C}DD were recorded from 50 µM samples in 0.2-mm-path-length quartz cuvettes using a Jasco J-810 CD spectrometer equipped with a Jasco PTC-423S temperature control system. The buffer, 50 mM sodium phosphate buffer (pH 6.5) with 100 mM NaCl, was identical to that used to record NMR spectra. Data were collected at 0.5 nm increments from 260 to 195 nm at 293K.

HR-HPLC-ESI-MS

Purified DDs were analyzed via high resolution (HR)-HPLC-ESI-UV-MS using a Dionex Ultimate 3000 LC system (Thermo Fisher) coupled to an Impact II electrospray ionization mass spectrometer (Bruker) and a DAD-3000 RS UV-detector (Thermo Fisher). The protein samples were separated on a C3 column (Zorbax 300SB-C3 300Å, 150 mm x 3.0 mm x 3.5 µm, Agilent). ACN and H_2O w/ 0.1% (v/v) formic acid were used as mobile phases at a flow rate of 0.6 mL min^{-1}. HPLC was

performed with 30% ACN equilibration (0–1.5 min), followed by a gradient from 30-65% ACN (1.5–27 min) and a further elution step with 95% ACN (27–30 min). For internal mass calibration an ESI-L Mix (Agilent) was injected. The HPLC-MS analysis was set to positive mode with a mass range of *m/z* 50–2000 and an UV at 190–800 nm. For Data analysis of UV-MS-chromatograms Compass DataAnalysis 4.3 (Bruker) was used. The theoretical average masses of proteins were calculated using Compass IsotopePattern 3.0 (Bruker).

REFERENCES

1. Weiner, M. P.; Anderson, C.; Jerpseth, B.; Wells, S.; Johnson-Browne, B.; Vaillancourt, P. (**1994**) Studier pET system vectors and hosts. *Strateg. Mol. Biol. 7*.
2. Fuchs, S. W.; Proschak, A.; Jaskolla, T. W.; Karas, M.; Bode, H. B. (**2011**) Structure elucidation and biosynthesis of lysine-rich cyclic peptides in Xenorhabdus nematophila. *Org. Biomol. Chem. 9*, 3130–3132.
3. Chaston, J. M.; Suen, G.; Tucker, S. L.; Andersen, A. W.; Bhasin, A.; Bode, E.; Bode, H. B.; Brachmann, A. O.; Cowles, C. E.; Cowles, K. N.; Darby, C.; Léon, L. de; Drace, K.; Du, Z.; Givaudan, A.; Herbert Tran, E. E.; Jewell, K. A.; Knack, J. J.; Krasomil-Osterfeld, K. C.; Kukor, R.; Lanois, A.; Latreille, P.; Leimgruber, N. K.; Lipke, C. M.; Liu, R.; Lu, X.; Martens, E. C.; Marri, P. R.; Médigue, C.; Menard, M. L.; Miller, N. M.; Morales-Soto, N.; Norton, S.; Ogier, J.-C.; Orchard, S. S.; Park, D.; Park, Y.; Qurollo, B. A.; Sugar, D. R.; Richards, G. R.; Rouy, Z.; Slominski, B.; Slominski, K.; Snyder, H.; Tjaden, B. C.; van der Hoeven, R.; Welch, R. D.; Wheeler, C.; Xiang, B.; Barbazuk, B.; Gaudriault, S.; Goodner, B.; Slater, S. C.; Forst, S.; Goldman, B. S.; Goodrich-Blair, H. (**2011**) The entomopathogenic bacterial endosymbionts Xenorhabdus and Photorhabdus: convergent lifestyles from divergent genomes. *PloS one 6*, e27909.
4. Hacker, C.; Cai, X.; Kegler, C.; Zhao, L.; Weickhmann, A. K.; Wurm, J. P.; Bode, H. B.; Wöhnert, J. (**2018**) Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nat. Commun. 9*, 4366.
5. Bhattacharya, A.; Tejero, R.; Montelione, G. T. (**2007**) Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct., Funct., Bioinf. 66*, 778–795.
6. Edgar, R. C. (**2004**) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics 5*, 113.
7. Edgar, R. C. (**2004**) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.32*, 1792–1797.
8. Broadhurst, R.W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. (**2003**) The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol. 10*, 723–731.
9. Sambrook, J.; Fritsch, E.; Maniatis, T. (**1989**) *Molecular cloning. A laboratory manual : Vol. 2*, 2. ed.; CSHL Press, New York.
10. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. (**2009**) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343–345.
11. Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. (**1998**) Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *Eur. J. Biochem. 256*, 1–15.
12. Sattler, M. (**1999**) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc. 34*, 93–158.
13. Favier, A.; Brutscher, B. (**2011**) Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR 49*, 9–15.
14. Keller, R. L. J. (**2004**) The Computer Aided Resonance Assignment Tutorial [Online].
15. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. (**2005**) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct., Funct., Bioinf. 59*, 687–696.

16. Mulder, F. A.; Schipper, D.; Bott, R.; Boelens, R. (**1999**) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol. 292*, 111–123.

17. Williamson, M. P. (**2013**) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc. 73*, 1–16.

18. Shen, Y.; Bax, A. (**2013**) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR 56*, 227–241.

19. Herrmann, T.; Güntert, P.; Wüthrich, K. (**2002**) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol. 319*, 209–227.

20. Güntert, P. (**2008**) Automated structure determination from NMR spectra. *Eur. Biophys. J. 38*, 129.

21. López-Méndez, B.; Güntert, P. (**2006**) Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc. 128*, 13112–13122.

22. Koradi, R.; Billeter, M.; Güntert, P. (**2000**) Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun. 124*, 139–147.

23. Ponder, J. W.; Case, D. A. (**2003**) Force fields for protein simulations. *Adv. Protein Chem. 66*, 27–85.

24. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. (**2004**) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res. 32*, 665-7.

25. Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. (**2017**) Improvements to the APBS biomolecular solvation software suite. *Protein Sci. 27*, 112–128.

2. Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases

2.1. Author contribution statements

Status: in revision

Journal: -

Authors: Kenan A. J. Bozhueyuek (KAJB)*, Jonas Watzel (JoWa)*, Nadya Abbood (NA)*, Helge B. Bode (HBB)

* these authors contributed equally

(1) Concept and design

KAJB (40 %), JoWa (15 %), NA (15 %), HBB (30 %)

(2) Conducting tests and experiments

JoWa (50 %): cloning of plasmids, heterologous expression, HPLC-MS, peptide quantification, chemical synthesis; NA (50 %): cloning of plasmids, heterologous expression, HPLC-MS, peptide quantification

(3) Compilation of data sets and figures

KAJB (10 %): protein sequence and structure analysis; JoWa (45 %): compilation of main figures, C/A-NRPS splits, compilation of data sets; NA (45 %): sequence alignment of linker sequences, table of content figure design, A/T- and T/C-NRPS splits, production of peptide library, compilation of data sets

(4) Analysis and interpretation of data

KAJB (10 %): protein sequence and structure analysis; JoWa (45 %): peptide production of C/A-NRPS splits; NA (45 %): peptide production of A/T- and T/C-NRPS splits, production of peptide library

(5) Drafting of manuscript

KAJB (70 %), JoWa (10 %), NA (10 %), HBB (10 %)

place and date | signature PhD student

place and date | signature supervisor

2.2. Manuscript

Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases

Kenan A. J. Bozhueyuek[1,†], Jonas Watzel[1,†], Nadya Abbood[1,2,†], Helge B. Bode[1,2,3,*]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.
2 Max-Planck-Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043, Marburg, Germany.
3 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

† equal contribution

* Corresponding author

Abstract

Non-ribosomal peptide synthetases (NRPSs) are the origin of a wide range of natural products, including many clinically used drugs. Efficient engineering of these often-giant biosynthetic machineries to produce novel non-ribosomal peptides (NRPs) is an ongoing challenge. Here we describe a high throughput cloning and co-expression strategy to functionally combine NRPS fragments of Gram-negative and -positive origin, synthesising novel peptides at titres up to 220 mg l^{-1}. Extending from the recently introduced definition of e*X*change *U*nits (XUs), we inserted synthetic zippers (SZs) to split single protein NRPSs into up to three independently expressed and translated polypeptide chains. These synthetic type of NRPS (type S) enables easier access to engineering, overcomes cloning limitations, and provides a simple and rapid approach to building peptide libraries via the combination of different NRPS subunits. Ultimately, we show how the novel type S NRPSs presented will revolutionise NRPS research in terms of simplicity and biocombinatorial approaches to mega-synthetases.

Introduction

Non-ribosomal peptide synthetases (NRPSs) are multifunctional enzymes, producing a broad range of structural complex and valuable compounds with applications in medicine and agriculture[1] making them key targets for bioengineering. The structural diversity of non-ribosomal peptides (NRPs) arises from the assembly line architecture of their biosynthesis. According to their biosynthetic logic, known NRPS systems are classified into three groups, linear (type A), iterative (type B), and non-linear NRPSs (type C)[2]. Type A NRPSs are composed of sequential catalytically active domains organised in modules, each responsible for the incorporation and modification of one specific amino acid (AA). The catalytic activity of a canonical module is based upon the orchestrated interplay of an adenylation (A) domain for AA selection and activation, a condensation (C) domain to catalyse peptide bond formation, and a thiolation (T) domain/peptidyl-carrier protein (PCP) onto which the AAs or intermediates are covalently tethered[3]. In addition, tailoring domains, including epimerization (E), methylation, and oxidation domains can be part of a module, or a heterocyclization (Cy) domain instead of a C domain can be present. Finally, most NRPS termination modules harbour a TE domain, usually responsible for the release of linear, cyclic or branched cyclic peptides[4].

Type A NRPSs (Fig. S1) follow the collinearity rule, *i.e.*, the number of NRPS modules corresponds directly to the number of monomers incorporated into the associated product, and the arrangement of the modules directly follows the primary sequence of the peptides[5]. Whereas in *in cis* type A NRPSs all modules are arranged on a single polypeptide chain (*e.g.* ACV synthetase[6]), *in trans* assembly lines comprise a number of individual proteins (daptomycin synthetase[7]). Mutual protein-protein interactions of the latter are mediated by specialized *C*- (donor) and *N*-terminally (acceptor) attached ~30 AAs long α-helical structural elements, so called communication-mediating (COM) or docking domains (DDs)[8]. DDs typically are located in between two modules and only interact with weak affinities (4-25 μM)[9–13], but are crucial to ensure biosynthesis of the desired product(s)[8,11,14,15]. Despite recent progress on applying DD substitutions to program new assembly lines, in most cases structural information is lacking to effectively apply DDs for general engineering purposes[11,16,17].

Although early engineering attempts, including the exchange of DDs, the targeted modification of the A domains substrate specificity conferring AA residues, and the

substitution of domains as well as whole modules, gave mixed results, several notable advances have been published recently[17–20]. To give but one example, we comprehensively analysed structural data as well as inter-domain linkers in NRPSs to define novel fusion sites and to provide guidelines for exchanging A-T-C units, denoted as e*X*change *U*nits (XUs)[21], as opposed to canonical modules (C-A-T)[22]. By combining XUs from 15 NRPSs *in cis*, it was possible to reconstitute naturally available peptides, peptide derivatives, and to generate new-to-nature peptides *de novo* in high yields.

However, cloning and engineering modular enzymatic assembly lines, due to their mere size (up to 1.81 million Da (kolossin-producing synthetase (KolS[23])) and conserved repetitive sequence stretches, naturally comes with a series of difficulties. Up to now, our engineering efforts and success in generating chimeric one protein type A NRPSs relied on transformation associated homologous recombination (TAR) based yeast cloning strategies[21,24,25] (Fig. 1a). Although TAR cloning even enabled us to generate NRP libraries[25], compared to *in vitro* cloning methods (*e.g.* Gibson, HiFi, Hot Fusion cloning) we were facing a much longer and arduous workflow (Fig. 1).

Herein, we explored the ability of synthetic zippers (SZs[26]) to manipulate collinear type A NRPSs by introducing artificial *in trans* regulation. SZs interact with high affinity (K_D<10 nM) via a coiled-coil structural motif, enabling the specific association of two proteins. Such a strategy not only would allow creating a synthetic type of *in trans* regulated mega-synthetases (type S), by combining NRPSs with high-affinity SZs[26] (Fig. 2a), but to overcome cloning and protein size limitations associated with heterologous NRP production. Eventually, this would provide a means to 'reuse' already cloned protein encoding sequences by building up plasmid collections while reducing the workload - paving the way for novel high-throughput biocombinatorial approaches at an unprecedented speed (Fig. 1b & c).

Results

Seeking to overcome present limitations of mega-synthetase cloning and bioengineering, we explored possibilities to reduce the complexity of targeted BGCs, via functionally splitting them into separately co-expressible subunits (Fig. 1b). Extending from the recently published XU concept[21] (Fig. 1), initially we decided to split NRPSs into two subunits between consecutive XUs at the previously defined W]-[NATE motif of the conformationally flexible C-A linker[21,27,28] region (Fig. 1b & 2). As

already known[21,27], this splicing position bears several advantages[27,28]. Yet, in depth structural analysis of the crystallised termination module SrfA-C (PDB-ID: 2VSQ) of the surfactin biosynthesis cluster indicated that major parts of the C-A linker must be removed to meet the distance-criteria set out by the WT C-A inter-domain linker to ensure correct C-A di-domain contacts before SZs could be introduced (Fig. 2b). With the aim of keeping the introduced steric hindrance as minimal as possible[27,28], we chose the shortest readily available anti-parallel interacting SZ pair 17 & 18[26] (Fig. 2c & d), as well as removed ~10 AAs from the unstructured *N*-terminus of resulting subunits 2, carrying the modules and domains downstream of the splicing position (Fig. 2b)[27,28].

Proof of Concept (I): Splitting NRPS in between XUs (A-T-C)

To assess the general suitability of SZ pairs to *in trans* connect two NRPS proteins and mediate biosynthetically functional protein-protein interface interactions, we targeted the xenotetrapeptide (**1**)-producing NRPS (XtpS; Fig. S1) from the Gram-negative entomopathogenic bacterium *Xenorhabdus nematophila* HGB081[29]. We split XtpS into two subunits in between XUs 2 and 3. Four artificial two component type S NRPSs (Fig. 3a) were constructed and heterologously produced in *E. coli* DH10B::*mtaA*[30] – either with SZs fused to both subunits (NRPS-1: subunit 1-SZ17, SZ18-subunit 2); only fused to subunit 1 (NRPS-2: subunit 1-SZ17, subunit 2) or subunit 2 (NRPS-3: subunit 1, SZ18-subunit 2), and without SZs (NRPS-4: subunit 1, subunit 2).

NRPS-2 and NRPS-4 showed no detectable peptide production, whereas NRPS-1 led to the production of **1** with ~30% (28 mg l^{-1}) yield compared to WT XtpS (Fig. 3a, Fig. S2), confirming that SZs indeed can be used to functionally mediate new-to-nature *in trans* regulation of NRP biosynthesis. NRPS-3 with SZ18 fused to subunit 2, but lacking SZ17 on subunit 1, showed moderate yields of **1**. Despite lacking SZ17, the *C*-terminus of XtpS subunit 1 might form a leucine-rich α-helical structure (*cf.* Fig. 2b & PDB-ID: 2VSQ) that could be able to interact with SZ18 of subunit 2 and mediate an impaired but catalytically active C-A interface[21,27,28,31].

Additionally, SZ17:18 were used to split the GameXPeptide A-D (**2**-**5**)-producing NRPS (GxpS[32,33], Fig. 3b: NRPS-5, Fig. S3) and the recombinant thiazole-peptide (**6**)-producing NRPS (RtpS[25], Fig. 3b: NRPS-6, Fig. S4). Whereas GxpS originates from the Gram-negative bacterium *Photorhabdus luminescens* TT01, RtpS was constructed previously[25] from building blocks (BBs) of Gram-positive origin (using NRPSs for the

production of bacitracin[34] and surfactin[35]). Both resulting type S NRPSs (Fig. 3b) showed good to very good titres of the desired peptides. NRPS-5 produced **2** (Fig. S3) with yields of ~64 % (4.9 mg l^{-1}) compared to WT GxpS and NRPS-6 produced **6** (Fig. S4) at WT RtpS level (~20 mg l^{-1}).

All aforementioned and following product structures and yields were confirmed by tandem mass spectrometry (MS/MS) analysis and comparison of the retention times with synthetic standards.

Bio-combinatoric Potential of Bipartite Type S NRPSs

Having a tool at hand to generate type S NRPSs that mimic WT behaviour, *inter alia*, the following questions (Q) arise: (Q1) Can type S NRPSs also be leveraged efficiently for NRPS bio-engineering purposes to produce natural product like peptide derivatives as it was shown for the XU concept; and (Q2) how the impact of SZs on peptide production is compared to covalently fused recombinant NRPSs?

To answer Q1 & Q2, we generated and co-expressed four sets of synthetic NRPSs (Fig. 4a, NRPS-7 – -14) with building blocks from XtpS[29], GxpS[32,33], RtpS[25], and the szentiamide[36]-producing synthetase (SzeS[37]), and analysed the culture extracts by HPLC-MS/MS. Each set consisted of two NRPSs: one type S NRPS, co-expressing two non-cognate subunits (NRPS-7, -9, -11, -13); and a corresponding covalently fused NRPS version (NRPS-8, -10, -12*XU, -14), constructed according to the XU approach and largely in line with the XUs' specificity rule to prevent potential substrate specificity issues[21,38] at the upstream C domains' acceptor site.

In sum, six out of eight recombinant NRPSs were functional (NRPS-7 – -12; Fig. 4a). Whereas type S NRPS-7 showed moderately decreased (~47%) yields of **1** compared to the covalently fused NRPS-8, type S NRPS-9 & -11 even showed 1.6 to 9-fold increased productivity of peptide derivatives **7** – **9** and **10** – **11** (Fig. 4d, Fig. S5 – S10), respectively, compared to their covalent counterparts NRPS-10 & -12. Only in case of type S NRPS-13, but also for its covalent version NRPS-14, no production was detected, suggesting a reason other than the SZs for their inactivity, likely being an issue of XtpS TE domain substrate specificity (*cf.* Fig. S17).

Therefore, to a first approximation, it can be concluded that type S NRPSs can indeed be used for bioengineering and -combinatorial purposes (Q1) without impairing NRP biosynthesis more than recombinant *in cis* NRPSs (Q2) do anyway. A potential

advantage is highlighted by type S NRPS-9 (subunit 1: XtpS; subunit 2: RtpS) and its *in cis* variant NRPS-10 (XU 1-2: XtpS; XU 3-5: RtpS). Both, NRPS-9 and -10, are producing peptides **7** – **9** and are composed of BBs from Gram-negative (*X. nematophila* HGB081) and -positive (*B. licheniformis* ATCC 10716, *B. subtilis* ATCC 21332) origin. Whereas NRPS-10 exhibits the typical impaired biosynthesis when XUs from Gram-negative and -positive origin are combined[21,25], *i.e.*, synthesising **7** (linear *v*LL*l*L; *D*-AAs in italics and lowercase throughout this work) in yields of 0.3 mg l^{-1}, type S NRPS-9 shows a 9-fold increased titre of **7** (2.7 mg l^{-1}), indicating that observed impairments might be caused rather at the level of translation than on protein level. Nonrelated bacterial phyla, like *Proteobacteria* and *Firmicutes*, adopted varying codon usages imposing dares during protein translation (distribution of rare codons to control pace of translation and proper protein folding), even aggravated by the respective heterologous host. Separating translation of non-related building blocks by introducing SZs mediated *in trans* protein-protein communication may have minimized the effects of divergent codon usage on protein translation, steadying catalytic activity.

Reuse of NRPS-encoding Plasmids Leads to Rapid Generation of Synthetic NRPSs

Type S NRPS not only open the possibility to convert single protein type A NRPSs into separately expressible subunits (Fig. 1b, 2, & 3), but also to the reuse (Fig. 4a: NRPS-7, -9, & -13) of already generated NRPS encoding plasmids. The reuse of these plasmids therefore opens the door to quickly generate a plethora of artificial BGCs from a small set of subunits by simple combinatorics. Thus, type S NRPS might provide a means to overcome one of the major limiting factors in NRPS research and engineering efforts, namely the substantial amount of lab work involved in generating artificial BGCs (Fig. 1a). This in turn would allow us to rapidly study and characterise WT and chimeric NRPSs as well as associated domains, *e.g.* with respect to substrate specificities of C and TE domains, or the compatibility of bacterial NRPS BBs from Gram-positive and -negative origin.

To showcase possible applications, in a first step we created three additional type S subunits 1 (Fig. 4b, NRPS-17 – -19) from XUs 1-2 of NRPSs producing ambactin (AmbS[30], *X. miraniensis* DSM 17902), bicornutin (BicA[39], *X. budapestensis* DSM 16342), and xenolindicin (XldS[30], *X. indica* DSM 17382). These subunits 1 (Fig. 4b)

were selected to provide a certain variety of substrate specificity at the second (*C*-terminal) XUs' C domain acceptor site. Next, all subunits 1 were co-expressed with all subunits 2. In total, 18 type S NRPSs were generated and culture extracts analysed via HPLC-MS/MS (Fig. 4a: NRPS-7, -9, -11, -13; Fig. 4b: NRPS-15 – -19; Fig. 4c: NRPS-20; Fig. S17: NRPS-48 – -55).

In brief, 9 out of 18 type S NRPSs showed catalytic activity (Fig. 4a & b, Fig. S17), synthesising 21 different linear- (**7**-**12**, **16**-**21**, Fig. S7, S9, S12, S15, S16), cyclic- (**1**-**5**, **13**-**15**, Fig. S5, S11, S13, S14), lipo- (**16**-**19**, Fig. S15), formyl- (**10**, **11**, Fig. S9), and thiazoline (**12**, Fig. S12) containing peptides (**1**-**21**; Fig. 3c & 4c) in yields ranging from ~0.1 – 220 mg l^{-1}. From these results it immediately became apparent that a plethora of chimeric NRPSs and associated NRPs could be generated in no time, and with only a minimum of necessary wet lab work. Moreover, from the results gained (Fig. 4, Fig. S17), it was possible to gather the following interesting insights.

All type S NRPSs sharing the same subunit 2 originating from GxpS (NRPS-5) were functional (Fig. 4a: NRPS-11, Fig. 4b: NRPS-15 – -19, Fig. 4c: NRPS-20), independent of subunit 1's origin (Gram-negative/-positive) and/or C domains' acceptor site substrate specificity. For instance, NRPS-18 (BicA subunit 1 + GxpS subunit 2) and -19 (XldS subunit 1 + GxpS subunit 2), not complying with the XUs' specificity rules (Fig. 4b), produced peptides **15** (~40 mg l^{-1}, Fig. S14) and **16**-**19** (0.1 – 5.5 mg l^{-1}, Fig. S15), respectively. The latter peptides (**16**-**19**) only differ in the *N*-terminal acyl starter unit, originating from the *E. coli* fatty-acid pool, as also observed in the original xenolindicins[30]. Especially NRPS-18 was expected to be inactive, as previous studies have shown that the BicA C3 domain's acceptor site is highly specific for Arg and cannot process Phe or Leu when covalently fused to subunit 2[25]. This might indicate that splitting *in cis* NRPSs in between C and A domains potentially decreases C domains' acceptor site specificity by introducing more geometric flexibility and minimizing potentially restrictive effects on A domain movements[40]. This finding in turn supports the idea that C domains indeed do not exhibit intrinsic substrate specificities, as suggested also by a recently published study[41]. Nevertheless, the effects of C domains on the substrate activation profile of A domains can be observed using NRPS-16 as an example. NRPS-16 produced the linear thiazoline containing peptide IC*/*IL (**12**; ~53 mg l^{-1}; Fig. S12). In its natural NRPS context as well as *in vitro*, the A3 domain of GxpS prefers Phe over Leu[25]. In case of NRPS-16, the terminal C domain of subunit

1, expecting Leu at its acceptor site, either prevents the incorporation of Phe due to its gatekeeping activity or rather fine tunes the downstream A domain specificity. Similar effects of engineered NRPSs, exhibiting chimeric C-A interfaces[21,40] or C domains[25], have been described.

In contrast, all except one (NRPS-7) type S NRPSs sharing subunit 2 from XtpS (NRPS-1), did not produce detectable amounts of any peptide (Fig. S17). In light of all co-expression experiments (Fig. 3, 4, S17), providing evidence that NRPS-1 subunit 2 and all subunits 1 are functional, only one possible explanation remains – the respective TE domains high specificity for peptide length and/or amino acid composition. Catalytic activity of NRPS-7 (Fig. 4a) might serve as proof of this. Subunits 1 of NRPS-1 and NRPS-7 possess synonymous A domain specificities, leading to the production of **1** in both cases (NRPS-1 & -7), and thus preventing TE domain substrate specificity issues.

A further interesting finding of this bio-combinatoric approach via reusing type S subunits, has been the possibility to functionally co-express BBs from Gram-negative and -positive origin. Albeit only 3 (NRPS-9, -16, -20) out of 8 combinations (Fig. S17) showed catalytic activity, type S NRPS represent a first very quick strategy to co-express different subunits of various origin with each other to identify functional combinations. For example, NRPS-16 (subunit 1: Gram-positive; subunit 2: Gram-negative) produced **12** in yields of 54 mg l^{-1} and even exceeded WT production rates of GxpS as well as NRPS-1.

Unpaired Activity of GxpS Subunit 2

All type S NRPS split in between C-A domains and sharing GxpS subunit 2 (NRPS-5, -11, and NRPS-15 – -19) showed an unexpected behaviour, producing a range of tripeptides (**35/36** and **37/38**) at high titre up to 86 mg l^{-1} related to the unpaired activity of GxpS subunit 2 (Fig. S18). Due to the promiscuous GxpS A3 domain, **35/36** and **37/38** differ from each other at position one, either carrying Phenylalanine or Leucine. In addition, **35** and **37** show a *D-D-L* configuration, whereas **36** and **38** have a *L-D-L* configuration. Apart from being unfavourable when it comes to bio-synthetic production of specific bioactive compounds for pharmaceutical application, this finding suggests that it is possible to repurpose elongation modules to initiate peptide biosynthesis – at least under certain conditions – as most recently reported in an *in vitro* study[42].

Proof of Concept (II): Optimization of Bipartite NRPSs Interactions

To elicit the general applicability of SZs other than SZ17:18, we also applied the SZ pair 19 & 18. In comparison to the anti-parallel SZ pair 17 (42 AAs) & 18 (41 AAs), the parallel interacting SZs 19 & 18 (45 AAs) are orientated closely to each other (Fig. 2c). Heterologous production of constructed NRPS-21 (Fig. 5a, Fig. S19a) resulted in yields of **1** with ~17% (32 mg l^{-1}) compared to WT XtpS. On the one hand indicating that indeed both, parallel as well as anti-parallel SZ interaction types can be applied, and on the other hand that short not closely positioned SZs as SZ pair 17 & 18 possess a more beneficial mode of interaction that introduces less impairments to forming a functional *in trans* C-A interface.

However, to minimize conformational/spatial constraints during C-A interface formation caused by the insertion of the SZs, as observed from NRPS-1 and NRPS-5, we tried to optimize SZ17:18 interactions (NRPS-22 – -24). Thus, to enhance the spatial mobility of domains and promote native domain-domain interactions, flexible synthetic stretches of Gly-Ser (GS) linkers[43] varying in length of 4–10 AAs were introduced in between the *C*-terminus of XtpS (NRPS-1) subunit 1 and SZ17 (Fig. 5b). All resulting chimeric NRPSs (NRPS-22 – -24) showed ~3-fold increased yields of **1** (Fig. S19b) compared to NRPS-1 (Fig. 3a), raising titres back at WT level (~230 mg l^{-1}) – and therefore indicating that the introduced flexible unstructured GS linkers did have the desired positive effect on forming correct C-A didomain contacts without impairing domain-domain interactions.

Proof of concept (III): Alternative splicing positions

Driven by these results and to further explore and extend functionality of type S NRPSs, we again targeted XtpS to introduce SZ17:18 within the T-C (Fig. 5c: NRPS-25, Fig. S20a) and A-T linker-regions (Fig. 5c: NRPS-26, Fig. S20b). Both, NRPS-25 and -26 (Fig. 5c, Fig. S19c) synthesised **1** with titres at ~86 % (174 mg l^{-1}) compared to WT XtpS level. While catalytic activity of NRPS-25 was not surprising, as the introduced SZs are mimicking natural DDs[44], the observed good activity of NRPS-26 was unexpected. The A-T linker sequence, consisting of ~15 AAs, represents the shortest inter-domain linker in the context of NRPS elongation modules, which is conformationally also the most flexible one. Structural insights indicate that T domains and A subdomains (A_{sub}) adopt alternative conformations to shuffle reaction intermediates among catalytic domains[28]. Thus, it was assumed that the additional

rigidity, inserted by the structured α-helical AA stretches, would result in loss of function. The most recently gathered structural data of large constructs of the linear gramicidin synthesising NRPS (LgrA)[44], might serve as an explanation for the observed activity. There a very high structural flexibility was reported, potentially bringing closely together domains that are far apart in protein sequence and therefore facilitating synthetic cycles with inserted tailoring domains, unusual domain arrangements like A-C-T[45], module skipping[46], and presumably also SZs.

Further bipartite type S NRPSs (NRPS-56 – -58) aiming at reconstituting WT behaviour are depicted in Fig. S21a. In brief, RtpS split within: (I) the A-T linker region (NRPS-56) led to the production of **6** in yields of 120% (20.4 mg l^{-1}, S21b – d) compared to the WT assembly line; (II) the T-C linker region (NRPS-57) resulted in no detectable peptide production; and (III) the C domain (NRPS-58), following the XUC concept[25], also did not result in any detectable peptide amounts.

Proof of Concept (IV): Tripartite NRPS

As depicted in Figure 1, the possibility of rationally splitting *in cis* regulated type A NRPSs to generate bipartite type S assembly lines comes with a series of benefits, *i.e.*, an increased recombination potential along with reduced workload. These benefits would be further enhanced by the ability to functionally split one protein NRPSs into three or more polypeptide chains. Therefore, again targeting XtpS, we created three orthogonal interaction networks (Fig. 5d) by introducing the anti-parallel and parallel interacting SZ pairs SZ17:18 and SZ1:2, respectively, in between XU2-3 and XU3-4 (NRPS-27), in between module 2-3 and module 3-4 (NRPS-28), as well as within the A-T linker regions of module 2 and 3 (NRPS-29), respectively (Fig. S19d).

All resulting tripartite type S NRPSs (NRPS-27 – 30, Fig. S19d) produced **1** with titres of 24 – 104 mg l^{-1} but with decreased yields compared to their bipartite counterparts (*cf.* NRPS-1, -25 & -26). In addition to the culminative effect of inserted impairments, caused by a higher degree of engineering, we hypothesized that SZ1:2 also contributed to the reduced production titres of **1**. As the SZ pair SZ1:2 is ~8 AAs longer than SZ17:18, NRPS-30 was recreated with truncated/adapted SZ[47] 1 & 2 versions as used in NRPS-29. Heterologous expression of NRPS-30 led to production of **1** with yields of 170% compared to NRPS-29, once again highlighting the great potential for developing optimized SZ pairs for NRPS cloning and co-expression purposes.

To demonstrate the potential of artificial *in trans* NRPSs, we designed and cloned a small library of type S BBs (Fig. 6), placing SZ17:18 and SZ1:2 within the A-T linker regions to perform co-expression experiments in a quick plug-and-play manner. The A-T linker region was targeted because in this case C domain specificities presumably do not represent a limitation of recombination and because T-C-A tri-domains as a catalytically active unit to reprogram NRPSs are underrepresented[48,49].

In brief, the created plasmid library, expressing 11 different type S NRPS BBs from XtpS, GxpS, SzeS, XldS and GarS. Overall, 18 (NRPS-29, NRPS-31 – -47) from 22 co-expressions of three plasmids each yielded detectable amounts (0.1 – 38 mg l^{-1}) of 18 different peptides, 13 of which were new (Fig. 6, S22 – S38). Despite the method's general simplicity, the overall efficacy or recombination potential of T-C-A units compared to XUs (A-T-C units) appears to be slightly more restricted or bares yet to be resolved caveats. For example, neither co-expression of all type S BBs to reconstitute SzeS, nor any combination involving the Ser and Thr specifying BBs from XldS and GarS, yielded any detectable peptide, respectively. These results therefore either indicate an incompatibility of formed chimeric A-T interfaces, substrate incompatibilities at the respective C domains donor site, or the loss of catalytic activity of respective subunits. In light of previous results[50,51], concerning C domain specificities, and our inability to reconstitute szentiamide biosynthesis (SzeS), the latter seems to be the most likely explanation. Due to the sequence and structural flexibility of the targeted A-T linker regions, key interactions within protein-protein interfaces that must be maintained are hard to predict. Therefore, it is likely that the insertion of SZ pairs structurally affected these subunits, resulting in a loss of function and/or their ability to 'communicate' with downstream subunits.

Conclusion

Recently the successful application of SZs to replace naturally present DDs in polyketide synthases (PKS) as a tool to create chimeric PKSs was published[43]. Here we reported the use of high-affinity SZs to split native single protein NRPSs into two and three individual subunits. Generating artificially *in trans* regulated assembly lines not only represents a new NRPS architecture, referred to as type S, but also proves to be extremely easy to handle, productive with WT level yields and provides an unprecedented degree of recombinability.

In addition, we would like to mention that we also tried to introduce SZs in between the *N*- and *C*-lobe of C domains (Fig. S21a: NRPS-58), following our recently published XUC concept[25]. Expectedly, this strategy did not lead in any detectable peptide amounts, as introduction of SZs most likely disrupt both T domain binding sites, at the C domain's acceptor as well as donor site. Yet, in principle it was possible to introduce SZs in between any di-domain (A-T, C-A, T-C).

Having SZs at hand, not only peptide libraries quickly can be constructed with high success rates, but now it also should be conceivable to combine different biosynthetic pathways *in situ*, by introducing SZs at the genomic level – *i.e.*, applying CRISPR/Cas9 based genetic engineering. We are convinced that further research into this direction, like elucidating structures of SZ connected NRPS domains, eventually will bring up even more versatile artificial DDs as it is already suggested by NRPS-22 – -24 and NRPS-30, constructed with synthetic $(GS)_x$ C-A linker regions and a truncated version of SZ1:2, respectively.

Finally, we believe that SZs will greatly accelerate NRPS research – not only for engineering purposes, but also to *in vivo* characterize domains and modules. In particular, the ability to build plasmid libraries holds enormous potential and will take future bio-combinatoric approaches to the next level, *e.g.*, for early drug discovery and high-throughput lead identification approaches.

Acknowledgements

This work was supported by the LOEWE research cluster MegaSyn, the LOEWE center TBG, both funded by the state of Hesse and an ERC advanced grant (grant agreement number 835108).

Competing interests

Goethe University filed a patent application for SZ technology in NRPSs. The patent is currently pending.

386 **References**

387 [1] R. D. Süssmuth, A. Mainz, *Angew. Chem., Int. Ed.* **2017**, *56*, 3770–3821.
388 [2] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490–504.
389 [3] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715–738.
390 [4] F. Kopp, M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735–749.
391 [5] K. A. Bozhüyük, J. Micklefield, B. Wilkinson, *Curr. Opin. Microbiol.* **2019**, *51*, 88–
392 96.
393 [6] M. F. Byford, J. E. Baldwin, C.-Y. Shiau, C. J. Schofield, *Chem. Rev.* **1997**, *97*,
394 2631–2650.
395 [7] R. H. Baltz, V. Miao, S. K. Wrigley, *Nat. Prod. Rep.* **2005**, *22*, 717–741.
396 [8] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15585–
397 15590.
398 [9] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H.
399 Sherman, J. L. Smith, *Chem. Biol.* **2013**, *20*, 1340–1351.
400 [10] J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B.
401 Chagot, K. J. Weissman, *J. Am. Chem. Soc.* **2016**, *138*, 4155–4167.
402 [11] C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B.
403 Bode, J. Wöhnert, *Nat. Commun.* **2018**, *9*, 4366.
404 [12] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, D. H.
405 Sherman, *ACS Chem. Biol.* **2009**, *4*, 41–52.
406 [13] J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *ACS Chem.*
407 *Biol.* **2020**, *15*, 982–989.
408 [14] C. Kegler, H. B. Bode, *Angew. Chem., Int. Ed.* **2020**.
409 [15] S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L.
410 C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea et al., *Nat. Chem.* **2019**, *11*,
411 913–923.
412 [16] X. Cai, S. Nowak, F. Wesche, I. Bischoff, M. Kaiser, R. Fürst, H. B. Bode, *Nat.*
413 *Chem.* **2017**, *9*, 379–386.
414 [17] X. Cai, L. Zhao, H. B. Bode, *Org. Lett.* **2019**, *21*, 2116–2120.
415 [18] H. Kries, D. L. Niquille, D. Hilvert, *Chem. Biol.* **2015**, *22*, 640–648.
416 [19] K. Zhang, K. M. Nelson, K. Bhuripanyo, K. D. Grimes, B. Zhao, C. C. Aldrich,
417 J. Yin, *Chem. Biol.* **2013**, *20*, 92–101.
418 [20] C. Beck, J. F. G. Garzón, T. Weber, *Biotechnol Bioproc E* **2020**, *25*, 886–894.
419 [21] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P.
420 Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275–281.
421 [22] H. D. Mootz, M. A. Marahiel, *Curr. Opin. Biotechnol.* **1999**, *10*, 341–348.
422 [23] H. B. Bode, A. O. Brachmann, K. B. Jadhav, L. Seyfarth, C. Dauth, S. W.
423 Fuchs, M. Kaiser, N. R. Waterfield, H. Sack, S. H. Heinemann et al., *Angew.*
424 *Chem., Int. Ed.* **2015**, *54*, 10352–10355.
425 [24] R. D. Gietz, R. H. Schiestl, *Nat. Protoc.* **2007**, *2*, 31–34.
426 [25] K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F.
427 Fleischhacker, Y.-N. Shi, P. Grün, H. B. Bode, *Nat. Chem.* **2019**, *11*, 653–661.
428 [26] K. E. Thompson, C. J. Bashor, W. A. Lim, A. E. Keating, *ACS Synth. Biol.*
429 **2012**, *1*, 118–129.
430 [27] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G.
431 Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235–238.
432 [28] A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science* **2008**, *321*,
433 659–663.
434 [29] C. Kegler, F. I. Nollmann, T. Ahrendt, F. Fleischhacker, E. Bode, H. B. Bode,
435 *ChemBioChem* **2014**, *15*, 826–828.

436 [30] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290–1294.
438 [31] K. Bloudoff, T. M. Schmeing, *Biochim. Biophys. Acta, Proteins Proteomics* **2017**, *1865*, 1587–1604.
440 [32] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün, *Chemistry* **2012**, *18*, 2342–2348.
442 [33] F. I. Nollmann, C. Dauth, G. Mulley, C. Kegler, M. Kaiser, N. R. Waterfield, H. B. Bode, *ChemBioChem* **2015**, *16*, 205–208.
444 [34] D. Konz, A. Klens, K. Schörgendorfer, M. A. Marahiel, *Chem. Biol.* **1997**, *4*, 927–937.
446 [35] P. Cosmina, F. Rodriguez, F. de Ferra, G. Grandi, M. Perego, G. Venema, D. van Sinderen, *Mol. Microbiol.* **1993**, *8*, 821–831.
448 [36] B. Ohlendorf, S. Simon, J. Wiese, J. F. Imhoff, *Natural product communications* **2011**, *6*, 1247–1250.
450 [37] E. Bode, A. K. Heinrich, M. Hirschmann, D. Abebew, Y.-N. Shi, T. D. Vo, F. Wesche, Y.-M. Shi, P. Grün, S. Simonyi et al., *Angew. Chem., Int. Ed.* **2019**, *58*, 18957–18963.
453 [38] U. Linne, M. A. Marahiel, *Biochemistry* **2000**, *39*, 10439–10447.
454 [39] S. W. Fuchs, C. C. Sachs, C. Kegler, F. I. Nollmann, M. Karas, H. B. Bode, *Anal. Chem.* **2012**, *84*, 6948–6955.
456 [40] S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth, E. Dittmann, *Cell Chem. Biol.* **2016**, *23*, 462–471.
458 [41] M. J. Calcott, J. G. Owen, D. F. Ackerley **2020**, bioRxiv preprint bioRxiv:2020.02.28.970632.
460 [42] M. Kaniusaite, R. J. A. Goode, J. Tailhades, R. B. Schittenhelm, M. J. Cryle, *Chem. Sci.* **2020**, *11*, 9443–9458.
462 [43] M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla, M. Grininger, *ACS Chem. Biol.* **2019**, *14*, 426–433.
464 [44] J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt, T. M. Schmeing, *Science* **2019**, *366*.
466 [45] T. A. Keating, C. G. Marshall, C. T. Walsh, *Biochemistry* **2000**, *39*, 15522–15530.
468 [46] F. Yan, C. Burgard, A. Popoff, N. Zaburannyi, G. Zipf, J. Maier, H. S. Bernauer, S. C. Wenzel, R. Müller, *Chem. Sci.* **2018**, *9*, 7510–7519.
470 [47] A. W. Reinke, R. A. Grant, A. E. Keating, *J. Am. Chem. Soc.* **2010**, *132*, 6025–6031.
472 [48] T. Duerfahrt, S. Doekel, T. Sonke, P. J. L. M. Quaedflieg, M. A. Marahiel, *Eur. J. Biochem.* **2003**, *270*, 4555–4563.
474 [49] M. J. Calcott, D. F. Ackerley, *BMC Microbiol.* **2015**, *15*, 162.
475 [50] P. J. Belshaw, C. T. Walsh, T. Stachelhaus, *Science* **1999**, *284*, 486–489.
476 [51] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L.-O. Essen, *Structure* **2007**, *15*, 781–792.
478

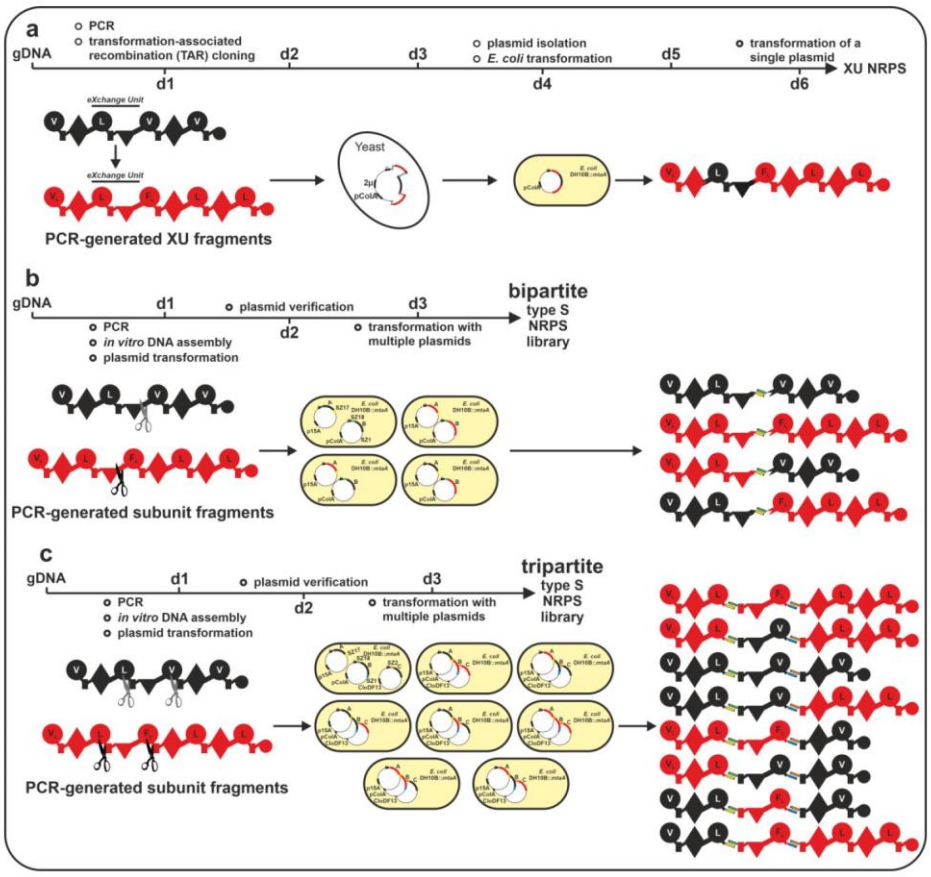


Figure 1. *Comparison of workflows to generate chimeric NRPSs.* **(a)** 'Classic' Yeast-based TAR cloning approach following the XU concept. Ideally cloning of **one** recombinant BGC takes 5 days (d). Here, cloning of one BGC leads to the expression of one assembly line. **(b)** *In vitro* cloning approach to generate bipartite type S NRPS takes about 3 days. Here, BGC complexity is reduced by turning them into smaller separately expressible subunits. Cloning of two bipartite type S NRPSs (four subunits) leads to four co-expression possibilities of recombinant BGCs. **(c)** *In vitro* cloning approach to generate tripartite type S NRPSs. Cloning of two tripartite type S NRPSs (six subunits) leads to eight co-expression possibilities of recombinant BGCs. 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, small circle.

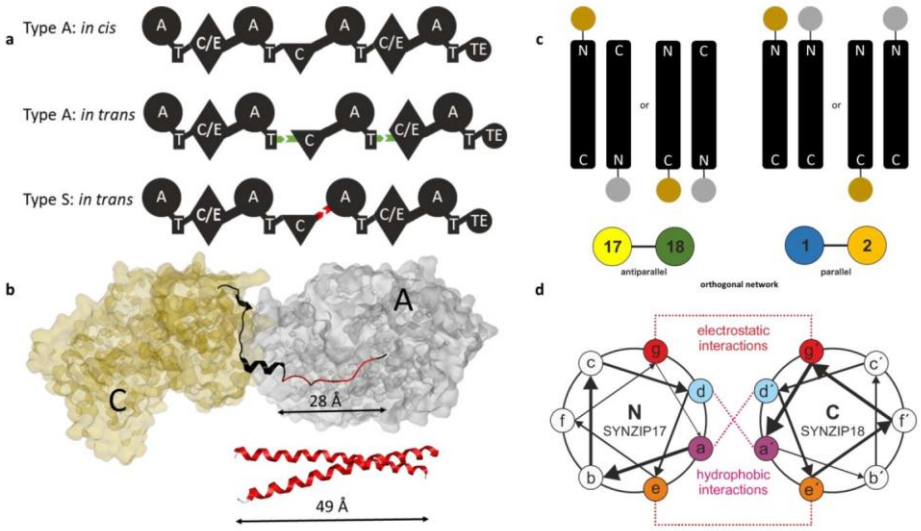


Figure 2. *Introduction to SZ mediated in trans protein-protein interaction*. **(a)** Schematic overview of type A and type S NRPSs. Natural DDs (*in trans* type A NRPS) are shown in green and artificial SZs in red (*in trans* type S NRPS). **(b)** Top: excised C-A di-domain and linker region (ribbon representation) from the SrfA-C termination module (PDB-ID: 2VSQ). Removed area of the C-A linker region to introduce SZs is highlighted red. Bottom: modelled 41 AAs comprising SZ pair. **(c)** Top: antiparallel (left) and parallel (right) interacting hetero-specific SZs. Bottom: SZ17:18 and SZ1:2 are forming an orthogonal interaction network. **(d)** SZ interactions. SZ17:18 are predicted to be electrostatic complementary at adjacent interfacial *e* and *g* positions.

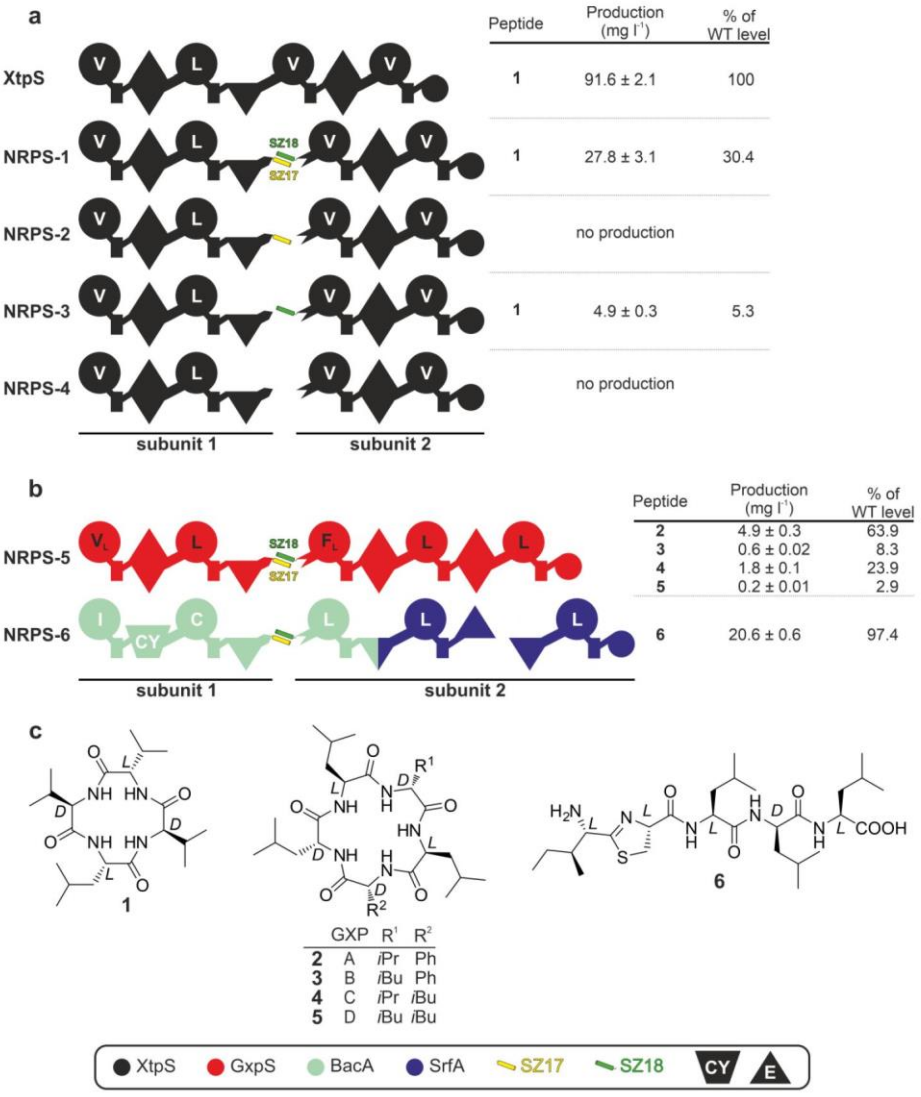
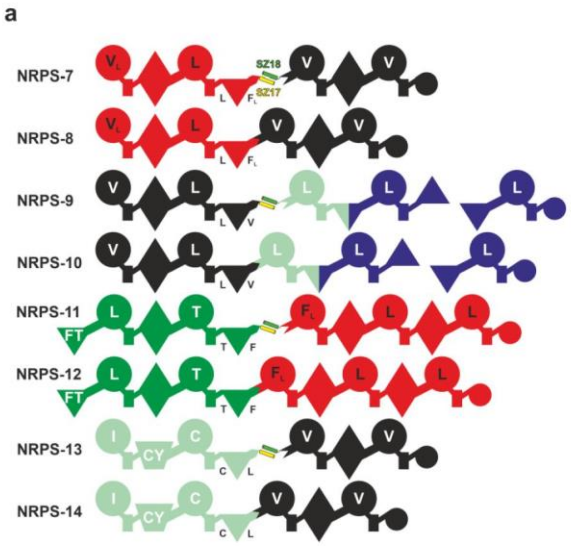
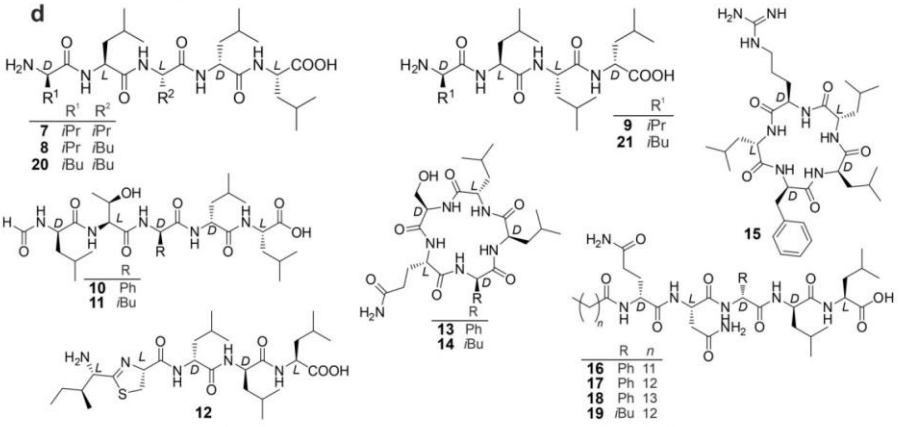


Figure 3. *Splitting NRPS in between XUs.* **(a)** Type S **NRPS-1** – **-4**, as well as corresponding peptide yields obtained from triplicate experiments. **(b)** Type S **NRPS-5** and **-6**, where GxpS and RtpS are split in two subunits. **(c)** Structures of **1**–**6** produced from **NRPS-1** to **NRPS-6** expressed in *E. coli*. See Fig. 1 for assignment of the domain symbols; further symbols: CY, heterocyclization domain; E, epimerization domain. Boxed are the colour coded NRPSs used as building blocks and the used SZ pairs.



Peptide	Production (mg l^{-1})	% of XU level
1	48.3 ± 2.8	46.8
1	103.3 ± 18.40	-
7	2.7 ± 0.7	900.0
8	1.2 ± 0.3	171.4
9	1.3 ± 0.2	325.0
7	0.3 ± 0.1	-
8	0.7 ± 0.1	-
9	0.4 ± 0.1	-
10	23.8 ± 2.6	161.9
11	4.1 ± 1.1	241.2
10	14.7 ± 0.5	-
11	1.7 ± 0.7	-
no production		
no production		

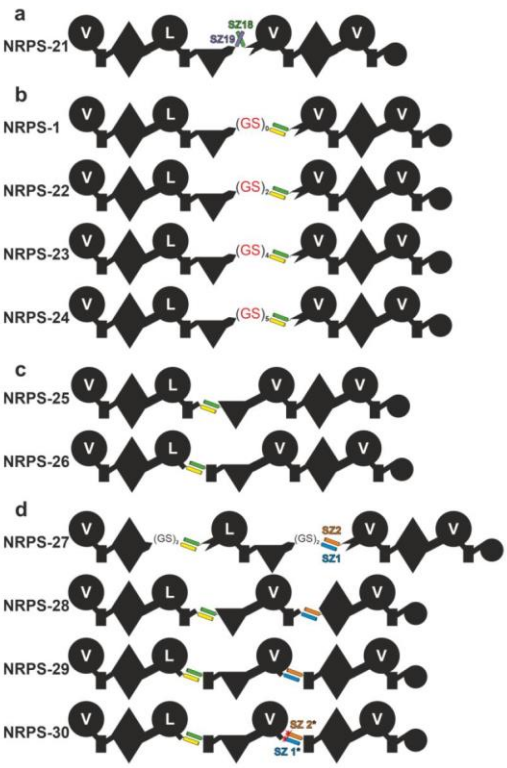
Peptide	Production (mg l^{-1})	% of WT level
2	9.3 ± 0.6	170.2[a]
3	-	-
4	4.3 ± 0.2	79.5[a]
5	-	-
12	53.5 ± 5.0	387.6[b]
13	218.6 ± 21.3	-
14	46.2 ± 3.3	-
15	39.9 ± 1.4	-
16	0.1 ± 0.03	-
17	5.5 ± 0.2	-
18	0.4 ± 0.04	-
19	0.9 ± 0.3	-
7	0.5 ± 0.1	-
8	0.2 ± 0.02	-
9	0.8 ± 0.1	-
20	0.7 ± 0.1	-
21	0.3 ± 0.03	-



513

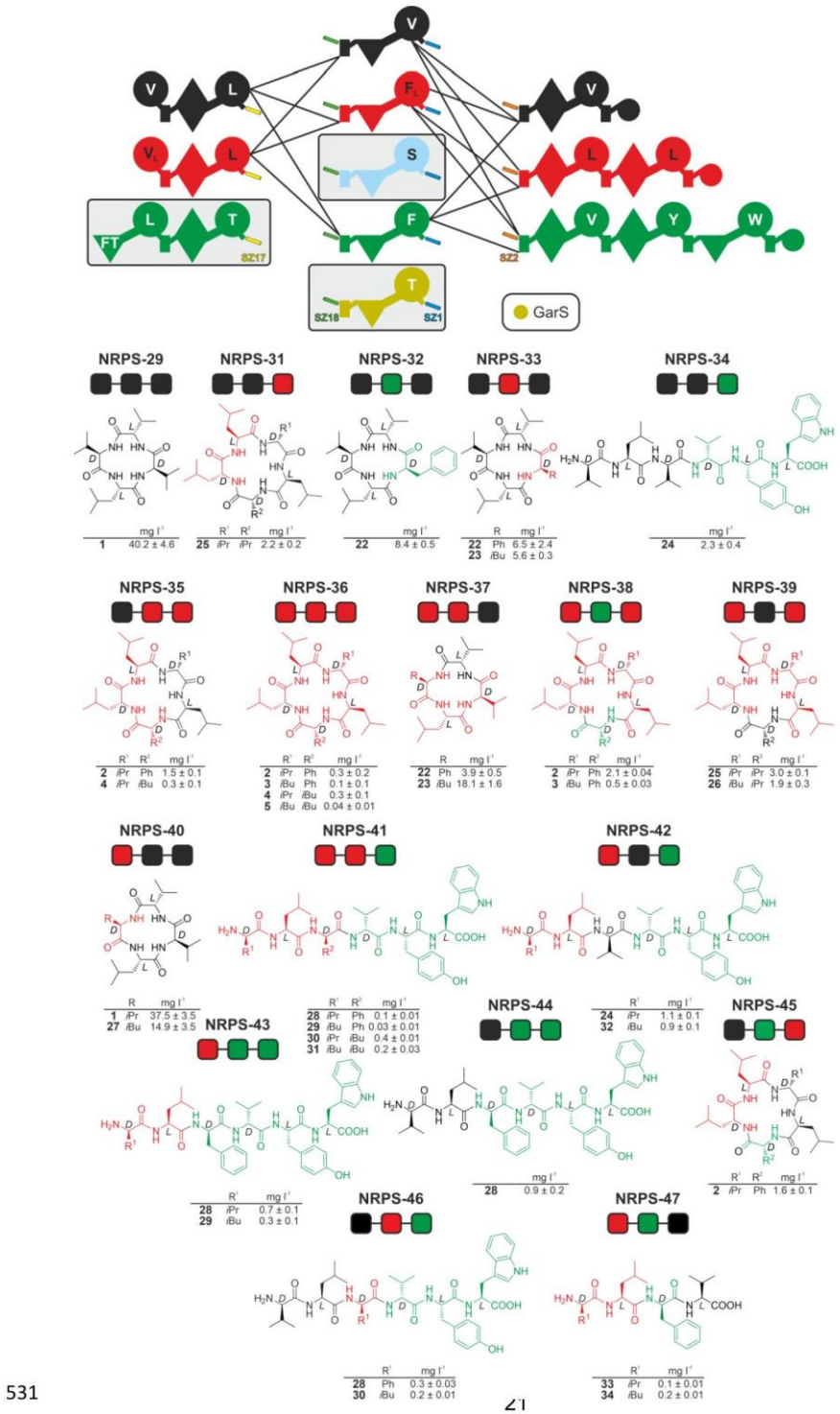
514 **Figure 4.** *Bio-combinatoric potential of bipartite type S NRPS.* **(a)** Comparison of
515 production yields of homologous *in cis* interacting XU and *in trans* interacting bipartite
516 type S NRPSs (**NRPS-7** to **NRPS-14)** and **(b/c)** further type S NRPSs generated by
517 the recombination of XU subunits (**NRPS-15** – **NRPS-20**). The WT production levels
518 are given in comparison to covalent GxpS ([a]) and RtpS ([b]). For additional type S NRPS
519 including non-functional examples see Fig. S17. **(d)** The structures of **7**–**21** produced
520 from **NRPS-7** to **NRPS-20** expressed in *E. coli*. See Fig. 1 and 3 for assignment of the
521 domain symbols; further symbols: FT, formyl-transferase domain.

522



Peptide	Production (mg l^{-1})	% of WT level
1	31.8 ± 4.4	17.8
1	82.5 ± 7.5	35.7
1	224.6 ± 7.8	99.9
1	216.0 ± 4.4	93.6
1	230.5 ± 6.7	97.3
1	174.3 ± 12.9	86.1
1	173.6 ± 9.2	85.9
1	23.8 ± 2.0	17.3
1	103.9 ± 10.2	69.9
1	33.3 ± 3.4	30.5
1	56.6 ± 0.2	51.7

Figure 5. *Schematic representation of type S XtpS.* **(a)** using parallel interacting SZ18:19 (**NRPS-21**), **(b)** $(GS)_x$-elongated C-A linker sequences (**NRPS-22** to **NRPS-24**), **(c)** using different split positions in between the T-C (**NRPS-25**) and A-T domains (**NRPS-26**), **(d)** using the SZ1:2 and SZ17:18 pairs splitting in between the C-A (**NRPS-27**), T-C (**NRPS-28**) and A-T linker region (**NRPS-29**). SZ2/SZ1 was *N*-terminally shortened by 9/4 AAs (**NRPS-30**). See Fig. 1 for assignment of the domain symbols.

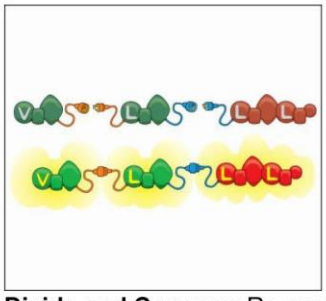

V
V
L
F
L
V
V
L
S
L
L
L
T
F
V
Y
W
FT
SZ17
SZ2
T
SZ18
SZ1
GarS
NRPS-29
NRPS-31
NRPS-32
NRPS-33
NRPS-34
NRPS-35
NRPS-36
NRPS-37
NRPS-38
NRPS-39
NRPS-40
NRPS-41
NRPS-42
NRPS-43
NRPS-44
NRPS-45
NRPS-46
NRPS-47

531

532 **Figure 6.** *Tripartite type S NRPS.* Schematic representation of recombinant, tripartite
533 type S NRPSs based on three subunit combinations (**NRPS-29** and **NRPS-31** – **-47**)
534 using the A-T linker region as split position leading to a structurally diverse peptide
535 library of **22**–**34** all expressed in *E. coli*. The co-expression of the encircled NRPS
536 subunits led in all cases to non-functional type S NRPS. See Fig. 3 and 4 for
537 assignment of the domain symbols.

538

Table of Contents



Divide and Conquer: Re-engineering of non-ribosomal peptide synthetases (NRPSs) is challenging due to their protein size up to a few megadalton. The artificial splitting of NRPSs to produce building blocks with moderate size is achieved by the introduction of synthetic zippers, which can be recombined in a rapid and simple plug and play manner to form an *in trans* regulated synthetic type of NRPSs to produce even peptide libraries.

2.3. Supporting information

Supplementary Information - Table of Contents

1 Material and methods

1.1 Cultivation of strains

All *E. coli* DH10B::*mtaA*, *Xenorhabdus szentirmaii, Xenorhabdus nematophila* and *Photorhabdus luminescens* cells were cultivated in liquid or on solid LB-medium (pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl). Solid media contained 1% (w/v) agar. Kanamycin (50 µg/ml), chloramphenicol (34 µg/ml) and spectinomycin (50 µg/ml) were used as selection markers. All *E. coli* cells cultures were cultivated at 37 °C and at 22 °C for peptide production purposes. *Xenorhabdus* and *Photorhabdus* strains were grown at 30 °C.

1.2 Cloning of biosynthetic gene clusters

Genomic DNA of selected *Xenorhabdus* and *Photorhabdus* strains were isolated using the Qiagen Gentra Puregene Yeast/Bact Kit. All PCRs were performed with oligonucleotides obtained from Eurofins Genomics (Supplementary Table 4). NRPS fragments for Hot Fusion cloning[1] were amplified with primers coding for the respective homology arms (20-30 bp) in a two-step PCR program. The coding sequences for the SYNZIPs were also attached upstream or downstream to the NRPS genes by PCR. In the following, the cloning procedure for the basic vectors is explained. pJW61/62 was obtained by the following steps: First, the SYNZIP17/18 coding sequences (pENTR-SYNZIP17/18[2] were a gift from Amy Keating, Addgene plasmids #80671/80672; RRID:Addgene_80671/80672) were inserted into the plasmids pCOLA_ara/tacI and pCK_0402 by oligonucleotides KB-pACYC-FW/RV or KB-pCOLA-FW/RV in two-step polymerase chain reactions (PCRs) combined with Hot Fusion Cloning[1]. Second, these plasmids were linearized by single-step PCRs with the help of the oligonucleotides KB-pCOLA-II-FW/RV or KB-pACYC-II-FW/RV, which further allowed us to introduce NRPS fragments by Hot Fusion cloning. Therefore, the respective NRPS coding sequences were amplified again in two-step PCRs, using oligonucleotides with additional coding regions for homology arms (20-30 bp). pJW63/64 coding for subunits of the XtpS without attached SYNZIPs were generated by amplifying pJW61/62 with a single phosphorylated [phos.] oligonucleotide pair excluding the SYNZIP coding region followed by T4 DNA ligation (following Thermo Fisher manufacturers' instructions). The control plasmids pCOLA_ara_xtpS/gxpS_tacI_JW coding for the native single protein xtpS/ gxpS were

created by Hot Fusion Cloning. Therefore, the plasmid pCOLA_ara/tacI was linearized by PCR using the oligonucleotides AL-XtpS-2-1 and AD64 and the insert *xtpS* was PCR amplified with the oligonucleotides jw0136_FW and jw0137_RV.

The plasmid pJW102 coding for NRPSs with two attached SNYZIPs were created by Hot Fusion cloning. Before this final cloning step a pCOLA_ara/tacI plasmid carrying the SYNZIP18 sequence downstream the P_{BAD} promoter was linearized. This linearization step by PCR was done twice and allowed us to incorporate the SYNZIP1 coding region (na28_FW, na29_FW) upstream the stop codon (pQLinkHD-SYNZIP1 was a gift from Amy Keating[2], Addgene plasmid #80647; RRID:Addgene_80647).

The starting point for plasmid pJW103 was vector pCDF_ara/tacI. This vector was generated by digesting the plasmids pCOLA_ara/tacI and pCDFDuet™-1 (Novagen) with the enzymes XbaI and NdeI. The fragment of pCDFDuet™-1 carrying the pCloDF13 replicon and streptomycin/spectinomycin resistance marker was then T4 ligated with the compatible pCOLA_ara/tacI fragment. Then, this plasmid pCDF_ara/tacI was linearized in two cycles including in parallel the sequence of SYNZIP2 (na32_RV, na33_RV) (pQLinkHD-SYNZIP2 was a gift from Amy Keating[2], Addgene plasmid #80658; RRID:Addgene_80658) downstream the P_{BAD} promoter, followed by the incorporation of the respective NRPS coding regions by Hot Fusion Cloning.

The plasmid pCOLA_ara_ gxpS_tacI_JW was generated in two Hot Fusion Cloning steps. First, the pCOLA_ara/tacI was linearized by PCR using the primers JW_tacI_PstI_FW2 and jw0064_RV and second the first part of *gxpS* was amplified using the oligonucleotides jw0124_FW/jw0160_RV. This intermediate plasmid was then opened with PstI and the second *gxpS* part, amplified with jw0151_FW/ jw0161_RV by PCR, was then integrated into the cleaving site by Hot Fusion Cloning. In all PCRs the S7 Fusion High-Fidelity DNA Polymerase (Mobidiag) was used according to the manufacturers' instructions. The amplified DNA was purified with the Invisorb Fragment CleanUp or MSB Spin PCRapace Kits (stratec molecular). The basic cloning of all new generated plasmids (Supplementary Table 3) was performed in *E. coli* DH10B. Each NRPS (subunit) was under the control of a P_{BAD} promotor. Plasmid isolation from *E. coli* was achieved with the Invisorb Spin Plasmid Mini Two Kit (stratec molecular). Restriction enzyme digests and the partial sequencing of essential plasmid regions, especially upstream and downstream of the NRPS genes,

where the SYNZIP coding sequences were located, confirmed the correct plasmid construction.

1.3 Heterologous expression of NRPS templates and LC-MS analysis

Constructed plasmids were transformed into *E. coli* DH10B::*mtaA*. Cells were grown overnight in LB medium containing the necessary antibiotics (50 µg/ml kanamycin, 34 µg/ml chloramphenicol, 50 µg/ml spectinomycin). 100 µl of an overnight culture were used for inoculation of 10 ml LB-cultures supplemented with the respective antibiotics as selection markers and additionally containing 0.002 mg/ml L-arabinose and 2 % (v/v) XAD-16. After incubation for 72 h at 22 °C the XAD-16 was harvested. One culture volume methanol was added and incubated for 60 min at 22 °C. The organic phase was filtrated and a sample was taken of the cleared extract. After centrifugation (17,000 x *g*, 20 min) the methanol extracts were used for LC-MS analysis. All measurements were performed by using an Ultimate 3000 LC system (Dionex) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml min^{-1} using acetonitrile (ACN) and water containing 0.1% formic acid (v/v) in a gradient ranging from 5-95% of ACN over 16 min (40 °C) coupled to an AmaZonX (Bruker) electron spray ionization mass spectrometer. The BPC spectra were recorded in positive ion mode with the range from 100-1200 *m/z* and ultraviolet (UV) at 200-600 nm. The software Compass DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

1.4 Peptide quantification

The absolute production titers of selected peptides were calculated with calibration curves based on pure synthetic **1** (for quantification of **1**, **22**, **23**, **27**), **2** (for quantification of **2**-**5**, **25**, **26**), **6** (for quantification of **6** and **12**), **7** (for quantification of **7, 8** and **20**), **9** (for quantification of **9**, **21**, **33** and **34**), **10** (for quantification of **10** and **11**), **13** (for quantification of **13** and **14**), **15**, **17** (for quantification of **16**-**19**), **24** (for quantification of **24**, **30**, **31**, **32**), **28** (for quantification of **28**, **29**) and **37** (for quantification of **35/36** and **37/38**). Therefore, the pure compounds were prepared at different concentrations (100, 50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, 0.195 and 0.0195 µg/mL) and measured by LC-MS using HPLC/MS measurements as described above. The peak area for each compound at different concentrations was calculated

using Compass Data Analysis and used for the calculation of a standard curve passing through the origin. Triplicates of all *in vivo* experiments were measured. The pure peptide standards **1**, **2**, **6, 13**, **17** and **35** were synthesized in-house[3,4] and the further pure synthetic **15**, **19**, **24** and **28** were produced by Synpeptide.

1.5 Chemical synthesis

Chemical synthesis of all peptides was performed as described previously[3].

2 Supplementary Tables

Table S1. ESI-MS data of all produced peptides.

Peptide (#)	theoretical mass-to-charge ratio (*m/z*) $[M+H]^+$	Molecular formula	Reference
1	410.29	$C_{21}H_{38}N_4O_4$	[5]
2	586.40	$C_{32}H_{51}O_5N_5$	[6]
3	600.41	$C_{33}H_{53}O_5N_5$	[6]
4	552.41	$C_{29}H_{53}O_5N_5$	[6]
5	566.43	$C_{30}H_{55}O_5N_5$	[6]
6	556.35	$C_{27}H_{49}N_5O_5S$	-
7	556.41	$C_{28}H_{53}N_5O_6$	-
8	570.42	$C_{29}H_{55}N_5O_6$	-
9	457.34	$C_{23}H_{44}N_4O_5$	-
10	634.38	$C_{32}H_{51}N_5O_8$	[3]
11	600.40	$C_{29}H_{53}N_5O_8$	[3]
12	556.35	$C_{27}H_{49}N_5O_5S$	-
13	589.33	$C_{29}H_{44}N_6O_7$	-
14	555.35	$C_{26}H_{46}N_6O_7$	-
15	643.43	$C_{33}H_{54}N_8O_5$	-
16	830.54	$C_{43}H_{71}N_7O_9$	-
17	844.55	$C_{44}H_{73}N_7O_9$	-
18	858.57	$C_{45}H_{75}N_7O_9$	-
19	810.57	$C_{41}H_{75}N_7O_9$	-
20	584.44	$C_{30}H_{57}N_5O_6$	-
21	471.35	$C_{24}H_{46}N_4O_5$	-
22	459,30	$C_{25}H_{38}N_4O_4$	-
23	425,31	$C_{22}H_{40}N_4O_4$	-
24	778,45	$C_{41}H_{59}N_7O_8$	-
25	538.40	$C_{28}H_{51}N_5O_5$	-
26	552.41	$C_{29}H_{53}N_5O_5$	-
27	425.31	$C_{22}H_{40}N_4O_4$	-
28	826.45	$C_{45}H_{59}N_7O_8$	-
29	840.47	$C_{46}H_{61}N_7O_8$	-
30	792.47	$C_{42}H_{61}N_7O_8$	-
31	806.48	$C_{43}H_{63}N_7O_8$	-
32	792.47	$C_{42}H_{61}N_7O_8$	-
33	476.30	$C_{25}H_{40}N_4O_5$	
34	490.32	$C_{26}H_{42}N_4O_5$	
35	358.27	$C_{18}H_{35}N_3O_4$	-
36	358.27	$C_{18}H_{35}N_3O_4$	-
37	392.25	$C_{21}H_{33}N_3O_4$	-
38	392.25	$C_{21}H_{33}N_3O_4$	-

Table S2. Strains used in this work.

Strain	Genotype/NRPS	Reference
E. coli DH10B	F_mcrA (*mrr-hsd*RMS-*mcr*BC), 80*lacZ*Δ, M15, Δ*lac*X74 *rec*A1 *end*A1 *ara*D 139Δ(*ara*, *leu*)7697 *gal*U *gal*K λ *rps*L (*Strr*) *nup*G	[7]
E. coli DH10B::*mtaA*	DH10B with *mtaA* from pCK_*mtaA*Δ*entD*	[8]
P. luminescens TTO1	*gxpS*[6]	DSMZ
X. bovienii SS-2004	*garS*/*xfpS*[9]	[10]
X. nematophila ATCC 19061	*xtpS*[5]	ATCC
X. budapestensis DSM 16342	*bicA*[11]	DSMZ
X. miraniensis DSM 17902	*ambS*[8]	DSMZ
X. szentirmaii DSM16338	*szeS*[12]	DSMZ
X. indica DSM 17382	*xldS*[8]	DSMZ
B. licheniformis ATCC 10716	*bacA*[13]	M. A. Marahiel / ATCC
B. subitlis MR 168	*srfA*[14]	ATCC

Table S3. Plasmids used in this work.

Plasmids	Genotype	Reference
pFF1_22A_szeS_gxpS	ori 2µ, kanMX4, ori ColA, kan^R, P_{BAD} *szeS*_Ft$A_1T_1C/E_2A_2T_2C_3$-*gxpS*_$A_3T_3C/E_4A_4T_4C/E_5A_6T_5$TE, Ypet-Flag	[3]
pFF1_NRPS_6 (NRPS-12)	ori 2µ, kanMX4, *araC*-P_{BAD}, ori ColA, Ypet-Flag, kan^R, *bacA*-A1T1CyA2T2C3A3T3CD_{sub4}_*sfrA-BC*-C_{Asub6}A6T6E6C7A7T7TE	[4]
pCOLA_ara/tacI	ori ColA, kan^R, *araC*-P_{BAD} and *tac*I	[15]
pCK_0402	ori p15A, cm^R, *araC*-P_{BAD} and *tacI*-*araE*	unpublished
pCDF_ara/tacI	ori CloDF13, spec^R, *araC*-P_{BAD} and *tac*I	this study
pCOLA_ara_xtpS_tacI_JW	ori ColA, kan^R, *araC*-P_{BAD} *xtpS* and *tac*I	this study
pCOLA_ara_gxpS_tacI_JW	ori ColA, kan^R, *araC*-P_{BAD} *gxpS* and *tac*I	this study
pJW61	ori p15A, cm^R, *araC*-P_{BAD} *xtpS*_$A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW62	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*xtpS*_$A_3T_3C/E_4A_4T_4$TE and *tac*I	this study
pJW63	ori p15A, cm^R, *araC*-P_{BAD} *xtpS*_$A_1T_1C/E_2A_2T_2C_3$ and *tacI*-*araE*	this study
pJW64	ori ColA, kan^R, *araC*-P_{BAD} *xtpS*_$A_3T_3C/E_4A_4T_4$TE, *tac*I	this study
pJW75	ori p15A, cm^R, *araC*-P_{BAD} *gxpS*_$A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW76	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*gxpS*_$A_3T_3C/E_4A_4T_4C/E_5A_6T_5$TE and *tac*I	this study
pJW77	ori p15A, cm^R, *araC*-P_{BAD} *bicA*_$A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW91	ori p15A, cm^R, *araC*-P_{BAD} *ambS*_ $A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW92	ori p15A, cm^R, *araC*-P_{BAD} *szeS*_Ft$A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW93	ori p15A, cm^R, *araC*-P_{BAD} *xldS*_$C_1A_1T_1C/E_2A_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW100	ori p15A, cm^R, *araC*-P_{BAD} *xtpS*_A_1T_1C/E_2-$(GS)_2$-SYNZIP17 and *tacI*-*araE*	this study
pJW102	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*xtpS*_$A_2T_2C_3$-$(GS)_2$-SYNZIP1 and *tac*I	this study
pJW103	ori CloDF13, spec^R, *araC*-P_{BAD} *xtpS*_SYNZIP2-$A_3T_3C/E_4A_4T_4$TE and *tac*I	this study
pJW114	ori p15A, cm^R, *araC*-P_{BAD} *bacA*_$A_1T_1CyA_2T_2C_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW116	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*bacA*_$A_3T_3C_{Dsub4}$-*sfrA-BC*_$C_{Asub6}A_6T_6E_6C_7A_7T_7$TE and *tac*I	this study
pJW118	ori p15A, cm^R, *araC*-P_{BAD} *bacA*_$A_1T_1CyA_2T_2C_3A_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW120	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*bacA*_T_3C_{Dsub4}-*sfrA-BC*_$C_{Asub6}A_6T_6E_6C_7A_7T_7$TE and *tac*I	this study
pJW122	ori p15A, cm^R, *araC*-P_{BAD} *bacA*_$A_1T_1CyA_2T_2C_3A_3T_3$-SYNZIP17 and *tacI*-*araE*	this study
pJW124	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*bacA*_C_{Dsub}4-*sfrA-BC*_$C_{Asub6}A_6T_6E_6C_7A_7T_7$TE and *tac*I	this study
pJW126	ori p15A, cm^R, *araC*-P_{BAD} *bacA*_$A_1T_1CyA_2T_2C_3A_3T_3C_{Dsub4}$-SYNZIP17 and *tacI*-*araE*	this study
pJW128	ori ColA, kan^R, *araC*-P_{BAD} SYNZIP18-*sfrA-BC*_$C_{Asub6}A_6T_6E_6C_7A_7T_7$TE and *tac*I	this study
pJW159	ori ColA, kan^R, *araC*-P_{BAD} *xtpS*_ $A_1T_1C/E_2A_2T_2C_3$-*sfrA-BC*_$C_{Asub6}A_6T_6E_6C_7A_7T_7$TE and *tac*I	this study
pJW160	ori ColA, kan^R, *araC*-P_{BAD} *bacA*_$A_1T_1CyA_2T_2C_3$-*xtpS*_$A_3T_3C/E_4A_4T_4$TE and *tac*I	this study
pJW161	ori ColA, kan^R, *araC*-P_{BAD} *gxpS*_$A_1T_1C/E_2A_2T_2C_3$-*xtpS*_$A_3T_3C/E_4A_4T_4$TE and *tac*I	this study
pNA1	ori p15A, cm^R, *araC*-P_{BAD} *xtpS*_ $A_1T_1C/E_2A_2T_2C_3$-SYNZIP19 and *tacI*-*araE*	this study
pNA2	ori p15A, cm^R, *araC*-P_{BAD} *xtpS*_ $A_1T_1C/E_2A_2T_2$-SYNZIP17 und *tacI*-*araE*	this study

pNA3	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xtpS*_ $C_3A_3T_3C/E_4A_4T_4TE$ und *tac*I	this study
pNA4	ori p15A, cm^R, *araC-P$_{BAD}$ xtpS*_ A_1T_1C/E_2A_2-SYNZIP17 und *tac*I-*araE*	this study
pNA5	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xtpS*_ $T_2C_3A_3T_3C/E_4A_4T_4TE$ und *tac*I	this study
pNA8	ori p15A, cm^R, *araC-P$_{BAD}$ xtpS*_$A_1T_1C/E_2A_2T_2C_3$-$(GS)_5$-SYNZIP17 and *tac*I-*araE*	this study
pNA9	ori p15A, cm^R, *araC-P$_{BAD}$ xtpS*_$A_1T_1C/E_2A_2T_2C_3$-$(GS)_4$-SYNZIP17 and *tac*I-*araE*	this study
pNA10	ori p15A, cm^R, *araC-P$_{BAD}$ xtpS*_$A_1T_1C/E_2A_2T_2C_3$-$(GS)_2$-SYNZIP17 and *tac*I-*araE*	this study
pNA15	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xtpS*_$T_2C_3A_3$-SYNZIP1 and *tac*I	this study
pNA16	ori CloDF13, spec^R, *araC-P$_{BAD}$* SYNZIP2-*xtpS*_$T_3C/E_4A_4T_4TE$ and *tac*I	this study
pNA17	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xtpS*_$C_3A_3T_4$-SYNZIP1 and *tac*I	this study
pNA18	ori CloDF13, spec^R, *araC-P$_{BAD}$* SYNZIP2-*xtpS*_$C/E_4A_4T_4TE$ and *tac*I	this study
pNA26	ori p15A, cm^R, *araC-P$_{BAD}$ gxpS*_A_1T_1C/E_2A_2-SYNZIP17 and *tac*I-*araE*	this study
pNA27	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*gxpS*_$T_2C_3A_3$-SYNZIP1 and *tac*I	this study
pNA28	ori CloDF13, spec^R, *araC-P$_{BAD}$* SYNZIP2-*gxpS*_$T_3C/E_4A_4T_4$ $C/E_5A_5T_5TE$ and *tac*I	this study
pNA29	ori p15A, cm^R, *araC-P$_{BAD}$ sze*S_ $C_1A_1T_1C/E_2A_2$-SYNZIP17 and *tac*I-*araE*	this study
pNA30	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*szeS*_$T_2C_3A_3$-SYNZIP1 and *tac*I	this study
pNA31	ori CloDF13, spec^R, *araC-P$_{BAD}$* SYNZIP2-*szeS*_$T_3C/E_4A_4T_4$ $C/E_5A_5T_5C_6A_6T_6TE$ and *tac*I	this study
pNA35	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*garS*_$T_2C_3A_3$-SYNZIP1 and *tac*I	this study
pNA59	ori p15A, cm^R, *araC-P$_{BAD}$ xfpS*_C_1A_1-SYNZIP17 and *tac*I-*araE*	this study
pNA60	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xfpS*_$T_1E_1C_2A_2T_2C_3A_3T_3TE$ and *tac*I	this study
pNA125	ori ColA, kan^R, *araC-P$_{BAD}$* SYNZIP18-*xtpS*_$T_2C_3A_3$-SYNZIP1_shortened and *tac*I	this study
pNA126	ori CloDF13, spec^R, *araC-P$_{BAD}$* SYNZIP2_shortened-*xtpS*_$T_3C/E_4A_4T_4TE$ and *tac*I	this study

Table S4. Oligonucleotides used in this work. Correlations of plasmids to figures from the main text and supplementary information are represented in brackets.

Plasmids	Oligo-nucleotide	Sequence (5´→3´; overlapping ends)	Template
pJW61 (NRPS-1, NRPS-2, NRPS-9, NRPS-15)	KB-pACYC-FW	GAACAGTTAAAACAGAAGCGTGAACAATTAAAGCAAAAGATCGCCAATCTGCGTAA GGAGATCGAAGCCTACAAGTGACAATTAATCATCGGCTCG	pCK_0402
	KB-pACYC-RV	TTCACGCTTCTGTTTTAACTGTTCGATGCGATTACGCAATTCAGCCTTTTTCGATTTT AATTCCTCCTTCTCGTTCATGGAATTCCTCCTGTTAGC	pCK_0402
	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	-
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	-
	KB-P1-FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	KB-P1-RV	CGATTTTAATTCCTCCTTCTCGTTCCAGGTTTTTAACAACAATGTGC	*X. nematophila* ATCC 19061
pJW62 (NRPS-1, NRPS-3, NRPS-7, NRPS-13, NRPS-49, NRPS-50, NRPS-51)	KB-pCOLA-FW	CATTGACAAAGAGCTGCGTGCCAACGAAAACGAACTTCGCGCCCTTGATAACGAGC TGACTGCAGCTATCTCATGACAATTAATCATCGGCTCG	pCOLA_ara/tacI
	KB-pCOLA-RV	TTGGCACGCAGCTCTTTGTCAATGGCATTTAACTCGCGGTCCAAGGCTTTCAGTTCA CGCTCTTCAGCATAGAACATGGAATTCCTCCTGTTAGC	pCOLA_ara/tacI
	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	-
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	-
	KB-P2-FW	AACGAGCTGACTGCAGCTATCTCATTATGTATTCATCAACTTTTTGAACAGC	*X. nematophila* ATCC 19061
	KB-P2-RV	ATACGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pJW63 (NRPS-3, NRPS-4)	jw0061_FW	[phos.] TGACAATTAATCATCGGCTCG	pJW61
	jw0062_RV	CCAGGTTTTTAACAACAATGTGC	pJW61
pJW64 (NRPS-2, NRPS-4)	jw0063_FW	[phos.] TTATGTATTCATCAACTTTTTGAACAGC	pJW62
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pJW62
pJW75 (NRPS-5, NRPS-7, NRPS-20)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw0124_FW	GGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGAAATTATC	*P. luminescens* TTO1
	jw0125_RV	TCGATTTTAATTCCTCCTTCTCGTTCCAATTTTCCAGTAATAACTCCCG	*P. luminescens* TTO1
pJW76 (NRPS-5, NRPS-11, NRPS-15, NRPS-16, NRPS-17, NRPS-18, NRPS-19)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	jw0172_FW	GGCTAACAGGAGGAATTCCATGTTCTATGCTGAAGAGCGTGAAC	*P. luminescens* TTO1
	jw0127_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTC	*P. luminescens* TTO1
pJW114 (NRPS-6, NRPS-13, NRPS-16)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw208_FW	GCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATGGG	pFF1_NRPS_6[4]
	jw209_RV	CGATTTTAATTCCTCCTTCTCGTTCTTTGTATGGTTAAAGGACTCTAAAAGTGTC	pFF1_NRPS_6[4]
pJW116 (NRPS-6, NRPS-9, NRPS-10, NRPS-20, NRPS-52, NRPS-53, NRPS-54, NRPS-55)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	jw0211_FW	CGAGCTGACTGCAGCTATCTCAAAAGCAATCCACCAGCTGTTT	pFF1_NRPS_6[4]
	jw0212_RV	CGAGCCGATGATTAATTGTCATGAAACCGTTACGGTTTGTGTATTA	pFF1_NRPS_6[4]
pJW77 (NRPS-18, NRPS-50, NRPS-54)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw0128_FW	GGGCTAACAGGAGGAATTCCATGAAAGATAACATTGCTACAGTGGCAAATAG	*X. budapestensis* DSM 16342
	jw0129_RV	CGATTTTAATTCCTCCTTCTCGTTCCAAGTTTTCAGCAACAACTGG	*X. budapestensis* DSM 16342
pJW91 (NRPS-17, NRPS-48, NRPS-52)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw0162_FW	GCTAACAGGAGGAATTCCATGAAAAATGATAAGGTGATGACTCTG	*X. miraniensis* DSM 17902
	jw0163_RV	TCGATTTTAATTCCTCCTTCTCGTTCCACGTTTCCAGCAATAACC	*X. miraniensis* DSM 17902
pJW92 (NRPS-11, NRPS-49, NRPS-53)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw0164_FW	GCTAACAGGAGGAATTCCATGAAAGGTAGTATTGCTAAAAAGGGAG	*X. szentirmaii* DSM16338
	jw0165_RV	TCGATTTTAATTCCTCCTTCTCGTTCCAGCTTTCCAGCAATAACC	*X. szentirmaii* DSM16338
pJW93 (NRPS-19, NRPS-51, NRPS-55)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61

	jw0166_FW	GCTAACAGGAGGAATTCCATGAAACTTTGGAACTATAAAATGAATATGAC	*X. indica* DSM 17382
	jw0167_RV	TCGATTTTAATTCCTCCTTCTCGTTGAAATCCACCAACAGTTGTTGAC	*X. indica* DSM 17382
pJW100 (NRPS-27)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	KB-P1-FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	jw0179_RV	CGATTTTAATTCCTCCTTCTCGTTGGATCCAGACCCCCAGGTTTTCAGCAATAACGTG	*X. nematophila* ATCC 19061
pJW102 (NRPS-27)	na29_FW	AACCTGGTTGCGCAGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAACCCTGAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTAC	pNA15
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pNA15
	jw0180_FW	AACGAGCTGACTGCAGCTATCTCACTGTGTATCCATCAGTTAATTGAACAACAG	*X. nematophila* ATCC 19061
	jw0182_RV	CGTTTTCGAGCTGCGCAACCAGGTTGGATCCAGACCCCCAGGTTTTTAACAACAATGTGCG	*X. nematophila* ATCC 19061
pJW103 (NRPS-27)	na34_FW	TGACAATTAATCATCGGCTCG	pCDF_ara/tacI
	na32_RV	GTTCTGTTCATCACGTTCCAGCTGCAGGTTGTCTTTTTTCAGACGTGCGATTTTCTTACGCAGATACGCGTTACGCGCCATGGAATTCCTCCTGTTAGCC	pCDF_ara/tacI
	na33_RV	CTGTTCGTGAGACGCAACTTCGTTTTCGAGACGCGCGATTTCGTCACGCAGGTTCGCGATGATTTTTTCCAGGTTCTGTTCATCACGTTCCAGC	pCDF_ara/tacI
	na84_RV	CTGTTCGTGAGACGCAACTTC	-
	jw0183_FW	AAACGAAGTTGCGTCTCACGAACAGTTATGTATTCATCAACTTTTTGAACAGC	*X. nematophila* ATCC 19061
	jw0188_RV	GCCTAAACCAATACGCCGT	*X. nematophila* ATCC 19061
	jw0189_FW	CGGCGTATTGGTTTAGGCCTGT	*X. nematophila* ATCC 19061
	na07_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pJW118 (NRPS-56)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw208_FW	GCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATGGG	pFF1_NRPS_6[4]
	jw0214_RV	CGATTTTAATTCCTCCTTCTCGTTGTAGCGGCGATCCATTGT	pFF1_NRPS_6[4]
pJW120 (NRPS-56)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	jw0216_FW	CGAGCTGACTGCAGCTATCTCAGAAGCGCCGCGGG	pFF1_NRPS_6[4]
	jw0212_RV	CGAGCCGATGATTAATTGTCATGAAACCGTTACGGTTTGTGTATTA	pFF1_NRPS_6[4]
pJW122 (NRPS-57)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw208_FW	GCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATGGG	pFF1_NRPS_6[4]
	jw0218_RV	CGATTTTAATTCCTCCTTCTCGTTTCGGTAATATGGTTTTTCTTCGG	pFF1_NRPS_6[4]
pJW124 (NRPS-57)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	jw0220_FW	CGAGCTGACTGCAGCTATCTCA TTGTCTTCAGCGCAAAAAGG	pFF1_NRPS_6[4]
	jw0212_RV	CGAGCCGATGATTAATTGTCA TGAAACCGTTACGGTTTGTGTATTA	pFF1_NRPS_6[4]
pJW126 (NRPS-58)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	jw208_FW	GCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATGGG	pFF1_NRPS_6[4]
	jw0222_RV	CGATTTTAATTCCTCCTTCTCGTTGGCATGGCTATTTTCCCATT	pFF1_NRPS_6[4]
pJW128 (NRPS-58)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	jw0224_FW	CGAGCTGACTGCAGCTATCTCACAAAAAGAACGGATGAAGGAGC	pFF1_NRPS_6[4]
	jw0212_RV	CGAGCCGATGATTAATTGTCATGAAACCGTTACGGTTTGTGTATTA	pFF1_NRPS_6[4]
pJW159 (NRPS-10)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pJW62
	na_0194_FW	GGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGG	pJW61
	jw0284_RV	GCCGTCTTTGTATGGTTAAAGGTTTTTAACAACAATGTGCGTTC	pJW61
	jw0285_FW	TTTAACCATACAAAGACGGCATATCCAAAAGGAAAAGCAATCCACCAGCTGTTT	pJW116
	jw0212_RV	CGAGCCGATGATTAATTGTCATGAAACCGTTACGGTTTGTGTATTA	pJW116
pJW160 (NRPS-14)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pJW62

	jw0208_FW	GCTAACAGGAGGAATTCCATGGTTGCTAAACATTCATTAGAAAATGGG	pJW114
	jw0286_RV	ACGGTTTCAGTGGCATTCCAGGACTCTAAAAGTGTCCGTTTTTCTTGAC	pJW114
	jw0287_FW	TGGAATGCCACTGAAACCGTGTATCCTGAATCGTTATGTATTCATCAACTTTTTGAA CAGC	pJW62
	jw0188_RV	GCCTAAACCAATACGCCGT	pJW62
	jw0189_FW	CGGCGTATTGGTTTAGGCCTGT	pJW62
	na07_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	pJW62
pJW161 (NRPS-8)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pJW62
	jw0124_FW	GGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGAAATTATC	pJW75
	jw0288_RV	ACGGTTTCAGTGGCATTCCAATTTTCCAGTAATAACTCCCGCTCAG	pJW75
	jw_0287_FW	TGGAATGCCACTGAAACCGTGTATCCTGAATCGTTATGTATTCATCAACTTTTTGAA CAGC	pJW62
	jw0188_RV	GCCTAAACCAATACGCCGT	pJW62
	jw0189_FW	CGGCGTATTGGTTTAGGCCTGT	pJW62
	na07_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	pJW62
pNA1 (NRPS-21)	na01_FW	CGTGAACAGCTGAAACAGAAACTGGCGGCTCTGCGTAACAAACTGGACGCGTACA AAAACCGTCTG TGACAATTAATCATCGGCTCG	pCK_0402
	na02_FW	AACGAACTGGAATCTCTGGAGAACAAAAAAGAAGAACTGAAGAACCGTAACGAAGA GCTGAAGCAGAAA CGTGAACAGCTGAAACAGAAAC	pCK_0402
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pCK_0402
	na03_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na04_RV	CTCCAGAGATTCCAGTTCGTTCCAGGTTTTTAACAACAATGTGC	*X. nematophila* ATCC 19061
pNA2 (NRPS-25, NRPS-28)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na03_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na05_RV	CGATTTTAATTCCTCCTTCTCGTTAACACGATCACGGGATATTG	*X. nematophila* ATCC 19061
pNA3 (NRPS-25)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	na06	AACGAGCTGACTGCAGCTATCTCATTGCCTTTATCGTTTGGTCAAC	*X. nematophila* ATCC 19061
	na07	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pNA4 (NRPS-26, NRPS-29, NRPS-31, NRPS-32, NRPS-33, NRPS-34, NRPS-35, NRPS-44, NRPS-45, NRPS-46)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na03_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na13_RV	CGATTTTAATTCCTCCTTCTCGTTATAAATCTGGCGGGCGAA	*X. nematophila* ATCC 19061
pNA5 (NRPS-26)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	na14_FW	AACGAGCTGACTGCAGCTATCTCAGTTGCGCCACAAGGAGAA	*X. nematophila* ATCC 19061
	na07_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pNA8 (NRPS-24)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na3_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na17_RV	CGATTTTAATTCCTCCTTCTCGTTTGATCCCGAACCTGAGCCGGATCCAGACCCCCA GGTTTTTAACAACAATGTGC	*X. nematophila* ATCC 19061
pNA9 (NRPS-23)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na3_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na19_RV	CGATTTTAATTCCTCCTTCTCGTTCGAACCTGAGCCGGATCCAGACCCCCAGGTTTT TAACAACAATGTGC	*X. nematophila* ATCC 19061
pNA10 (NRPS-22)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na3_FW	TGGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGG	*X. nematophila* ATCC 19061
	na20_RV	CGATTTTAATTCCTCCTTCTCGTTGGATCCAGACCCCCAGGTTTTTAACAACAATGT G	*X. nematophila* ATCC 19061
pNA15 (NRPS-29, NRPS-31, NRPS-34, NRPS-39, NRPS-40, NRPS-42)	na29_FW	AACCTGGTTGCGCAGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAAC CCTGAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTAC	pJW61
	na30_RV	TGAGATAGCTGCAGTCAGCTCG	pJW61
	na22_FW	GGCTAACAGGAGGAATTCCATGGTTGCGCCACAAGGAGAA	*X. nematophila* ATCC 19061

	na41_RV	CGAGCCGATGATTAATTGTCAATAGACCTGCCGGGCAAAC	*X. nematophila* ATCC 19061
pNA16 (NRPS-29, NRPS-32, NRPS-33, NRPS-37, NRPS-40, NRPS-47)	na33_RV	CTGTTCGTGAGACGCAACTTCGTTTTCGAGACGCGCGATTTCGTCACGCAGGTTCG CGATGATTTTTTCCAGGTTCT GTTCATCACGTTCCAGC	pCDF_ara/tacI
	na34_FW	TGACAATTAATCATCGGCTCG	pCDF_ara/tacI
	na42_FW	TTGGGCTAACAGGAGGAATTCC ATGGCGGCTCCGCAGGG	*X. nematophila* ATCC 19061
	na7_RV	CGAGCCGATGATTAATTGTCA CAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pNA17 (NRPS-28)	na28_FW	CACAAAAAAGACCTGATCGCGTACCTGGAGAAAGAAATCGCGAATCTGCGTAAGAA AATCGAAGAATGACAATTAATCATCGGCTCG	pJW61
	na29_FW	AACCTGGTTGCGCAGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAAC CCTGAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTAC	pJW61
	na30_RV	TGAGATAGCTGCAGTCAGCTCG	pJW61
	na06_FW	AACGAGCTGACTGCAGCTATCTCATTGCCTTTATCGTTTGGTCAAC	*X. nematophila* ATCC 19061
	na37_RV	CGTTTTCGAGCTGCGCAACCAGGTTCATGGCTGGCGTTAGTACCG	*X. nematophila* ATCC 19061
pNA18 (NRPS-28)	na32_RV	GTTCTGTTCATCACGTTCCAGCTGCAGGTTGTCTTTTTTCAGACGTGCGATTTTCTT ACGCAGATACGCGTTACGCGCCATGGAATTCCTCCTGTTAGCC	pCDF_ara/tacI
	na33_RV	CTGTTCGTGAGACGCAACTTCGTTTTCGAGACGCGCGATTTCGTCACGCAGGTTCG CGATGATTTTTTCCAGGTTCTGTTCATCACGTTCCAGC	pCDF_ara/tacI
	na34_FW	TGACAATTAATCATCGGCTCG	pCDF_ara/tacI
	na38_FW	AAACGAAGTTGCGTCTCACGAACAGTTGCCGCTGATTGATCTCAC	*X. nematophila* ATCC 19061
	na7_RV	CGAGCCGATGATTAATTGTCACAGCGCCTCCACTTCG	*X. nematophila* ATCC 19061
pNA26 (NRPS-36, NRPS-37, NRPS-38, NRPS-39, NRPS-40, NRPS-41, NRPS-42, NRPS-43, NRPS-47)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na51	GCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGAAAT	*P. luminescens* TTO1
	na52	CGATTTTAATTCCTCCTTCTCGTTATAAATTTGGCGAGCAAAAGC	*P. luminescens* TTO1
pNA27 (NRPS-33, NRPS-35, NRPS-36, NRPS-37, NRPS-41, NRPS-46)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	na53	CGAGCTGACTGCAGCTATCTCAGTCGCGCCACAGGGAG	*P. luminescens* TTO1
	na54	CGTTTTCGAGCTGCGCAACCAGGTTGTAAGCTTGGCGAGCAAAGG	*P. luminescens* TTO1
pNA28 (NRPS-31, NRPS-35, NRPS-36, NRPS-38, NRPS-39, NRPS-45)	na33_RV	CTGTTCGTGAGACGCAACTTCGTTTTCGAGACGCGCGATTTCGTCACGCAGGTTCG CGATGATTTTTTCCAGGTTCTGTTCATCACGTTCCAGC	pCDF_ara/tacI
	na34_FW	TGACAATTAATCATCGGCTCG	pCDF_ara/tacI
	na55	GAAGTTGCGTCTCACGAACAGCAAGCGCCACAAGGGGA	*P. luminescens* TTO1
	na56	CGAGCCGATGATTAATTGTCACAGCGCCTCCGCTTCAC	*P. luminescens* TTO1
pNA29	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na57	GCTAACAGGAGGAATTCCATGAAAGGTAGTATTGCTAAAAAGGGAGATG	*X. szentirmaii* DSM16338
	na58	CGATTTTAATTCCTCCTTCTCGTTATAATGCTGACGGGCAAATG	*X. szentirmaii* DSM16338
pNA30 (NRPS-32, NRPS-38, NRPS-43, NRPS-44, NRPS-45, NRPS-47)	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	na59	CGAGCTGACTGCAGCTATCTCAGAATCCCCACAAGGGGAGA	*X. szentirmaii* DSM16338
	na60	TCGAGCTGCGCAACCAGGTTATAATGCTGACGGGCAAACG	*X. szentirmaii* DSM16338
pNA31 (NRPS-34, NRPS-41, NRPS-42, NRPS-43, NRPS-44, NRPS-46)	KB-pACYC-II-FW	AACGAGAAGGAGGAATTAAAATCG	pJW61
	KB-pACYC-II-RV	CATGGAATTCCTCCTGTTAGC	pJW61
	na61	AAAACGAAGTTGCGTCTCACGAACAGGAGTTGCCACAAGGTGAAA	*X. szentirmaii* DSM16338
	na62	CGAGCCGATGATTAATTGTCAAATATTATTTACTATATTGTATTCCTCTGTACCA	*X. szentirmaii* DSM16338
pNA35	KB-pCOLA-II-FW	TGACAATTAATCATCGGCTCG	pJW62
	KB-pCOLA-II-RV	TGAGATAGCTGCAGTCAGCTCG	pJW62
	na68	TAACGAGCTGACTGCAGCTATCTCAGAAACTCCGGTCGGTAAAGTAG	*X. bovienii* SS2004
	na69	TTCGTTTTCGAGCTGCGCAACCAGGTTATAGCCGCGCACCACTAC	*X. bovienii* SS2004
pNA125 (NRPS-30)	na285	GTTTGCCCGGCAGGTCTATGAAAACGAAGTTGCGTCTCTGG	pNA15
	na286	ATAGACCTGCCGGGCAAAC	pNA15
pNA126 (NRPS-30)	na141	ATCGCACGTCTGAAAAAAGAC	pNA16
	na43	CATGGAATTCCTCCTGTTAGCC	pNA16
pCOLA_ara_xtpS _tacI_JW	jw0136_FW	CGCTGCTGGTTCTGGCGATTGACAATTAATCATCGGCTCG	pCOLA_ara/tacI
	jw0137_RV	AACGGGTATGGAGAAACAGTAGAGAGTTGCGATAAAAAGCG	pCOLA_ara/tacI
	AL-GxpS-2-1	ACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTCCATGAAAGATAGC ATGGCTAAAAAGG	*X. nematophila* ATCC 19061

pCOLA_ara_gxpS_tacI_JW	AD64	TCGCCAGAACCAGCAGCGGAGCCAGCGGATCCGGCGCGCCTTACAGCGCCTCCAC	*X. nematophila* ATCC 19061
	JW_tacI_PstI_FW2	CTGCAGGAGCTGTTGACAAT	pCOLA_ara/tacI
	jw0064_RV	CATGGAATTCCTCCTGTTAGCC	pCOLA_ara/tacI
	jw0124_FW	GGGCTAACAGGAGGAATTCCATGAAAGATAGCATGGCTAAAAAGGAAATTATC	*P. luminescens* TTO1
	jw0160_RW	GATTAATTGTCAACAGCTCCTGCAGGCAGATAGAGACGTTTGTTGGC	*P. luminescens* TTO1
	jw0151_FW/	GCCAACAAACGTCTCTATCTGCTGGATGAACACCG	*P. luminescens* TTO1
	jw0161_RV	GATTAATTGTCAACAGCTCCTGCAGTCACAGCGCCTCCGCTTCAC	*P. luminescens* TTO1

3 Supplementary Figures

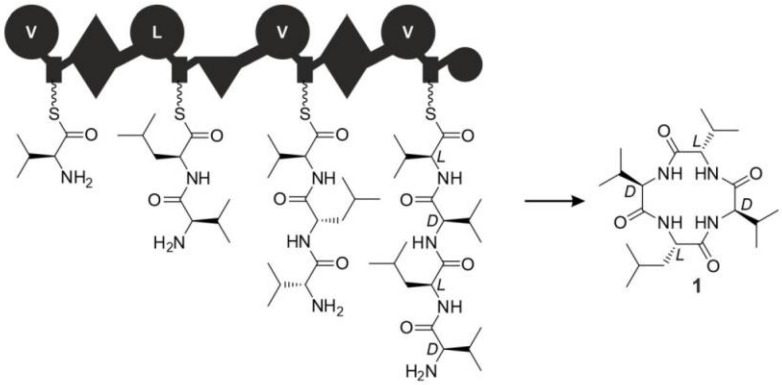


Figure S1. A schematic representation of the xenotetrapeptide (**1**) producing type A NRPS (XtpS). For domain assignment the following symbols are used: A, adenylation domain, large circles; T, thiolation domain, rectangle; C, condensation domain, triangle; C/E, dual condensation/epimerization domain, diamond; TE, thioesterase domain, small circle.

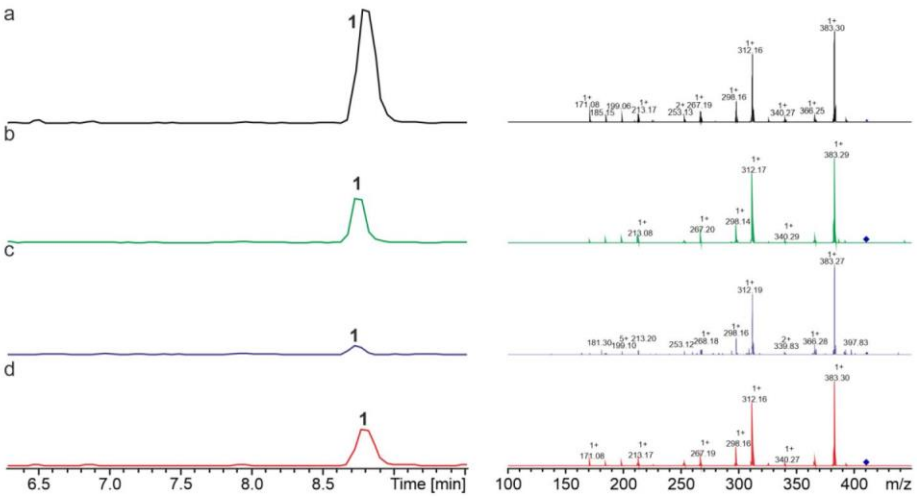


Figure S2. HPLC/MS data refers to Figure 3a (WT XtpS, NRPS-1 and -3) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) Extracted ion chromatogram (EIC)/MS^2 of **1** (*m/z* $[M+H]^+$ = 411.30; WT XtpS). (b) EIC/MS^2 of **1** (*m/z* $[M+H]^+$ = 411.30; NRPS-1). (c) EIC/MS^2 of **1** (*m/z* $[M+H]^+$ = 411.30; NRPS-3). EICs (a–c) are displayed with the same intensity range. (d) EIC/MS^2 data of synthetic **1** (*m/z* $[M+H]^+$ = 411.30).

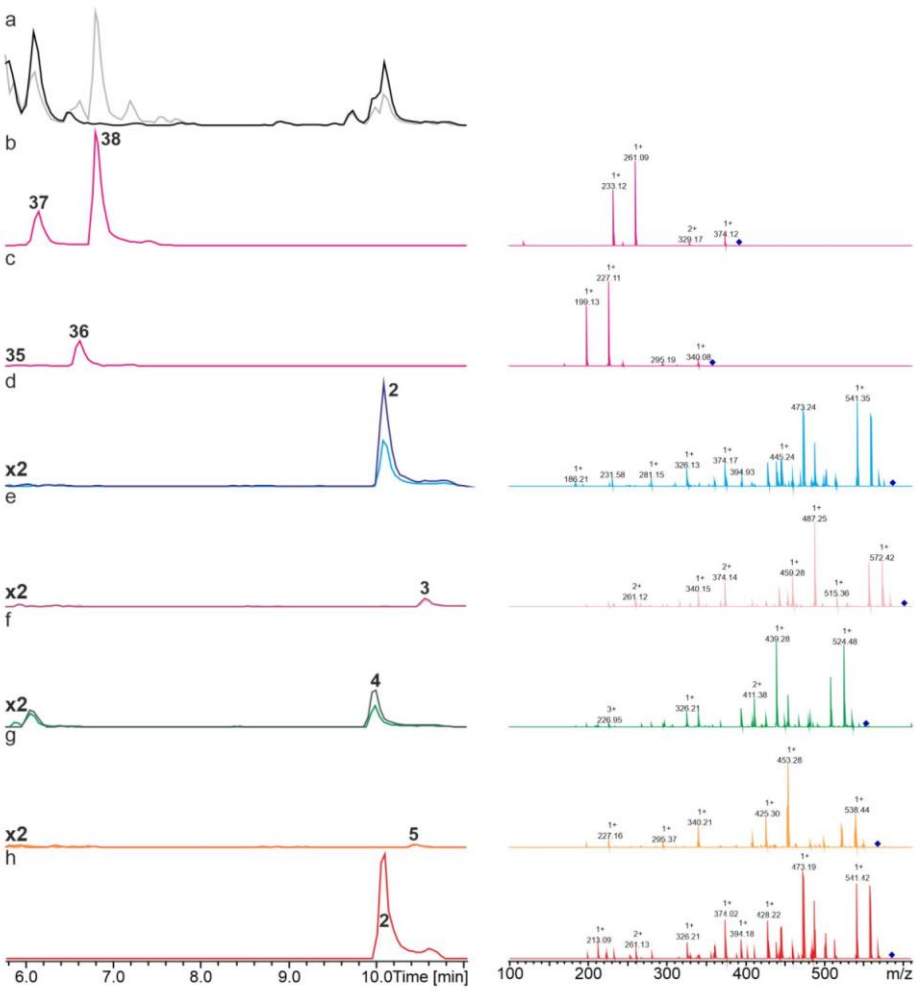


Figure S3. HPLC/MS data refers to Figure 3b (WT GxpS: dark colours, NRPS-5: pale colours) of compounds **2**–**5** (WT GxpS/NRPS-5), **35**/**36** (NRPS-5) and **37**/**38** (NRPS-5) produced in *E. coli* DH10B::*mtaA*. (a) Base Peak Chromatogram (BPC) of an exemplary culture extract. (b) EIC/MS^2 of **37/38** (*m/z* $[M+H]^+$ = 392.25). (c) EIC/MS^2 of **35/36** (*m/z* $[M+H]^+$ = 358.27). (d) EIC/MS^2 of **2** (*m/z* $[M+H]^+$ = 586.40). (e) EIC/MS^2 of **3** (*m/z* $[M+H]^+$ = 600.41). (f) EIC/MS^2 of **4** (*m/z* $[M+H]^+$ = 552.41). (g) EIC/MS^2 **5** (*m/z* $[M+H]^+$ = 566.43). BPCs/EICs (a–c) are displayed with the same intensity range, whereas EICs (d–g) of **2**-**5** are depicted with 2-fold increased intensity. (h) EIC/MS^2 of synthetic **2** (*m/z* $[M+H]^+$ = 586.40).

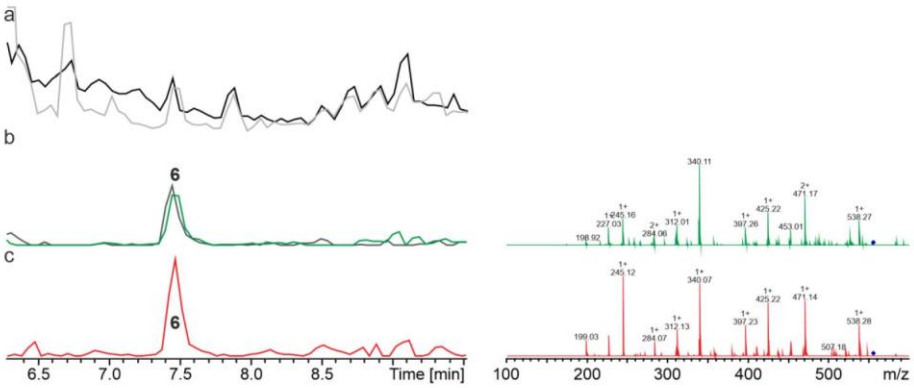


Figure S4. HPLC/MS data refers to Figure 3b (WT RtpS: dark colours, NRPS-6: pale colours) of compound **6** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **6** (*m/z* [M+H]$^+$ = 556.35). BPCs/EICs (a/b) are displayed with the same intensity range. (c) EIC/MS^2 of synthetic **6** (*m/z* [M+H]$^+$ = 556.35).

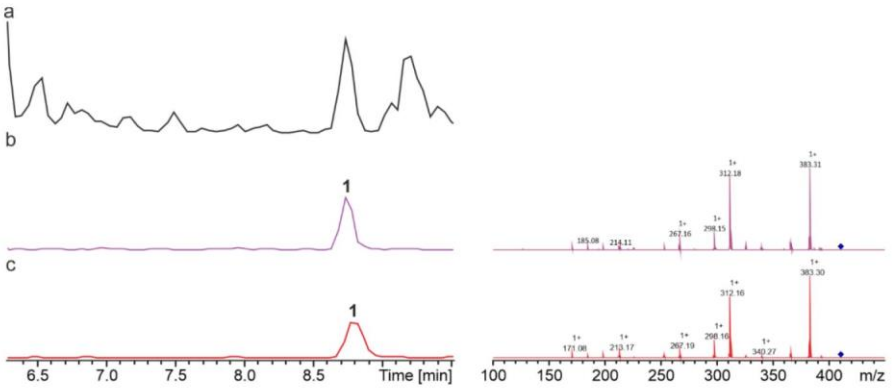


Figure S5. HPLC/MS data refers to Figure 4a (NRPS-7) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **1** (*m/z* [M+H]$^+$ = 411.30). (c) EIC/MS^2 of synthetic **1** (*m/z* [M+H]$^+$ = 411.30).

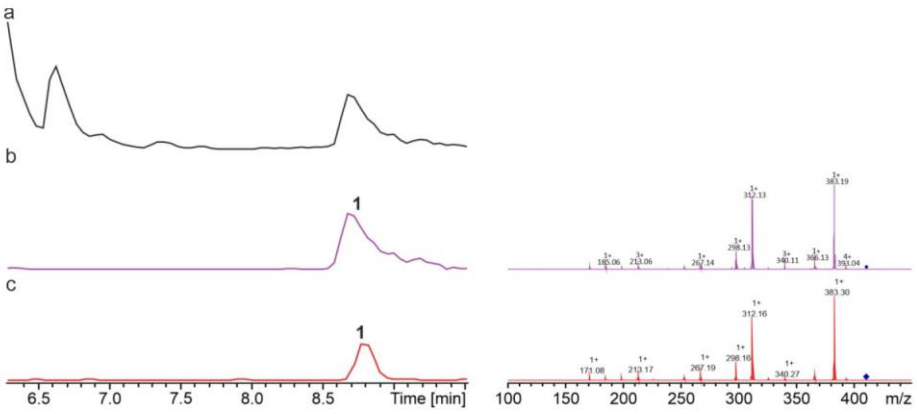


Figure S6. HPLC/MS data refers to Figure 4a (NRPS-8) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **1** (*m/z* $[M+H]^+$ = 411.30). (c) EIC/MS^2 of synthetic **1** (*m/z* $[M+H]^+$ = 411.30).

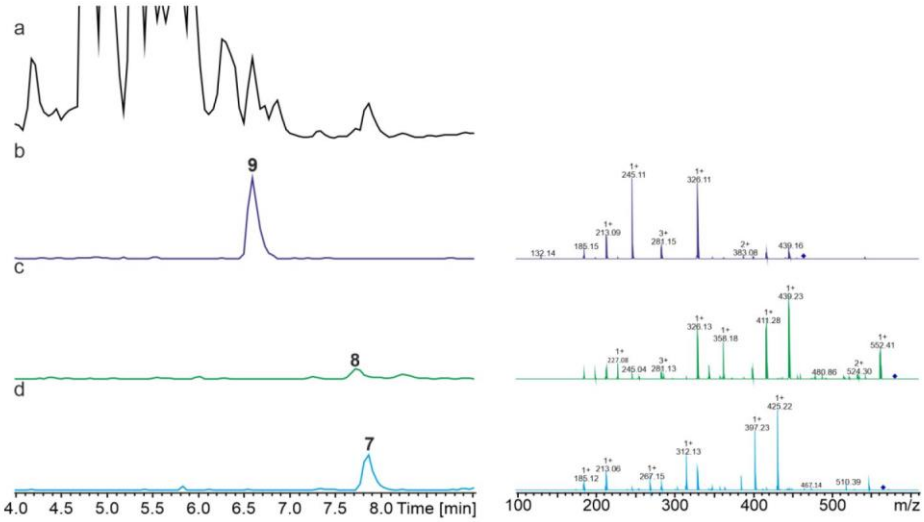


Figure S7. HPLC/MS data refers to Figure 4a (NRPS-9) of compounds **7**, **8** and **9** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **9** (*m/z* $[M+H]^+$ = 457.34). (c) EIC/MS^2 of **8** (*m/z* $[M+H]^+$ = 570.42). (d) EIC/MS^2 of **7** (*m/z* $[M+H]^+$ = 556.41).

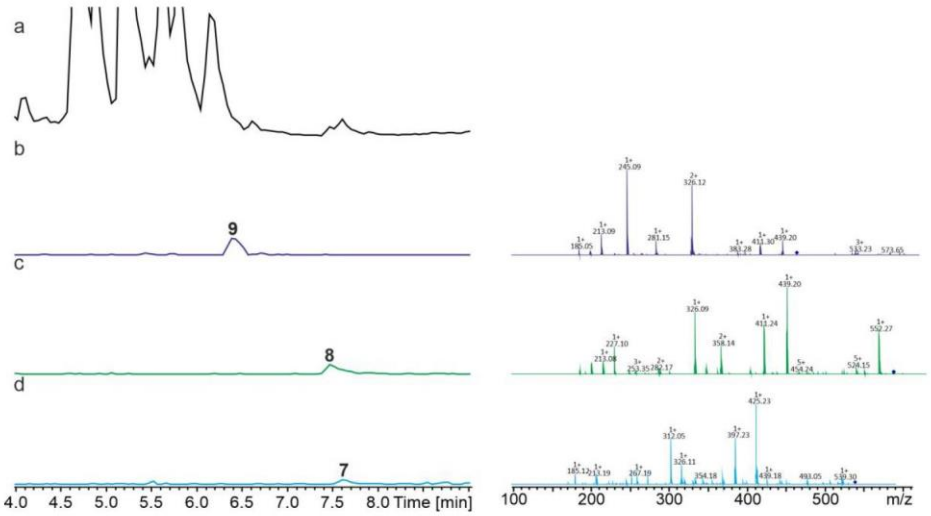


Figure S8. HPLC/MS data refers to Figure 4a (NRPS-10) of compounds **7**, **8** and **9** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **9** (*m/z* $[M+H]^+$ = 457.34). (c) EIC/MS^2 of **8** (*m/z* $[M+H]^+$ = 570.42). (d) EIC/MS^2 of **7** (*m/z* $[M+H]^+$ = 556.41).

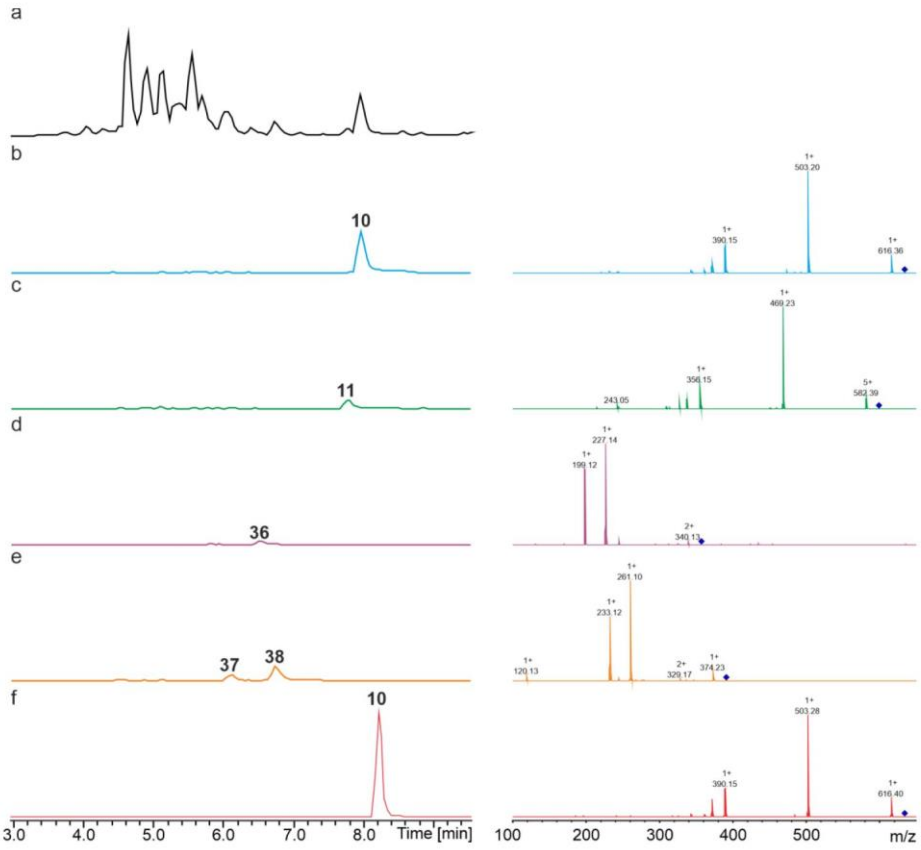


Figure S9. HPLC/MS data refers to Figure 4a (NRPS-11) of compounds **10**, **11**, **36** and **37**/**38** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **10** (*m/z* $[M+H]^+$ = 634.38). (c) EIC/MS^2 data of **11** (*m/z* $[M+H]^+$ = 600.40). (d) EIC/MS^2 data of **36** (*m/z* $[M+H]^+$ = 358.27). (e) EIC/MS^2 data of **37**/**38** (*m/z* $[M+H]^+$ = 392.25). (f) EIC/MS^2 data of synthetic **10** (*m/z* $[M+H]^+$ = 634.38).

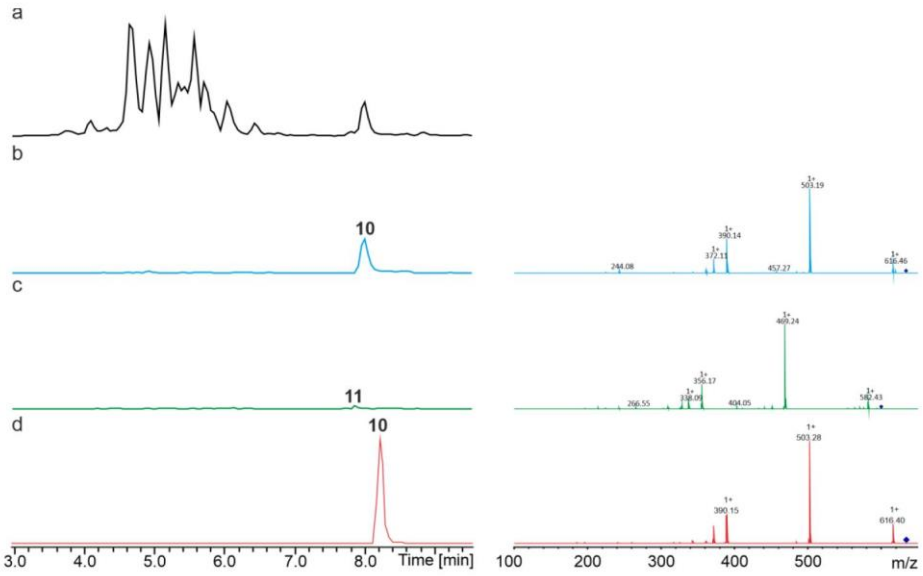


Figure S10. HPLC/MS data refers to Figure 4a (NRPS-12) of compounds **10** and **11** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **10** (*m/z* $[M+H]^+$ = 634.38). (c) EIC/MS^2 data of **11** (*m/z* $[M+H]^+$ = 600.40). (d) EIC/MS^2 data of synthetic **10** (*m/z* $[M+H]^+$ = 634.38).

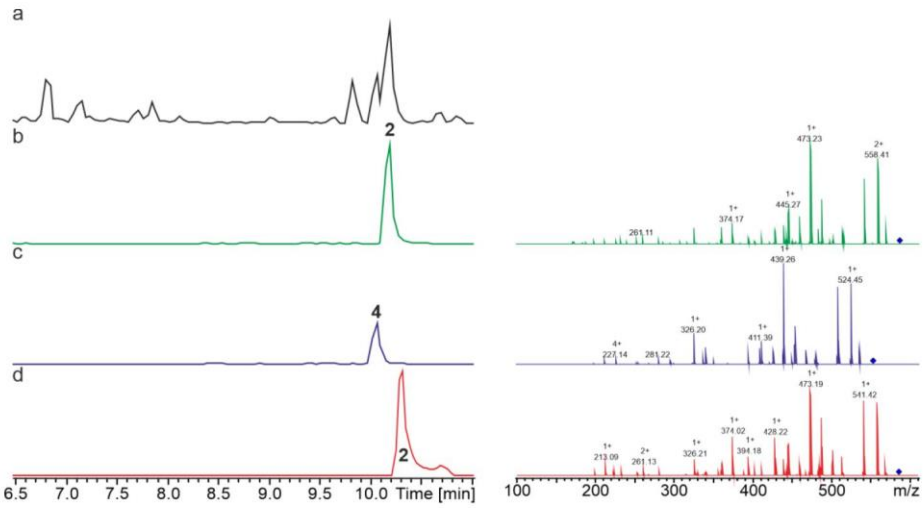


Figure S11. HPLC/MS data refers to Figure 4b (NRPS-15) of compounds **2** and **4** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **2** (*m/z* $[M+H]^+$ = 586.40). (c) EIC/MS^2 of **4** (*m/z* $[M+H]^+$ = 552.41). (d) EIC/MS^2 of synthetic **2** (*m/z* $[M+H]^+$ = 586.40).

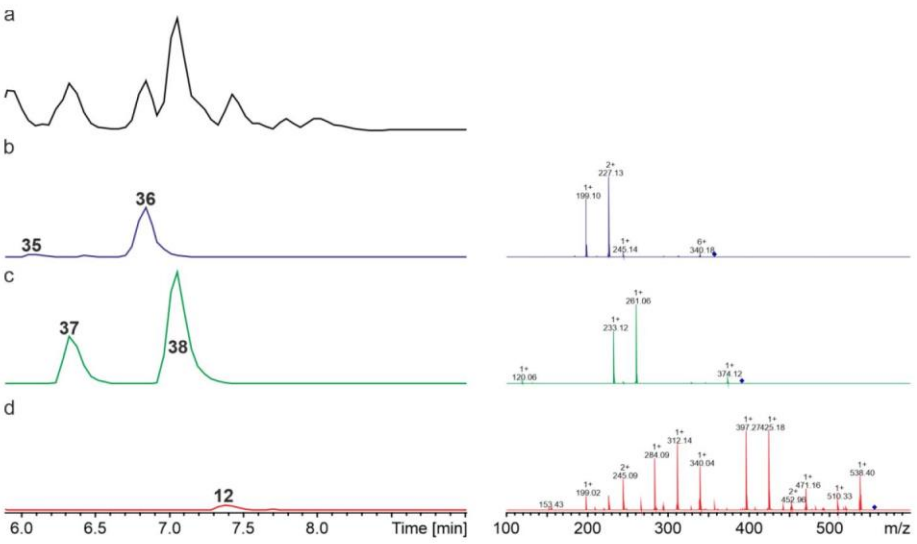


Figure S12. HPLC/MS data refers to Figure 4b (NRPS-16) of compounds **35**/**36**, **37**/**38** and **12** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data data of **35**/**36** (*m/z* $[M+H]^+$ = 358.27). (c) EIC/MS^2 data of **37**/**38** (*m/z* $[M+H]^+$ = 392.25). (d) EIC/MS^2 of **12** (*m/z* $[M+H]^+$ = 556.35).

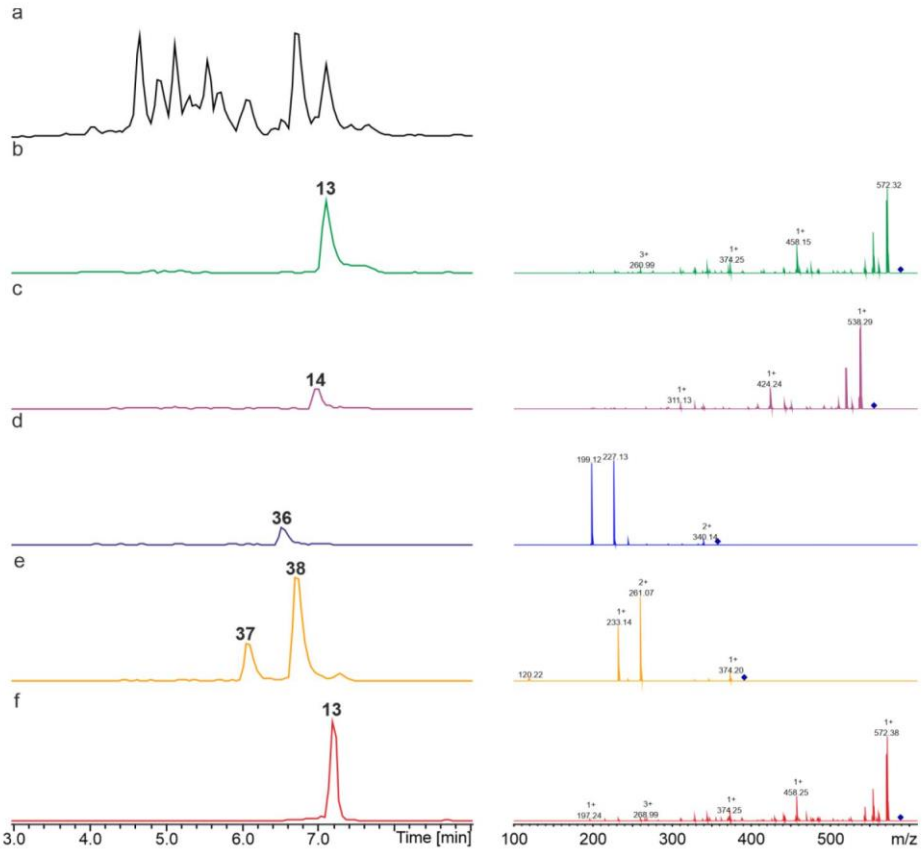


Figure S13. HPLC/MS data refers to Figure 4b (NRPS-17) of compounds **13**, **14**, **36** and **37**/**38** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **13** (*m/z* $[M+H]^+$ = 589.33). (c) EIC/MS^2 data of **14** (*m/z* $[M+H]^+$ = 555.35). (d) EIC/MS^2 data data of **36** (*m/z* $[M+H]^+$ = 358.27). (e) EIC/MS^2 data of **37/38** (*m/z* $[M+H]^+$ = 392.25). (f) EIC/MS^2 data of synthetic **13** (*m/z* $[M+H]^+$ = 589.33).

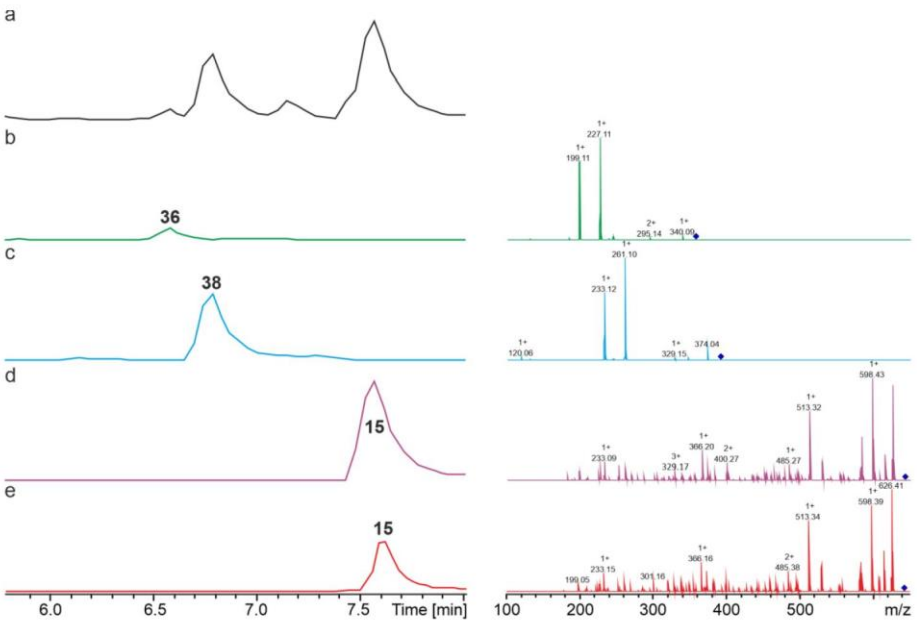


Figure S14. HPLC/MS data refers to Figure 4b (NRPS-18) of compounds **36**, **38** and **15** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **36** (*m/z* $[M+H]^+$ = 358.27). (c) EIC/MS^2 data of **38** (*m/z* $[M+H]^+$ = 392.25). (d) EIC/MS^2 data of **15** (*m/z* $[M+H]^+$ = 643.43). (e) EIC/MS^2 data of synthetic **15** (*m/z* $[M+H]^+$ = 643.43).

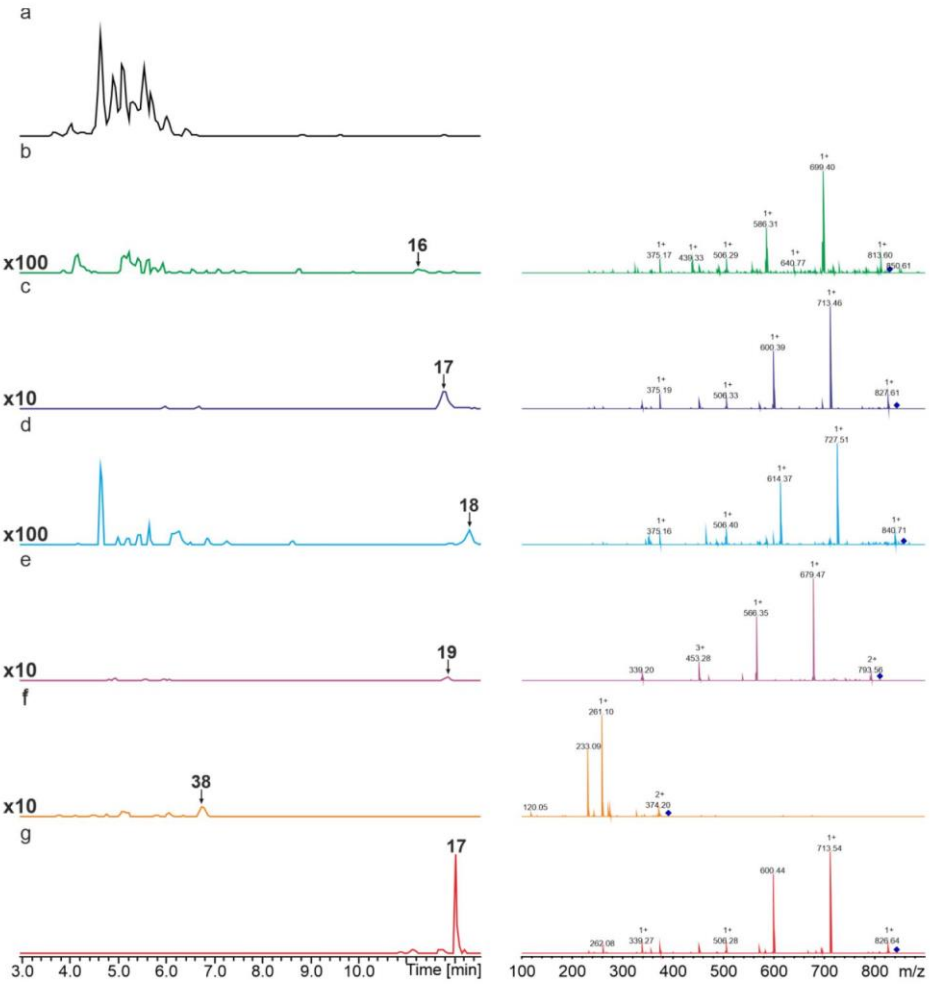


Figure S15. HPLC/MS data refers to Figure 4b (NRPS-19) of compounds **16**–**19** and **38** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **16** (*m/z* [M+H]$^+$ = 830.54). (c) EIC/MS^2 data of **17** (*m/z* [M+H]$^+$ = 844.55). (d) EIC/MS^2 data of **18** (*m/z* [M+H]$^+$ = 858.57). (e) EIC/MS^2 data of **19** (*m/z* [M+H]$^+$ = 810.57). (f) EIC/MS^2 data of **38** (*m/z* [M+H]$^+$ = 392.25). (g) EIC/MS^2 data of synthetic **17** (*m/z* [M+H]$^+$ = 844.55).

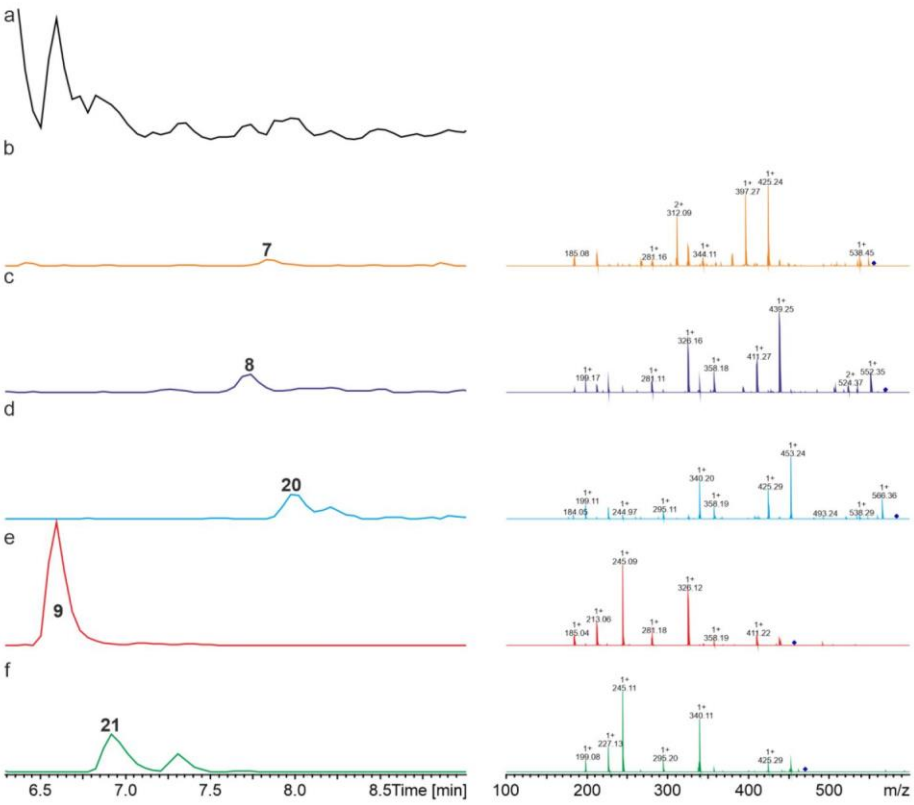
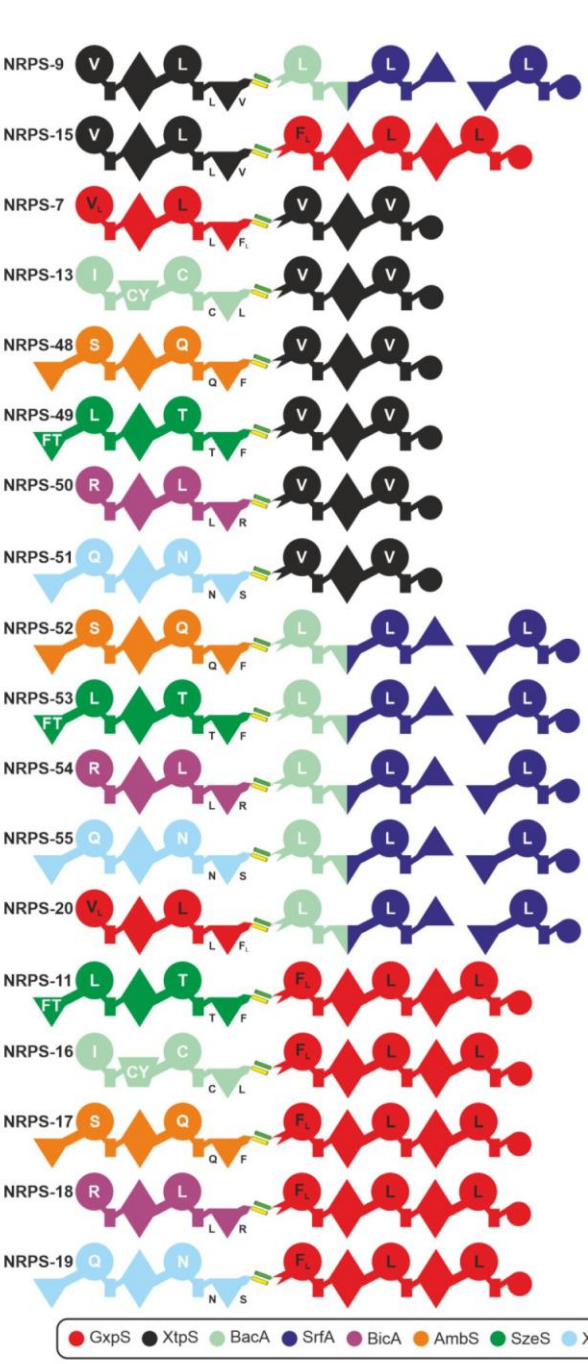
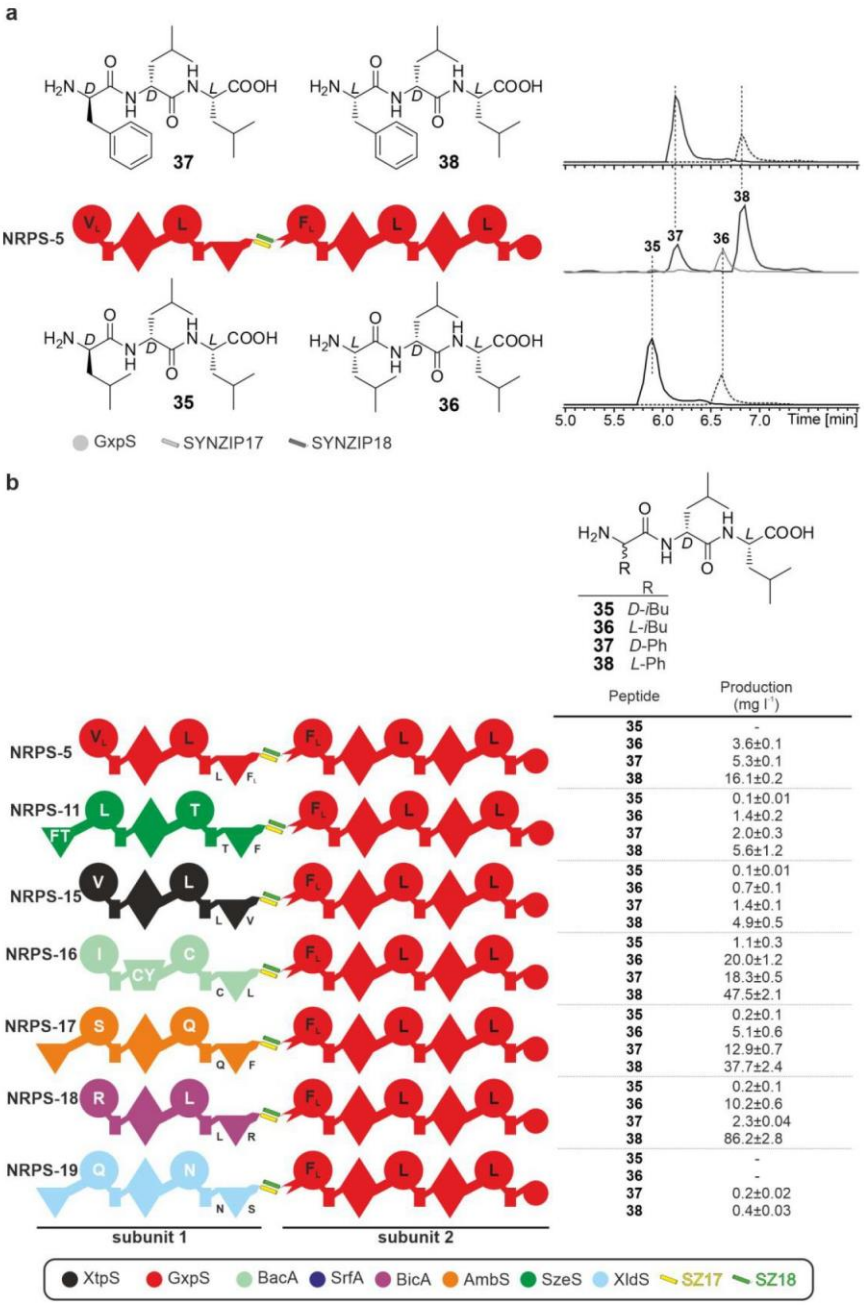


Figure S16. HPLC/MS data refers to Figure 4c (NRPS-20) of compounds **7**, **8**, **20**, **9** and **11** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 of **7** (*m/z* [M+H]$^+$ = 556.41). (c) EIC/MS^2 of **8** (*m/z* [M+H]$^+$ = 570.42). (d) EIC/MS^2 of **20** (*m/z* [M+H]$^+$ = 584.44). (e) EIC/MS^2 of **9** (*m/z* [M+H]$^+$ = 457.34). (f) EIC/MS^2 of **21** (*m/z* [M+H]$^+$ = 471.35).



Peptide	Production (mg l^{-1})
7 8 9	2.7 ± 0.7 1.2 ± 0.3 1.3 ± 0.2
2 3 4 5	9.3 ± 0.6 - 4.3 ± 0.2 -
1	48.3 ± 2.8
no production	
no production	
no production	
no production	
no production	
no production	
no production	
no production	
no production	
7 8 9 20 21	0.5 ± 0.1 0.2 ± 0.02 0.8 ± 0.1 0.7 ± 0.1 0.3 ± 0.03
10 11	23.8 ± 2.6 4.1 ± 1.1
12	53.5 ± 5.0
13 14	218.6 ± 21.3 46.2 ± 3.3
15	39.9 ± 1.4
16 17 18 19	0.1 ± 0.03 5.5 ± 0.2 0.4 ± 0.04 0.9 ± 0.3

Figure S17. A schematic representation of all bipartite type S NRPSs (NRPS-7, -9, -11, -13, -15 – -19, -20 and -48 – -55) using subunit 1 building blocks from GxpS, XtpS, BacA, AmbS XldS, SzeS and BicA combined with subunit 2 building blocks from GxpS, XtpS and RtpS.



Peptide	Production (mg l^{-1})
35	-
36	3.6±0.1
37	5.3±0.1
38	16.1±0.2
35	0.1±0.01
36	1.4±0.2
37	2.0±0.3
38	5.6±1.2
35	0.1±0.01
36	0.7±0.1
37	1.4±0.1
38	4.9±0.5
35	1.1±0.3
36	20.0±1.2
37	18.3±0.5
38	47.5±2.1
35	0.2±0.1
36	5.1±0.6
37	12.9±0.7
38	37.7±2.4
35	0.2±0.1
36	10.2±0.6
37	2.3±0.04
38	86.2±2.8
35	-
36	-
37	0.2±0.02
38	0.4±0.03

Figure S18. (a) Production of *D*/*L*-tripeptides exemplary of NRPS-5. The tripeptide production is related to the unpaired activity of GxpS subunit 2 resulted in the production of peptides **35/36** and **37/38**. The different epimers could be identified by their retention times. (b) Tripeptide **35/36** and **37/38** amounts and yields (determined in triplicates (n=3)) are given. The colour code of the NRPS subunits is depicted at the bottom of the figures. The domain assignment is as described in Figure 3 and Figure 4.

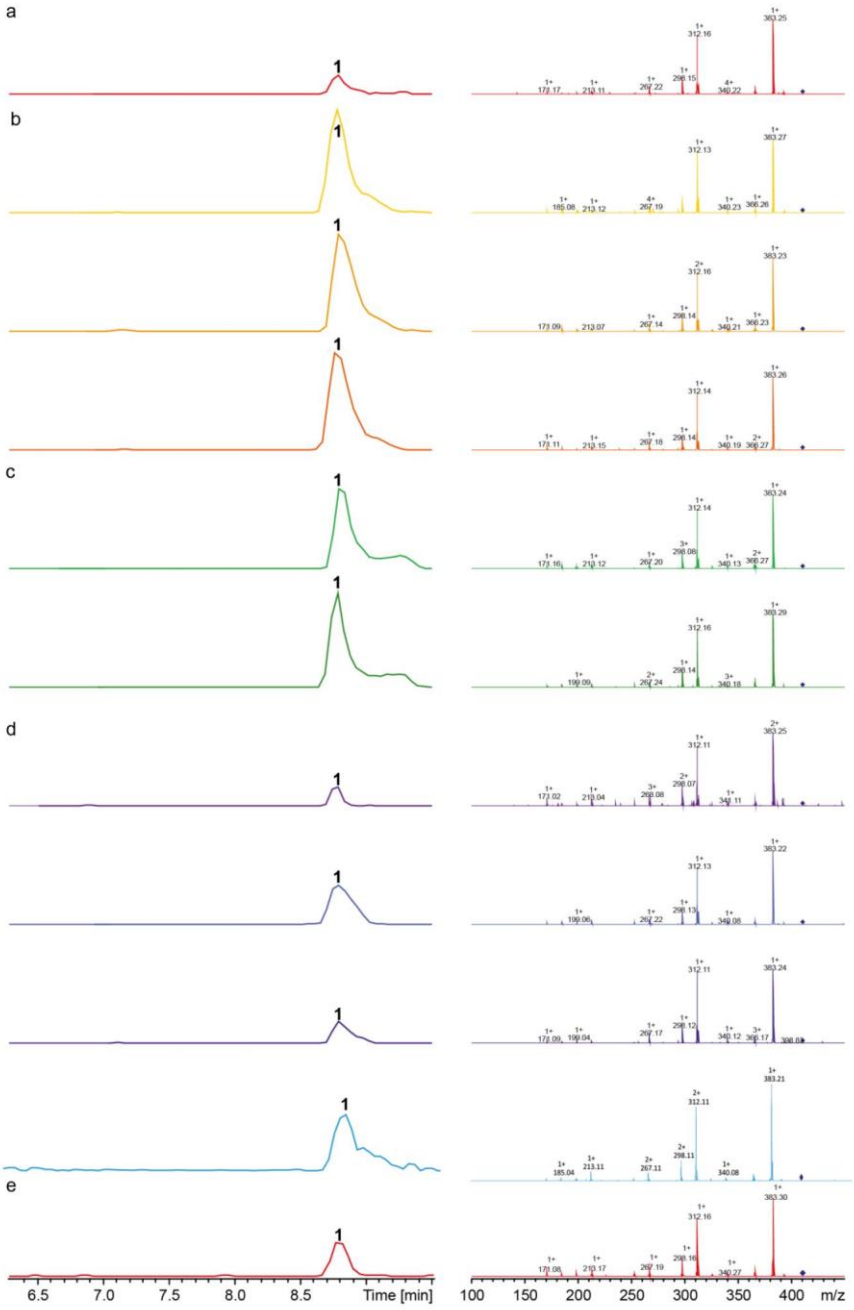


Figure S19. HPLC/MS data refers to Figure 5a (NRPS-21), Figure 5b (NRPS-22–24), Figure 5c (NRPS-25 and NRPS-26) and Figure 5d (NRPS-27–29) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) EIC/MS^2 (NRPS-21) of **1** (*m/z* [M+H]$^+$ = 411.30). (b) EIC/MS^2 (NRPS-22–24) of **1** (*m/z* [M+H]$^+$ = 411.30). (c) EIC/MS^2 (NRPS-25 and NRPS-26) of **1** (*m/z* [M+H]$^+$ = 411.30). (d) EIC/MS^2 (NRPS-27–29) of **1** (*m/z* [M+H]$^+$ = 411.30). (e) EIC/MS^2 data of synthetic **1** (*m/z* [M+H]$^+$ = 411.30).

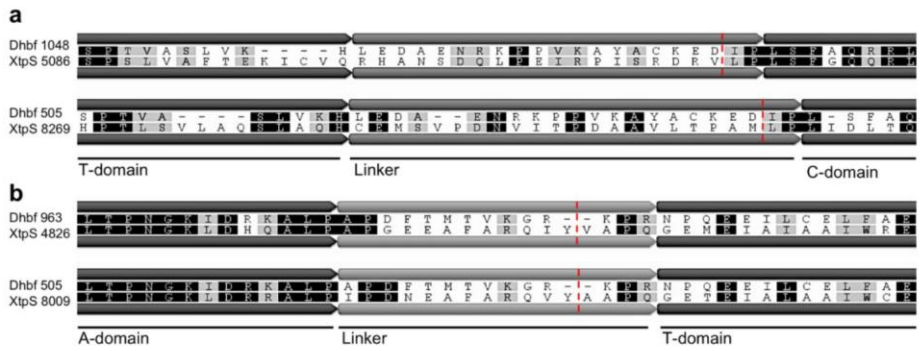
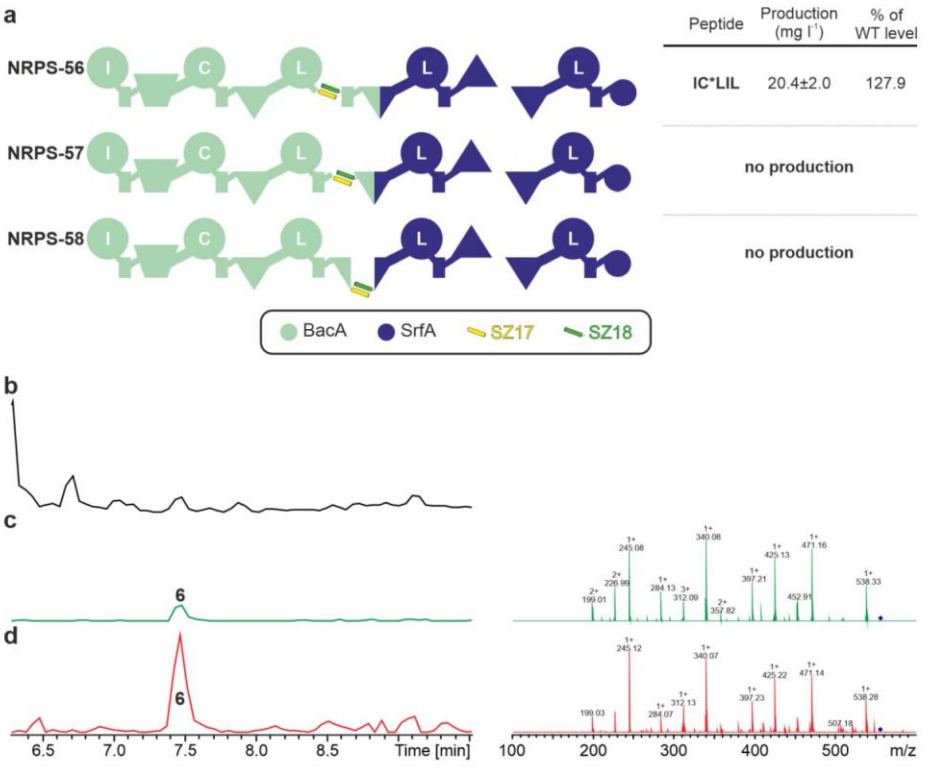


Figure S20. Sequence alignments of XtpS linker sequences. (a) XtpS T_2-C_2 (XtpS 5086) and T_3-C_3 (XtpS 8269) linker were aligned to the linker region excised from Dhbf crystal structure (Protein Database ID: 5U89). (b) XtpS A_2-T_2 (XtpS 4826) and A_3-T_3 (XtpS 8009) linker were aligned to the linker region excised from Dhbf crystal structure. Red dashed lines indicate SYNZIP insertion points.



Peptide	Production (mg l^{-1})	% of WT level
IC*LIL	20.4±2.0	127.9
no production		
no production		

Figure S21. (a) Bipartite split of RtpS in-between and within the modules. HPLC/MS data refers to Supplementary Figure 21 (NRPS-56) of compound **6** produced in *E. coli* DH10B::*mtaA*. (b) BPC of an exemplary culture extract. (c) EIC/MS^2 of **6** (*m/z* $[M+H]^+$ = 556.35). (d) EIC/MS^2 of synthetic **6** (*m/z* $[M+H]^+$ = 556.35).

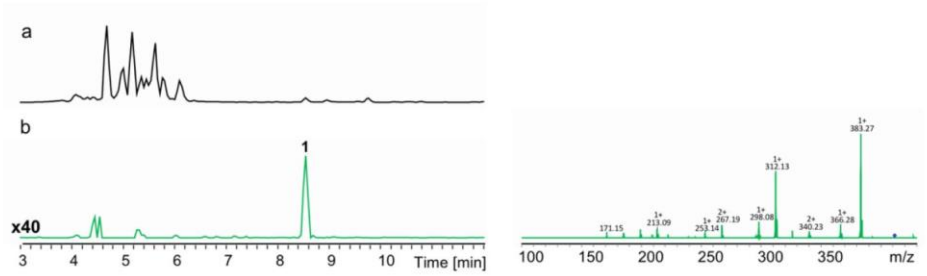


Figure S22. HPLC/MS data refers to Figure 6 (NRPS-29) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **1** (*m/z* $[M+H]^+$ = 411.30).

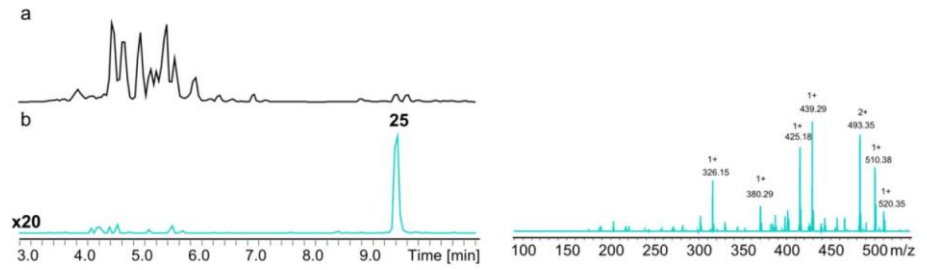


Figure S23. HPLC/MS data refers to Figure 6 (NRPS-31) of compound **25** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **25** (*m/z* $[M+H]^+$ = 538.40).

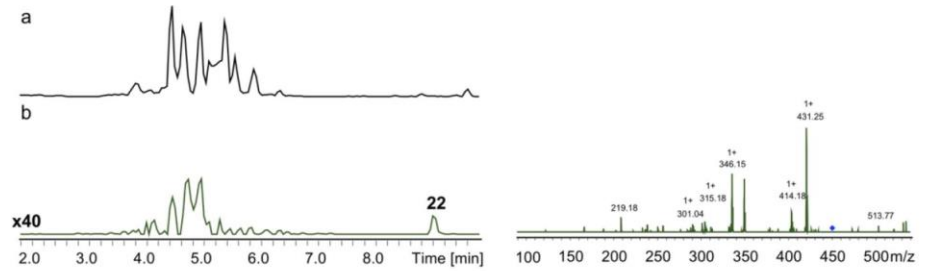


Figure S24. HPLC/MS data refers to Figure 6 (NRPS-32) of compound **22** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **22** (*m/z* $[M+H]^+$ = 459.30).

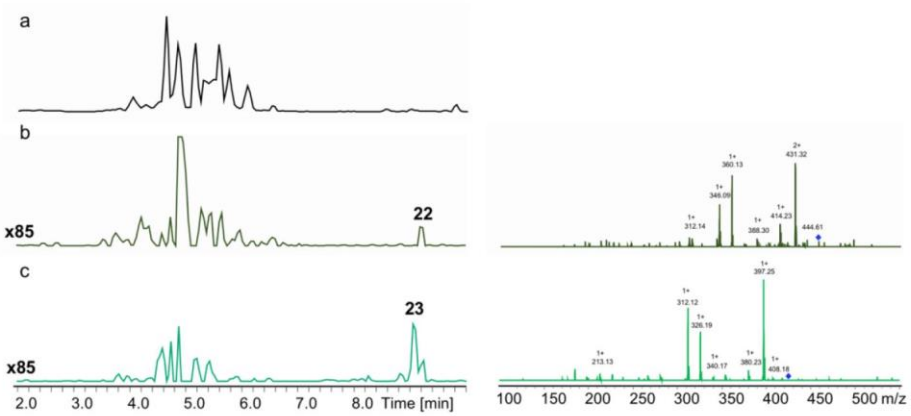


Figure S25. HPLC/MS data refers to Figure 6 (NRPS-33) and (NRPS-37) of compounds **22** and **23** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **22** (m/z $[M+H]^+$ = 459.30). (c) EIC/MS^2 data of **23** (m/z $[M+H]^+$ = 425.31).

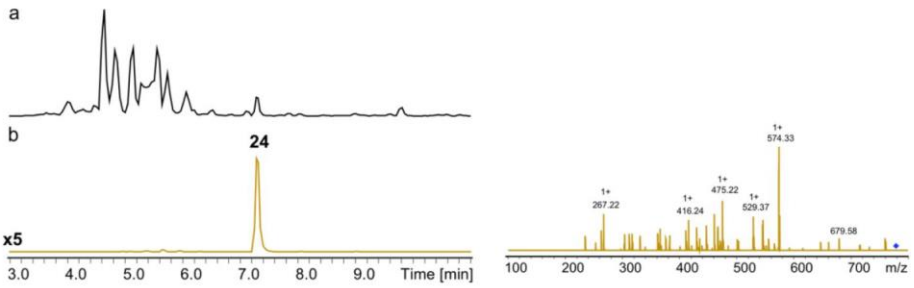


Figure S26. HPLC/MS data refers to Figure 6 (NRPS-34) of compound **24** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **24** (*m/z* $[M+H]^+$ = 778.45).

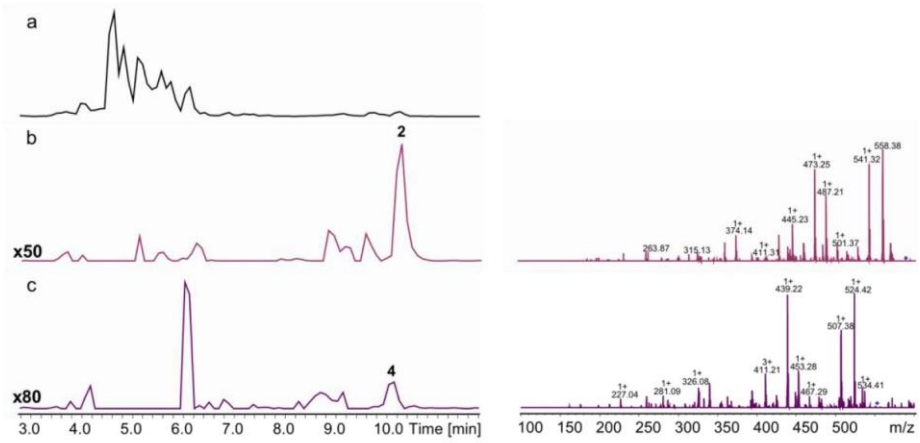


Figure S27. HPLC/MS data refers to Figure 6 (NRPS-35) of compounds **2** and **4** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **2** (*m/z* [M+H]$^+$ = 586.40). (c) EIC/MS^2 data of **4** (*m/z* [M+H]$^+$ = 552.41).

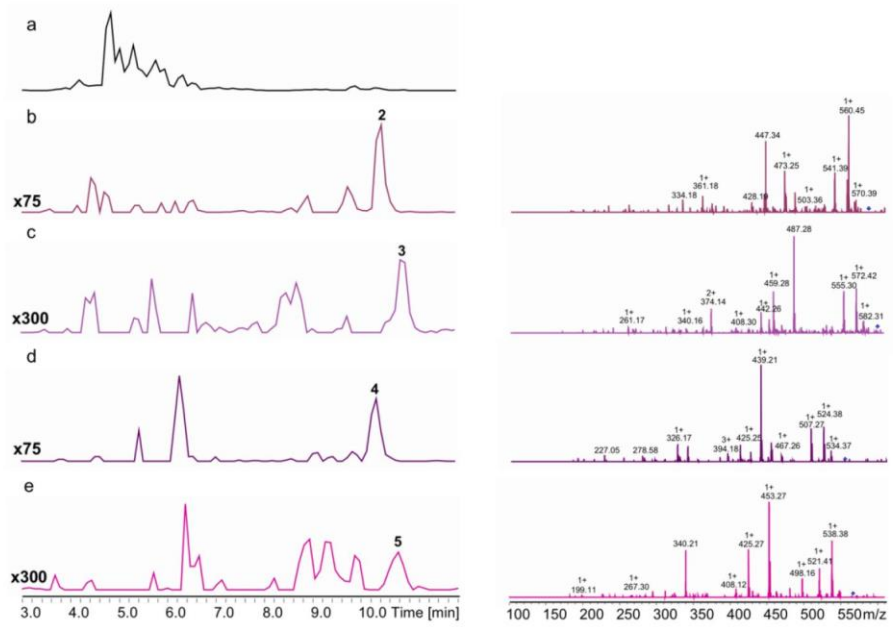


Figure S28. HPLC/MS data refers to Figure 6 (NRPS-36) of compounds **2**, **3**, **4** and **5** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **2** (m/z [M+H]$^+$ = 586.40). (c) EIC/MS^2 data of **3** (m/z [M+H]$^+$ = 600.41). (d) EIC/MS^2 data of **3** (m/z [M+H]$^+$ = 552.41). (e) EIC/MS^2 data of **3** (m/z [M+H]$^+$ = 566.43).

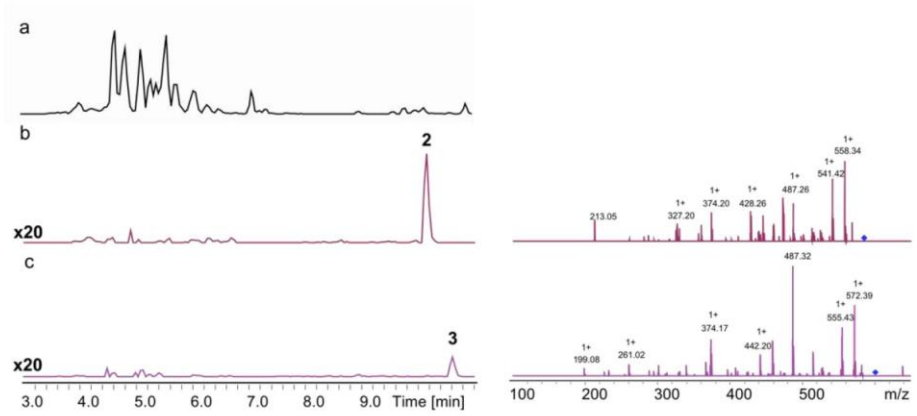


Figure S29. HPLC/MS data refers to Figure 6 (NRPS-38) of compounds **2** and **3** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **2** (m/z $[M+H]^+$ = 586.40). (c) EIC/MS^2 data of **3** (m/z $[M+H]^+$ = 600.41).

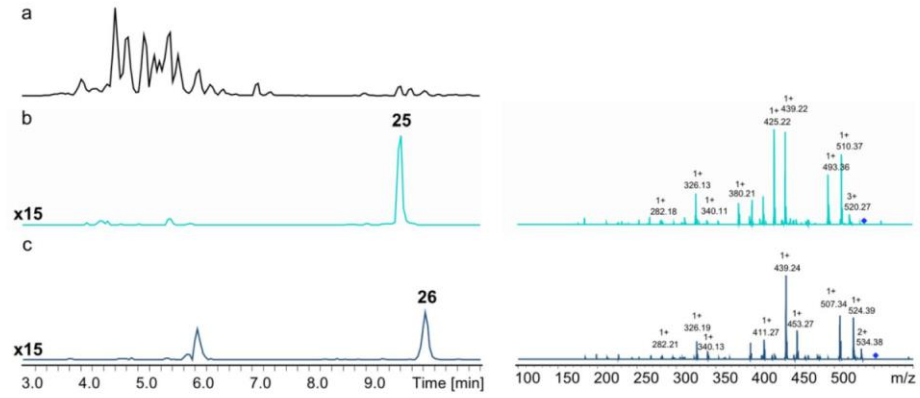


Figure S30. HPLC/MS data refers to Figure 6 (NRPS-39) of compounds **25** and **26** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **25** (m/z $[M+H]^+$ = 588.40). (c) EIC/MS^2 data of **26** (m/z $[M+H]^+$ = 552.41).

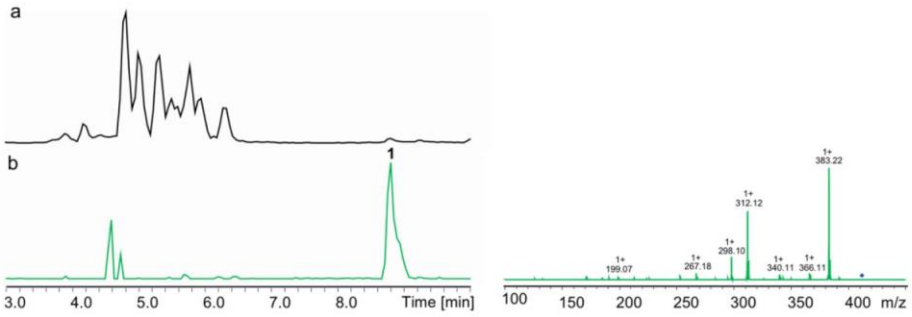


Figure S31. HPLC/MS data refers to Figure 6 (NRPS-40) of compound **1** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **1** (m/z $[M+H]^+$ = 411.29).

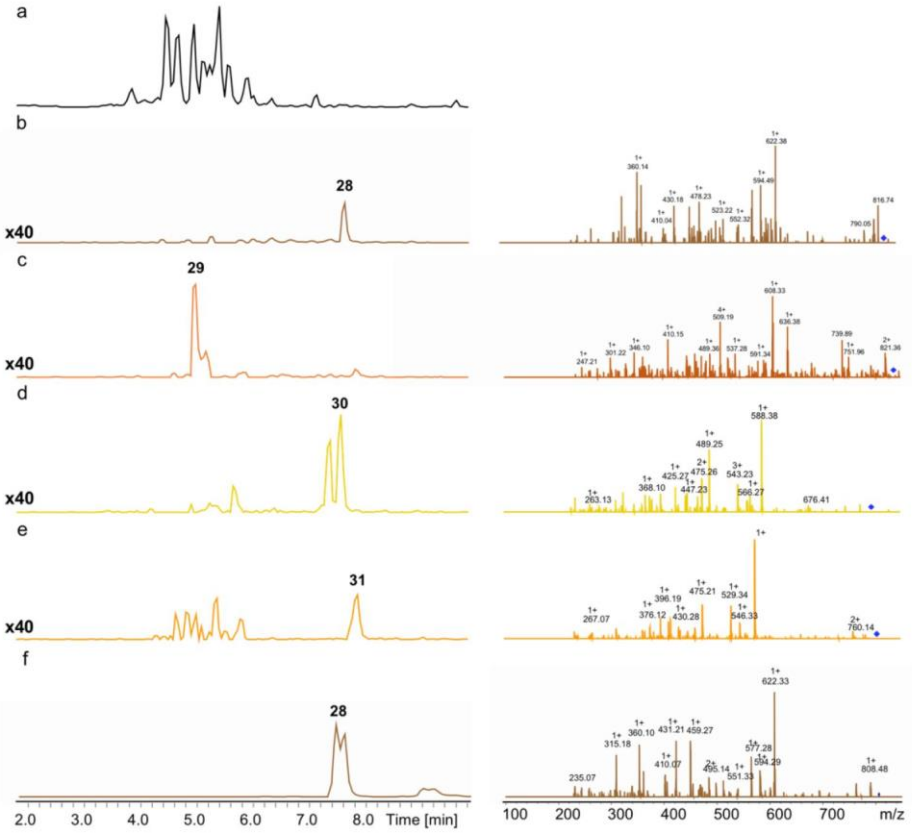


Figure S32. HPLC/MS data refers to Figure 6 (NRPS-41) of compounds **28**, **29**, **30** and **31** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **28** (m/z $[M+H]^+$ = 826.45). (c) EIC/MS^2 data of **29** (m/z $[M+H]^+$ = 840.47). (d) EIC/MS^2 data of **30** (m/z $[M+H]^+$ = 792.47). (e) EIC/MS^2 data of **31** (m/z $[M+H]^+$ = 806.48). (f) EIC/MS^2 data of synthetic **28** (m/z $[M+H]^+$ = 826.45).

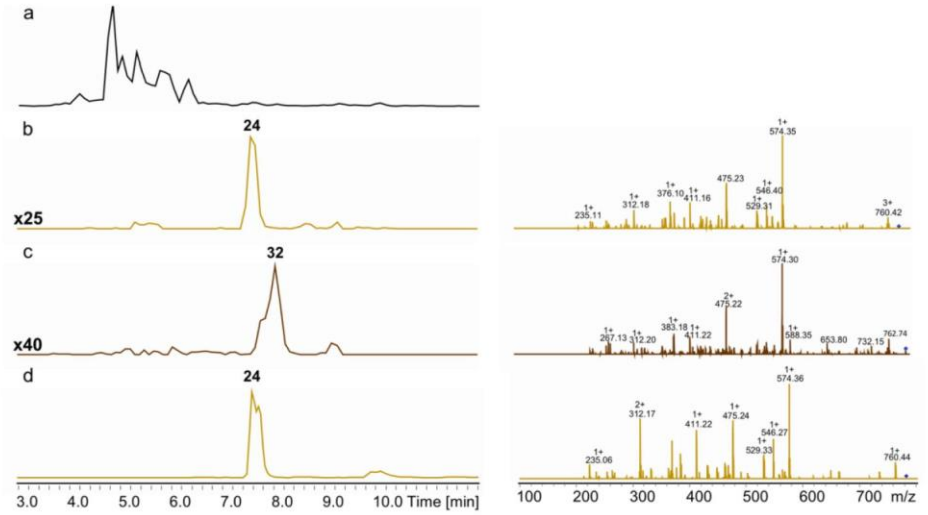


Figure S33. HPLC/MS data refers to Figure 6 (NRPS-42) of compounds **24** and **32** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **24** (m/z $[M+H]^+$ = 778.45). (c) EIC/MS^2 data of **32** (m/z $[M+H]^+$ = 792.47). (d) EIC/MS^2 data of synthetic **24** (m/z $[M+H]^+$ = 778.45).

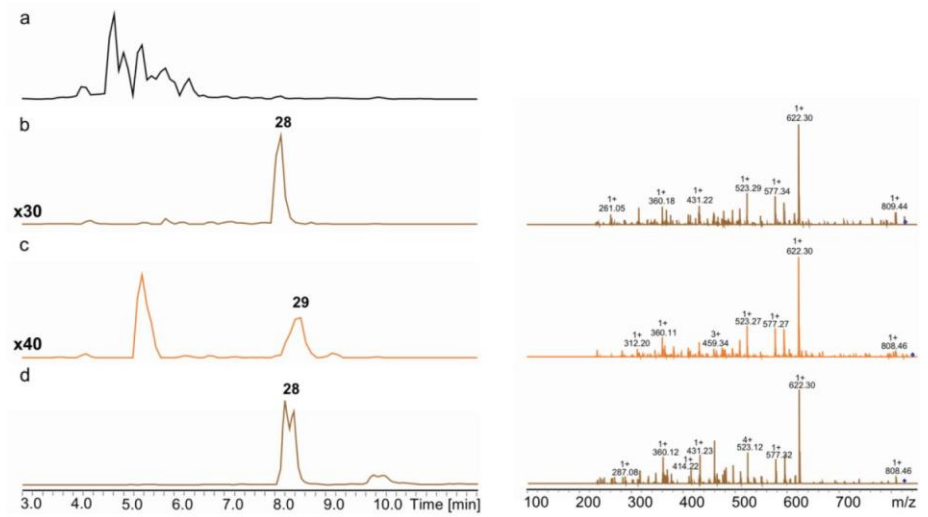


Figure S34. HPLC/MS data refers to Figure 6 (NRPS-43) of compounds **28** and **29** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **28** (m/z $[M+H]^+$ = 826.45). (c) EIC/MS^2 data of **29** (m/z $[M+H]^+$ = 840.47). (d) EIC/MS^2 data of synthetic **28** (m/z $[M+H]^+$ = 826.45).

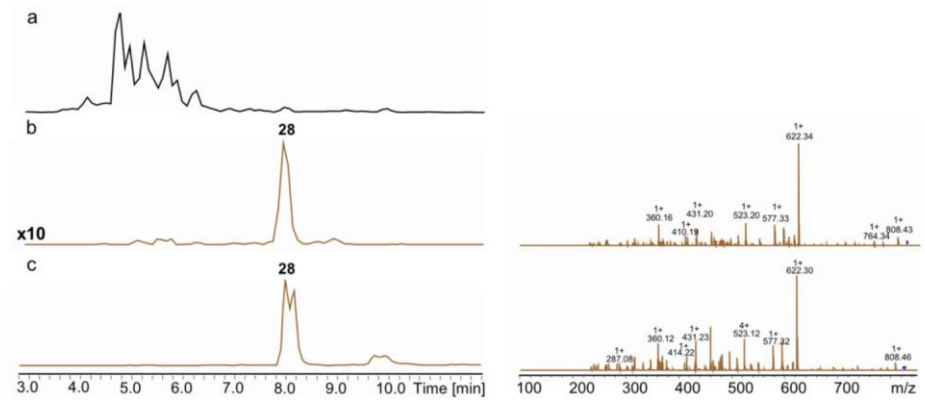


Figure S35. HPLC/MS data refers to Figure 6 (NRPS-44) of compound **28** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **28** (m/z $[M+H]^+$ = 826.45). (c) EIC/MS^2 data of synthetic **28** (m/z $[M+H]^+$ = 826.45).

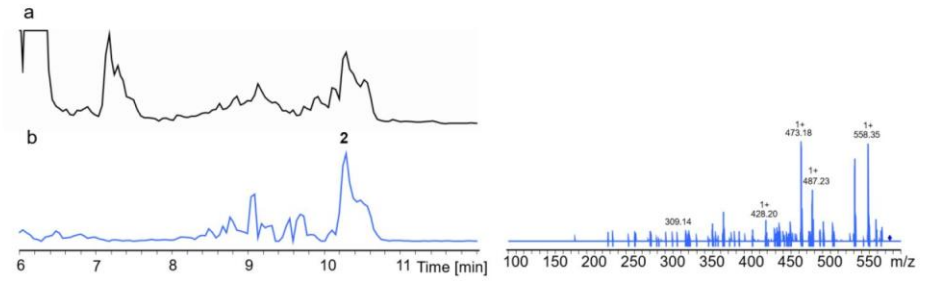


Figure S36. HPLC/MS data refers to Figure 6 (NRPS-45) of compound **2** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **2** (m/z $[M+H]^+$ = 586.40).

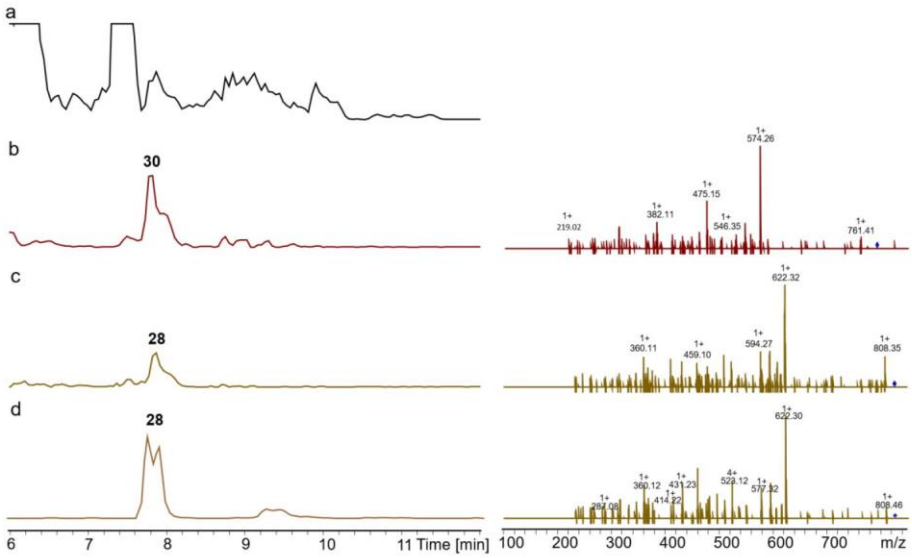


Figure S37. HPLC/MS data refers to Figure 6 (NRPS-46) of compounds **30** and **28** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **30** (m/z $[M+H]^+$ = 792.47). (c) EIC/MS^2 data of **28** (m/z $[M+H]^+$ = 826.45). (d) EIC/MS^2 data of synthetic **28** (m/z $[M+H]^+$ = 826.45).

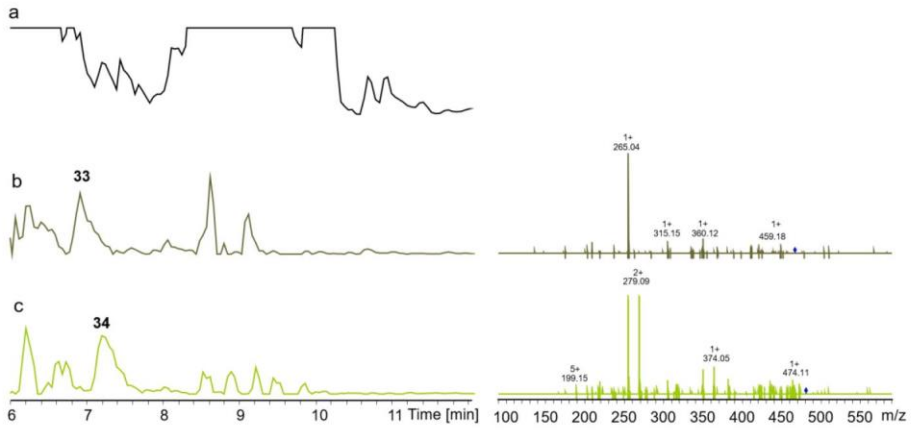


Figure S38. HPLC/MS data refers to Figure 6 (NRPS-47) of compounds **33** and **34** produced in *E. coli* DH10B::*mtaA*. (a) BPC of an exemplary culture extract. (b) EIC/MS^2 data of **33** (m/z $[M+H]^+$ = 476.30). (c) EIC/MS^2 data of **34** (m/z $[M+H]^+$ = 490.32).

4 References

[1] C. Fu, W. P. Donovan, O. Shikapwashya-Hasser, X. Ye, R. H. Cole, *PLoS One* **2014**, *9*, e115318.
[2] K. E. Thompson, C. J. Bashor, W. A. Lim, A. E. Keating, *ACS Synth. Biol.* **2012**, *1*, 118–129.
[3] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275–281.
[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] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün, *Chemistry* **2012**, *18*, 2342–2348.
[7] D. Hanahan, *J. Mol. Biol.* **1983**, *166*, 557–580.
[8] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290–1294.
[9] N. J. Tobias, H. Wolff, B. Djahanschiri, F. Grundmann, M. Kronenwerth, Y.-M. Shi, S. Simonyi, P. Grün, D. Shapiro-Ilan, S. J. Pidot et al., *Nat. Microbiol.* **2017**, *2*, 1676–1685.
[10] J. M. Chaston, G. Suen, S. L. Tucker, A. W. Andersen, A. Bhasin, E. Bode, H. B. Bode, A. O. Brachmann, C. E. Cowles, K. N. Cowles et al., *PLoS One* **2011**, *6*, e27909.
[11] S. W. Fuchs, C. C. Sachs, C. Kegler, F. I. Nollmann, M. Karas, H. B. Bode, *Anal. Chem.* **2012**, *84*, 6948–6955.
[12] E. Bode, A. K. Heinrich, M. Hirschmann, D. Abebew, Y.-N. Shi, T. D. Vo, F. Wesche, Y.-M. Shi, P. Grün, S. Simonyi et al., *Angew. Chem., Int. Ed.* **2019**, *58*, 18957–18963.
[13] D. Konz, A. Klens, K. Schörgendorfer, M. A. Marahiel, *Chem. Biol.* **1997**, *4*, 927–937.
[14] P. Cosmina, F. Rodriguez, F. de Ferra, G. Grandi, M. Perego, G. Venema, D. van Sinderen, *Mol. Microbiol.* **1993**, *8*, 821–831.
[15] W. Lorenzen, T. Ahrendt, K. A. J. Bozhüyük, H. B. Bode, *Nat. Chem. Biol.* **2014**, *10*, 425–427.

3. NMR resonance assignments for a docking domain pair with an attached thiolation domain from the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii*

3.1. Author contribution statements

Status: published

Journal: *Biomol. NMR Assignments* (2021), 10.1007/s12104-021-10010-1

Authors: Jonas Watzel (JoWa), Sepas Sarawi (SS), Elke Duchardt-Ferner (EDF), Helge B. Bode (HBB) and Jens Wöhnert (JW)

(1) Concept and design

JoWa (65 %), JW (35 %)

(2) Conducting tests and experiments

JoWa (85 %): cloning of plasmids, protein expression and purification, NMR experiments; SS (5 %): protein expression and purification; EDF (10 %): NMR measurements

(3) Compilation of data sets and figures

JoWa (100 %): compilation of data sets and figures (NMR spectra, chemical shift index)

(4) Analysis and interpretation of data

JoWa (100 %): analysis and interpretation of NMR data, NMR resonance assignment

(5) Drafting of manuscript

JoWa (70 %), JW (25 %), HBB (5 %)

place and date	signature PhD student
place and date	signature supervisor

3.2. Publication

Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-021-10010-1

ARTICLE

NMR resonance assignments for a docking domain pair with an attached thiolation domain from the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii*

Jonas Watzel[1] · Sepas Sarawi[1,2] · Elke Duchardt-Ferner[2] · Helge B. Bode[1,3,4,5] · Jens Wöhnert[2]

Received: 12 December 2020 / Accepted: 16 February 2021

Abstract
Non-ribosomal peptide synthetases (NRPSs) are large multienzyme machineries. They synthesize numerous important natural products starting from amino acids. For peptide synthesis functionally specialized NRPS modules interact in a defined manner. Individual modules are either located on a single or on multiple different polypeptide chains. The "peptide-antimicrobial-*Xenorhabdus*" (PAX) peptide producing NRPS PaxS from *Xenorhabdus* bacteria consists of the three proteins PaxA, PaxB and PaxC. Different docking domains (DDs) located at the N-termini of PaxB and PaxC and at the C-termini of PaxA and BaxB mediate specific non-covalent interactions between them. The N-terminal docking domains precede condensation domains while the C-terminal docking domains follow thiolation domains. The binding specificity of individual DDs is important for the correct assembly of multi-protein NRPS systems. In many multi-protein NRPS systems the docking domains are sufficient to mediate the necessary interactions between individual protein chains. However, it remains unclear if this is a general feature for all types of structurally different docking domains or if the neighboring domains in some cases support the function of the docking domains. Here, we report the ^{1}H, ^{13}C and ^{15}N NMR resonance assignments for a C-terminal di-domain construct containing a thiolation (T) domain followed by a C-terminal docking domain (^{C}DD) from PaxA and for its binding partner – the N-terminal docking domain (^{N}DD) from PaxB from the Gram-negative entomopathogenic bacterium *Xenorhabdus cabanillasii* JM26 in their free states and for a 1:1 complex formed by the two proteins. These NMR resonance assignments will facilitate further structural and dynamic studies of this protein complex.

Keywords NMR assignments · Peptide-antimicrobial-*Xenorhabdus* (PAX) peptide · Non-ribosomal peptide synthetase (NRPS) · Docking domain · Thiolation domain

✉ Jonas Watzel
watzel@bio.uni-frankfurt.de

✉ Jens Wöhnert
woehnert@bio.uni-frankfurt.de

[1] Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[2] Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[3] Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[4] Senckenberg Gesellschaft Für Naturforschung, 60325 Frankfurt am Main, Germany

[5] Department of Natural Products in Organismic Interactions, Max-Planck-Institute for Terrestrial Microbiology, 35043 Marburg, Germany

Biological context

Non-ribosomal peptide synthases (NRPS) use amino acids as building blocks for the synthesis of complex natural products. Each amino acid is incorporated into the final product by an individual specialized functional module of the NRPS. In turn, each individual module contains a number of enzymatically distinct catalytic domains. A typical NRPS elongation module contains at least an adenylation (A), a thiolation (T) and a condensation (C) domain. The adenylation (A) domain uses ATP to activate a specific amino acid as aminoacyl adenylate and then transfers the activated amino acid to the neighbouring thiolation (T) domain. The T domain in its *holo* form contains a phosphopantetheinyl (Ppant) moiety derived from coenzyme A covalently bound to a conserved serine side chain. The activated amino acid reacts with the terminal thiol group

Published online: 05 March 2021

of the Ppant moiety to form a thioester. The T domain is followed by a condensation domain (C) which catalyses the formation of peptide bonds between the amino acid bound to the T domain of its own module and an amino acid or a peptide chain bound to the T domain of the downstream module. Additional domains with enzymatic activities such as methylation or epimerization can be included in individual modules and may further modify the peptide product. Finally, the nascent peptide is released by a thioesterase (TE) domain localized at the C-terminus of a specialized termination module. TE domains can release either linear, cyclic or branched cyclic peptides (Süssmuth and Mainz 2017).

The individual functional modules necessary for the stepwise incorporation of each amino acid into the final product can be located either on a single long protein chain or on multiple proteins. When all modules are arranged on a single protein chain their linear order directly predicts (Mootz et al. 2002) the order of amino acid building blocks in the final product. In multi-protein NRPS systems specific non-covalent interactions between the individual protein chains determine the functional assembly of the NRPS complex and thereby the order of amino acids in the synthesized peptide. Hahn and Stachelhaus (Hahn and Stachelhaus 2004) were first in identifying short amino acid sequences at the N- and C-termini of the individual protein chains in a multi-protein NRPS that were capable of mediating specific non-covalent interactions between these protein chains. They referred to these sequences as "communication-mediating (COM) domains". Functionally similar structural elements, which were named "docking domains (DD)" (Broadhurst et al. 2003), were identified in polyketide synthase (PKS) complexes which possess a multi-modular architecture similar to NRPS systems. The term "docking domain" is now more commonly used for domains that enable non-covalent specific interactions between individual protein chains in both PKS and NRPS complexes. Multiple structurally diverse families of DD architectures have been described so far (Broadhurst et al. 2003; Buchholz et al. 2009; Dorival et al. 2016; Hacker et al. 2018; Watzel et al. 2020; Whicher et al. 2013). Many DD structural families are dominated by α-helical secondary structure elements (Buchholz et al. 2009; Watzel et al. 2020; Whicher et al. 2013). The binding affinities even for specific interactions between DDs are relatively weak with typical dissociation constants for the DD complexes in the range between 5 μM and 25 μM (Dorival et al. 2016; Hacker et al. 2018; Watzel et al. 2020; Whicher et al. 2013). Nevertheless, for many types of DD interactions it has been experimentally demonstrated that the DDs act independently from the other functional domains in an NRPS or PKS in order to mediate the non-covalent interactions between protein chains needed for the functional assembly of functional megasynthase complexes (Dorival et al. 2016; Hacker et al. 2018). However, it is not yet clear if this is true for all types of docking domains.

The PaxS NRPS found in Gram-negative entomopathogenic bacteria from the genus *Xenorhabdus* is a prototypical example for a multi-protein assembly line. It consists of the three proteins PaxA, PaxB and PaxC. The three proteins interact with each other non-covalently in a unidirectional manner where PaxA is bound by PaxB and PaxB is bound by PaxC. The DD pair mediating the specific interaction between the C-terminus of PaxB and the N-terminus of PaxC has been characterized functionally, structurally and biophysically in previous work (Kegler and Bode 2020; Watzel et al. 2020). A structural basis for the specific interaction between the C-terminus of PaxA and the N-terminus of PaxB has not yet been established. Notably, the putative C-terminal docking domain (^{C}DD) in PaxA was predicted to be very short in comparison to other structurally characterized types of docking domains. Thus, in this case the T domain directly preceding the ^{C}DD of PaxA might play a supporting role in establishing a specific non-covalent interaction with the N-terminal docking domain (^{N}DD) of PaxB. As a prerequisite for investigating a putative role of the T domain in this DD interaction and for establishing a structural basis for the specific interaction between PaxA and PaxB in the PaxS NRPS from *Xenorhabdus cabanillasii* JM26 we present here the NMR resonance assignments for a PaxA T_1-^{C}DD di-domain construct and for the PaxB ^{N}DD in their free states and in a 1:1 non-covalent complex. The T_1-^{C}DD (amino acids 981–1084 of PaxA) comprises 104 amino acids (12 kDa), whereas the ^{N}DD (amino acids 1–30 of PaxB) contains only 30 amino acids (3.6 kDa).

Methods and experiments

Cloning, expression and purification

The DNA coding sequences for PaxA T_1-^{C}DD and PaxB ^{N}DD were generated by PCR amplification using the genomic DNA from *Xenorhabdus cabanillasii* JM26 as the template. All protein sequences referred to in this work are based on the UniProt Archive (UniParc) entries for PaxA (UPI0003E57C57) and PaxB (UPI000C04EDD1) based on a genome assembly for *Xenorhabdus cabanillasii* JM26 produced in our group (NCBI: ASM263290v1; GenBank: NJGH00000000 (Tobias et al. 2017)). For protein concentration measurements a codon for a tyrosine residue was added downstream of the PaxB ^{N}DD coding sequence via the used primer. The inserts were cloned into a modified pET-11a vector (Hacker et al. 2018) containing the sequence for an N-terminal hexahistidine (His_6)-tag followed by a SUMO (SMT3)-tag which also functions as a cleavage site for the ULP1 protease (Malakhov et al. 2004). The PaxA T_1-^{C}DD

was expressed in its *apo* form by using *Escherichia coli* BL21(DE3)Δ*entD* (Owen et al. 2012) cells preventing post-translational modification of the T domain by the covalent addition of a phosphopantetheinyl moiety, whereas the PaxB ^{N}DD was expressed in *E. coli* BL21(DE3) Gold (Agilent Technologies/Stratagene) cells.

Protein expression was induced at OD_{600} ~ 0.7 with 1 mM IPTG at 20 °C for ~ 18 h in media supplemented with ampicillin (100 µg/ml). The expression of uniformly ^{15} N- and ^{15} N,^{13}C-labelled proteins was achieved by using M9 minimal medium supplemented with 1 g/l $^{15}NH_4Cl$ or 1 g/l $^{15}NH_4Cl$ and 2.5 g/l $^{13}C_6$-*D*-glucose (Cambridge Isotope Laboratories). For the stereospecific NMR assignment of $\gamma^{1/2}$ and $\delta^{1/2}$ methyl groups of valine and leucine, the PaxA T_1-^{C}DD as well as the PaxB ^{N}DD were expressed as fractionally ^{13}C-labelled proteins in M9 minimal medium containing a mixture of 0.25 g L^{-1} $^{13}C_6$-*D*-glucose and 2.25 g L^{-1} unlabeled glucose as the sole carbon source (Neri et al. 1989).

Cell lysis of *E. coli* cells was accomplished by sonication in a buffer containing 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 1 mM EDTA, 10 mM $MgCl_2$, 2 mM β-mercaptoethanol, Benzonase (Merck) and cOmplete protease inhibitor (Roche). The cell debris was cleared from the lysate by centrifugation (8000 × *g*, 4 °C, 30 min) and the supernatant was run through a HisTrap HP column (GE Healthcare). With the help of a recombinantly produced His_6-tagged ULP1 protease the His_6-tagged SUMO (SMT3)-tag was cleaved off from the recombinant fusion proteins, leading to native target protein sequences. Both His_6-tagged ULP1 and the SUMO-tag were separated from the protein of interest by a second immobilized metal ion affinity chromatography using a HisTrap HP column, followed by gel filtration chromatography with a HiPrep 16/60 Sephacryl S-100 high resolution column (GE Healthcare). For NMR measurements, the samples of the individual proteins (300 µM protein concentration) and the PaxA T_1-^{C}DD:PaxB ^{N}DD/PaxB ^{N}DD:PaxA T_1-^{C}DD complexes (300 µM protein A concentration: 360 µM protein B concentration) were prepared in a buffer containing 50 mM sodium phosphate, pH 6.5, 100 mM NaCl, 2 mM β-mercaptoethanol and 5% (v/v) D_2O.

NMR spectroscopy

NMR experiments were recorded at 293 K on Bruker AVANCE III HD 600, 700 and 800 MHz spectrometers, each of them equipped with 5 mm cryogenic triple resonance probes. ^{1}H chemical shifts were internally referenced to DSS, whereas the heteronuclear ^{13}C and ^{15} N chemical shifts were indirectly referenced with the appropriate conversion factors (Markley et al. 1998). Spectra were processed using TOPSPIN 3.6.2 (Bruker) and analysed with CARA (Keller 2004). The secondary structure of the unbound and bound PaxA T_1-^{C}DD and PaxB ^{N}DD was derived from TALOS-N (Shen and Bax 2013) based on the chemical shift assignments.

For the backbone assignment of the free PaxA T_1-^{C}DD a uniformly ^{13}C,^{15} N-labelled sample was used and the following triple resonance experiments were recorded: HNCO, HNCA, HNCACB, HBHA(CO)NH, CBCA(CO)NH (Sattler 1999). The backbone and side chain assignments of the unbound ^{13}C,^{15} N-labelled PaxB ^{N}DD were derived from 3D HNCO, HNCACB, HBHA(CO)NH, H(CCO)NH (mixing time 12 ms) and (H)C(CO)NH (mixing time 12 ms) experiments (Sattler 1999). The backbone resonances of ^{13}C,^{15} N-labelled PaxA T_1-^{C}DD/PaxB ^{N}DD in complex with a 1.2-fold excess of unlabelled PaxB ^{N}DD/PaxA T_1-^{C}DD were assigned on the basis of 3D HNCO, HNCA, HNCACB, and CBCA(CO)NH experiments (Sattler 1999). Side chain assignments were obtained from 3D HBHA(CO)NH, H(C)CH-/(H)CCH-TOCSY (mixing times 12 ms) (Sattler 1999) and H(C)CH-COSY (Bax et al. 1990) experiments. Stereospecific assignments of valine $\gamma^{1/2}$ and leucine $\delta^{1/2}$ methyl groups of the bound PaxA T_1-^{C}DD and free and bound PaxB ^{N}DD were determined in ^{1}H,^{13}C-HSQC experiments with a resolution in the ^{13}C dimension of ~ 28 Hz, ~ 26 Hz and ~ 23 Hz, respectively. This allowed the unambiguous discrimination between the signals for the γ^2/ δ^2 CH_3 groups of valine and leucine which appear as singlets in the ^{13}C-dimension and the signals of the γ^1/ δ^1 CH_3 groups which appear as doublets separated by the $^{1}J_{13C,13C}$ coupling constant of ~ 33 Hz (Neri et al. 1989).

Assignment and data deposition

The protein construct PaxA T_1-^{C}DD contains amino acids 981 to 1084 of the PaxA protein from *Xenorhabdus cabanillasii* JM26. It has a molecular weight of 12 kDa and consists of 104 amino acid residues. Most of the backbone amid signals are well dispersed in the ^{1}H,^{15} N-HSQC spectrum of the unbound PaxA T_1-^{C}DD as is typical for a well-folded protein. The ^{1}H,^{15} N-HSQC is expected to contain 101 backbone amide signals since there is no observable amide signal for the N-terminal residue D981 and there are two proline residues (P993, P1054) present in the sequence. However, only 99 backbone amide signals were observed and assigned (99/101, 98.0%; Fig. 1a, top). No backbone amide signals were detectable for residues H982 and S1027 most likely due to fast exchange of the respective amide protons with the solvent or due to conformational exchange. S1027 is the residue that would be posttranslationally modified by addition of the Ppant-arm in the native environment. Furthermore, 99% of all Cα (103/104), 98% of all Cβ (98/100), 95% of all CO (99/104) and Hα (99/104) and 94% of all Hβ (94/100) chemical shifts were assigned. The backbone chemical shifts were used to derive the putative secondary structure of the

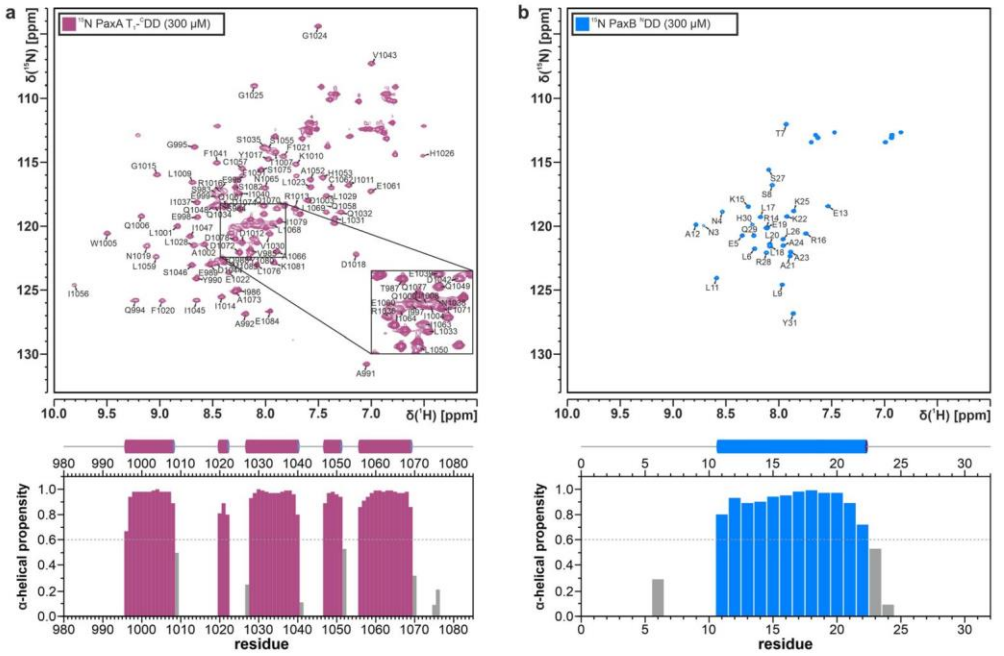


Fig. 1 (**a**, top) ^{1}H,^{15} N-HSQC spectra and assigned backbone amide signals for the free PaxA T$_1$-^{C}DD and (**b**, top) the free PaxB ^{N}DD. The spectra were recorded at 293 K on a 600 MHz Bruker Avance III spectrometer. The assigned NMR signals are labelled with their respective residue names and numbers and the central spectral region of the PaxA T$_1$-^{C}DD with increased peak overlap is enlarged for a better overview (box). (**a**, **b**, bottom) TALOS-N based secondary structure analysis

free PaxA T$_1$-^{C}DD. Based on these chemical shifts TALOS-N (Shen and Bax 2013) identified five amino acid stretches with high α-helical propensity (Fig. 1a, bottom). This secondary structure fits well to that of previously described T domain structures, which typically feature a four-helix bundle (Weber et al. 2000). Compared to the canonical four-helix bundle T domain fold the PaxA T$_1$ domain contains an additional very short (3 residues) α-helix in the linker region between the canonical helices α1 and α2. Such an additional helix is also present at this position in some other carrier protein structures (Lohman et al. 2014). According to TALOS-N the C-terminus including the predicted ^{C}DD of the PaxA T$_1$-^{C}DD construct is unstructured.

The PaxB ^{N}DD construct contains amino acids 1–30 of the PaxB protein followed by a non-native tyrosine at the C-terminus and has a molecular weight of 3.6 kDa. Surprisingly, the backbone amide signals of the free PaxB ^{N}DD are well dispersed in the ^{1}H,^{15} N-HSQC spectrum (Fig. 1b, top) indicating a folded protein. Overall, 27 backbone amide signals were assigned since the sequence contains one proline residue (P10) and no backbone amide signals were observable for M1, N2 and the non-native Y31. Additionally, 93% of all CO (28/30) and all Cα (30/30), Cβ (30/30), Hα (30/30) and Hβ (30/30) chemical shifts were assigned. The overall side chain assignments were completed to 99.2% for aliphatic protons and to 99.0% for aliphatic carbons. No aromatic side chain NMR signals were assigned since the PaxB ^{N}DD contains only a single native aromatic residue (H30) and the non-native Y31. The TALOS-N analysis of the secondary structure suggests the presence of a single continuous α-helix including residues L11 to K22 in the free PaxB ^{N}DD (Fig. 1b, bottom). The ^{13}Cγ and ^{13}Cβ chemical shifts of proline P10 are in agreement with a *trans* conformation for this residue (Schubert et al. 2002). In addition, all valine $\gamma^{1/2}$ and leucine $\delta^{1/2}$ methyl groups of the unbound PaxB ^{N}DD were stereospecifically assigned.

PaxA T$_1$-^{C}DD and PaxB ^{N}DD form a stable 1:1 complex according to analytical gel filtration and NMR titration experiments which is in slow exchange on the NMR time scale. Assignments for both proteins in their bound states were made using samples containing one binding partner in ^{13}C,^{15} N-labelled form and a 1.2-fold excess of

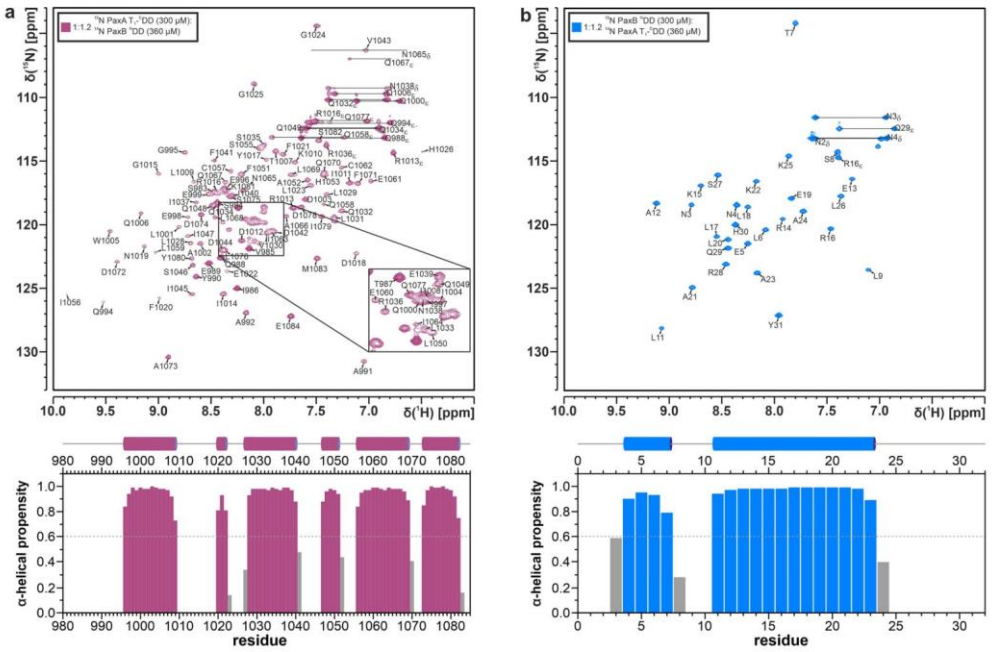


Fig. 2 (**a**, top) $^{1}H,^{15}$ N-HSQC spectra and assigned backbone amide signals for the bound PaxA T_1-^{C}DD and (**b**, top) the bound PaxB ^{N}DD. The spectra were recorded at 293 K on a 600 MHz Bruker Avance III spectrometer. The assigned NMR signals are labelled with their respective residue names and numbers. The central spectral region of the PaxA T_1-^{C}DD with increased peak overlap is enlarged for a better overview (box). Assigned sidechain NH_2 signals are indicated by horizontal bars and labelled with the respective residue names and numbers. The labelled ε imino groups of R1013, R1016 and R1036 in the PaxA T_1-^{C}DD and of R16 in the PaxB ^{N}DD are folded into the spectrum. (**a**, **b**, bottom) TALOS-N based secondary structure analysis

the unlabelled binding partner. The backbone resonances of PaxA T_1-^{C}DD bound to the PaxB ^{N}DD could be assigned with the same degree of completeness as for the free protein. All expected backbone amide signals except those for H982 and S1027 (99/101, 98.0%; Fig. 2a, top) and all Cα (104/104), 99% of all Cβ (99/100), 95% of all CO (99/104), all Hα (104/104) and 99% of all Hβ (99/100) chemical shifts were assigned. The overall side chain assignments were completed to 98.7% for the remaining aliphatic protons and to 98.2% for aliphatic carbons. All valine $\gamma^{1/2}$ and leucine $\delta^{1/2}$ methyl groups were stereospecifically assigned. In addition, 56.3% of all aromatic side chain proton-bound carbon and carbon-bound proton signals were assigned.

The analysis of the backbone chemical shifts of the PaxA T_1-^{C}DD with TALOS-N in its bound state suggests that its secondary structure is very similar to that in its free state but that there is now an additional α-helix present at the very C-terminus suggesting that the predicted docking domain becomes structured upon binding (Fig. 2a, bottom).

The backbone assignment of the bound PaxB ^{N}DD (Fig. 2b, top) includes 27 amide backbone signals, 93% of all CO (28/30) and all Cα (30/30), Cβ (30/30), Hα (30/30) and Hβ (30/30) chemical shifts. The overall side chain assignments were completed to 99.2% for aliphatic protons and to 99.0% for aliphatic carbons. All valine $\gamma^{1/2}$ and leucine $\delta^{1/2}$ methyl groups were stereospecifically assigned. In addition, 50% of all aromatic side chain proton-bound carbon and carbon-bound proton signals were assigned. The TALOS-N derived secondary structure for the bound PaxB ^{N}DD indicates the presence of two α-helices (Fig. 2b, bottom). The proline residue P10 remains in the *trans* conformation in the bound state according to its ^{13}Cγ and ^{13}Cβ chemical shifts (Schubert et al. 2002).

The assigned chemical shifts have been deposited in the BMRB under the accession numbers 50,594, 34,576 and 34,575 for the free PaxA T_1-^{C}DD construct, the free PaxB ^{N}DD and the complex, respectively.

Acknowledgements All NMR measurements were carried out at the Center for Biomolecular Magnetic Resonance (BMRZ) at the Goethe University Frankfurt which is generously supported by the State of Hesse. This work was further supported by the LOEWE program of the state of Hesse and was conducted within the framework of the MegaSyn Research Cluster in the laboratories of H.B.B. and J.W. Work in the Bode lab is also supported by an ERC Advanced Grant (grant agreement number 835108).

Funding Open Access funding enabled and organized by Projekt DEAL. The funding was supported by H2020 European Research Council (Grant No. 835108) and the LOEWE program of the state of Hesse (Grant No. MegaSyn Research Cluster).

References

Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE (1990) Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of ^{13}C-labeled proteins. J Magn Reson 87:620–627

Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, Weissman KJ (2003) The structure of docking domains in modular polyketide synthases. Chem Biol 10:723–731

Buchholz TJ, Geders TW, Bartley FE, Reynolds KA, Smith JL, Sherman DH (2009) Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. ACS Chem Biol 4:41–52

Dorival J, Annaval T, Risser F, Collin S, Roblin P, Jacob C, Gruez A, Chagot B, Weissman KJ (2016) Characterization of intersubunit communication in the virginiamycin trans-acyl transferase polyketide synthase. J Am Chem Soc 138:4155–4167

Hacker C, Cai X, Kegler C, Zhao L, Weickhmann AK, Wurm JP, Bode HB, Wöhnert J (2018) Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. Nat Commun 9:4366

Hahn M, Stachelhaus T (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc Natl Acad Sci U S A 101:15585–15590

Kegler C, Bode HB (2020) Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains. Angew Chem Int Ed Engl 59:13463–13467

Keller R (2004) The computer aided resonance assignment tutorial, 1st edn. Cantina Verlag, Goldau

Lohman JR, Ma M, Cuff ME, Bigelow L, Bearden J, Babnigg G, Joachimiak A, Phillips GN, Shen B (2014) The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. Proteins 82:1210–1218

Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Func Genom 5:75–86

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids – IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. J Biomol NMR 12:1–23

Mootz HD, Schwarzer D, Marahiel MA (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. ChemBioChem 3:490–504

Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry 28:7510–7516

Owen JG, Robins KJ, Parachin NS, Ackerley DF (2012) A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. Environ Microbiol 14:1198–1209

Sattler M (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. J Biomol NMR 24:149–154

Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56:227–241

Süssmuth RD, Mainz A (2017) Nonribosomal peptide synthesis – principles and prospects. Angew Chem Int Ed Engl 56:3770–3821

Tobias NJ, Wolff H, Djahanschiri B, Grundmann F, Kronenwerth M, Shi Y-M, Simonyi S, Grün P, Shapiro-Ilan D, Pidot SJ, Stinear TP, Ebersberger I, Bode HB (2017) Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. Nat Microbiol 2:1676–1685

Watzel J, Hacker C, Duchardt-Ferner E, Bode HB, Wöhnert J (2020) A new docking domain type in the peptide-antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*. ACS Chem Biol 15:982–989

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

Whicher JR, Smaga SS, Hansen DA, Brown WC, Gerwick WH, Sherman DH, Smith JL (2013) Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. Chem Biol 20:1340–1351

4. Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS

4.1. Author contribution statements

Status: submitted

Journal: -

Authors: Jonas Watzel (JoWa), Elke Duchardt-Ferner (EDF), Sepas Sarawi (SS), Helge B. Bode (HBB) and Jens Wöhnert (JW)

(1) Concept and design

JoWa (50 %), JW (25 %), HBB (25 %)

(2) Conducting tests and experiments

JoWa (70 %): cloning of plasmids, heterologous expression, protein expression and purification, NMR measurements, ITC and CD measurements, HPLC-MS analysis, peptide quantification; SS (20 %): cloning of plasmids, protein expression and purification, ITC and CD measurements; EDF (10 %): NMR measurements

(3) Compilation of data sets and figures

JoWa (100 %): compilation of data sets and figures (NMR spectra, protein structures, ITC thermograms, chemical shift index, HPLC-MS data)

(4) Analysis and interpretation of data

JoWa (85 %): structure calculation, analysis and interpretation of NMR and HPLC-MS data; JW (15 %): structure analysis and interpretation

(5) Drafting of manuscript

JoWa (40 %), JW (40 %), HBB (15 %), EDF (5 %)

place and date	signature PhD student
place and date	signature supervisor

4.2. Manuscript

Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS

Jonas Watzel[1], Elke Duchardt-Ferner[2], Sepas Sarawi[1,2], Helge B. Bode[1,3,4], Jens Wöhnert[2]

1 Molecular Biotechnology, Institute of Molecular Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

2 Institute of Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438, Frankfurt am Main, Germany. woehnert@bio.uni-frankfurt.de.

3 Max-Planck-Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043, Marburg, Germany. helge.bode@mpi-marburg.mpg.de

4 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany

Abstract

Non-ribosomal peptide synthetases (NRPS) produce natural products from amino acid building blocks. They often consist of multiple polypeptide chains which assemble in a specific linear order via specialized N- and C-terminal docking domains ($^{N/C}$DDs). Typically, docking domains function independently from other domains in NRPS assembly. Thus, docking domain replacements enable the assembly of “designer” NRPS from proteins that normally do not interact. The multiprotein “peptide-antimicrobial-*Xenorhabdus*” (PAX) peptide-producing PaxS NRPS is assembled from the three proteins PaxA, PaxB and PaxC. Here, we show that the small ^{C}DD of PaxA cooperates with its preceding thiolation (T_1) domain to bind the ^{N}DD of PaxB with very high affinity, establishing a structural and thermodynamical basis for this unprecedented docking interaction, and test its functional importance *in vivo* in a truncated PaxS assembly line. Similar docking interactions are apparently present in other NRPS systems.

Introduction

Non-ribosomal peptide synthetases (NRPSs) are able to manufacture a very diverse range of natural products starting from amino acids. Many of these products such as daptomycin, vancomycin, bleomycin are highly relevant for clinical applications[1]. The biosynthesis of these peptide products is accomplished by the orchestrated interplay of a number of functionally distinct catalytic domains, which are grouped into modules. Each module is responsible for incorporating one specific amino acid building block into the nascent product. A typical elongation module in an NRPS contains an adenylation (A) domain, a thiolation (T) domain which is also sometimes referred to as a peptidyl-carrier protein (PCP) domain and a condensation (C) domain[2]. The A domain selects a specific amino acid and uses ATP to chemically activate this amino acid by forming a reactive aminoacyl adenylate. The adjacent T domain in its catalytically active *holo* form contains a phosphopantetheinyl (Ppant) arm derived from coenzyme A which is covalently bound to a conserved serine residue. The reactive thiol group of the Ppant arm bound to the T domain reacts with the aminoacyl intermediate activated by the preceding A domain and forms a thioester with the amino acid while releasing AMP. The subsequent condensation domain (C) then catalyzes the formation of a peptide bond between the amino acid bound to its preceding T domain and an amino acid bound to the T domain of the downstream module. Additional domains with specific tailoring functions such as methyltransferases or epimerization domains can be included in a functional module and may further modify the peptide product. The final product is cleaved off from the Ppant arm of the T domain in the last elongation module by a thioesterase (TE) domain localized at the C-terminus of a termination module, which releases either linear, cyclic or branched cyclic peptides. Multiple modules can be located on a single protein. An extreme example in this regard is the NRPS synthesizing kollosin A which consists of 15 modules on a single polypeptide chain[3]. However, often the individual modules of an NRPS are located on multiple proteins giving rise to multiprotein NRPS or multienzyme complexes.

In single protein NRPS the linear order of modules from the N- to the C-terminus directly determines the linear order of amino acid building blocks in the final product. This relationship is known as the collinearity rule[4]. In multiprotein NRPS a specific linear arrangement of the individual proteins and thereby of the different modules is established by non-covalent interactions between specialized structural elements or

domains of the individual proteins. These domains are often localized at the N- and C-termini of the respective proteins of a multiprotein NRPS. They were initially referred to as “communication-mediating domains”[5] in NRPS or as “inter-polypeptide linkers”[6] or “docking domains (DD)”[7] in the architecturally related polyketide synthases (PKSs) and are now generally referred to as docking domains (DDs) in megasynthase systems. Since they are able to direct the correct linear assembly of multiprotein megasynthases[5,8,9] and are portable between different megasynthase systems often without diminishing the activity of the neighboring catalytic domains[5,10,11], they are ideal tools to enable combinatorial biosynthesis by connecting protein chains of non-related megasynthases in a predictable, functional and specific way in order to synthesize structurally novel and diverse “designer” products not found in nature[10].

The structural analysis of discrete docking domains and docking domain complexes revealed a large structural diversity[11–15]. However, a majority of docking domain structures is dominated by α-helical secondary structure elements. Typically docking domains are rather small protein domains (~30–65 amino acids (AA)). They bind their cognate docking domains with medium affinities with dissociation constants in the range between ~5–25 μM[8,11,12,14]. The successful transplantation of docking domains between different functionally unrelated megasynthase systems[9,10,16] already suggests that they act independently from the other functional domains in directing specific protein-protein interactions between megasynthase components. Affinity measurements for cognate docking domain pairs of megasynthases that included flanking functional domains such as the acyl carrier proteins (ACP)[14] in PKS, T domains or C domains[7] in NRPS showed that the affinity between the docking domains did not increase in the presence of the flanking domains in comparison to the isolated docking domains. Together, these results reinforced the notion that docking domains are functionally independent in mediating the specificity of the assembly in multiprotein megasynthases and not supported by the other domains in the protein chain.

Here, our focus is on the multiprotein NRPS PaxS which consists of the three NRPS proteins PaxA, PaxB and PaxC (Figure 1a). PaxS occurs in bacterial species belonging to the genus *Xenorhabdus* and produces “peptide-antimicrobial-*Xenorhabdus* (PAX)” peptides[17]. PAX peptides such as **1** and **2** (Figure 1b) are cyclic peptides with predominantly basic residues and N-terminally attached acyl chains that protect the

bacterial producer cell against insect-derived basic antimicrobial peptides[18]. PaxA contains a specialized starter C domain transferring an acyl chain to the amino group of a serine which is activated by the A domain and stored on the downstream T domain (T_1). PaxB consists of three canonical elongation modules with a preference for arginine/lysine or lysine. PaxC contains three additional canonical elongation modules and a C-terminal thioesterase (TE) domain which also catalyzes the cyclization between a lysine side chain amino group and the C-terminal carboxylate group of the peptide. Based on the structure of the PAX peptides and the collinearity rule PaxA, PaxB and PaxC form a unidirectional assembly line where the C-terminus of PaxA supposedly interacts with the N-terminus of PaxB and the C-terminus of PaxB with the N-terminus of PaxC.

We recently established the structural basis for the specific interaction between the PaxB C-terminus and the PaxC N-terminus in the PaxS NRPS of *Xenorhabdus bovienii* SS-2004[19]. PaxB contains a specific C-terminal docking domain (^{C}DD) following the last T domain in its sequence while PaxC features a specific N-terminal docking domain (^{N}DD) preceding the first C domain in this protein chain. The structure of the PaxB ^{C}DD/PaxC ^{N}DD complex defined a novel type of docking domain interactions and identified key amino acids contributing to the specificity of complex formation.[12] In line with observations for other cognate pairs of docking domains the PaxB ^{C}DD/PaxC ^{N}DD pair can be grafted onto other unrelated multiprotein NRPS systems and induces functional and specific interchain interactions in these systems[10].

Here, we investigate the interaction between PaxA and PaxB and find that this interaction is mediated by an unprecedented docking interface not only involving the ^{C}DD of PaxA and the ^{N}DD of PaxB as expected in analogy to previously described docking interactions but requiring the cooperation of the T domain and the ^{C}DD of PaxA to bind the PaxB ^{N}DD with a surprisingly high affinity. We also show that this type of interaction in the PaxS NRPS is functionally important for efficient product formation *in vivo* and also occurs in other NRPS systems in bacteria not related to *Xenorhabdus* species.

Results and Discussion

While the structural basis for the specific interaction between PaxA and PaxB has not yet been established, a bioinformatics analysis showed that the C-terminal T_1 domain of PaxA is followed by a short stretch of only ~19 amino acids predicted to be

unstructured while the N-terminal C domain of PaxB is preceded by ~36 amino acids predicted to contain a region with elevated α-helical propensity. While these two stretches of amino acids could represent putative C- and N-terminal docking domains, they show no clear sequence homologies to known docking domain pairs in general and to the PaxB ^{C}DD/PaxC ^{N}DD pair in particular. Furthermore, the putative PaxA ^{C}DD is significantly shorter than other known types of docking domains.

In order to test if the putative PaxA ^{C}DD and PaxB ^{N}DD and their interactions are important for efficient product formation in the *Xenorhabdus bovienii* SS-2004 PaxS NRPS *in vivo*, we created a truncated version of the PaxS NRPS consisting of only two proteins. This artificial NRPS contained full-length native PaxA and a modified PaxB where the thioesterase domain of the PaxC termination module (PaxB-TE$_{PaxC}$, Figure 1a) replaces the native ^{C}DD of PaxB. It was expressed successfully in the heterologous host *E. coli* DH10B::*mtaA*[20] and shown to efficiently produce two shortened linear PAX tetra-peptides **3** (SR*k*K; *D*-AAs in italics and lower case throughout this work (Figure 1b)) and **4** (SK*k*K, Figure 1b) with a defined N-terminally attached (3*R*)-3-hydroxytetradecanoyl fatty acid moiety as expected. Besides these main products two additional side products (*; (3*R*,7*Z*)-3-hydroxytetradec-7-enoyl-SR*k*K (**5**)) and ((3*R*)-3-hydroxytetradecanoyl-SK (**6**))) were detected (Figure 1c). The amino acid and acyl chain composition of **3**–**6** was confirmed by feeding experiments followed by HR-HPLC/MS analysis (Figure S1/S2) and the comparison of the MS^2 fragmentation pattern with those of synthetic standards (**3**, **4**) (Figure S3). In this truncated system product formation should only rely on a productive non-covalent interaction between PaxA and PaxB. Deletion of the putative PaxB ^{N}DD or of the PaxB ^{N}DD and the PaxA ^{C}DD in this system almost completely abolished product formation *in vivo* while a deletion of the PaxA ^{C}DD alone significantly reduced product formation to 25% for **3** and to 50% for **4** (Figure 1d). This shows that the putative docking domains play an important role for mediating productive non-covalent interactions between PaxA and PaxB *in vivo*. However, the observation that the deletion of the short PaxA ^{C}DD alone did not completely abolish product formation indicates that other structural elements in PaxA might also contribute to the docking interaction *in vivo*.

To test the interaction of the putative PaxA ^{C}DD and PaxB ^{N}DD from the *Xenorhabdus bovienii* SS-2004 PaxS *in vitro*, we overexpressed a PaxA$_{1071–1089}$ (PaxA ^{C}DD) and a PaxB$_{1–36}$ (PaxB ^{N}DD) peptide as SUMO fusion proteins and obtained the native peptide

sequences with an additional N-terminal tyrosine residue in PaxA ^{C}DD and an additional C-terminal tyrosine residue in PaxB ^{N}DD for concentration measurements upon cleavage with Ulp protease[21] and gel filtration (Figure S4/S5a). Additionally, the integrity of all purified proteins was confirmed by HPLC/MS (Table S1). Surprisingly, the interaction between the PaxA ^{C}DD and the PaxB ^{N}DD was too weak to be reliably quantified by isothermal titration calorimetry (ITC; Figure S5b) even though we have successfully used ITC to quantify interactions with K_Ds in the mid-micromolar range for other docking domain pairs[8,12]. However, when we titrated PaxA$_{986–1089}$, a di-domain protein construct containing both the entire T_1 domain in its *apo* state and the putative ^{C}DD of PaxA, with the PaxB ^{N}DD peptide in ITC experiments we measured a K_D of 201±20 nM (Figure S5c) for this interaction with a 1:1 stoichiometry (n: 0.74±0.08). In contrast, in a titration of the isolated T_1 domain of PaxA (PaxA$_{986–1076}$) with the PaxB ^{N}DD peptide the interaction was again too weak to be reliably quantified (Figure S5d). Thus, in the PaxS NRPS from *Xenorhabdus bovienii* SS-2004 the non-covalent interaction between the PaxA and PaxB proteins is dependent on the presence of both the T_1 and the ^{C}DD of PaxA which apparently have to cooperate to bind the PaxB ^{N}DD with an affinity that is remarkably high in comparison to the affinities for other characterized docking domain pairs[8,12–15].

Unfortunately, despite its very high affinity the complex between the PaxA T_1-^{C}DD di-domain construct and the PaxB ^{N}DD peptide from the PaxS NRPS from *Xenorhabdus bovienii* SS-2004 did not crystallize and was not suitable for determining a high-resolution or NMR structure due to unfavorable relaxation properties very likely caused by transient aggregation. The closely related species *Xenorhabdus cabanillasii* JM26 also contains the PaxS NRPS with PaxA and PaxB having overall sequence identities of 75 % and 71 % to its counterparts in *Xenorhabdus bovienii* SS-2004 and sequence identities of 83% and 86% for its ^{C}DD and ^{N}DD, respectively (Figure S4). More importantly, the *X. cabanillasii* PaxA T_1-^{C}DD di-domain (PaxA$_{981–1084}$) construct bound a slightly length-optimized *X. cabanillasii* PaxB ^{N}DD construct (PaxB$_{1–30}$) with a similarly high affinity (248±18 nM) and a 1:1 stoichiometry (n: 0.88±0.01) according to ITC (Figure 2/Figure S6) as observed for *X. bovienii* SS-2004. The affinity of this interaction was not modulated by the presence of the Ppant moiety in the PaxA T_1 domain (Figure S7). No quantifiable interactions were observed by ITC between the isolated PaxA ^{C}DD (PaxA$_{1066–1084}$) and the PaxB ^{N}DD as well as between the isolated PaxA T_1 domain (PaxA$_{981–1071}$) and the PaxB ^{N}DD, respectively (Figure 2/Figure S6).

Thus, the interaction between PaxA and PaxB from the PaxS NRPS of *X. cabanillasii* has biophysical properties very similar to the one from the *X. bovienii* system and turned out to be amenable to structural characterization by NMR spectroscopy.

The secondary structure of the PaxA T_1-^{C}DD di-domain construct in its *apo* state in the absence of the PaxB ^{N}DD was derived from our previously published NMR resonance assignments[22] using the TALOS-N[23] derived chemical shift index (CSI) and {^{1}H},^{15}N-hetNOE experiments (Figure S8). According to the CSI and the hetNOE data the T domain is compactly folded and features the characteristic four α-helices (α1–α4) typical of the canonical carrier protein fold[24]. The long loop between helix α1 and helix α2 is interrupted by a short, single-turn α-helix as previously observed in the structure of other carrier proteins[25,26]. Importantly, residues Q1070–E1084 corresponding to the PaxA ^{C}DD are unstructured and flexible both in the framework of the di-domain construct based on the NMR data as well as in the absence of the T_1 domain according to CD spectroscopy (Figure S9). The N-terminal residues D981–G995 of the PaxA T_1-^{C}DD di-domain construct which correspond to the linker between the A_1 and the T_1 domain of PaxA are also unstructured and flexible (Figure S8).

The free PaxB ^{N}DD showed a surprisingly high α-helical content in CD spectroscopy experiments (Figure S9) and according to the CSI (Figure S10) in contrast to the intrinsically disordered protein (IDP) character typically observed for other docking domains in the absence of their binding partner[14,15,27]. This is borne out in our NOE-based solution NMR structure of free PaxB (Figure 3a, Table S2, Protein Data Bank (PDB) ID: 7B2F) where residues L11–K22 form a continuous three-turn α-helix (α2). Residues L6–S8 adopt an additional α-helical turn (α1) which is separated from the longer α-helix by a sharp kink (~83°). The kink is induced by P10 which adopts a *trans* peptide bond conformation based on its ^{13}Cγ and ^{13}Cβ chemical shifts[28]. According to the {^{1}H},^{15}N-hetNOE data for the free PaxB ^{N}DD the helix from L11 to K22 is stable while the helical turn from L6 to S8 forms only transiently (Figure S10). The stability of the long helix can be attributed to the presence of two pairs of residues with oppositely charged side chains – K15/E19 and E13/R16 – spaced in a manner to enable favorable intra-helical electrostatic interactions. Interestingly, the structure of the free PaxB ^{N}DD reveals a very uneven charge distribution for the long α-helix. One side of the helix is highly positively charged while the other surface is mostly hydrophobic with interspersed negative charges (Figure 3b).

Starting from our previously reported NMR assignments including stereospecific methyl group assignments and using complementary isotope labelled samples in conjunction with standard and isotope-filtered NOESY experiments[29] we determined the solution structure of an 1:1 complex between the PaxA T_1-^{C}DD di-domain construct and the PaxB ^{N}DD based on a large number of intra- and unambiguously identified intermolecular NOEs (Table S2) and characterized its dynamics by {^{1}H},^{15}N-hetNOE and T_1/T_2-relaxation experiments (Figure S11). The structural bundle representing the NMR solution structure of this complex (PDB ID: 7B2B) is shown in Figure 4. The structure shows that the PaxB ^{N}DD interacts extensively with the C-terminal half of helix α4 from the T_1 domain of PaxA as well as with the ^{C}DD of PaxA. Importantly, the ^{C}DD is no longer unstructured as in its unbound form but forms a stable three-turn α-helix (α5) involving residues A1073–S1082. Compared to its structure in the free state the PaxB ^{N}DD in the complex now features an extended stable N-terminal α-helix spanning residues N3–S8 (α1) while the sharp kink centered around P10 in the *trans* conformation and the C-terminal α-helix spanning residues L11 to A24 (α2) are preserved. The pairs of interacting helices from PaxA (α4 and α5) and PaxB (α1 and α2) each form a V-shape and the two V shapes interlock in an antiparallel orientation (Figure 4). Helices α1 to α3 and the N-terminal half of α4 from the T_1 domain of PaxA are not involved in the interaction which is also in agreement with the chemical shift perturbation data for complex formation (Figure S11a). Overall the fold of the T_1 domain in the bound state closely corresponds to the classical right-handed up-and-down four-helical bundle fold observed in many other T domain structures in particular and carrier protein structures in general[24–26]. However, it is notable that helix α4 is longer than in many other previously reported T domain structures[30] (Figure S12) and that the C-terminal helix extension is the part of the T domain that interacts with the PaxB ^{N}DD. Importantly, the docking interface between PaxA and PaxB is located on the site of the T domain that is opposite to the site for the attachment of the Ppant arm to residue S1027 (Figure 4a). Thus, the docking interaction does not interfere with substrate loading and transfer mediated by the Ppant arm of the T_1 domain. This also explains why the presence or absence of the Ppant moiety does not influence the affinity of the docking interaction in contrast to observations for the interaction of T domains with other NRPS domains which are often promoted by the presence of the Ppant arm[26,31–33]. The high affinity for the docking interaction between PaxA T_1-^{C}DD

and PaxB ^{N}DD is the result of both electrostatic complementarity (Figure 4b) and favorable matching of extensive hydrophobic surfaces (Figure 4c, Figure S13). Salt bridges are formed for instance between the negatively charged residues E1060 and E1061 in helix α4 of the PaxA T_1 domain and the positively charged K15 and R14 residues in helix α2 of the PaxB ^{N}DD, respectively, and between E1084 of the PaxA ^{C}DD and R16 in helix α2 of the PaxB ^{N}DD. The side chain of Q1067 in helix α4 of the PaxA T_1 domain is involved in an additional polar interaction with the side chain of K22 in helix α2 of the PaxB ^{N}DD which also forms a salt bridge with E996 in the T_1 domain. Residues from both helices of the PaxB ^{N}DD (L6, T7, L9, L11, L17, L18, L20, A21; Figure S13a) show hydrophobic interactions with amino acids in the C-terminal part of helix α4 in the PaxA T_1 domain (I1064, L1068, L1069, F1071, A1073, L1076, I1079; Figure S13b). Residues L1076, I1079 and Y1080 from the PaxA ^{C}DD (helix α5) add further hydrophobic interactions with both helices of the PaxB ^{N}DD (L6, L9, L17, L20, A24) to the docking interface (Figure 4c).

We tested the importance of key structural features of the PaxA/PaxB interaction for their affinity by measuring the K_Ds for variants by ITC. Our structure of the complex suggested that the extensive intermolecular contacts contributed by residues in helix α2 in the PaxB ^{N}DD might be dominant for mediating the PaxA/PaxB interaction and helix α1 might be less important. However, ITC titrations with a PaxB ^{N}DD variant where residues 1–8 were deleted showed that the absence of helix α1 lead to a complete loss of the interaction (Figure S14a) suggesting that the full-length PaxB ^{N}DD is required for a high-affinity docking interaction.

We also asked if the pronounced kink between helix α1 and α2 in the PaxB ^{N}DD which is induced by the presence of P10 in the sequence and already preformed in the structure of the free PaxB ^{N}DD is important for the binding affinity. Therefore, we replaced P10 by a leucine which promotes and stabilizes α-helical conformations. An NMR analysis of the secondary structure of the unbound P10L mutant of the PaxB ^{N}DD using the chemical shift index and backbone NMR signal assignments (^{H}N, ^{N}H, C^{α}, C^{β}, and C′) for the mutant showed that the P10L variant now adopts a conformation with a continuous straight α-helix spanning residues L6–K25 (Figure S15a). In agreement with the NMR data the CD spectrum of the P10L mutant showed an increased ellipticity at ~222 nm compared to the CD spectrum of the wild type (WT) (Figure 5a). According to ITC data the affinity of the P10L mutant for the PaxA T_1-^{C}DD di-domain construct is

diminished ~16-fold (K_D: 3.9±0.1 µM; Figure 5b, Figure S15b) compared to the WT. Thus, the preformed kink induced by P10 is important for high affinity binding. Interestingly, chemical shift comparisons between the WT complex and the P10L complex under saturating conditions show that the chemical shifts in both complexes are almost identical both for the bound PaxA T_1-^{C}DD di-domain (Figure S16) and the bound PaxB ^{N}DD (Figure S15c) despite the presence of the mutation. This suggests that the conformation of both binding partners in the WT and the mutant complex are very similar. In particular, the continuous α-helix in the free mutant ^{N}DD P10L apparently breaks and the kinked conformation is adopted upon binding. In agreement with such a scenario, particularly large chemical shift changes are observed for residues T7–E13 upon binding of the PaxB ^{N}DD P10L mutant to the PaxA T_1-^{C}DD di-domain (Figure S15a). Thus, P10 preorganizes the conformation of the PaxB ^{N}DD in a binding-competent conformation. The conformational preorganization in the PaxB ^{N}DD is apparently also important *in vivo* for efficient product formation. When we introduced the equivalent P10L mutation in our truncated *Xenorhabdus bovienii* SS-2004 PaxS NRPS (PaxA, PaxB-TE_{PaxC}) and monitored product formation upon overexpression in *E. coli* the amount of product is reduced to ~35% for **3** and ~20 % for **4** in comparison to the WT (Figure 5c, Figure S17).

The importance of the intermolecular salt bridges for the binding affinity in the PaxA T_1-^{C}DD/PaxB ^{N}DD complex was investigated in ITC titration experiments with charge reversing point mutations. A PaxB ^{N}DD R14E/K15E double mutant intended to break the salt bridges to E1060 and E1061 in helix α4 of the PaxA T_1 domain lead to a complete loss of binding (Figure 6a/b, Figure S14b). The PaxB ^{N}DD K22E mutant, which abolishes the salt bridge with E996 and the hydrogen bond with Q1067 of the PaxA T_1 domain, diminished the affinity ~15-fold (K_D: 3.8±0.2 µM; Figure 6a/c, Figure S14c) while the R16E mutant, destroying the salt bridge with E1084 in the PaxA ^{C}DD, decreased the binding affinity ~11-fold (K_D: 2.8±0.3 µM; Figure 6a/d, Figure S14d). Importantly, all PaxB ^{N}DD mutants discussed here showed an α-helical content similar to the WT according to their CD spectra confirming that the mutations did not interfere with the structure of the unbound PaxB ^{N}DD (Figure S18). Overall, the pronounced effects of the salt bridge abrogating mutations on binding demonstrated that these interactions are important for establishing the high affinity interaction between PaxA and PaxB *in vitro*. The introduction of the PaxB ^{N}DD R14E/K15E double mutant which completely abolished binding *in vitro* also strongly inhibited product formation for the

truncated *Xenorhabdus bovienii* SS-2004 PaxS NRPS (PaxA, PaxB-TE_{PaxC}) *in vivo*. Only ~5 %/~10 % of the amount of **3**/**4** formed by the WT was detected for this mutant (Figure 6e). The effect of introducing the K22E mutant in this cluster on product formation *in vivo* was rather limited. The mutant NRPS still produced ~90 %/~85 % of **3**/**4** compared to the WT (Figure 6e). However, it should be noted that even if the K22E mutant reduced the affinity between PaxA and PaxB ~15-fold *in vitro* the K_D for the interaction is still in the low μM range (K_D: 3.8 μM). Thus, even for the mutant the affinity is higher than for many other previously described functional docking domain interactions [8,12,14]. This might explain why still significant amounts of product are formed particularly under conditions of heterologous protein overexpression.

The extended docking interface observed here for the PaxS NRPS between PaxA and PaxB where the T domain and a minimal ^{C}DD must cooperate to bind with high affinity to the PaxB ^{N}DD is not limited to PaxS-related NRPS systems from *Xenorhabdus* species. A simple BlastP search of NCBI´s non-redundant protein sequence database[34] using the sequences of the PaxA T_1-^{C}DD and the PaxB ^{N}DD from *X. cabanillasii* and *X. bovienii* as input revealed a number of hits to putative multiprotein NRPS systems in different classes of the proteobacteria as well as in cyanobacteria (Figure S19). In these hits, the predicted C-terminal T domains feature long α4 helices (~4-turns) in one protein chain of the NRPS gene cluster which precede short ^{C}DDs while a ~35 AA potential ^{N}DD with α-helical secondary structure propensity in another protein chain precedes a C domain. Furthermore, the α4 helix regions of the T domains and the putative ^{C}DD regions are in general rich in negatively charged residues whereas the predicted ^{N}DDs contain a number of positively charged residues (Figure S19a). Unfortunately, none of the putative NRPS systems found in this search were functionally characterized or have known peptide products (Figure S19b). For a hit from the β-proteobacterium *Chromobacterium violaceum* Bergonzini we expressed and purified the PaxA T_1-^{C}DD di-domain analog and the PaxB ^{C}DD analog and tested their interaction by NMR. In an NMR titration experiment the addition of the unlabeled PaxB ^{N}DD analog to the ^{15}N-labeled PaxA T_1-^{C}DD di-domain analog caused widespread chemical shift perturbations indicative of binding (Figure S20a). In contrast, NMR titration experiments of an isolated ^{15}N-labelled PaxA T_1 domain analog (Figure S20b) or an isolated PaxA ^{C}DD analog (Figure S20c) with the PaxB ^{N}DD analog showed no obvious evidence for an interaction. Thus, also in this system the T domain

and the ^{C}DD must cooperate to bind to their cognate ^{N}DD suggesting that this type of composite docking interface is widespread in NRPS and NRPS-PKS systems from different classes of proteobacteria and also occurs in additional bacterial phyla such as the cyanobacteria (Figure S19).

Conclusions

Previous structural and functional studies of docking domains in megasynthase systems have not only uncovered their rich structural diversity but also established them as interesting building blocks for megasynthase engineering since they are able to mediate non-covalent protein-protein interactions in multiprotein NRPS or PKS systems independently from other domains and can be transplanted between systems without loss of function[9–11]. While the docking domain interactions are specific for cognate pairs of N-terminal and C-terminal docking domains the affinity of these interactions are often limited to K_Ds in the mid- and low-micromolar range. Here, we have identified a novel type of docking interaction in a multiprotein NRPS system where the C-terminal T domain and a short C-terminal docking domain of one NRPS protein cooperate to create a binding interface for the N-terminal docking domain of another NRPS protein to non-covalently assemble a productive NRPS complex. In comparison to previously described docking interactions the affinity of this interaction is significantly higher with a K_D in the nanomolar range. The T domain takes part in the docking interface based on the presence of an elongated helix α4 where the C-terminal part of this extended helix contributes to the docking interaction. Since this helix is part of the canonical carrier protein fold it is structurally pre-organized already prior to docking. The ^{N}DD also shows a high degree of structural preorganization in its free form and only the rather short ^{C}DD behaves as an IDP in contrast to previous observations in 4 α-helix bundle (4HB) docking domains[14,35]. This high degree of structural pre-organization is a likely reason for the high affinity of the interaction described here. Importantly, the involvement of helix α4 in the docking interaction does not interfere with the other functions of the T domain in substrate shuttling and the structural elements often involved in interactions of T domains with other NRPS domains such as helices α2 and α3[30,36] are still accessible upon binding the ^{N}DD.

Docking interactions involving T domains such as the one described here are apparently not limited to the PaxS NRPS but also occur in NRPS and NRPS-PKS systems from organisms not related to *Xenorhabdus*. Due to the high affinities

observed for these systems they might be attractive tools for mediating non-native intermolecular interactions in “designer” NRPS systems.

Acknowledgements

All NMR measurements were carried out at the Center for Biomolecular Magnetic Resonance (BMRZ) at the Goethe University Frankfurt which is generously supported by the State of Hesse. This work was further supported by the LOEWE program of the state of Hesse and was conducted within the framework of the MegaSyn Research Cluster in the laboratories of H.B.B. and J.W. Work in the Bode lab is also supported by an ERC Advanced Grant (grant agreement number 835108).

References

[1] R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770.
[2] J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, *Current opinion in structural biology* **2018**, *49*, 104.
[3] H. B. Bode, A. O. Brachmann, K. B. Jadhav, L. Seyfarth, C. Dauth, S. W. Fuchs, M. Kaiser, N. R. Waterfield, H. Sack, S. H. Heinemann et al., *Angewandte Chemie (International ed. in English)* **2015**, *54*, 10352.
[4] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490.
[5] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15585.
[6] R. Gokhale, *Curr. Opin. Chem. Biol.* **2000**, *4*, 22.
[7] R. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, *Chem. Biol.* **2003**, *10*, 723.
[8] C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert, *Nat. Commun.* **2018**, *9*, 4366.
[9] C. Kegler, H. B. Bode, *Angew. Chem. Int. Ed.* **2020**, *59*, 13463.
[10] X. Cai, L. Zhao, H. B. Bode, *Org. Lett.* **2019**, *21*, 2116.
[11] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman, J. L. Smith, *Chem. Biol.* **2013**, *20*, 1340.
[12] J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *ACS Chem. Biol.* **2020**, *15*, 982.
[13] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, D. H. Sherman, *ACS Chem. Biol.* **2009**, *4*, 41.
[14] J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot, K. J. Weissman, *J. Am. Chem. Soc.* **2016**, *138*, 4155.
[15] S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea et al., *Nat. Chem.* **2019**, *11*, 913.
[16] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 275.
[17] S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas, H. B. Bode, *Org. Biomol. Chem.* **2011**, *9*, 3130.
[18] T. D. Vo, C. Spahn, M. Heilemann, H. B. Bode, *ACS Chem. Biol.* **2021**.
[19] J. M. Chaston, G. Suen, S. L. Tucker, A. W. Andersen, A. Bhasin, E. Bode, H. B. Bode, A. O. Brachmann, C. E. Cowles, K. N. Cowles et al., *PLoS One* **2011**, *6*, e27909.
[20] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290.
[21] M. P. Malakhov, M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks, T. R. Butt, *J. Struct. Funct. Genomics* **2004**, *5*, 75.
[22] J. Watzel, S. Sarawi, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *Biomol. NMR Assignments* **2021**, 1.
[23] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56*, 227.
[24] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407.
[25] A. C. Goodrich, B. J. Harden, D. P. Frueh, *J. Am. Chem. Soc.* **2015**, *137*, 12100.
[26] C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2012**, *51*, 3252.
[27] F. Risser, S. Collin, R. Dos Santos-Morais, A. Gruez, B. Chagot, K. J. Weissman, *J. Struct. Biol.* **2020**, *212*, 107581.
[28] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, *J. Biomol. NMR* **2002**, *24*, 149.

[29] W. Lee, M. J. Revington, C. Arrowsmith, L. E. Kay, *FEBS Letters* **1994**, *350*, 87.
[30] J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips, B. Shen, *Proteins: Struct. Funct. Bioinf.* **2014**, *82*, 1210.
[31] Y. Liu, T. Zheng, S. D. Bruner, *Chem. Biol.* **2011**, *18*, 1482.
[32] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235.
[33] K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süssmuth, M. J. Cryle, *Angew. Chem. Int. Ed.* **2014**, *53*, 8518.
[34] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389.
[35] J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2016**, *11*, 2466.
[36] J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, *Chem. Biol.* **2012**, *19*, 188.

Figures

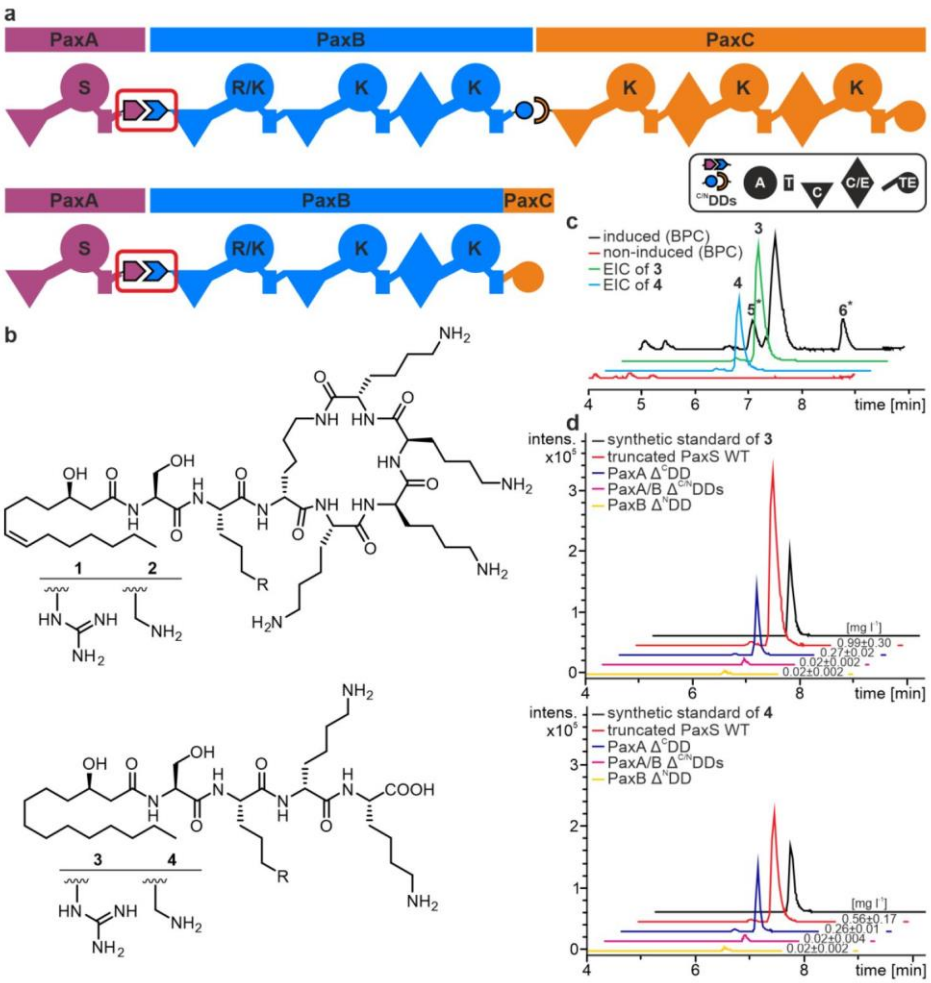


Figure 1. The "Peptide-antimicrobial-*Xenorhabdus*" (PAX) peptide-producing NRPS PaxS. a) Schematic representation of the full length and truncated PaxS producing the peptides **1**–**4**. For domain assignment the following symbols are used: adenylation (A) domain, large circles; thiolation (T) domain, rectangle; condensation (C) domain, triangle; dual condensation/epimerization (C/E) domain, diamond; thioesterase (TE) domain, small circle. For each adenylation domain its amino acid preference is given by the single letter code. The PaxA/B docking domain interface framed in red was investigated in this work. b) Main products **1**/**2** of the full length and **3**/**4** of the truncated PaxS. c/d) HPLC/MS analysis of the tetra-peptide (**3**/**4**) production of the (modified) truncated PaxS, which are shown as base peak chromatograms (BPC) and extracted ion chromatograms (EICs) of **3**–**6**.

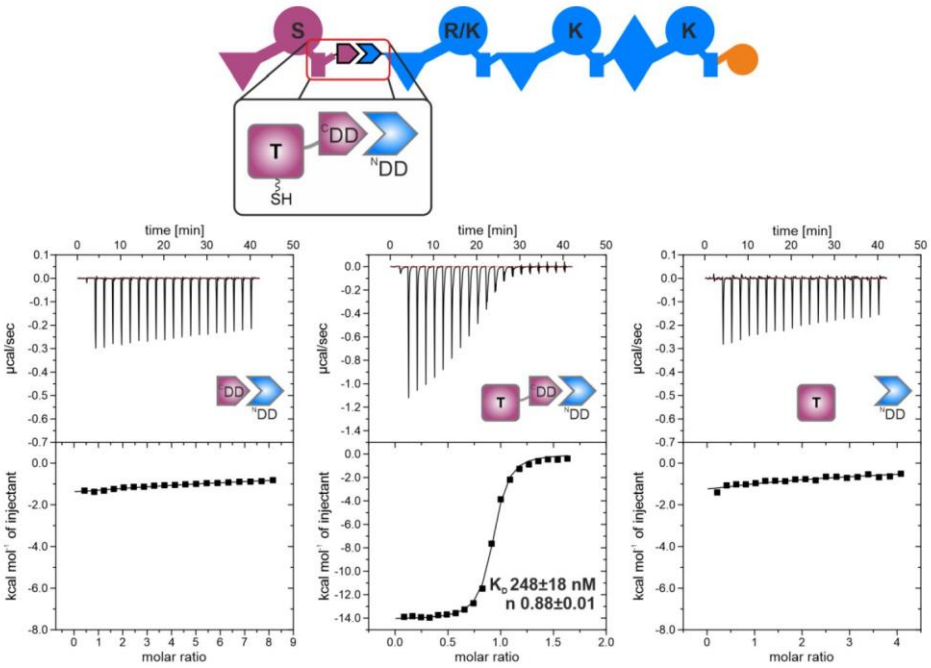


Figure 2. ITC analysis of the PaxA/B docking domain interface. Exemplary ITC thermograms and the derived binding curves for titrations between *X. cabanillasii* JM26 PaxA ^{C}DD/PaxB ^{N}DD, PaxA T_1-^{C}DD/PaxB ^{N}DD and PaxA T_1/PaxB ^{N}DD, respectively.

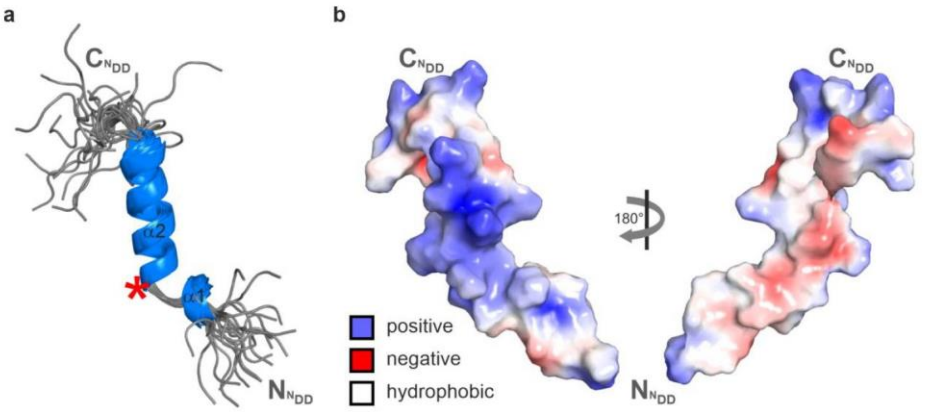


Figure 3. NMR solution structure of the free PaxB ^{N}DD a) Solution structure bundle of the 20 energy minimized conformers with the lowest CYANA target functions of the unbound PaxB ^{N}DD. The location of the proline residue (P10) is marked with a red asterisk. b) Electrostatic surface potentials mapped on the molecule surface of the free PaxB ^{N}DD. Negatively charged surface areas are colored in red, positively charged areas are colored in blue and white areas correspond to hydrophobic surfaces.

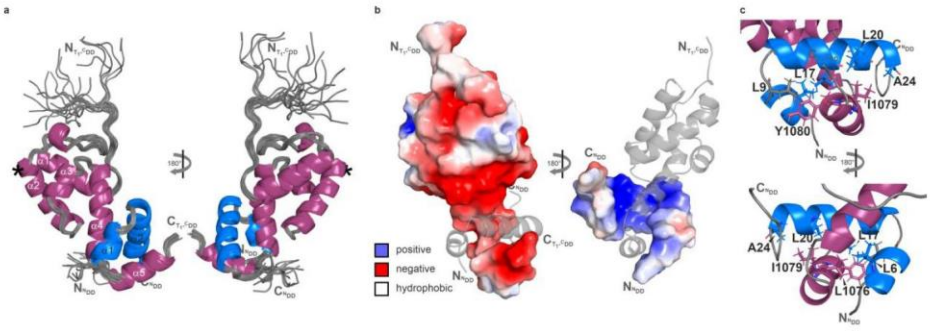


Figure 4. NMR solution structure of the PaxA T_1-^{C}DD/PaxB ^{N}DD complex. a) Solution structure bundle of the 20 energy minimized conformers with the lowest CYANA target functions of the PaxA T_1-^{C}DD/PaxB ^{N}DD complex. The location of the serine residue, where the Ppant moiety is post-translationally added, is marked with an asterisk. b) Electrostatic surface potentials are mapped on the molecule surface of the complex. Negatively charged surface areas are colored in red, positively charged areas are colored in blue and white areas correspond to hydrophobic surfaces. c) Molecular architecture of the hydrophobic core of the ^{C}DD/^{N}DD interface formed by the residues shown in stick representation.

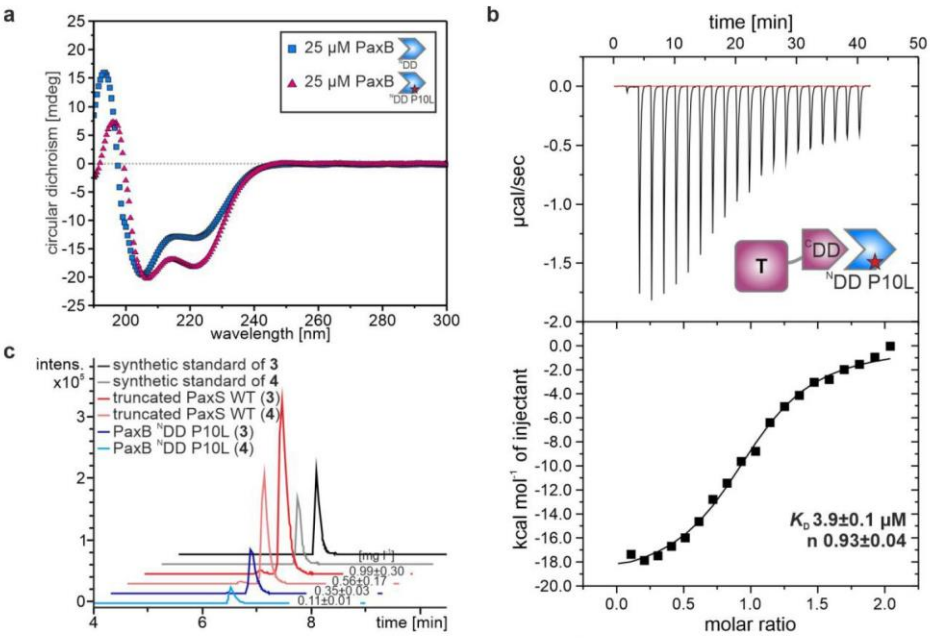


Figure 5. Structural and functional characterization of the PaxB ^{N}DD P10L variant. a) Circular dichroism spectra of the wild type PaxB ^{N}DD (blue) and the PaxB ^{N}DD P10L variant (purple). b) Exemplary ITC thermogram and the derived binding curve for a titration between the PaxA T_1-^{C}DD and the PaxB ^{N}DD P10L variant (n=3). c) HPLC/MS data for the characterization of product formation by the modified truncated PaxS carrying the P10L mutation in the PaxB ^{N}DD, producing peptides **3** (pale EICs) and **4** (bright EICs).

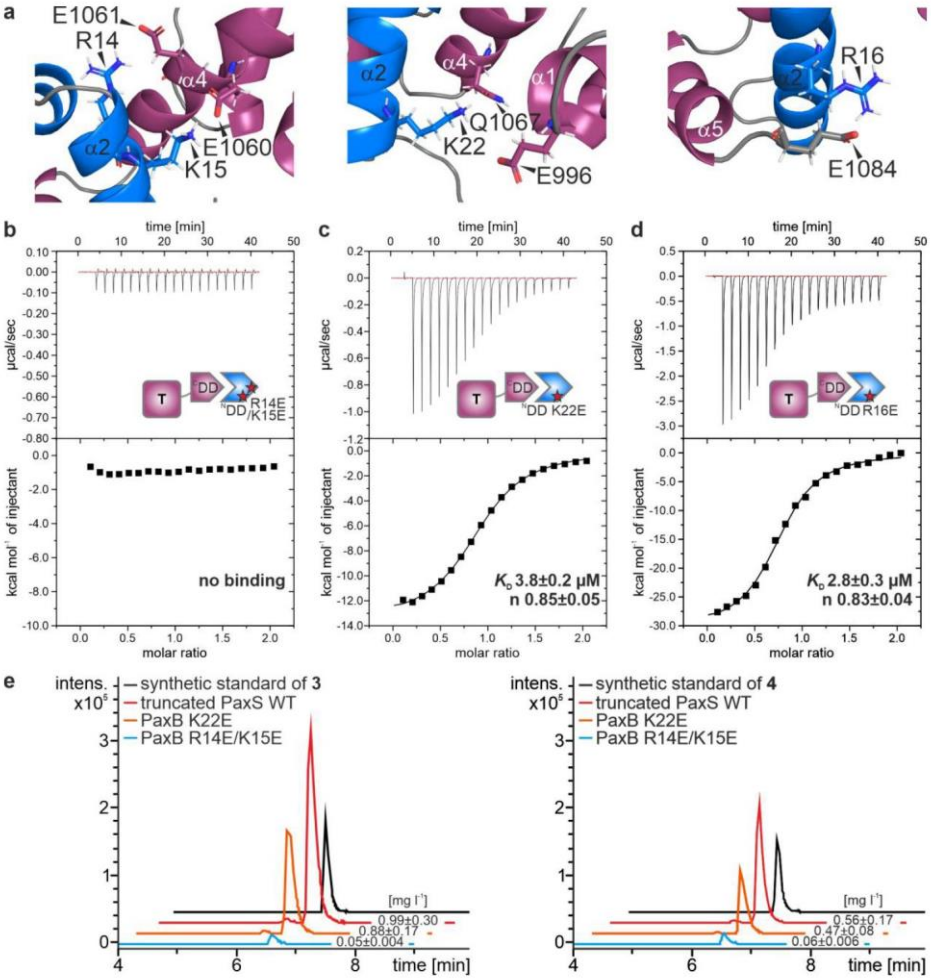
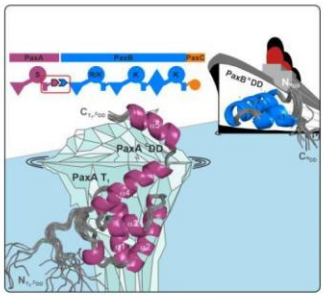


Figure 6. The importance of salt bridges for the PaxA T_1-^{C}DD/PaxB ^{N}DD interaction. *In vitro* and *in vivo* analysis of PaxB ^{N}DD variants. a) The molecular architecture of salt bridge-forming residues that stabilize the cooperative docking domain interface. The PaxA T1-^{C}DD (purple) and the PaxB ^{N}DD (blue) are shown in cartoon representation. Side chains of relevant residues are labeled and shown in stick representation. *In vitro* characterization of PaxB ^{N}DD b) R14E/K15E, c) K22E and d) R16E variants by ITC. Exemplary ITC thermograms and the derived binding curves for titration between PaxA T_1-^{C}DD and PaxB ^{N}DD variants. Resulting dissociation constants and binding stoichiometries are given (n=3). e) HPLC/MS data for the characterization of product formation by the modified truncated PaxS carrying the K22E or R14E/K15E mutation(s) in the PaxB ^{N}DD, producing peptides **3** and **4** (EICs).

TOC Figure



It´s so much more than a docking domain! The interaction between two non-ribosomal peptide synthetases (NRPSs) mediates the biosynthesis of PAX peptides and involves an extended docking interface, where a part of the thiolation (T) domain and a minimal C-terminal docking domain (^{C}DD) cooperate to bind to the N-terminal docking domain (^{N}DD) with a nanomolar affinity, which is the highest docking domain affinity yet observed in such megasynthases.

4.3. Supporting information

Table of contents

1 Material and methods

General molecular biology

Molecular biology techniques such as plasmid DNA preparation, transformation, restriction digestion and DNA gel electrophoresis, were adapted from standard protocols[1]. Isolation of genomic DNA was carried out according to the manufacturer's instructions (QIAGEN). S7 Fusion Polymerase (Mobidiag) was used for PCR amplifications following the guidelines of the producer. PCR primers (Eurofins Scientific) used in this study are listed in Supporting Table S4. All plasmids (Supporting Table S5) generated in this study were constructed via Hot Fusion cloning[2] and used for the transformation of *E. coli* DH10B cells (Thermo Fisher).

General cultivation conditions of bacteria

All wild type strains are listed in Supporting Table S3. *Xenorhabdus* strains were cultured in LB medium (pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl) on an orbital shaker or on LB agar (1 % (w/v) agar) plates at 30 °C. *E. coli* cells were cultured in LB medium on an orbital shaker or on LB agar plates at 37 °C supplemented with antibiotics in appropriate concentrations (ampicillin 100 µg/ml, chloramphenicol 34 µg/ml, 50 µg/ml spectinomycin).

Construction of protein expression plasmids and cloning of biosynthetic gene clusters

The coding sequences of all proteins of interest were cloned into a modified pET-11a vector[3] containing an N-terminal His_6-SUMO tag, which allows cleavage of the tagged SUMO protein by Ulp1 treatment. All protein sequences referred to in this work are based on the UniProt Archive (UniParc) entries for PaxA (UPI0003E57C57) and PaxB (UPI000C04EDD1) based on a genome assembly for *Xenorhabdus cabanillasii* JM26 produced in our group (NCBI: ASM263290v1; GenBank: NJGH00000000[4] and *Xenorhabdus bovienii* SS-2004[5].The modified pET-11a plasmid was linearized via PCR using the primer pair pET-11a_FW and pET-11a_smt3_RV. The resulting backbone fragment and the respective PCR amplified inserts (protein coding sequences) were assembled in a Hot Fusion cloning[2] step. Cloning of constructs with specific amino exchanges was accomplished by oligonucleotide-directed mutagenesis using primers that contain the desired mutation followed by a Hot Fusion assembly.

The insertion of the truncated peptide-antimicrobial-*Xenorhabdus* (PAX) producing NRPS (truncated PaxS - PaxA/PaxB-PaxC_TE) in plasmid pCK_0402 was achieved in a two-step cloning procedure. First, the plasmid pCK_0402 was linearized by digestion with the restriction enzyme PstI (New England BioLabs) followed by a Hot Fusion cloning step in which the genes *paxA-paxBΔ^{C}DD* were inserted. Second, this intermediate plasmid was reopened by restriction digest with the enzyme PstI followed by a Hot Fusion cloning step in which the gene *paxC_TE* was incorporated leading to a truncated version of PaxS (pJW81). To analyze the effect of single amino acid exchanges and the $^{C/N}$DD deletion in the framework of this truncated PaxS

on the production level the whole plasmid pJW81 was amplified in two fragments by PCR to delete the DD coding sequences or to insert the amino acid exchanges via the primer sequences. Both PCR fragments were assembled in a NEBuilder HiFi DNA assembly (New England BioLabs) approach following the manufacturer`s instructions.

Protein expression and purification

For structure elucidation and the thermodynamic characterization of protein-protein interactions, proteins from *Xenorhabdus cabanillasii* and *Xenorhabdus bovienii* were heterologously expressed in *E. coli* BL21-Gold(DE3), *E. coli* BL21-Gold(DE3)Δ*entD* or *E. coli* BAP1 under the control of a T7 promoter. All proteins containing the PaxA T_1 domain were expressed either in *E. coli* BL21-Gold(DE3)Δ*entD* or in *E. coli* BAP1 cells to generate uniform T domain species in their *apo* or *holo* state. For the uniform isotope labeling of the proteins of interest, *E. coli* cells harboring the respective protein expression plasmids were grown in ^{15}N and ^{15}N,^{13}C M9 minimal media containing 1 g L^{-1} $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) or 1 g L^{-1} $^{15}NH_4Cl$ and 2.5 g L^{-1} $^{13}C_6$-*D*-glucose (Cambridge Isotope Laboratories). The non-random stereospecific ^{13}C-labeling of valine and leucine methyl groups was accomplished in M9 minimal medium containing a mixture of 0.25 g L^{-1} $^{13}C_6$-*D*-glucose and 2.25 g L^{-1} unlabeled glucose as the sole carbon source[6]. For ITC measurements, proteins were expressed in LB medium. Protein expression was induced at an OD_{600} of 0.6–0.8 with 1 mM IPTG overnight at 20 °C. After expression, cells were lysed by sonication and purified as previous described in a three-step purification procedure including immobilized metal-affinity chromatographies (IMAC) combined with a final gel filtration chromatography step[7]. The buffer solution containing the purified proteins was composed of 50 mM sodium phosphate (pH 6.5), 100 mM NaCl and 2 mM β-mercaptoethanol.

NMR spectroscopy

NMR spectra were recorded at 20 °C on Bruker AVANCE III 600, 700, 800, and 950 MHz spectrometers equipped with cryogenic 5 mm triple resonance probes. The proton chemical shifts were internally referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid and the heteronuclear ^{13}C and ^{15}N chemical shifts were indirectly referenced with the appropriate conversion factors[8]. PaxB ^{N}DD (unbound) samples were at concentrations of 300 µM and PaxA T_1-^{C}DD/PaxB ^{N}DD complex samples (and *vice versa*) were at concentrations of 300 µM:360 µM in 50 mM sodium phosphate buffer (pH 6.5), 100 mM NaCl, 2 mM β-mercaptoethanol and 5 % D_2O. Sequential assignments of backbone amide signals and assignments of side chain atoms were obtained from standard triple resonance experiments as described previously[9]. The non-random ^{13}C-labeling of valine and leucine methyl groups enabled the stereospecific assignment for all $\gamma^{1/2}CH_3$ groups of Val and all $\delta^{1/2}CH_3$ groups of

Leu. All spectra were processed with Bruker TopSpin 3.6.2 and analyzed using the programs CARA[10] (www.nmr.ch) and CcpNmr Analysis[11].

{^{1}H},^{15}N-heteronuclear NOE[12], T_1 and T_2 data were recorded at 20 °C on a Bruker AVANCE II 600 MHz spectrometer equipped with a cryogenic 5 mm triple resonance probe for the *X. cabanillasii* ^{15}N-labeled PaxA T_1-^{C}DD and PaxB ^{N}DD in their unbound (only hetNOE data) and bound state using standard Bruker pulse sequences. {^{1}H},^{15}N-heteronuclear NOE experiments were run twice in an interleaved fashion with and without proton saturation during the recovery delay. In T_1 and T_2 experiments 8 delays were set with maximum delay lengths of 1.6 s for T_1 and 0.204 s for T_2, respectively. Signal intensities were extracted by using Bruker TopSpin 3.6.2. The ^{15}N T_1 and T_2 relaxation times were determined by a nonlinear fit of a two-parameter, mono-exponential decay function (Origin fitting function: Exp2Mod1) of the measured signal intensities using OriginPro 2020 SR1. The {^{1}H},^{15}N-heteronuclear NOE values were determined as the ratio of the signal intensities obtained from the NOE and reference experiments ($I = I_X/I_0$), respectively.

For NMR titration experiments ^{1}H,^{15}N-HSQC or ^{1}H,^{15}N best-TROSY-HSQC spectra were recorded after the stepwise addition of unlabeled protein to a ^{15}N-labeled protein sample with a concentration of 80 µM. To evaluate NMR titration experiments, the chemical shifts were determined using the peak picking function of CcpNmr Analysis[11]. The chemical shift differences were calculated using the following function[13]:

$$(1)\ \Delta\delta = \sqrt{\Delta\delta_{HN}{}^2 + \left(\frac{\Delta\delta_N}{6.5}\right)^2}$$

Structure calculation

3D ^{15}N-edited nuclear Overhauser spectroscopy (NOESY)-HSQC, 3D ^{13}C-edited NOESY-HSQC (aliphatic carbons) and 3D ^{13}C-edited NOESY- HSQC (aromatic carbons) experiments in H_2O with mixing times of 120 ms were recorded in H_2O and used to obtain distance restraints. 3D ^{13}C-edited NOESY-HSQC (aliphatic carbons) spectra for the PaxA T_1-^{C}DD/PaxB ^{N}DD complex were also recorded in D_2O with mixing times of 200 ms to identify additional NOEs that would be located near the water signal in the H_2O sample. Additional isotope-filtered 3D ^{13}C NOESY-HSQC experiments were performed for the PaxA T_1-^{C}DD/PaxB ^{N}DD complex to extract intermolecular NOE contacts in the molecular complex. Torsion angle restraints (ϕ and ψ) were generated from chemical shifts using TALOS-N[14]. All NOE peaks were picked manually from the 3D NOESY-HSQC spectra mentioned above and incorporated in the structure calculation procedure with the automated NOESY assignment tools of CYANA 3.98[15,16] according to the respective chemical shift lists and TALOS-N restraints. The tolerances (^{1}H, ^{1}H, ^{13}C/^{15}N) applied during NOE assignments were set to 0.03, 0.03, 0.45 ppm.

In each of the seven cycles 100 structures were calculated and evaluated by CYANA. In the final cycle ambiguous assignments were kept and an ensemble of the 20 conformers with the lowest target function was generated. Notably, the complex assembled in its final conformation without the manual predefinition of any intermolecular NOE assignments. The structural bundle was further refined by manually curating the upper distance limit file from wrongly assigned NOE distance restraints. Restrained energy refinement with OPALp[17] and the AMBER94 force field[18] of the final 20 structures with the lowest target function was carried out. This set of CYANA generated, energy minimized structures with the lowest target functions were validated with the Protein Structure Validation Software (Supporting Table S1) suite1.5.[19] Electrostatic surface potential calculations were conducted and visualized with the APBS electrostatics plugin[20] implemented in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.3 Schrödinger, LLC). The range for electrostatic potential shading was set from -3 k*T*/*e* to +3 kT/e (k = Boltzmann's constant, T = absolute temperature and e = electron charge. All structure figures were prepared with PyMOL.

Isothermal titration calorimetry

ITC measurements were performed at 20 °C in 50 mM sodium phosphate buffer (pH 6.5) and 100 mM NaCl using a MicroCal iTC200 (Malvern Instruments) calorimeter. In all experiments, 25 µM or 50 µM of the respective interaction partner were provided in the reference cell. The potential binding partners were prepared in suitable concentrations and added stepwise. ITC experiments started with an initial delay time of 120 s. The first injection of 0.2 µl was followed by 19 serial injections of 2 µl, separated by an interval of 120°s or 180 s. For each experiment, the reference power was set to 11 μcal^{-1}, the stirring speed to 750 rpm and the high feedback mode was selected. Two (if no binding was observed) or three independent titrations were performed for the analysis of each biomolecular interaction. The thermograms were processed using Origin7.0 (OriginLab) assuming a one site binding model. If in ITC measurements a saturation of the binding partner was observed as evidenced by a clear plateau, this plateau was used for baseline correction.

Circular dichroism

The circular dichroism spectra of the respective proteins were recorded from 25 µM samples in 1.0 mm path length quartz cuvettes using a Jasco J-810 CD spectrometer equipped with a Jasco PTC-423S temperature control system. The baseline correction and averaging of three measurements were performed automatically. The buffer, 50 mM sodium phosphate buffer (pH 6.5) with 100 mM NaCl, was identical to that used to record NMR spectra. Data were collected at 0.5 nm/1 nm increments from 300 to 190 nm at 293 K (bandwidth: 1 nm, scanning speed: 50 nm min^{-1}).

HR-HPLC-ESI-MS analysis of purified proteins

Purified proteins were analyzed via high resolution HR-HPLC-ESI-UV-MS using a Dionex UltiMate 3000 HPLC system (Thermo Fisher) coupled to an impact II electrospray ionization mass spectrometer (Bruker) and a DAD-3000 RS UV-detector (Thermo Fisher). The protein samples were separated on a C3 column (Zorbax 300SB-C3, 300 Å, 3.0 x 150 mm, 3.5 µm particle size; Agilent). ACN and ddH_2O w/ 0.1% (v/v) formic acid were used as mobile phases at a flow rate of 0.6 mL min^{-1}. HPLC was performed with 15% ACN equilibration (0–1.5 min), followed by a gradient from 15–65 % or 15–35 % ACN (1.5–27 min) and a further elution step with 95% ACN (27–30 min). For internal mass calibration an ESI-L Mix (Agilent) was injected. The HPLC/MS analysis was set to positive mode with a mass range of *m/z* 50–2000 and an UV-visible absorption wavelength range from 190–800 nm. For data analysis of UV-MS-chromatograms Compass DataAnalysis 4.3 (Bruker) was used. The theoretical average masses of proteins were calculated using Compass IsotopePattern 3.0 (Bruker).

Heterologous production of PAX tetra-peptides and HR-HPLC-ESI-MS analysis

Constructed PaxS plasmids were transformed together with plasmid pCDF_xb2154 (ABC transporter from the PaxS gene cluster) into *E. coli* DH10B::*mtaA*. Cells were grown overnight in LB medium containing the necessary antibiotics (34 µg/ml chloramphenicol; 50 µg/ml spectinomycin). 100 µl of overnight culture (≙ 1 % of total culture volume) were used to inoculate 10 ml XPPM[21] production cultures supplemented with chloramphenicol and spectinomycin as selection markers and additional 0.0002 mg/ml *L*-arabinose to induce expression. The protein expression was confirmed in all cases by SDS-PAGE analysis. After incubation for 72 h at 22 °C the cells were harvested. Therefore 1 mL of cell culture was centrifuged at 13.300 x g for 5 min at 20 °C and the cell pellet and the supernatant fraction were separated. The pelleted cells were sonicated for 15 min, freeze-dried and re-suspended in 200 µL of a methanol/ddH_2O mixture (v/v 1:1) acidified with 1 % formic acid. The cell suspension was treated again for 15 min with sonication followed by a centrifugation step at 13.300 x g for 15 min at 20 °C. The soluble fraction was analyzed by HR-HPLC-ESI-MS. All measurements were performed by using a Dionex UltiMate 3000 HPLC system (Thermo Fisher) with an C18 column (ACQUITY UPLC BEH C18, 130 Å, 2.1 x 50 mm, 1.7 µm particle size; Waters) at a flow rate of 0.4 ml min^{-1} using acetonitrile (ACN) and water containing 0.1% formic acid (v/v) in a gradient ranging from 5–95% of ACN over 16 min (40 °C) coupled to an impact II electron spray ionization mass spectrometer (Bruker). The base peak chromatograms (BPC) were recorded in positive ion mode with the range from 100–1200 *m/z* and UV-visible absorption spectra with the wavelength range from 200–600 nm. The software Compass DataAnalysis 4.3 (Bruker) was used to evaluate the measurements.

MS-based structure elucidation of PAX tetra-peptides

In order to define the number of carbon and nitrogen atoms of the peptide-antimicrobial-*Xenorhabdus* (PAX) peptide-producing truncated synthetase consisting of the proteins PaxA/B-TE_{PaxC}, *E. coli* DH10B::*mtaA* cells were grown in ISOGRO-^{13}C- and ISOGRO-^{15}N-growth medium. By comparison of the mass-to-charge ratio (*m/z*) of the unlabeled to the fully ^{15}N or ^{13}C single charged molecule ions in the MS^1 spectra, the observed shifts to higher *m/z* ratios indicated the exact number of the respective carbon or nitrogen atoms. To further confirm the assumed amino acid composition of the detected tetra-peptides a reversed labelling approach was used. To this end *E. coli* DH10B::*mtaA* cells were cultivated in ISOGRO-^{13}C-, or ^{15}N,^{13}C-growth medium supplemented with 3 mM of unlabeled arginine, lysine or serine. By comparison of the *m/z* ratios of the single charged molecule ions in the MS^1 spectra, shifts to lower *m/z* ratios due to the incorporation of the unlabeled amino acid in the otherwise fully ^{13}C- or ^{15}N,^{13}C-labeled peptide, clearly identified the amino acid composition of peptides **3**–**6**. The resulting sum formulae of the respective peptides are listed in Supporting Table S6.

Peptide quantification

The absolute production titers of peptide **3** and **4** were calculated with calibration curves based on pure synthetic standards synthesized by WuXi AppTec. Therefore, the pure compounds were prepared at different concentrations (50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, 0.195 and 0.0195 µg/mL) and measured by HPLC/MS using the HPLC/MS measurement methods described above. The peak area for each compound at different concentrations was calculated using Compass DataAnalysis 4.3 and used for the calculation of a standard curve. Triplicates of all *in vivo* experiments were measured.

2 Supporting Information Tables

Supporting Table S1. Proteins used in this work.

protein ID	strain	protein	molecular weight [Da]	sum formula	charge state	theoretical mass [*m/z*]	detected mass [*m/z*]
JW28	*Xenorhabdus bovienii* SS-2004	**PaxA Y-^{C}DD**	2335	$C_{105}H_{160}N_{24}O_{34}S$	2^+	1168.0712	1168.0672
JW29	*Xenorhabdus bovienii* SS-2004	**PaxB ^{N}DD-Y**	4328	$C_{189}H_{325}N_{55}O_{58}S$	6^+	722.2397	722.2394
JW51	*Xenorhabdus bovienii* SS-2004	**PaxA T_1-^{C}DD_*apo***	12078	$C_{533}H_{835}{}^{15}N_{141}O_{166}S_2$	11^+	1098.9537	1098.9511
JW84	*Xenorhabdus bovienii* SS-2004	**PaxA T_1_*apo***	10486	$C_{471}H_{738}N_{126}O_{143}S$	8^+	1311.6852	1311.6917
JW140	*X. cabanillasii*	**PaxA T_1-^{C}DD_*apo***	11968	$C_{534}H_{829}N_{137}O_{169}S_3$	9^+	1330.7832	1330.7882
JW140	*X. cabanillasii*	**PaxA T_1-^{C}DD_*holo***	12308	$C_{545}H_{850}N_{139}O_{175}PS_4$	9^+	1368.5705	1368.5758
JW144	*X. cabanillasii*	**PaxA Y-^{C}DD**	2379	$C_{106}H_{160}N_{24}O_{36}S$	3^+	793.7132	793.7173
JW145	*X. cabanillasii*	**PaxA T_1_*apo***	10472	$C_{471}H_{732}N_{122}O_{144}S_2$	8^+	1309.9237	1309.9237
JW146	*X. cabanillasii*	**PaxB ^{N}DD-Y**	3609	$C_{157}H_{267}N_{49}O_{46}S$	6^+	602.5045	602.5023
JW146 _Δα1	*X. cabanillasii*	**PaxB ^{N}DD_Δα1-Y**	chemical synthesized by GenScript Biotech				
JW146 _P10L	*X. cabanillasii*	**PaxB ^{N}DD _P10L-Y**	3625	$C_{158}H_{271}N_{49}O_{46}S$	6^+	605.1764	605.1811
JW146 _R14E/ K15E	*X. cabanillasii*	**PaxB ^{N}DD _R14E/K15E-Y**	3583	$C_{155}H_{257}N_{45}O_{50}S$	6^+	598.1527	598.1566
JW146 _R16E	*X. cabanillasii*	**PaxB ^{N}DD _R16E-Y**	3582	$C_{156}H_{262}N_{46}O_{48}S$	6^+	597.9948	597.9997
JW146 _K22E	*X. cabanillasii*	**PaxB ^{N}DD _K22E-Y**	3610	$C_{156}H_{262}N_{48}O_{48}S$	6^+	602.6624	602.6829
pJW157	*C. violaceum*	**putative PaxA T_1-^{C}DD_*apo* analog**	12061	$C_{539}H_{841}{}^{15}N_{143}O_{167}S_2$	10^+	1221.1536	1221.1161
pJW158	*C. violaceum*	**putative PaxB ^{N}DD analog**	3606	$C_{155}H_{270}N_{48}O_{48}S$	6^+	601.8390	601.8286
pJW167	*C. violaceum*	**putative PaxA T_1_*apo* analog**	10314	$C_{460}H_{715}{}^{15}N_{123}O_{135}S_2$	9^+	1146.9549	1146.9316
pJW168	*C. violaceum*	**putative PaxA ^{C}DD analog**	2779	$C_{122}H_{189}{}^{15}N_{29}O_{43}$	3^+	927.1472	927.0743

Supporting Table S2. Structural statistics of the NMR solution structures. Structural statistics of the NMR solution structures of the unbound PaxB ^{N}DD and the PaxA T_1-^{C}DD/PaxB ^{N}DD complex.

	unbound PaxB ^{N}DD	PaxA T_1-^{C}DD:PaxB ^{N}DD complex
Conformational restricting constraints		
Total NOE distance constraints	209	2990
intraresidue |i = j|	83	716
sequential |i - j| = 1	74	734
medium-range 1 <|i - j| < 5	51	780
long-range |i - j| ≥ 5	1	760
NOE constraints per residue	7.5	22.5
Dihedral angle constraints (Talos-N)	38	220
total No. of constraints per residue	8.8	24.1
No. of long-range constraints per residue	0.0	5.7
Residual restraint violations[a]		
Average no. of distance violations per structure		
0.1-0.2 Å	0.15	13.6
0.2-0.5 Å	0	0
>0.5 Å	0	0
Average no. of dihedral angle violations per structure		
1-10°	0.75	15.3
>10°	0	0
Model quality (ordered residues)[a]		
RMSD backbone atoms (Å)	0.3	0.4
RMSD heavy atoms (Å)	1.2	0.9
RMSD bond lengths (Å)	0.010	0.011
RMSD bond angles (°)	1.9	2.1
Richardson Lab's Molprobity Ramachandran statistics[a]		
Most favored regions	100 %	96.3
Allowed regions	0.0 %	3.2
Disallowed regions	0.0 %	0.5
Global quality scores (raw score/Z-score)[a]		
Verify3D	-0.14/-9.63	0.12/-5.46
Prosall	0.02/-2.61	0.70/0.21
Procheck (φ-ψ)	0.77/3.34	0.02/0.39
Procheck (all)	0.29/1.71	-0.43/-2.54
MolProbity clashscore	0.18/1.49	4.32/0.78
Model contents		
Ordered residue ranges	11-23	4-29, 991-1083
Total no. of residues	31	135
BMRB accession number	34576	34575
PDB ID code	7B2F	7B2B

[a] calculated using PSVS 1.5[19] for using ordered residues (hetNOE values > 0.5). Average distance violations were calculated using the sum over r^{-6}

Supporting Table S3. Strains used in this work.

Strain	Genotype / NRPS	Reference
E. coli BL21-Gold(DE3)	*E. coli* B F⁻ *ompT hsdS*(r_B^-, m_B^-) *dcm*⁺ Tet^r *gal* λ(DE3) *endA* Hte / -	Agilent
E. coli BAP1	F⁻ *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3) Δ*prpRBCD*::$T7_{prom}$-*sfp* $T7_{prom}$-*prpE* / -	[22]
E. coli BL21(DE3)Δ*entD*	F⁻ *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3) Δ*prpRBCD*::$T7_{prom}$-*sfp* $T7_{prom}$-*prpE* Δ*entD* / -	[23]
E. coli DH10B	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *nup*G/ -	Invitrogen
E. coli DH10B::*mtaA*	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *nup*G *entD*::*mtaA*	[24]
Xenorhabdus bovienii SS-2004	wild type / *paxS*	[5]
Xenorhabdus cabanillasii JM26 (DSM 17905)	wild type / *paxS*	DSMZ
Chromobacterium violaceum Bergonzini (ATCC 53434)	wild type / putative *paxS* analog	ATCC

Supporting Table S4. Oligonucleotides used in this work.

plasmid	oligo-nucleotide	sequence (5´→3´; overlapping ends, *nulceotide exchanged to introduce amino acid exchange)	template
pCDF_xb2154	DUET_Gib_21_FW	CAGCTTAATTAACCTAGGCTG	pCDFDuet-1
	DUET_Gib_21_RV	CATGGAATTCCCTCCTGTGTG	pCDFDuet-1
	Xb2154_Gib_FW	TTTCACACAGGAGGGAATTCCATGACTCTAATAGCTTATCTATACCGCCA	*X. bovienii* SS-2004
	Xb2154_Gib_RV	AGCAGCCTAGGTTAATTAAGCTGTTTTCATGACATTCGAGAGGATG	*X. bovienii* SS-2004
pJW28	pET11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0035-FW	CATCGTGAACAGATTGGTGGTTATGCTCAGTTACTGCAATTTGAC	*X. bovienii* SS-2004
	jw0013-RV	TTTGTTAGCAGCCGGATCCTTAACCCATTGTTTTGTAGAGATCTT	*X. bovienii* SS-2004
pJW29	pET11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0014-FW	CATCGTGAACAGATTGGTGGTATGAATGATAATGAATTAATATCTTTACC	*X. bovienii* SS-2004
	jw0036-RV	TTTGTTAGCAGCCGGATCCTTAATAAATTTCTGTTTTTTGTGTCTGC	*X. bovienii* SS-2004
pJW51	pET11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0060-FW	CATCGTGAACAGATTGGTGGTGATCACTCGGCCGTCC	*X. bovienii* SS-2004
	jw0013-RV	TTTGTTAGCAGCCGGATCCTTAACCCATTGTTTTGTAGAGATCTT	*X. bovienii* SS-2004
pJW84	pET11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0060-FW	CATCGTGAACAGATTGGTGGTGATCACTCGGCCGTCC	*X. bovienii* SS-2004
	jw0145-RV	TTTGTTAGCAGCCGGATCCTTAAAATTGCAGTAACTGAGCATCAATAAT	*X. bovienii* SS-2004
pJW140	pET-11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET-11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0253-FW	CATCGTGAACAGATTGGTGGTGATCACTCGTCTGTGATCACACA	*X. cabanillasii*
	jw0252-RV	TTTGTTAGCAGCCGGATCCTTATTCCATTGATTTATAAATATCTTGTAGGGA	*X. cabanillasii*
pJW144	pET-11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET-11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0035-FW	CATCGTGAACAGATTGGTGGTTATGCTCAGTTACTGCAATTTGAC	*X. cabanillasii*
	jw0252-RV	TTTGTTAGCAGCCGGATCCTTATTCCATTGATTTATAAATATCTTGTAGGGA	*X. cabanillasii*
pJW145	pET-11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET-11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0253-FW	CATCGTGAACAGATTGGTGGTGATCACTCGTCTGTGATCACACA	*X. cabanillasii*
	jw0260-RV	TTTGTTAGCAGCCGGATCCTTAAAATTGCAGTAATTGAGCGTTGA	*X. cabanillasii*
pJW146	pET-11a-FW	TAAGGATCCGGCTGCTAAC	pET-11a modified
	pET-11a_smt3-RV	ACCACCAATCTGTTCACGA	pET-11a modified
	jw0258-FW	CATCGTGAACAGATTGGTGGTATGAATAATAATGAATTAACATCTTTACCATTAGC	*X. cabanillasii*
	jw0262-RV	CAGCCGGATCCTTAATAATGTTGACGAGATAGCTTGGC	*X. cabanillasii*
pJW146_P10L	jw0273-FW	GTATGAATAATAATGAATTAACATCTTTACT*T*TTAGCTGAACGG	pJW146
	jw0274-RV	GTAAAGATGTTAATTCATTATTATTCATACCACCAATCTGTTCACG	pJW146
pJW146 _R14E/K15E	Sepas6-FW	TACTTGAGTTAGCCAAAGCCGCCAAGCTATCTC	pJW146
	Sepas19-RV	GCTTTGGCTAACTCAAGTAGTCTTTC*CT*C*TTCAGCTAATGGTAAAGATGTTAATTC	pJW146
pJW146_R16E	Sepas6-FW	TACTTGAGTTAGCCAAAGCCGCCAAGCTATCTC	pJW146
	Sepas5-RV	GCTTTGGCTAACTCAAGTAGC*T*C*TTTCCGTTCAGCTAATGG	pJW146
pJW146_K22E	Sepas10-FW	GAACGGAAAAGACTACTTGAGTTAGCCG*AAGCCGC	pJW146
	Sepas9-RV	CTCAAGTAGTCTTTTCCGTTCAGCTAATGGTAAAGATGTTAATTCATT	pJW146
pJW157	gene sequence *CXB49_11435*_T-^{C}DD was synthesized and subcloned into pET-11a by GenScript Biotech		pET-11a
pJW158	gene sequence *CXB49_11430*_^{N}DD was synthesized and subcloned into pET-11a by GenScript Biotech		pET-11a

pJW167	jw0299_FW	TAAGGATCCGGCTGCTAACAAAGC	pJW157
	JW1_201201_RV	GTTAGCAGCCGGATCCTTAAAATTGCAGCAGCTGCGC	pJW157
pJW168	CV_Cdd_FW	GTGAGCAGATCGGTGGCTATGCGCAGCTGCTGCAATTTG	pJW157
	pET_SMT3_rev2	GCCACCGATCTGCTCACGGTGC	pJW157
pJW81	jw0061-FW	TGACAATTAATCATCGGCTCG	pCK_0402
	jw0064-RV	CATGGAATTCCTCCTGTTAGCC	pCK_0402
	jw0117-FW	CGTTTTTTTGGGCTAACAGGAGGAATTCCATGAACCACCCTGAAAAGTTGA	*X. bovienii* SS-2004
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	*X. bovienii* SS-2004
	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	*X. bovienii* SS-2004
	jw0146-RV	GATTAATTGTCAACAGCTCCTGCAGTTAATGTCCTCTTGGGGTCG	*X. bovienii* SS-2004
pJW162	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW81
	jw0290-RV_mod.	GCTAACTCAAGTAGTCTTTTGCGTTCTGCTAAA*A*GTAAAGATATTAATTC	pJW81
	jw0291-FW	GCAGAACGCAAAAGACTACTTGAGTTAGC	pJW81
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW81
pJW164	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW81
	jw0296-RV_mod.	GCTTTGCTGCTTC*AGCTAACTCAAGTAGTCTTTTGCGTTCTGC	pJW81
	jw0295-FW_mod.	CCATTAGCAGAACGCAAAAGACTACTTGAGTTAGCTG*AAG	pJW81
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW81
pJW166	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW81
	jw0298-RV	GCTTTGCTGCTTTAGCTAACTCAAGTAGTCTTTC*C*T*C*TTCTGC	pJW81
	jw0293-FW_mod.	CTACTTGAGTTAGCTAAAGCAGCAAAGCTAACTCGTCAG	pJW81
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW81
pJW169	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW81
	jw0303_RV	ATCATTCATTTTTTAAGTACCATTCAAAGCCAG	pJW81
	jw0302_FW	CTGGCTTTGAATGGTACTTAAAAAATGAATGATATTCATGCACAACCCCGTGATG	pJW81
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW81
pJW170	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW81
	jw0305_RV	AGCGTCAAATTGCAGTAACTGAGCATCAAT	pJW81
	jw0304_FW	ATTGATGCTCAGTTACTGCAATTTGACGCTATGGGTTAATTCTGGTTTTTAACTGGCTTTG AATGG	pJW81
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW81
pJW171	jw0140-FW	CACTGTATAAACTTGCTGAACTTATTCAAGCGAAAAATTCATTGAGTGC	pJW169
	jw0305_RV	AGCGTCAAATTGCAGTAACTGAGCATCAAT	pJW169
	jw0304_FW	ATTGATGCTCAGTTACTGCAATTTGACGCTATGGGTTAATTCTGGTTTTTAACTGGCTTTG AATGG	pJW169
	jw0138-RV	GGGCACTCAATGAATTTTTCGCTTGAATAAGTTCAGCAAGTTTATACAGTGTTGAGT	pJW169

Supporting Table S5. Plasmids used in this work.

plasmid	base pairs [bp]	genotype	reference
pET-11a modified	5938	ori pBR322 ,amp^R, *T7prom-his$_6$-smt3*, Ulp1 cleavage site	[3]
pCK_0402	5256	ori p15A, cm^R, *araC-P$_{BAD}$* and *tacI-araE*	[25]
pCDF_xb2154	5397	ori CloDF13, spec^R, *T7prom, xb2154*	this work
pJW28	5998	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xb2153*_Y-^{C}DD	this work
pJW29	6049	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xb2152*_^{N}DD-Y	this work
pJW51	6250	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xb2153*_T$_1$-^{C}DD	this work
pJW84	6211	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xb2153*_T$_1$	this work
pJW140	6250	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2000*_T$_1$-^{C}DD	this work
pJW145	6211	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2000*_T$_1$	this work
pJW144	5998	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2000*_^{C}DD	this work
pJW146	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2001*_^{N}DD-Y	this work
pJW146_P10L	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2001*_^{N}DD_P10L-Y	this work
pJW146_R14E/K15E	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2001*_^{N}DD_R14E/K15E-Y	this work
pJW146_R16E	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2001*_^{N}DD_R16E-Y	this work
pJW146_K22E	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, xcab2001*_^{N}DD_K22E-Y	this work
pJW157	6262	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, CXB49_11435*_T-^{C}DD	this work
pJW158	6031	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, CXB49_11430*_^{N}DD	this work
pJW167	6211	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, CXB49_11435*_T	this work
pJW168	6010	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3, CXB49_11435*_^{C}DD	this work
pJW81	19247	ori p15A, cm^R, *araC-P$_{BAD}$* *xb2153*_C$_1$A$_1$T$_1$_*xb2152*_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE and *tacI-araE*	this work
pJW162	19247	ori p15A, cm^R, *araC-P$_{BAD}$* *xb2153*_C$_1$A$_1$T$_1$_*xb2152*_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE (*xb2152*_PaxB_^{N}DD_P10L) and *tacI-araE*	this work
pJW164	19247	ori p15A, cm^R, *araC-P$_{BAD}$* *xb2153*_C$_1$A$_1$T$_1$_*xb2152*_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE (*xb2152*_PaxB_^{N}DD_K22E) and *tacI-araE*	this work
pJW166	19247	ori p15A, cm^R, *araC-P$_{BAD}$* *xb2153*_C$_1$A$_1$T$_1$_*xb2152*_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE (*xb2152*_PaxB_^{N}DD_R14E/K15E) and *tacI-araE*	this work
pJW169	19151	ori p15A, cm^R, *araC-P$_{BAD}$ xb2153*_C$_1$A$_1$T$_1$_*xb2152*_ Δ^{N}DD_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE and *tacI-araE*	this work
pJW170	19220	ori p15A, cm^R, *araC-P$_{BAD}$ xb2153*_C$_1$A$_1$T$_1$Δ^{C}DD_ *xb2152*_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE and *tacI-araE*	this work
pJW171	19124	ori p15A, cm^R, *araC-P$_{BAD}$ xb2153*_C$_1$A$_1$T$_1$Δ^{C}DD_*xb2152*_ Δ^{N}DD_C$_2$A$_2$T$_2$C$_3$A$_3$T$_3$C/E$_4$A$_4$T$_4$_*xb2151*_TE and *tacI-araE*	this work

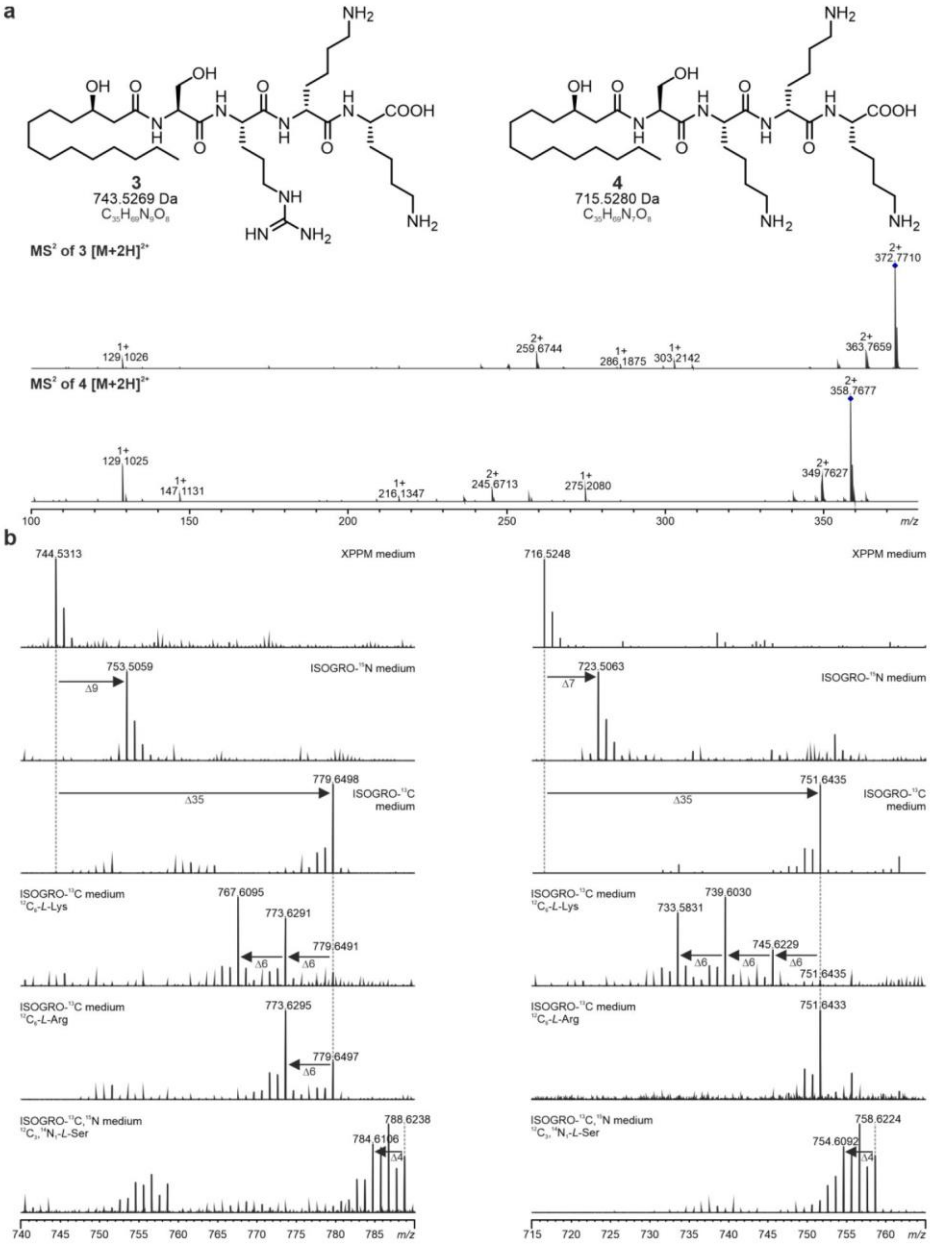
Supporting Table S6. ESI-MS data of all produced peptides.

peptide (#)	theoretical mass-to-charge ratio (*m/z*) $[M+H]^+/[M+2H]^{2+}$	molecular formula	reference
3	744.5342/372.7707	$C_{35}H_{69}N_9O_8$	this work
4	716.5280/358.7677	$C_{35}H_{69}N_7O_8$	this work
5	742.5185/371.7629	$C_{35}H_{67}N_9O_8$	this work
6	460.3381/-	$C_{23}H_{45}N_3O_6$	this work

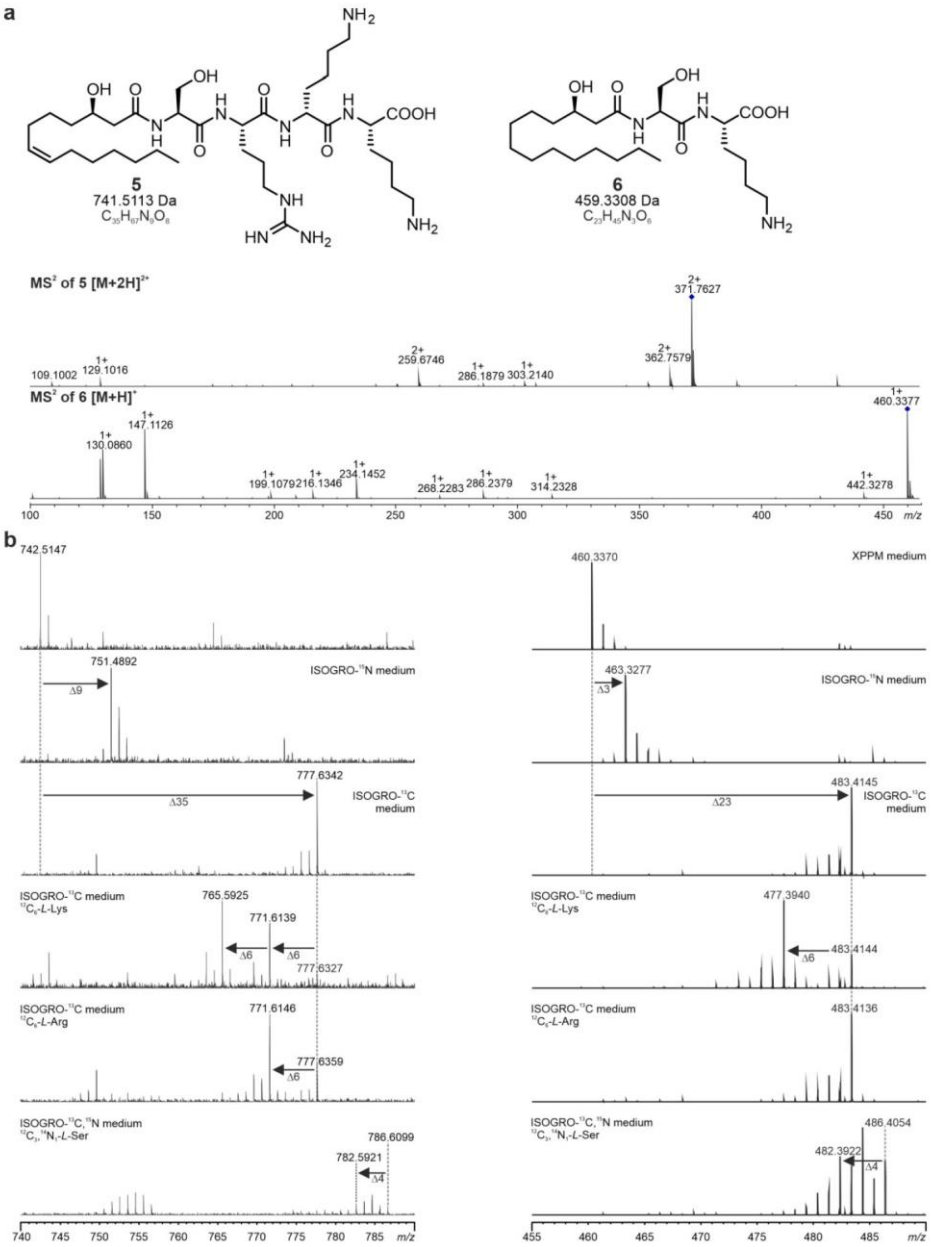
Supporting Table S7. ITC titration experiments of all PaxA (T_1-)^{C}DDs with PaxB ^{N}DDs.

Organism			K_D	n [sites]	ΔH [kcal/mol]	ΔS [cal/mol/deg]
Xenorhabdus bovienii SS-2004	**PaxA ^{C}DD**		no binding			
	PaxA T_1-^{C}DD (*apo*)	**PaxB ^{N}DD**	201±20 nM	0.74±0.08	-9.6±1.1	-2.0±3.8
	PaxA T_1		no binding			
Xenorhabdus cabanillasii JM26	**PaxA ^{C}DD**		no binding			
	PaxA T_1-^{C}DD (*apo*)	**PaxB ^{N}DD**	248±18 nM	0.88±0.01	-14.3±0.3	-18.6±0.9
	PaxA T_1-^{C}DD (*holo*)		220±23 nM	0.73±0.03	-16.8±0.2	-26.8±0.3
	PaxA T_1		no binding			
Xenorhabdus cabanillasii JM26	**PaxA T_1-^{C}DD**	**PaxB ^{N}DD Δα1**	no binding			
	PaxA T_1-^{C}DD	**PaxB ^{N}DD P10L**	3.9±0.1 µM	0.93±0.04	-20.3±0.6	-44.7±1.9
	PaxA T_1-^{C}DD	**PaxB ^{N}DD R14E/K15E**	no binding			
	PaxA T_1-^{C}DD	**PaxB ^{N}DD R16E**	2.8±0.3 µM	0.83±0.04	-30.1±0.2	-77.3±1.0
	PaxA T_1-^{C}DD	**PaxB ^{N}DD K22E**	3.8±0.2 µM	0.85±0.05	-13.8±0.3	-22.4±0.8

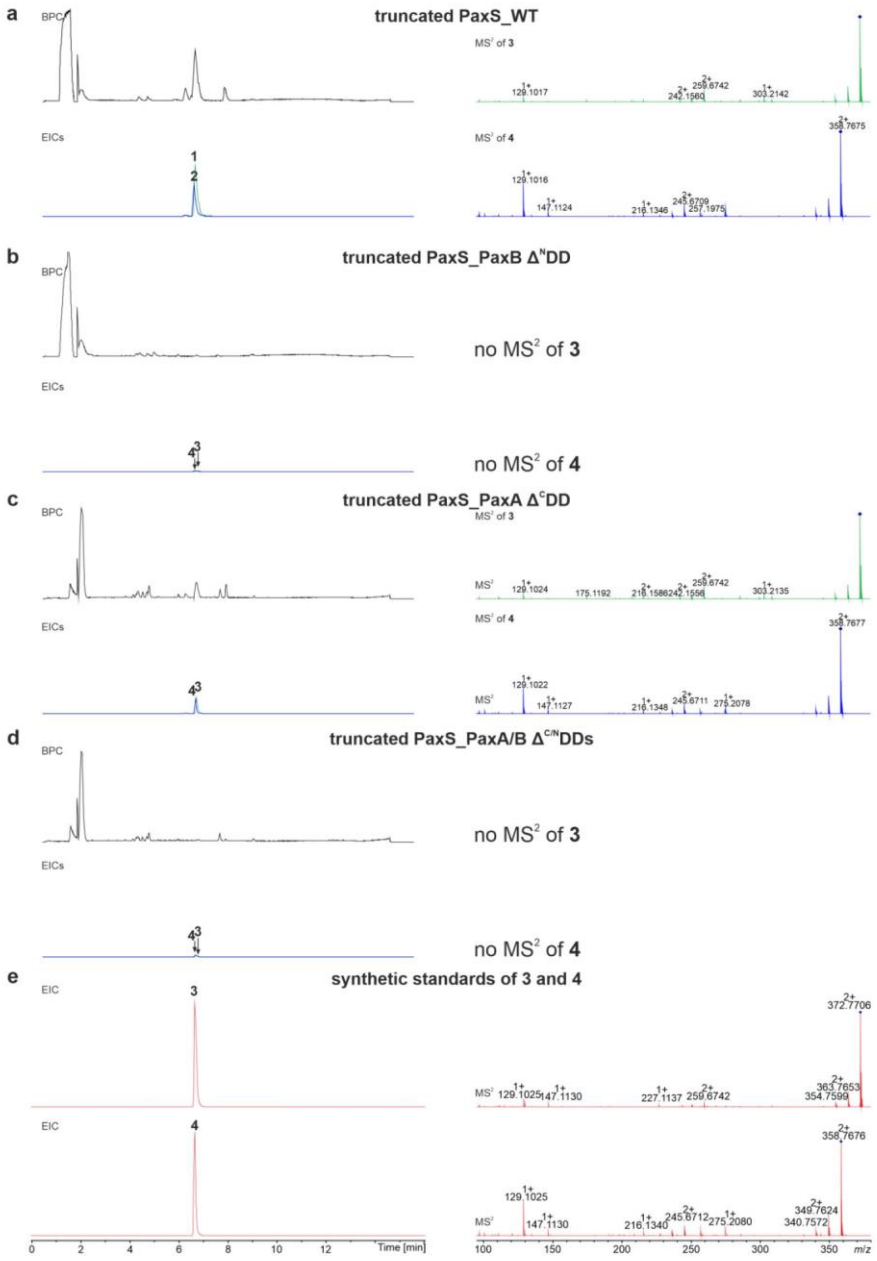
3 Supporting Information Figures



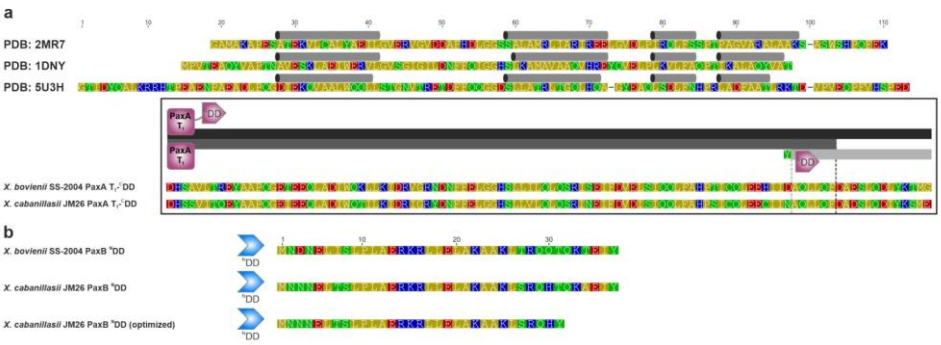
Supporting Figure S1. Characterization of the product spectrum of the truncated PaxS by HR-HPLC-ESI-MS. a) MS^2 fragmentation pattern of PAX tetra-peptides **3** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_9O_8$+2H$^+$: 372.7707 [*M*+2H]$^{2+}$) and **4** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_7O_8$+2H$^+$: 358.7677 [*M*+2H]$^{2+}$) produced by the truncated PaxS of *Xenorhabdus bovienii* SS-2004. b) MS^1-based structure elucidation of PAX tetra-peptides **3** and **4** by feeding experiments with stable isotopes.



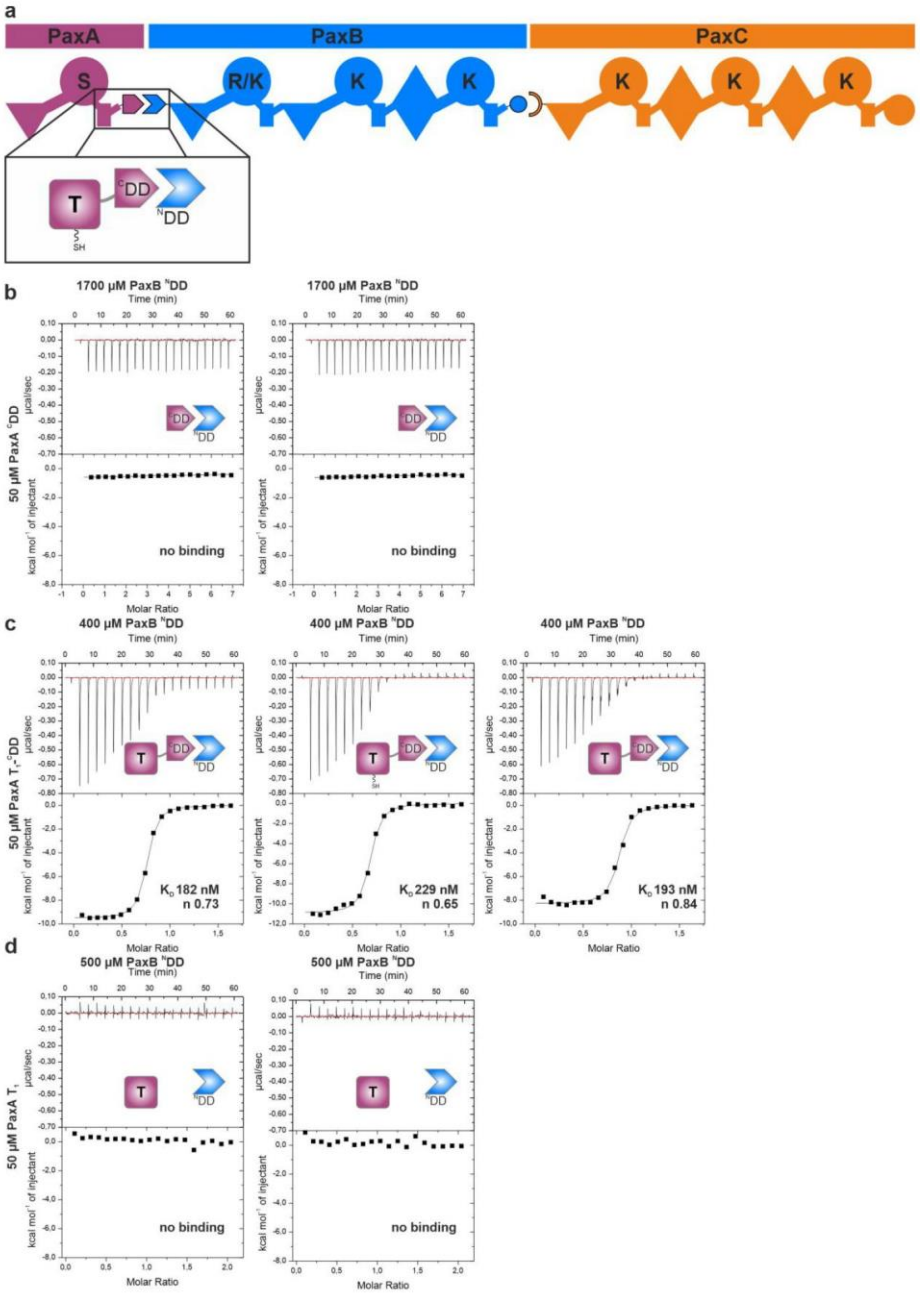
Supporting Figure S2. Characterization of the product spectrum of the truncated PaxS by HR-HPLC-ESI-MS. a) MS^2 fragmentation pattern of PAX tetra-peptide **5** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{67}N_9O_8$+2H$^+$: 371.7629 [*M*+2H]$^{2+}$) and di-peptide **6** (HRMS (ESI) *m*/*z* calcd for $C_{23}H_{45}N_3O_6$+H$^+$: 460.3381 [*M*+H]$^+$) produced by the truncated PaxS of *Xenorhabdus bovienii* SS-2004. b) MS^1-based structure elucidation of PAX tetra-/di-peptides **5** and **6** by feeding experiments with stable isotopes.



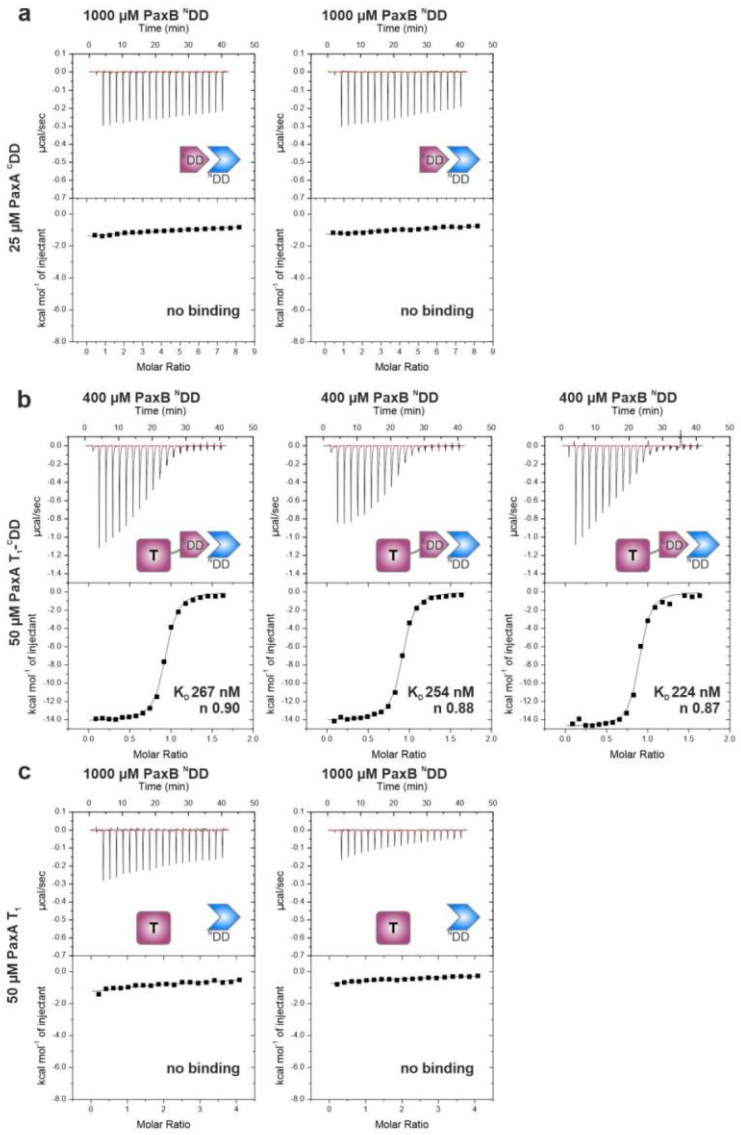
Supporting Figure S3. HPLC/MS data for compounds **3** and **4** produced by truncated PaxS variants in *E. coli* DH10B::*mtaA*. Exemplary base peak chromatograms (BPCs) of truncated a) PaxS_WT, b) PaxB Δ^{N}DD, c) PaxS_PaxA Δ^{C}DD and d) PaxS_PaxA/B $\Delta^{C/N}$DDs culture extracts and extracted ion chromatograms (EICs)/MS^2 fragmentation pattern of **3** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_9O_8$+2H$^+$: 372.7707 [*M*+2H]$^{2+}$) and **4** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_7O_8$+2H$^+$: 358.7677 [*M*+2H]$^{2+}$). e) Extracted ion chromatograms (EICs)/MS^2 fragmentation pattern of synthetic **3** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_9O_8$+2H$^+$: 372.7707 [*M*+2H]$^{2+}$) and **4** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_7O_8$+2H$^+$: 358.7677 [*M*+2H]$^{2+}$).



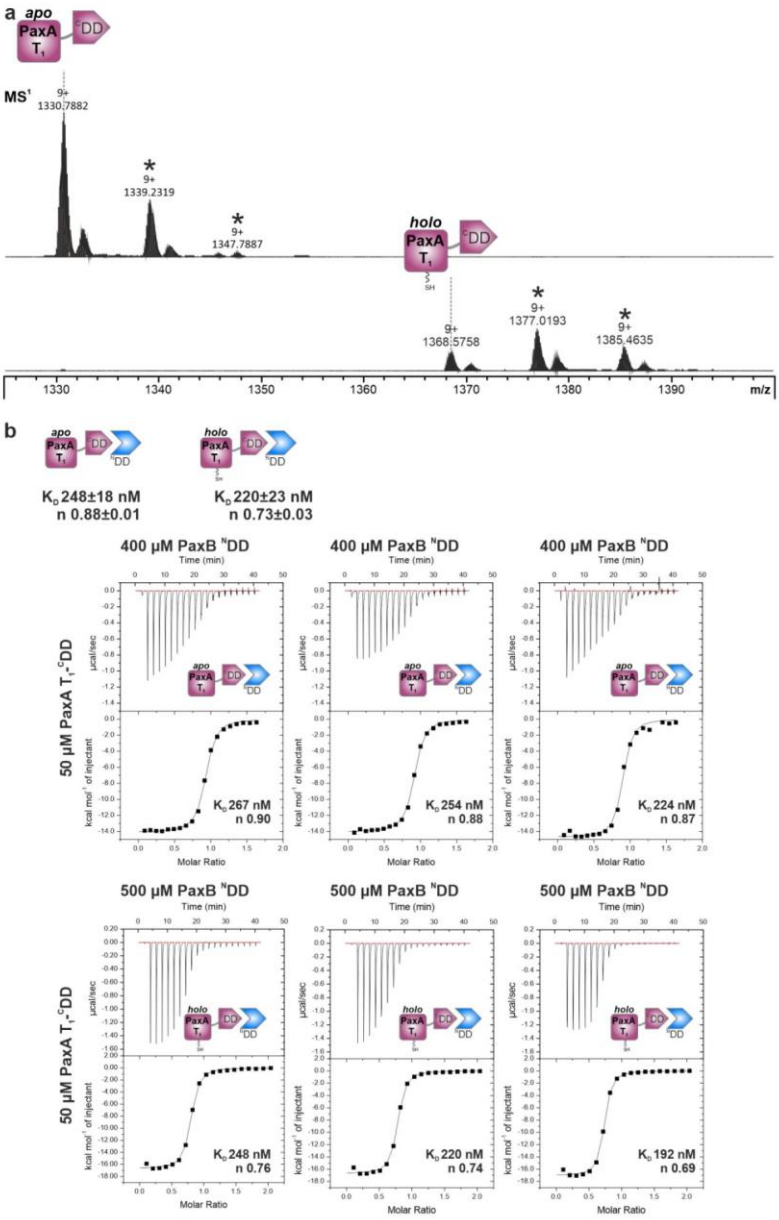
Supporting Figure S4. Design of PaxA T_1-^{C}DD and PaxB ^{N}DD constructs. a) Sequence alignment of *Xenorhabdus cabanillasii* JM26 and *Xenorhabdus bovienii* SS-2004 PaxA T_1-^{C}DDs to selected sequences of known T domain structures (Protein Data Bank (PDB) IDs: 7B2F: 2MR7, 1DNY, 5U3H; secondary structure depicted above each protein sequence). The alignment was performed using the multiple alignment program MUSCLE (default parameters)[26,27] and the amino acids are colored with respect to their polarity. The bars with different grey shades indicating the respective PaxA T_1-^{C}DD, T_1 and ^{C}DD sequences and the construct lengths is additionally highlighted by dashed lines. b) Sequence comparison of *X. bovienii* SS-2004 and *X. cabanillasii* JM26 PaxB ^{N}DDs. Additionally, the sequence of the length-optimized *X. cabanillasii* JM26 PaxB ^{N}DD is given. For domain assignment the following symbols are used: thiolation (T_1, square) domain and ^{C}DD (arrow) of PaxA (purple) and ^{N}DD (arrow) of PaxB (blue).



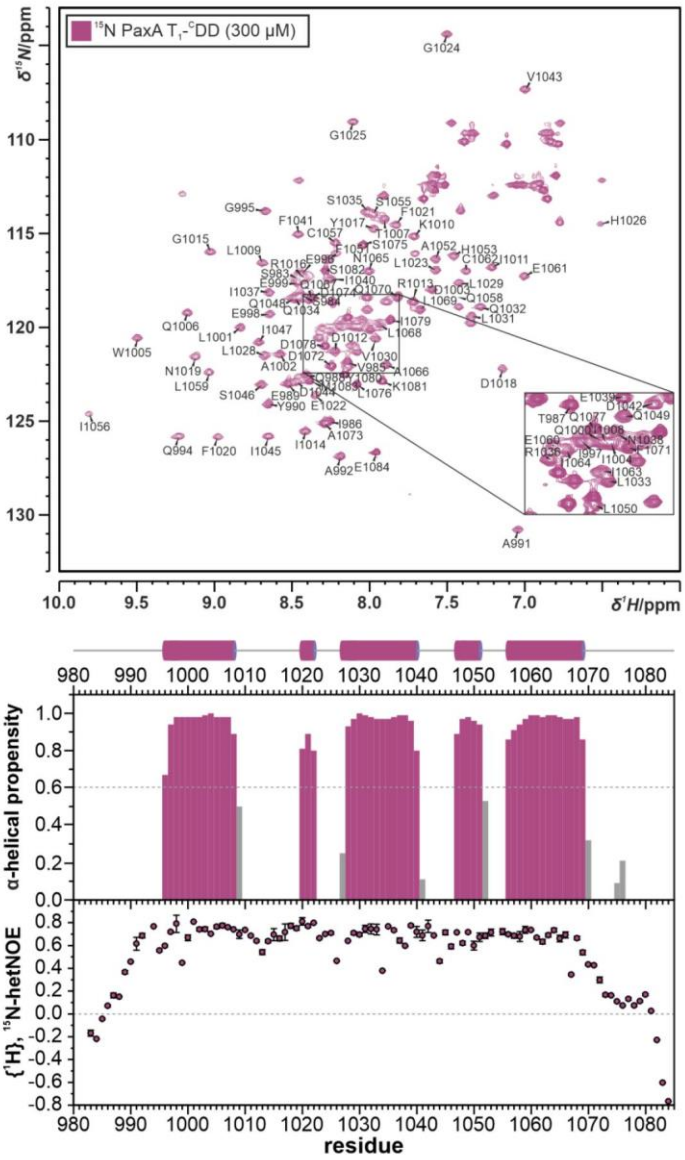
Supporting Figure S5. Thermodynamic characterization of the *X. bovienii* SS-2004 PaxA/B docking domain interface. a) Schematic representation of the PAX peptide-producing NRPS (PaxS). b) ITC experiments for PaxA ^{C}DD titrations with PaxB ^{N}DD, c) PaxA T$_1$-^{C}DD titrations with PaxB ^{N}DD and d) the PaxA T$_1$ domain titrated with PaxB ^{N}DD. See Figure 1 for assignment of the domain symbols.



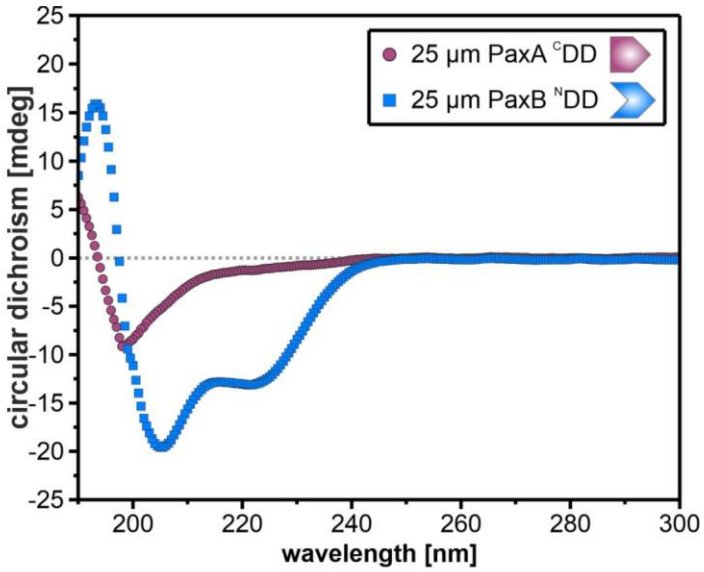
Supporting Figure S6. Thermodynamic characterization of the *X. cabanillasii* JM26 PaxA/B docking domain interface. a) Thermograms and resulting binding curves for all replications of the ITC titration experiments for PaxA ^{C}DD with PaxB ^{N}DD, b) PaxA T_1-^{C}DD with PaxB ^{N}DD and c) PaxA T_1 domain with PaxB ^{N}DD.



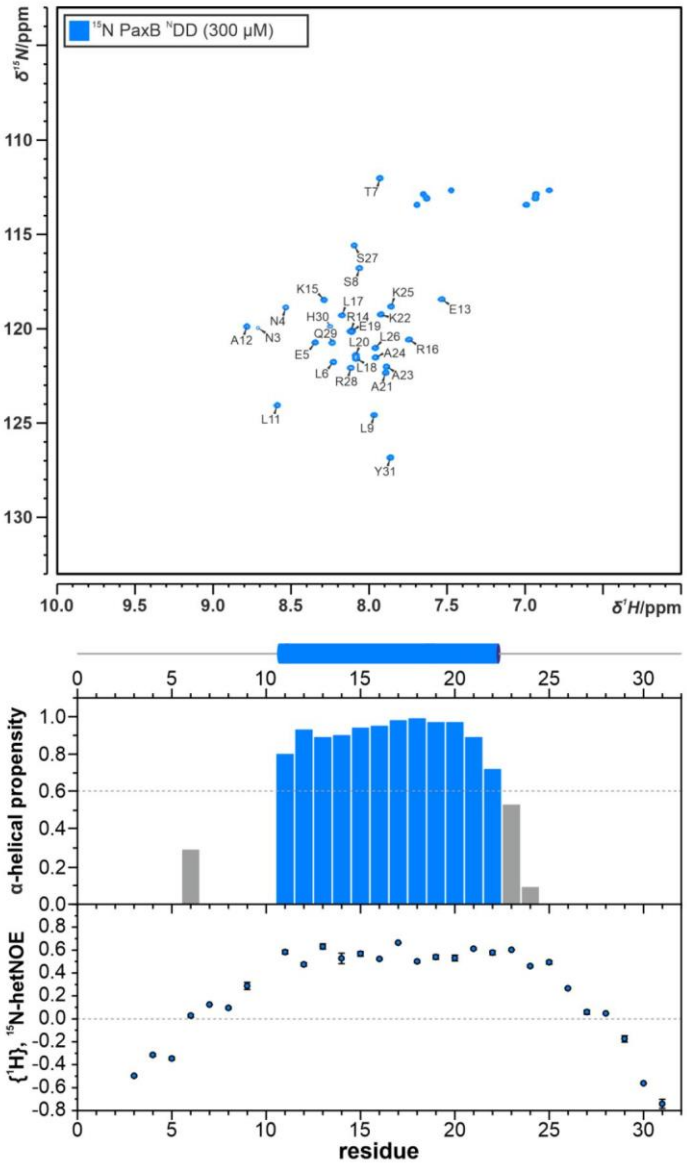
Supporting Figure S7. The presence or absence of the phosphopantetheinyl arm in the T_1 domain domain does not influence the affinity of the docking interaction. a) Protein HR-HPLC-ESI-MS analyses of purified *X. cabanillasii* JM26 PaxA T_1-^{C}DDs after protein expression either in *E. coli* BL21 BAP1 or *E. coli* BL21(DE3)Δ*entD* to evaluate the completeness of the post-translational modification. Displayed are the *m*/*z* values of the average protein masses of the 9^+ charge states (MS^1). *m*/*z* 1330.7882 corresponds to *apo*, *m*/*z* 1368.5758 to *holo* PaxA T_1-^{C}DD in MS^1, respectively. *m*/*z* 1339.2319/*m*/*z* 1347.7887 of *apo* PaxA T_1-^{C}DD and *m*/*z* 1377.0193/*m*/*z* 1385.4635 of *holo* PaxA T_1-^{C}DD (marked with an asterisk) are disulfide adducts (+76 Da) of β-mercaptoethanol[28]. b) ITC titration experiments for the *apo*/*holo* PaxA T_1-^{C}DD with the PaxB ^{N}DD in three replicates.



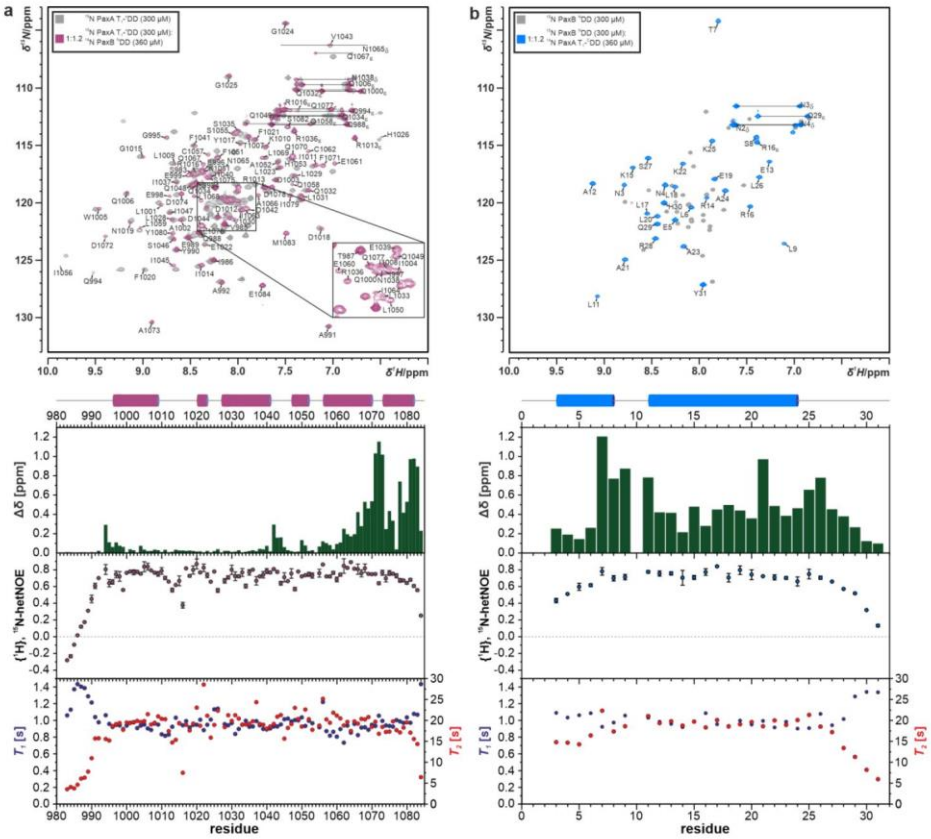
Supporting Figure S8. Structural characterization of the unbound *X. cabanillasii* JM26 PaxA T_1-^{C}DD di-domain by NMR. a) ^{1}H,^{15}N-HSQC spectrum (top) and TALOS-N-derived chemical shift index (middle) of the assigned *X. cabanillasii* JM26 PaxA T_1-^{C}DD in its unbound state. Additionally, values of backbone amide {^{1}H},^{15}N-heteronuclear NOEs for the unbound PaxA T_1-^{C}DD are plotted onto the sequence (bottom). The predicted secondary structure (confidence value ≥ 0.6) elements according to TALOS-N are depicted on top. The ^{C}DD region ($PaxA_{1070–1084}$) is unstructured and flexible.



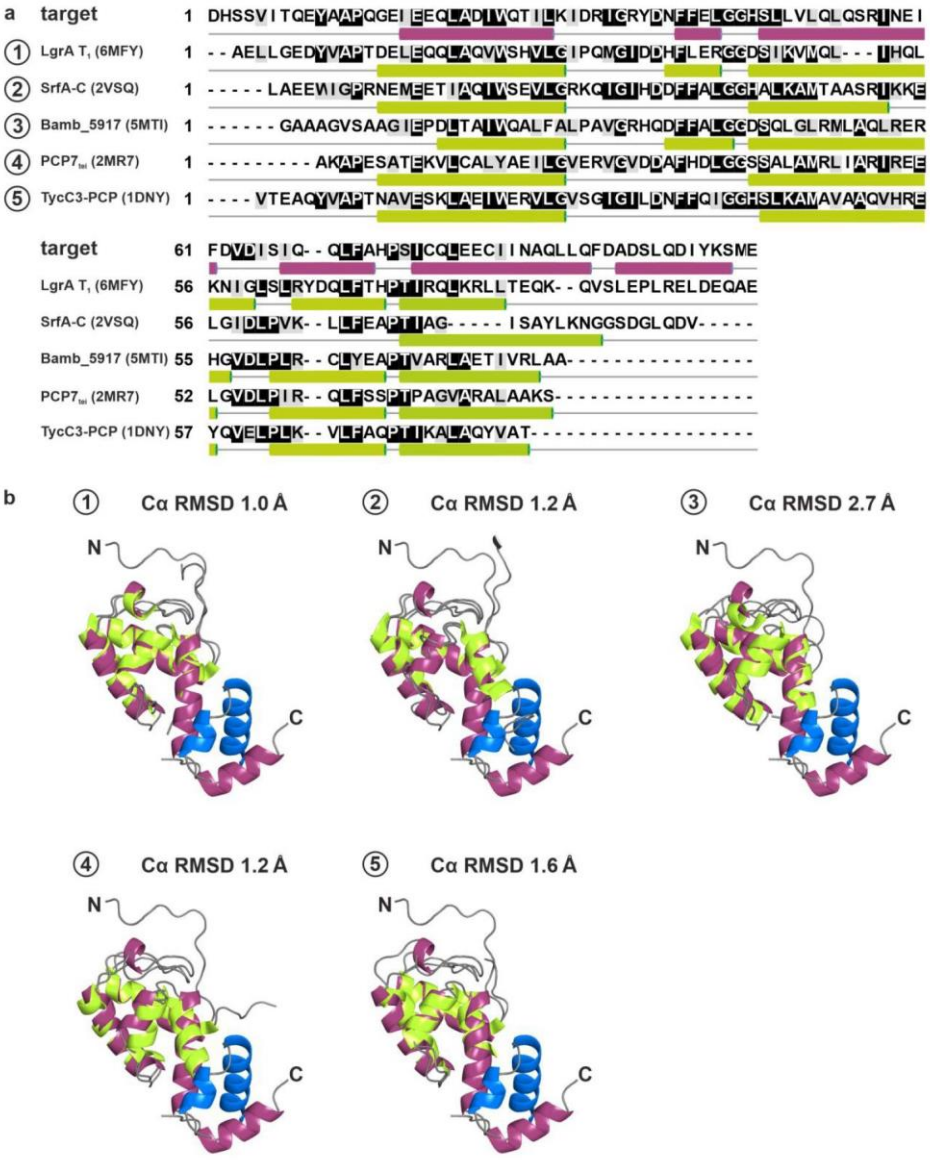
Supporting Figure S9. Comparison of the CD spectra for the isolated *X. cabanillasii* JM26 PaxA ^{C}DD peptide ($PaxA_{1066–1084}$, purple) and the PaxB ^{N}DD peptide ($PaxB_{1–30}$, blue).



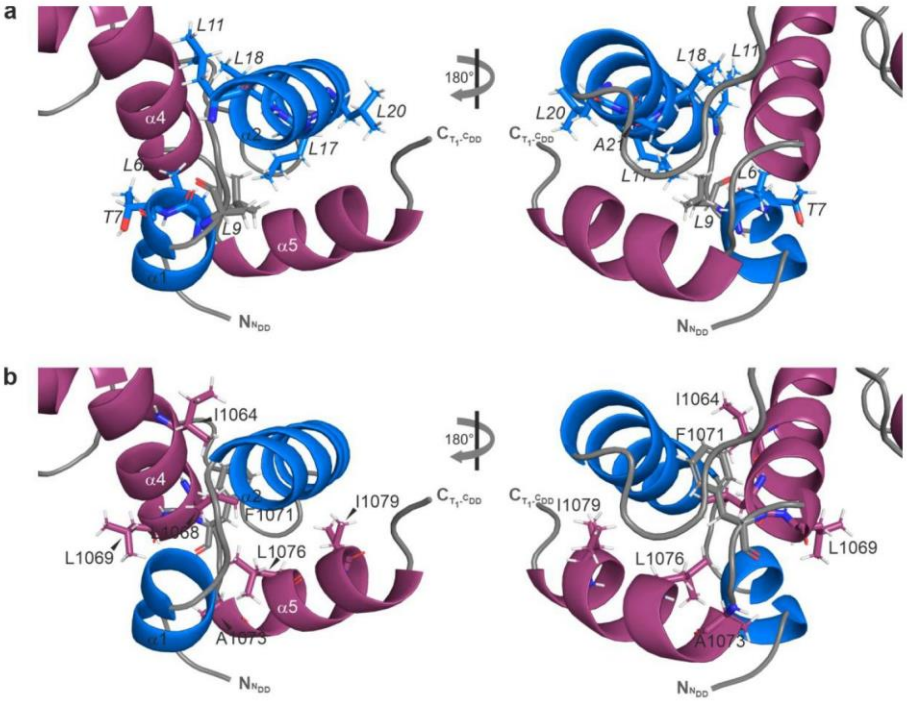
Supporting Figure S10. Structural characterization of the free *X. cabanillasii* JM26 PaxB ^{N}DD by NMR. ^{1}H,^{15}N-HSQC spectrum (top) and TALOS-N-derived chemical shift index (middle) of the assigned PaxB ^{N}DD in its unbound state. Additionally, values of backbone amide {^{1}H},^{15}N-heteronuclear NOEs of the unbound PaxB ^{N}DD are plotted onto the sequence (bottom). The predicted secondary structure (confidence value ≥ 0.6) elements according to TALOS-N are depicted on top..



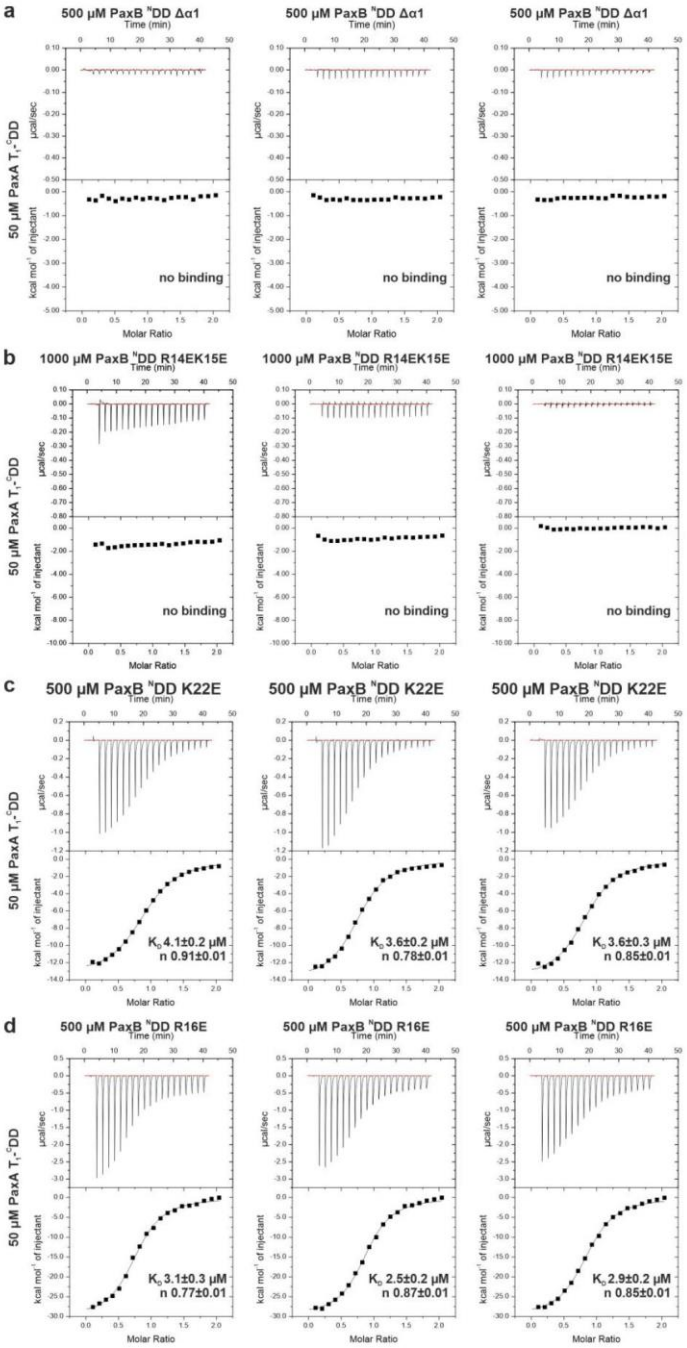
Supporting Figure S11. NMR chemical shift perturbation and dynamics data for the *X. cabanillasii* JM26 PaxA T_1-^{C}DD/PaxB ^{N}DD complex. a/b) Overlay of ^{1}H,^{15}N-HSQC spectra (top) of the assigned PaxA T_1-^{C}DD (purple) and the PaxB ^{N}DD (blue) in their bound and unbound (grey) states. In the lower panels chemical shift perturbations observed upon binding and the values of backbone amide {^{1}H},^{15}N-heteronuclear NOEs, ^{15}N T_1s and ^{15}N T_2s in the complex are plotted onto the sequence. The secondary structure elements in the complex according to the NMR solution complex structure are depicted on top.



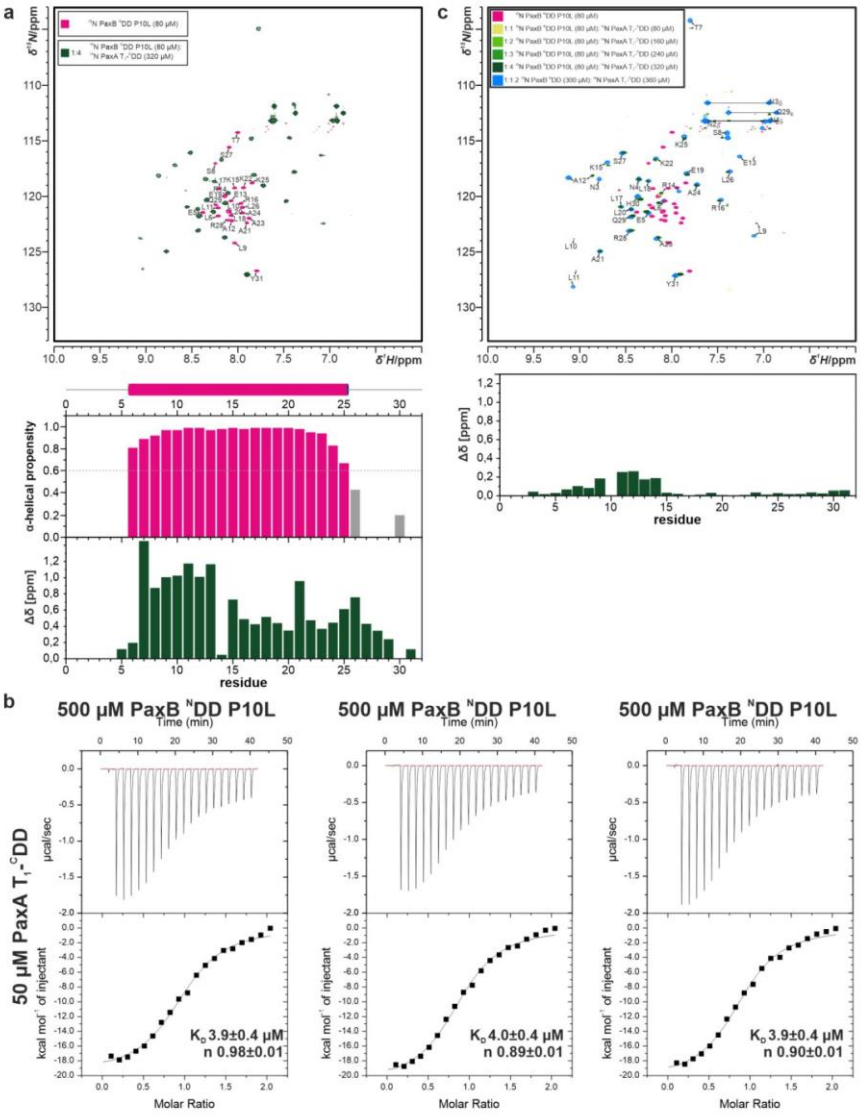
Supporting Figure S12. Comparison of the bound PaxA T_1-^{C}DD structure to previously described T domain structures. a) Sequence alignment of selected known carrier protein structures (PDB ID: 6MFY, 2VSQ, 5MTI, 2MR7, 1DNY; secondary structure (purple/green α-helices) are depicted above each sequence). The alignment was performed using the multiple alignment program MUSCLE (default parameters)[26,27] and the amino acids are colored with respect to their similarity (grey shades). b) Superimposition of selected structures with the structure of PaxA T_1-^{C}DD in the bound state based on Cα atoms in PyMOL.



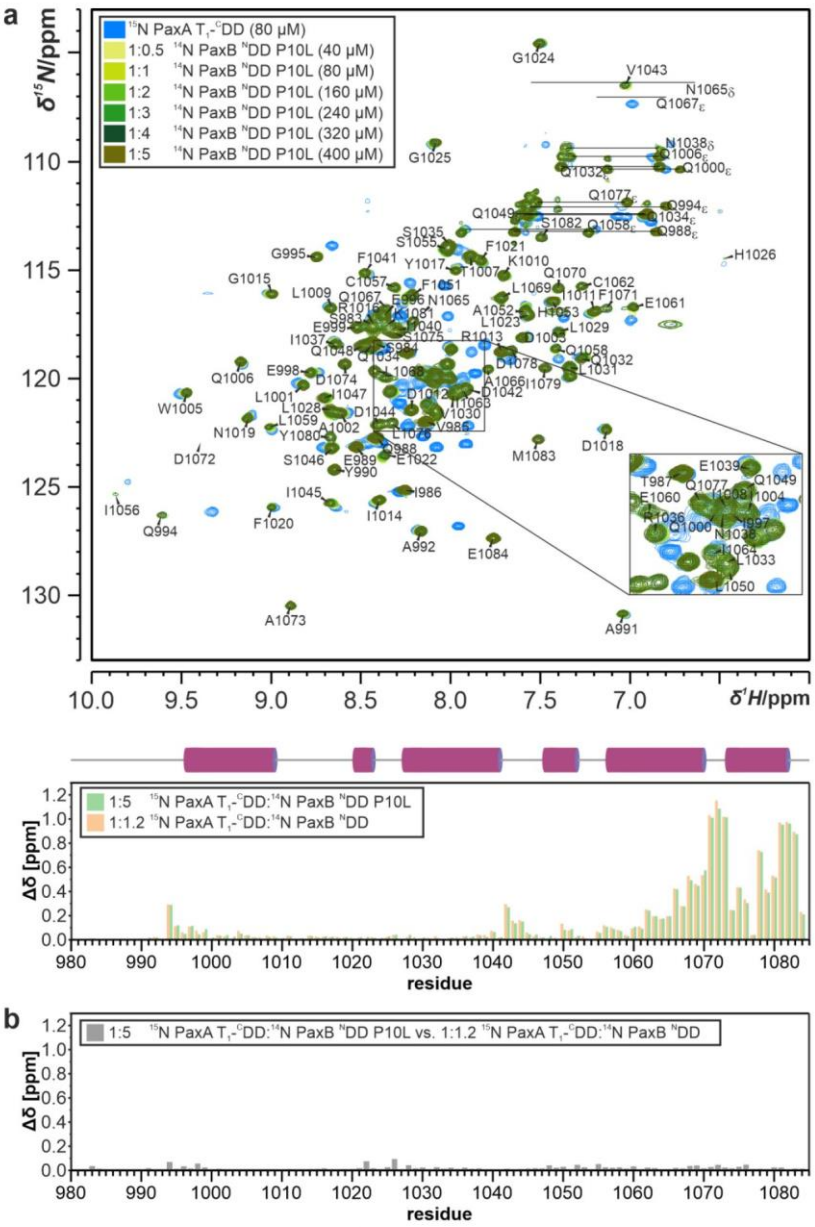
Supporting Figure S13. Stabilization of the docking interface by extensive hydrophobic interactions. a) PaxB ^{N}DD residues (italic) L6, T7, L9, L11, L17, L18, L20, A21 and b) PaxA T_1-^{C}DD residues I1064, L1068, L1069, F1071, A1073, L1076, I1079 shown as stick representations.



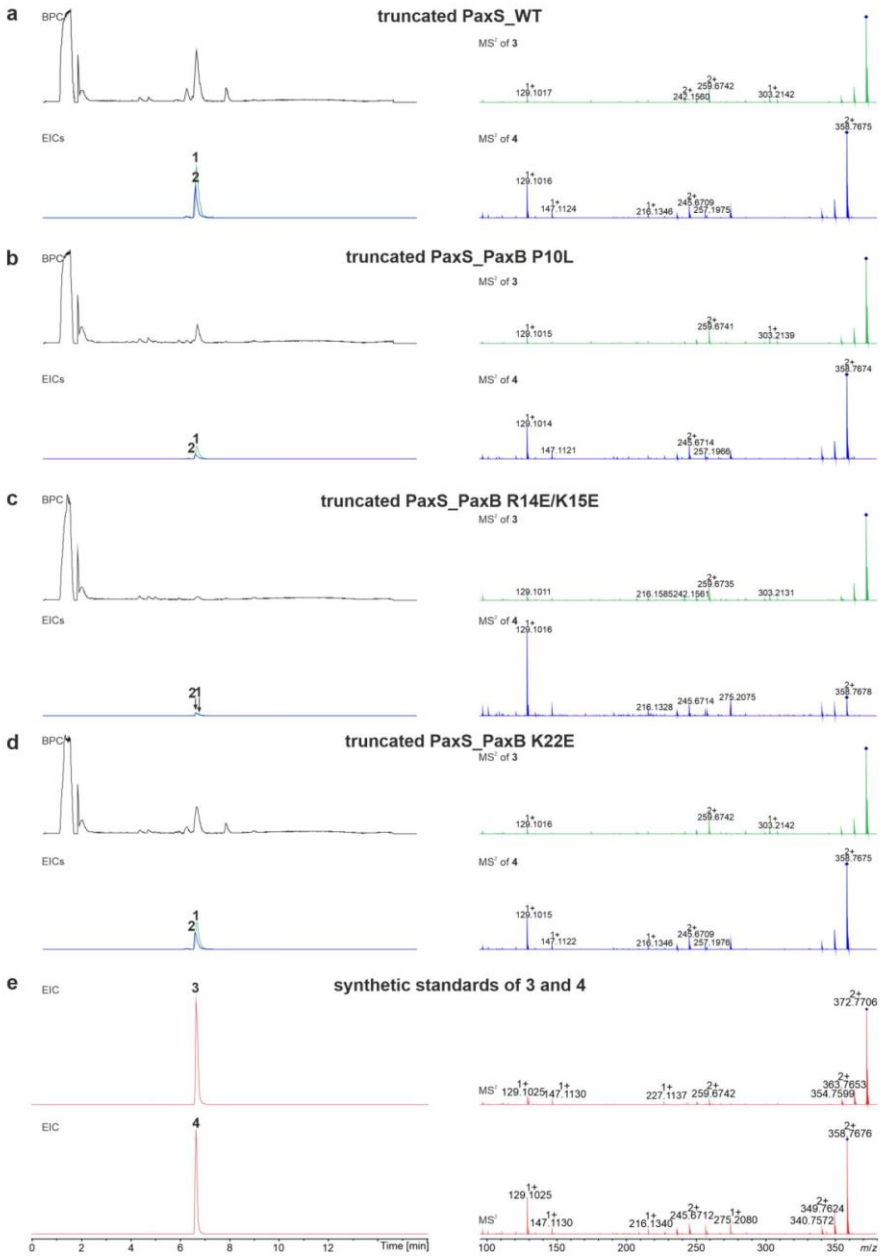
Supporting Figure S14. Binding of PaxB ^{N}DD variants to PaxA T_1-^{C}DD *in vitro*. Thermograms and the resulting binding curves for three replicates of ITC experiments for titrations of PaxA T_1-^{C}DD with the PaxB ^{N}DD variants a) $\Delta\alpha1$, b) R14E/K15E, c) K22E and d) R16E.



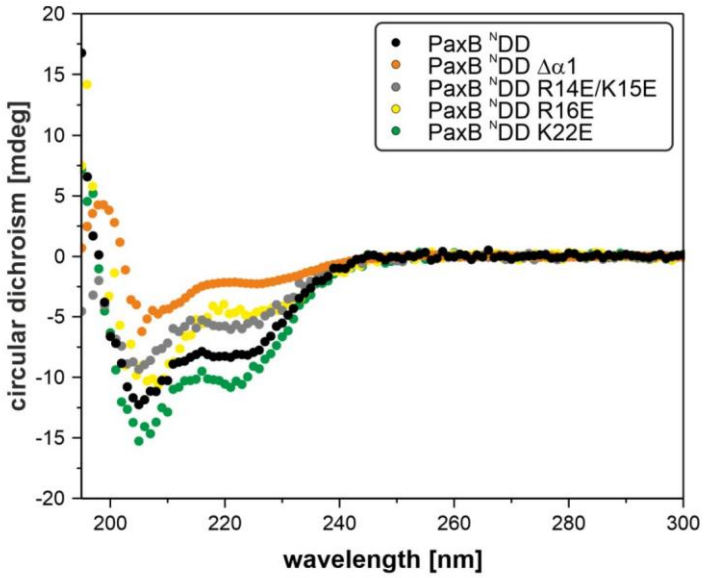
Supporting Figure S15. Structural characterization of the PaxB ^{N}DD P10L variant in its unbound and bound state. a) ^{1}H,^{15}N-HSQC spectra (top) of the PaxB ^{N}DD P10L mutant in its assigned unbound (pink) and bound (green) state. In the lower panel the TALOS-N-derived chemical shift index for the free PaxB ^{N}DD P10L mutant and the chemical shift changes observed upon binding of PaxA T$_1$-^{C}DD are plotted onto the sequence. The predicted secondary structure (confidence value ≥ 0.6) for free PaxB ^{N}DD P10L according to TALOS-N is depicted on top. b) ITC experiments for the PaxA T$_1$-^{C}DD with mutant PaxB ^{N}DD P10L in three replicates. c) Overlay of ^{1}H,^{15}N-HSQC spectra (top) of PaxB ^{N}DD P10L in the absence (pink) and presence of increasing amounts of unlabeled PaxA T$_1$-^{C}DD. The molar ratios of the two interaction partners are 1:1, 1:2, 1:3, 1:4 (different green shades). Binding is saturated at a fourfold excess of PaxA T$_1$-^{C}DD over PaxB ^{N}DD P10L. The assignment is given for the PaxA T$_1$-^{C}DD/PaxB ^{N}DD P10L complex and could be transferred from the bound wild type ^{N}DD. Below the chemical shift differences between the PaxA T$_1$-^{C}DD/PaxB ^{N}DD and PaxA T$_1$-^{C}DD/PaxB ^{N}DD P10L complex are plotted onto the ^{N}DD sequence of the bound wild type PaxB ^{N}DD and PaxB ^{N}DD P10L mutant are shown.



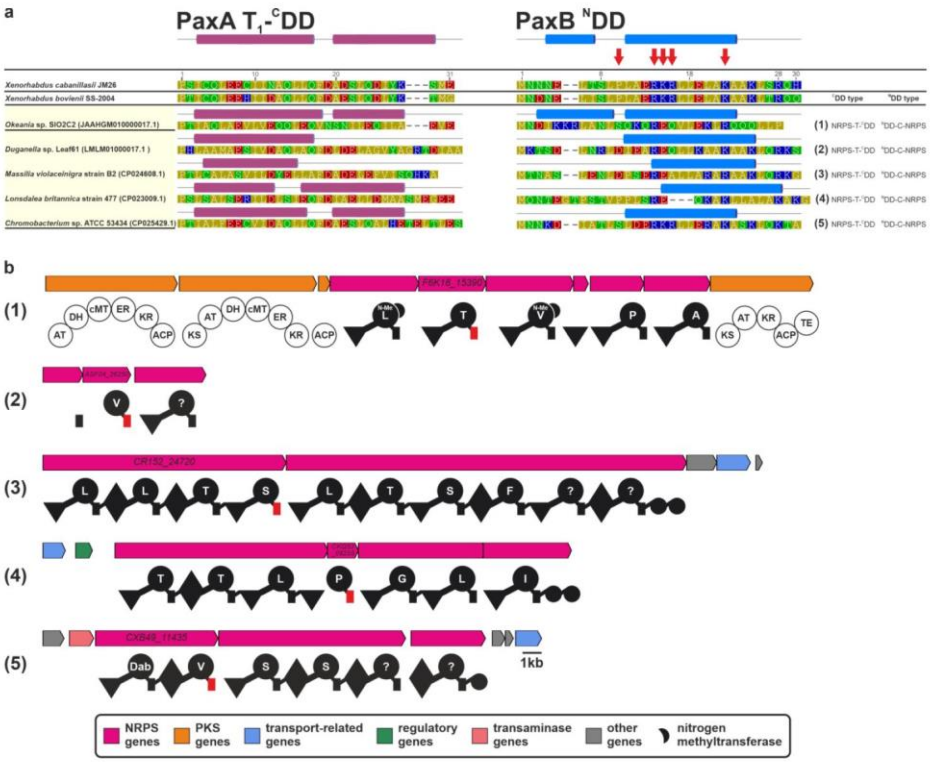
Supporting Figure S16. Structural characterization of PaxA T_1-^{C}DD bound to PaxB ^{N}DD P10L. a) Overlay of ^{1}H,^{15}N-HSQC spectra of PaxA T_1-^{C}DD in the absence (blue) and presence of increasing amounts of unlabeled PaxB ^{N}DD P10L. The molar ratios of the two interaction partners are 1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5 (different green shades, top). Comparison of chemical shift changes observed upon binding (saturated complex) of wild type PaxB ^{N}DD and PaxB ^{N}DD P10L to PaxA T_1-^{C}DD (bottom). b) Comparison of PaxA T_1-^{C}DD chemical shifts bound to either the WT PaxB ^{N}DD or to the PaxB ^{N}DD P10L mutant under saturating conditions plotted onto the PaxA T_1-^{C}DD sequence.



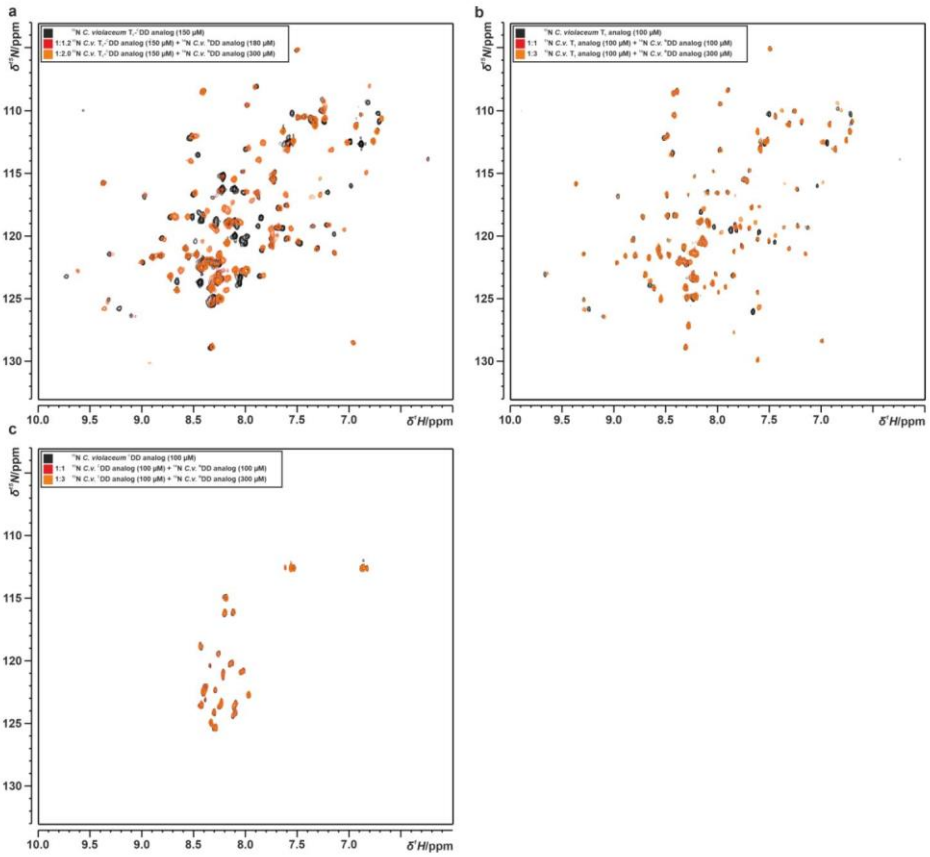
Supporting Figure S17. HPLC/MS data for compounds **3** and **4** produced by truncated PaxS variants in *E. coli* DH10B::*mtaA*. Exemplary base peak chromatograms (BPCs) of truncated a) PaxS_WT, b) PaxS_PaxB P10L, c) PaxS_PaxB R14E/K15E and c) PaxS_PaxB K22E culture extracts and extracted ion chromatograms (EICs)/MS^2 fragmentation pattern of **3** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_9O_8$+2H$^+$: 372.7707 [*M*+2H]$^{2+}$) and **4** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_7O_8$+2H$^+$: 358.7677 [*M*+2H]$^{2+}$). e) Extracted ion chromatograms (EICs)/MS^2 fragmentation pattern of synthetic **3** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_9O_8$+2H$^+$: 372.7707 [*M*+2H]$^{2+}$) and **4** (HRMS (ESI) *m*/*z* calcd for $C_{35}H_{69}N_7O_8$+2H$^+$: 358.7677 [*M*+2H]$^{2+}$).



Supporting Figure S18. CD spectra for *X. cabanillasii* JM26 PaxB NDD variant peptides measured at a concentration of 20 µM.



Supporting Figure S19. Multiprotein NRPS and NRPS-PKS systems with a similar composite docking interface. a) Alignment of selected NRPS T-^{C}DDs parts and ^{N}DDs identified by BLASTP search using *Xenorhabdus cabanillasii* JM26 PaxA/B T_1-^{C}DD/^{N}DDs as query sequences. Secondary structural elements are depicted above the table according to the NMR solution structure of the PaxA T_1-^{C}DD/PaxB ^{N}DD complex and red arrows indicate exchanged amino acids. The bacteria in the sequence alignment are grouped by their phylum including Cyanobacteria and Proteobacteria (from top to bottom). The alignment was performed using the multiple alignment program MUSCLE (default parameters)[26,27] and the amino acids are colored according to their polarity. Above the sequence of each entry the secondary structure consensus prediction with confidence values ≥6 of Jpred4[29] is given. b) This composite type of docking interface was exclusively found between NRPS thiolation (T) and condensation (C) domains in gene cluster (1)–(5). See Figure 1 for assignment of the NRPS domain symbols. PKS domains (circles): acyltransferase (AT), dehydratase (DH), methyltransferase (cMT), enoylreductase (ER), ketoreductase (KR), acyl carrier protein (ACP), thioesterase (TE).



Supporting Figure S20. NMR and ITC data of a similar cooperative docking domain pair. Overlay of $^{1}H,^{15}N$-HSQC spectra of the a) PaxA T_1-^{C}DD, b) PaxA T_1 and c) PaxA ^{C}DD analogs from *Chromobacterium violaceum* (*C.v.*) Bergonzini in the absence (black) and presence of increasing amounts of unlabeled PaxB ^{N}DD analog. The different molar ratios of the two interaction partners during the titration experiments are indicated by different colors (red and orange).

4 References

[1] J. Sambrook, E. Fritsch, T. Maniatis, *Molecular cloning. A laboratory manual: Vol. 2*, Cold Spring Harbor, S.l., **1989**.
[2] C. Fu, W. P. Donovan, O. Shikapwashya-Hasser, X. Ye, R. H. Cole, *PLoS One* **2014**, *9*, e115318.
[3] C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert, *Nat. Commun.* **2018**, *9*, 4366.
[4] N. J. Tobias, H. Wolff, B. Djahanschiri, F. Grundmann, M. Kronenwerth, Y.-M. Shi, S. Simonyi, P. Grün, D. Shapiro-Ilan, S. J. Pidot et al., *Nature microbiology* **2017**, *2*, 1676.
[5] J. M. Chaston, G. Suen, S. L. Tucker, A. W. Andersen, A. Bhasin, E. Bode, H. B. Bode, A. O. Brachmann, C. E. Cowles, K. N. Cowles et al., *PLoS One* **2011**, *6*, e27909.
[6] D. Neri, T. Szyperski, G. Otting, H. Senn, K. Wüthrich, *Biochemistry* **1989**, *28*, 7510.
[7] J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *ACS Chem. Biol.* **2020**, *15*, 982.
[8] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J. Biomol. NMR* **1998**, *12*, 1.
[9] J. Watzel, S. Sarawi, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *Biomol. NMR Assignments* **2021**, 1.
[10] R. L. J. Keller **2004**.
[11] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins: Struct. Funct. Bioinf.* **2005**, *59*, 687.
[12] N. A. Farrow, O. Zhang, J. D. Forman-Kay, L. E. Kay, *J. Biomol. NMR* **1994**, *4*, 727.
[13] F. A. Mulder, D. Schipper, R. Bott, R. Boelens, *J. Mol. Biol.* **1999**, *292*, 111.
[14] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56*, 227.
[15] P. Güntert, *Eur. Biophys. J.* **2009**, *38*, 129.
[16] B. López-Méndez, P. Güntert, *J. Am. Chem. Soc.* **2006**, *128*, 13112.
[17] R. Koradi, M. Billeter, P. Güntert, *Comput. Phys. Commun.* **2000**, *124*, 139.
[18] J. W. Ponder, D. A. Case in *Advances in Protein Chemistry: Protein Simulations*, Academic Press, **2003**, pp. 27–85.
[19] A. Bhattacharya, R. Tejero, G. T. Montelione, *Proteins: Struct. Funct. Bioinf.* **2007**, *66*, 778.
[20] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles et al., *Protein Sci.* **2018**, *27*, 112.
[21] E. Bode, A. K. Heinrich, M. Hirschmann, D. Abebew, Y.-N. Shi, T. D. Vo, F. Wesche, Y.-M. Shi, P. Grün, S. Simonyi et al., *Angew. Chem. Int. Ed.* **2019**, *58*, 18957.
[22] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790.
[23] J. G. Owen, K. J. Robins, N. S. Parachin, D. F. Ackerley, *Environ. Microbiol.* **2012**, *14*, 1198.
[24] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290.
[25] C. Kegler, H. B. Bode, *Angew. Chem. Int. Ed.* **2020**, *59*, 13463.
[26] R. C. Edgar, *BMC Bioinf.* **2004**, *5*, 113.
[27] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792.
[28] G. E. Begg, D. W. Speicher, *J. Biomol. Tech.* **1999**, *10*, 17.
[29] A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, *Nucleic Acids Res.* **2015**, *43*, W389.

5. Supporting information

5.1. Structural investigation of the extended PaxB/C docking domain interface in the PAX peptide-producing NRPS from *Xenorhabdus bovienii*

5.1.1. Material and methods

Strains

Cultivation of *X. bovienii* SS-2004 was performed in lysogeny broth (LB) medium (pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl) on an orbital shaker or on LB agar (1% (w/v) agar) plates at 30 °C. Cultivation of *E. coli* cells was performed in LB medium on an orbital shaker or on LB agar plates at 37 °C supplemented with antibiotics in appropriate concentrations (ampicillin 100 µg/ml). All strains that were used and generated in this work are summarized in Supporting Table 1.

Supporting Table 1. Strains used in this work.

Strain	Genotype / NRPS	Reference
E. coli BL21(DE3)Δ*entD*	F⁻ *omp*T *hsd*S_B (r_B^-, m_B^-) *gal dcm* (DE3) Δ*prpRBCD*::*T7*$_{prom}$-*sfp T7*$_{prom}$-*prpE* Δ*entD* / -	[154]
E. coli DH10B	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *nup*G/ -	Invitrogen
Xenorhabdus bovienii SS-2004	wild type / *paxS*	[155]

Isolation and purification of DNA

For the isolation of genomic DNA the Gentra Puragene Yeast/Bact Kit (QIAGEN) was used. The isolation of plasmids from *E. coli* cells was accomplished with help of the Invisorb Spin Plasmid Mini Two Kit (STRATEC Biomedical AG). DNA from polymerase chain reactions (PCRs) was purified with MSB SpinPCRapace (STRATEC Biomedical AG) or from 1% Tris-acetate-ethylenediaminetetraacetic acid (TAE) agarose gels using Invisorb Spin DNA Extraction (STRATEC Biomedical AG). An additional DpnI (Thermo Fisher Scientific) digestion was performed if the PCR template was plasmid-based. All steps of the plasmid isolation procedure and DNA fragment cleanup were carried out according to the manufacturers' protocols.

PCR, cloning of plasmids and transformation of cells

S7 Fusion Polymerase (Mobidiag) was used for DNA amplification by PCR following the instructions of the manufacturer. PCR primers (Eurofins Scientific) used in this

study are listed in Supporting Table 2 and the homology arms for cloning purposes were introduced via oligonucleotide design and a two-step PCR. The linearization of the plasmid pET11a modified was accomplished by PCR using the primer pair pET11a-FW/pET11a_smt3-RV.

Supporting Table 2. Oligonucleotides used in this work.

plasmid	oligo-nucleotide	sequence (5´→3´; overlapping ends)	template
pJW43	pET11a-FW	TAAGGATCCGGCTGCTAAC	pet11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pet11a modified
	jw0046-FW	CATCGTGAACAGATTGGTGGTACACTGTATAAACTTGCTGAACTTATTCTATC	*X. bovienii* SS-2004
	ck0045-RV	TTTGTTAGCAGCCGGATCCTTATTGTTGATCTCCATTTAACATGG	*X. bovienii* SS-2004
pJW44	pET11a-FW	TAAGGATCCGGCTGCTAAC	pet11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pet11a modified
	jw0047-FW	CATCGTGAACAGATTGGTGGTGATCGTTCCGCCAGCG	*X. bovienii* SS-2004
	ck0045-RV	TTTGTTAGCAGCCGGATCCTTATTGTTGATCTCCATTTAACATGG	*X. bovienii* SS-2004
pJW46	pET11a-FW	TAAGGATCCGGCTGCTAAC	pet11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pet11a modified
	ck0042-FW	CATCGTGAACAGATTGGTGGTATGAACATAAATGAACAAACTTTGG	*X. bovienii* SS-2004
	jw0051-RV	TTTGTTAGCAGCCGGATCCTTAGTTCGGATAGGCCACGGTAG	*X. bovienii* SS-2004
pJW52	pET11a-FW	TAAGGATCCGGCTGCTAAC	pet11a modified
	pET11a_smt3-RV	ACCACCAATCTGTTCACGA	pet11a modified
	jw0047-FW	CATCGTGAACAGATTGGTGGTGATCGTTCCGCCAGCG	*X. bovienii* SS-2004
	jw0067-RV	TTTGTTAGCAGCCGGATCCTTAAAAAGTTTCAATTTGTCTTGATAGAATAAGTTCAG	*X. bovienii* SS-2004

Cloning was realized with NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) according to the manufacturer's instructions and *E. coli* cells were transformed with the plasmids by electroporation. Plasmids were verified by sequencing (Eurofins Genomics). All plasmids generated in this study are listed in Supporting Table 3.

Supporting Table 3. Plasmids used in this work.

plasmid	base pairs [bp]	genotype	reference
pet11a modified	5938	ori pBR322 ,amp^R, *T7prom-his$_6$-smt3*, Ulp1 cleavage site	[11]
pJW30	6043	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2152*_Y-^{C}DD	[82]
pJW31	6037	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2151*_^{N}DD	[82]
pJW43	6076	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2152*_T_4 α-helix 4-^{C}DD	this work
pJW44	6298	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2152*_T_4-^{C}DD	this work
pJW46	7450	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2151*_^{N}DD-C_5	this work
pJW52	6211	ori pBR322 ,amp^R, T7*prom-his$_6$-smt3*, *xb2152*_T_4	this work

Protein expression and purification

For the structure elucidation by NMR, proteins from *Xenorhabdus bovienii* SS-2004 were heterologously expressed in *E. coli* BL21-Gold(DE3)Δ*entD* under the control of a T7 promoter. The use of *E. coli* BL21-Gold(DE3)Δ*entD* cells was necessary to

obtain a uniform T domain species in its *apo*-state, which was confirmed by high resolution (HR)-high performance liquid chromatography (HPLC)-coupled electrospray ionization mass spectrometry (ESI-MS) analysis. For the uniform isotope labeling of the proteins of interest, *E. coli* cells harboring the respective protein coding plasmids were grown in ^{15}N and ^{15}N,^{13}C M9 minimal media containing 1 g L^{-1} $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) or 1 g L^{-1} $^{15}NH_4Cl$ and 2.5 g L^{-1} $^{13}C_6$-*D*-glucose (Cambridge Isotope Laboratories). For selective labeling of specific amino acids, cells were grown in M9 medium complemented with appropriately isotope labeled amino acids and the remaining amino acids in the unlabeled form.[156] The following labeling schemes were applied: ^{15}N leucine (0.23 g/l), ^{15}N phenylalanine (0.13 g/l). For ITC measurements, proteins were expressed in LB medium. Protein expression was induced at an OD_{600} of 0.6–0.8 with 1 mM IPTG followed by an overnight incubation at 20 °C. After expression, cells were lysed by sonication and purified as previously described in a three-step purification procedure including immobilized metal-affinity chromatographies (IMAC) combined with a final gel filtration chromatography.[82] The buffer solution containing the purified proteins was composed of 50 mM sodium phosphate (pH 6.5) and 100 mM NaCl. In Supporting Table 4 all isolated proteins are listed.

Supporting Table 4. Proteins used in this work.

protein ID	strain	protein	molecular weight [Da]	sum formula	charge state	theoretical mass [*m*/*z*]	detected mass [*m*/*z*]
JW30	*Xenorhabdus bovienii* SS-2004	**PaxB Y-^{C}DD**	4103	$C_{179}H_{277}N_{43}O_{65}S$	3^+	1368.6562	1368.6549
JW31	*Xenorhabdus bovienii* SS-2004	**PaxC ^{N}DD**	3962	$C_{171}H_{295}N_{51}O_{54}S$	6^+	661.2019	661.1985
JW43	*Xenorhabdus bovienii* SS-2004	**PaxB T_4 α-helix 4-^{C}DD**	5342	$C_{236}H_{380}N_{58}O_{80}S_1$	4^+	1336.4386	1336.4445
JW44	*Xenorhabdus bovienii* SS-2004	**PaxB T_4-^{C}DD_*apo***	13734	$C_{610}H_{949}N_{163}O_{194}S_2$	10^+	1374.3979	1374.4058
JW46	*Xenorhabdus bovienii* SS-2004	**PaxC ^{N}DD-C_5**	56537	$C_{2526}H_{3999}N_{687}O_{764}S_{10}$	1^+	56537.3383	*
JW52	*Xenorhabdus bovienii* SS-2004	**PaxB T_4_*apo***	10556	$C_{469}H_{725}{}^{15}N_{127}O_{141}S$	10^+	1056.5699	1056.6039

* confirmed by MALDI-TOF MS (data not shown)

NMR spectroscopy

NMR spectra were recorded at 20 °C on Bruker AVANCE III 600 and 900 MHz spectrometers equipped with cryogenic 5 mm triple resonance probes. The proton chemical shifts were internally referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid and the heteronuclear ^{13}C and ^{15}N chemical shifts were indirectly referenced with the appropriate conversion factors.[157]

For backbone assignments of unbound PaxB T_4-^{C}DD (200 µM) the BEST-TROSY-based versions of the following 3D experiments with a uniformly ^{13}C,^{15}N-labeled sample were recorded: HNCO, HN(CA)CO, HNCA and HNCACB.[158–160] The backbone assignments of the bound state were obtained from HNCO and HNCA triple resonance experiments[128] (Bruker standard pulse sequences) with uniformly ^{13}C,^{15}N-labeled samples of PaxB T_4-^{C}DD (200 µM) in the presence of 4 molar excess of unlabeled PaxC ^{N}DD (800 µM). All NMR samples were supplemented with 10% D_2O and all spectra were processed with Bruker TopSpin 3.6.2 and analyzed with the program CARA[161] (www.nmr.ch). TALOS-N[143] was used to calculate the secondary structure probability for the bound and unbound state of the PaxB T_4-^{C}DD based on the chemical shift assignments.

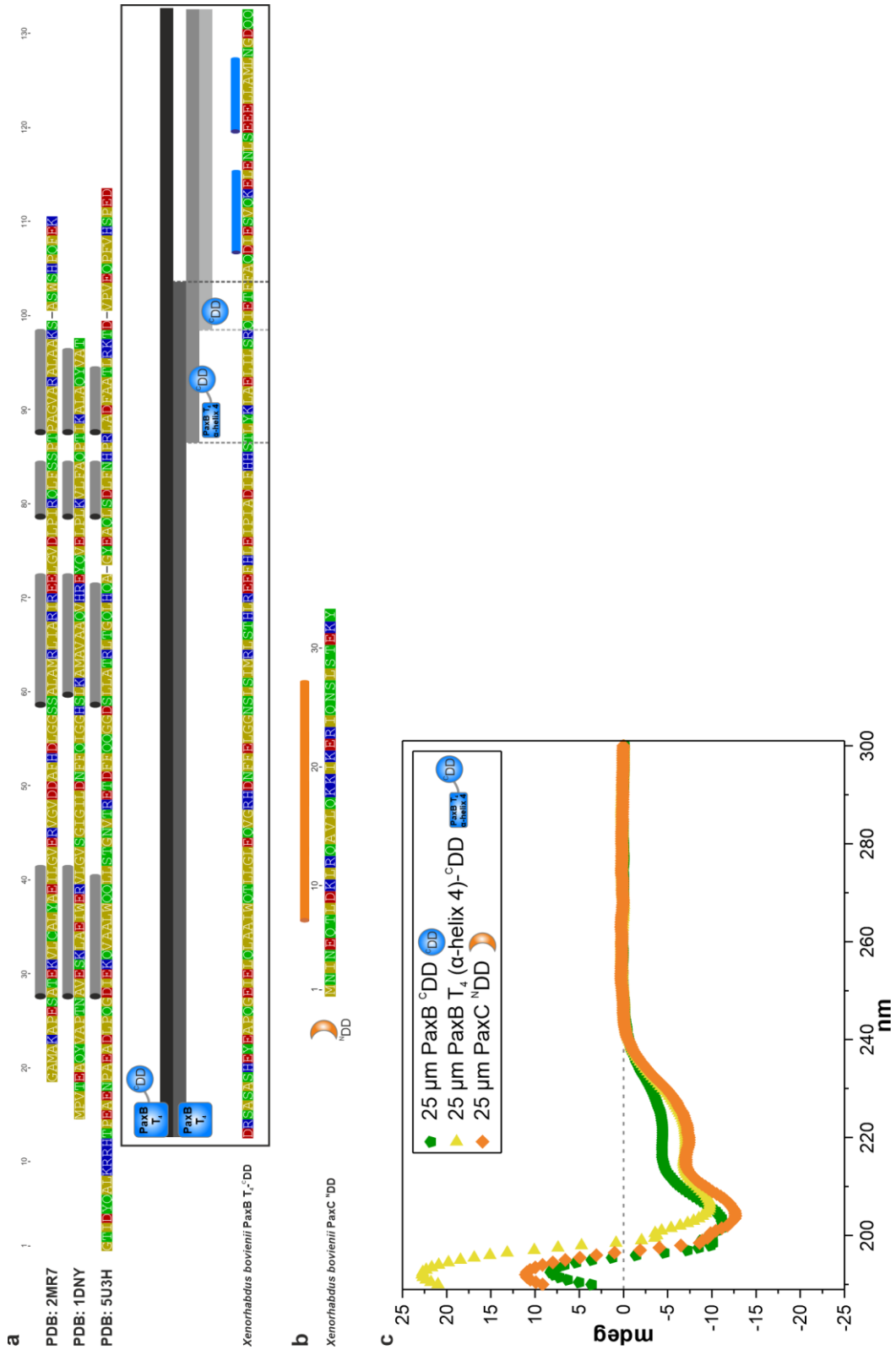
NMR titration experiments based on ^{1}H,^{15}N-HSQC spectra, which were recorded after the stepwise addition of unlabeled PaxC ^{N}DD to uniformly ^{15}N-labeled PaxB T_4-^{C}DD (80 µM). To evaluate the NMR titration experiments, the chemical shifts were determined using the peak picking function of CcpNmr Analysis[162] and with the following equation the chemical shift differences were calculated:[163]

$$(1)\ \Delta\delta = \sqrt{\Delta\delta_{HN}{}^{2} + \left(\frac{\Delta\delta_N}{6.5}\right)^{2}}$$

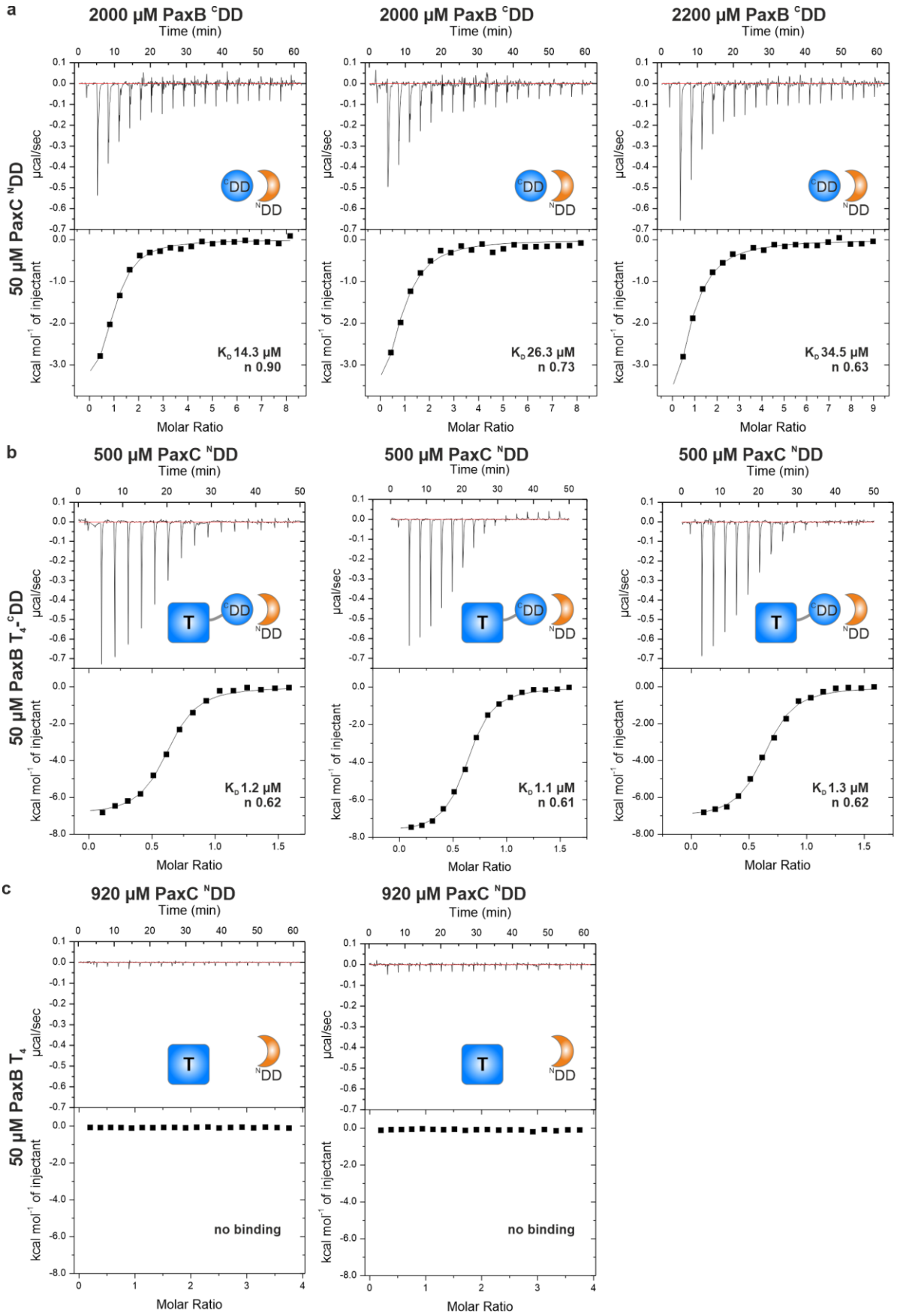
Isothermal titration calorimetry, circular dichroism and HR-HPLC-ESI-MS analyses

These analyses were performed as described in the supporting information of manuscript 2 (“Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS”).

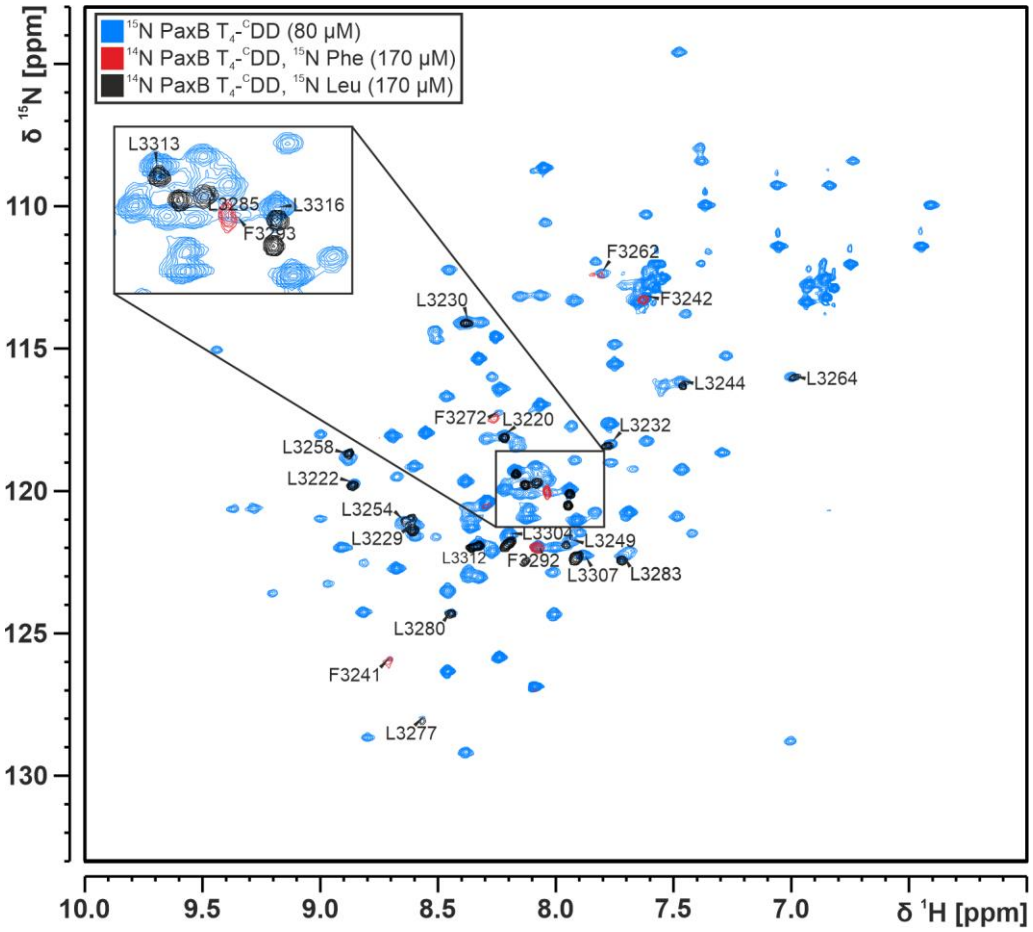
5.1.2. Supporting figures



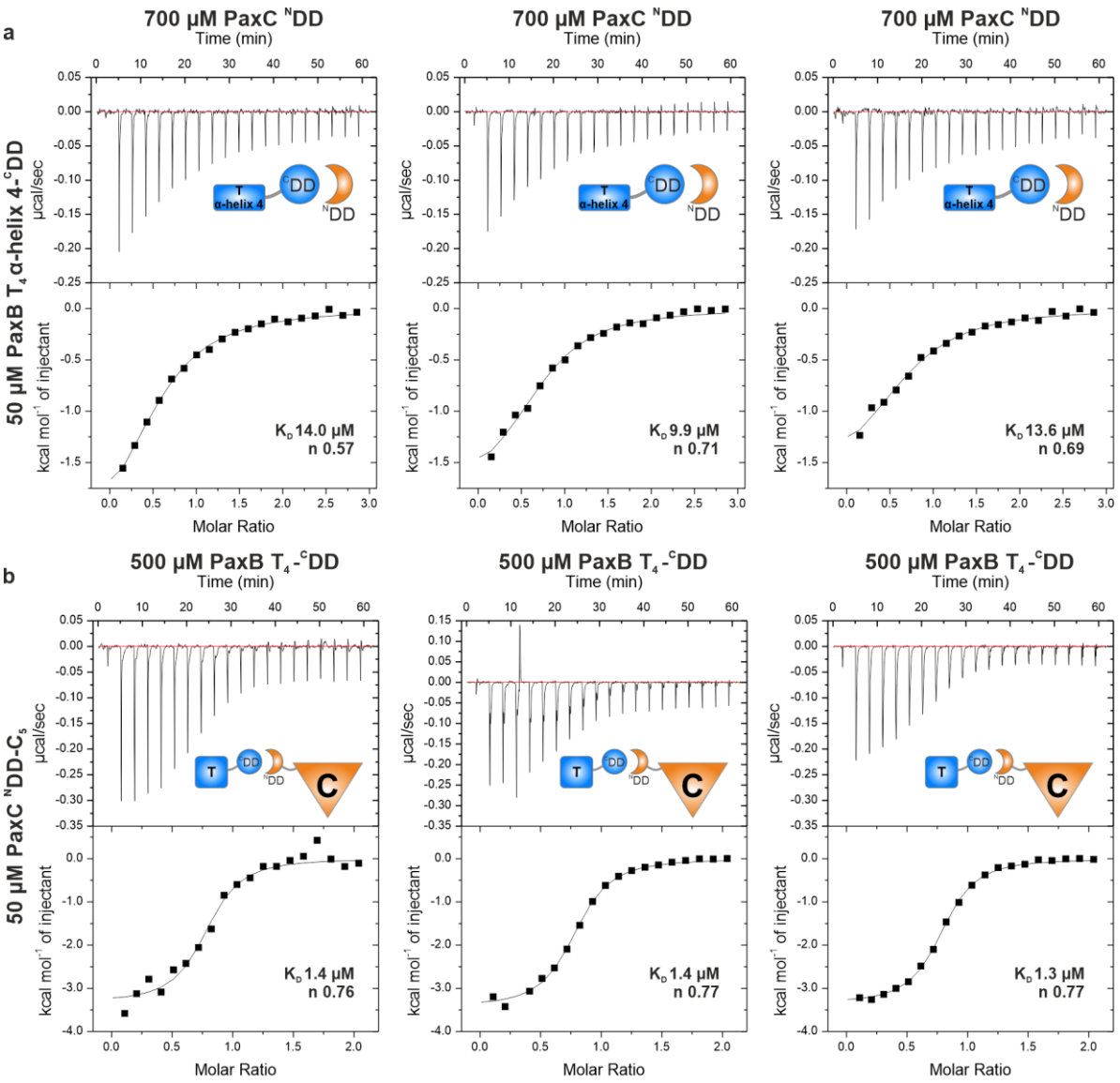
Supporting Fig. 1. Design of PaxB T_4-^{C}DD and PaxC ^{N}DD constructs. (a) Sequence alignment of *Xenorhabdus bovienii* PaxB T_4-^{C}DD to selected sequences of known T domain structures (PDB: 2MR7, 1DNY, 5U3H). The alignment was performed using the multiple alignment program MUSCLE (default parameters) and the amino acids are colored with respect to their polarity. (b) Sequence of *Xenorhabdus bovienii* PaxC ^{N}DD. (c) CD spectra of PaxB/C $^{C/N}$DD constructs.



Supporting Fig. 2. (a–c) ITC thermograms and the derived binding curves for titration between PaxB ^{C}DD:PaxC ^{N}DD, PaxB T$_4$-^{C}DD:PaxC ^{N}DD and PaxB T$_4$:PaxC ^{N}DD are shown.



Supporting Fig. 3. Overlay of ^{1}H,^{15}N-HSQC spectra of a fully ^{15}N-labeled and amino acid selective isotope labeled (^{15}N Phe (red), ^{15}N Leu (black)) PaxB T_4-^{C}DD samples.



Supporting Fig. 4. (a, b) ITC thermograms and the derived binding curves for titration between PaxB T_4 α-helix 4-^{C}DD:PaxC ^{N}DD and PaxB T_4-^{C}DD:PaxC ^{N}DD-C_5 are shown.

Curriculum Vitae

Personal Information

Name	Jonas Watzel
Date and place of birth	[redacted]
Nationality	
Contact	

Education

Goethe University Frankfurt

11/2016 – today	PhD in Biology Group of Prof. Helge B. Bode
09/2014 – 09/2016	Master of Science, Molecular Biotechnology (final grade: 1.1)

University of Applied Sciences Biberach

03/2011 – 08/2014	Bachelor of Science, Pharmaceutical Biotechnology (final grade: 1.5)

Bischöfliches Willigis-Gymnasium Mainz

08/2002 – 03/2011	Allgemeine Hochschulreife (final grade: 1.2)

Work Experience

09/2015 – 11/2015	**Internship at Sanofi** Dr. Udo Roth Bioanalytics & Formulation Frankfurt, Germany
07/2014 – 08/2014	**Working Student at Boehringer Ingelheim** Dr. Benedikt Greulich Biopharma Process Development & Strategy Biberach, Germany
02/2014 – 05/2014	**Bachelor Thesis at Boehringer Ingelheim** Dr. Benedikt Greulich Biopharma Process Development & Strategy Biberach, Germany
08/2013 – 02/2014	**Practical Semester at Boehringer Ingelheim** Dr. Benedikt Greulich Biopharma Process Development & Strategy Biberach, Germany

List of Publications and Manuscripts

A new docking domain type in the Peptide-Antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*

Jonas Watzel, Carolin Hacker, Elke Duchardt-Ferner, Helge B. Bode, Jens Wöhnert
ACS Chem. Biol. **2020**, *15*, 4, 982–989. 10.1021/acschembio.9b01022

Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases

Kenan A. J. Bozhueyuek*, Jonas Watzel*, Nadya Abbood*, Helge B. Bode
in revision 10.1101/2020.05.06.080655
* these authors contributed equally

NMR resonance assignments for a docking domain pair with an attached thiolation domain from the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii*

Jonas Watzel, Sepas Sarawi, Elke Duchardt-Ferner, Helge B. Bode, Jens Wöhnert
Biomol. NMR Assignments. **2021**. 10.1007/s12104-021-10010-1

Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS

Jonas Watzel, Elke Duchardt-Ferner, Sepas Sarawi, Helge B. Bode, Jens Wöhnert
submitted

Record of Conferences

MegaSyn Symposium, Bad Nauheim, Germany (29.09.2020 – 01.10.2020)

Poster presentation: T domain moonlight as docking domain: NMR solution structure of an extended docking domain interface in the PAX peptide-producing NRPS from *Xenorhabdus cabanillasii* (winner of the poster award)

Protein Engineering Congress EU, Frankfurt am Main, Germany (12.11.2019 – 13.11.2019)

Poster presentation: Plug-and-play with non-ribosomal peptide synthetases

3rd European Conference on Natural Products, Frankfurt am Main, Germany (02.09.2018 – 05.09.2018)

Poster presentation: Structure-function analysis of a new docking domain class identified in the PAX peptide-producing NRPS from *Xenorhabdus bovienii*

International VAAM Workshop on Biology of Bacteria Producing Natural Products, Frankfurt am Main, Germany (31.08 – 02.09.2018)

Oral presentation: Structure-function analysis of a new docking domain class identified in the PAX peptide-producing NRPS from *Xenorhabdus bovienii*

Erklärung und eidesstattliche Versicherung

Erklärung

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

Ort/ Datum

Jonas Watzel

Eidesstattliche Versicherung

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

Non-ribosomal peptide synthetase docking domains:
structure, function and engineering strategies

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

Jonas Watzel