

## Peptide Engineering

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 17531–17538

International Edition: doi.org/10.1002/anie.202102859

German Edition: doi.org/10.1002/ange.202102859

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

Kenan A. J. Bozhueyuek<sup>+</sup>, Jonas Watzel<sup>+</sup>, Nadya Abbood<sup>+</sup>, and Helge B. Bode\*

**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 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<sup>-1</sup>. Extending from the recently introduced definition of eXchange Units (XUs), we inserted synthetic zippers (SZs) to split single protein NRPSs into 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.

## 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<sup>[1]</sup> 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).<sup>[2]</sup>

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.<sup>[3]</sup> 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 thioesterase (TE) domain, usually responsible for the release of linear, cyclic or branched cyclic peptides.<sup>[4]</sup>

Type A NRPSs (Figure 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.<sup>[5]</sup> Whereas in *in cis* type A NRPSs all modules are arranged on a single polypeptide chain (e.g. ACV synthetase<sup>[6]</sup>), *in trans* assembly lines comprise a number of individual proteins (daptomycin synthetase<sup>[7]</sup>). 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).<sup>[8]</sup> DDs typically are located in between two modules and only interact with weak affinities (4–25 μM),<sup>[9–13]</sup> but are crucial to ensure biosynthesis of the desired product(s).<sup>[8,11,14,15]</sup> 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.<sup>[11,16,17]</sup>

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.<sup>[17,18]</sup> 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 eXchange Units (XUs),<sup>[19]</sup> as opposed to canonical modules (C-A-T).<sup>[20]</sup> The XU concept is based on a conserved W]-[NATE motif within the C-A linker region that can be leveraged as splicing point to functionally recombine type A NRPSs.<sup>[19]</sup> Based on previous biochemical characterizations of C domains<sup>[21,22]</sup> as well as our experimental observations,<sup>[19]</sup> a further requirement of the XU concept is compliance with the C domain

[\*] K. A. J. Bozhueyuek,<sup>[a]</sup> J. Watzel,<sup>[a]</sup> N. Abbood,<sup>[a]</sup> H. B. Bode  
Molecular Biotechnology, Institute of Molecular Biosciences,  
Goethe University Frankfurt  
60438 Frankfurt am Main (Germany)

N. Abbood,<sup>[a]</sup> H. B. Bode  
Max-Planck-Institute for Terrestrial Microbiology, Department of  
Natural Products in Organismic Interactions  
35043 Marburg (Germany)

H. B. Bode  
Senckenberg Gesellschaft für Naturforschung  
60325 Frankfurt am Main (Germany)

[+] These authors contributed equally to this work.

[\*\*] A previous version of this manuscript has been deposited on a preprint server (<https://doi.org/10.1101/2020.05.06.080655>).

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/anie.202102859>.

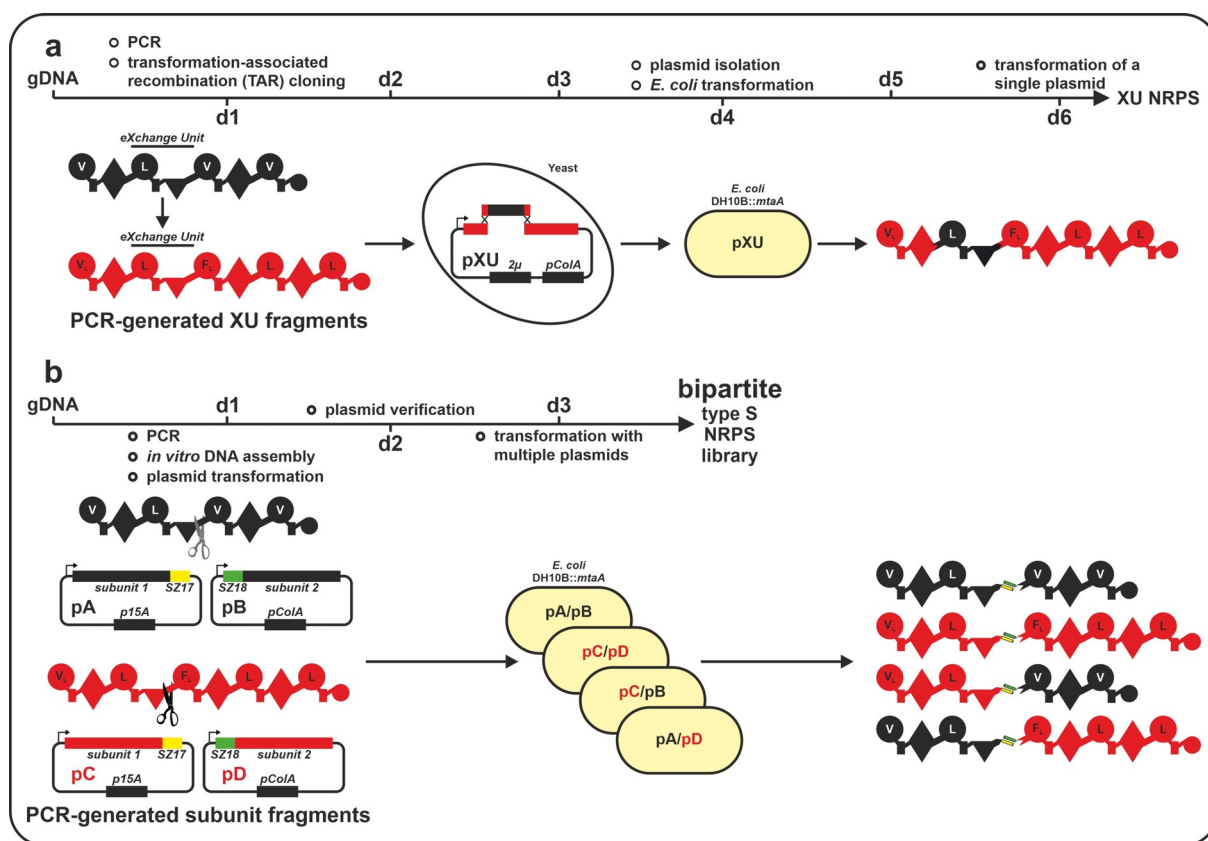
© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

specificity rule. Two XU building blocks should only be fused if the attributed C domains acceptor-site specificity of the upstream XU ( $XU_n$ ) matches the substrate specificity of the downstream A domain ( $XU_{n+1}$ ).<sup>[19,22]</sup> Although recent studies have questioned this gatekeeping function of the C domain,<sup>[23,24]</sup> by following these rules it was possible to covalently combine XUs from 15 different NRPSs *in cis* to reconstitute natural and generate new peptides *de novo* with unprecedented efficiency.<sup>[19]</sup>

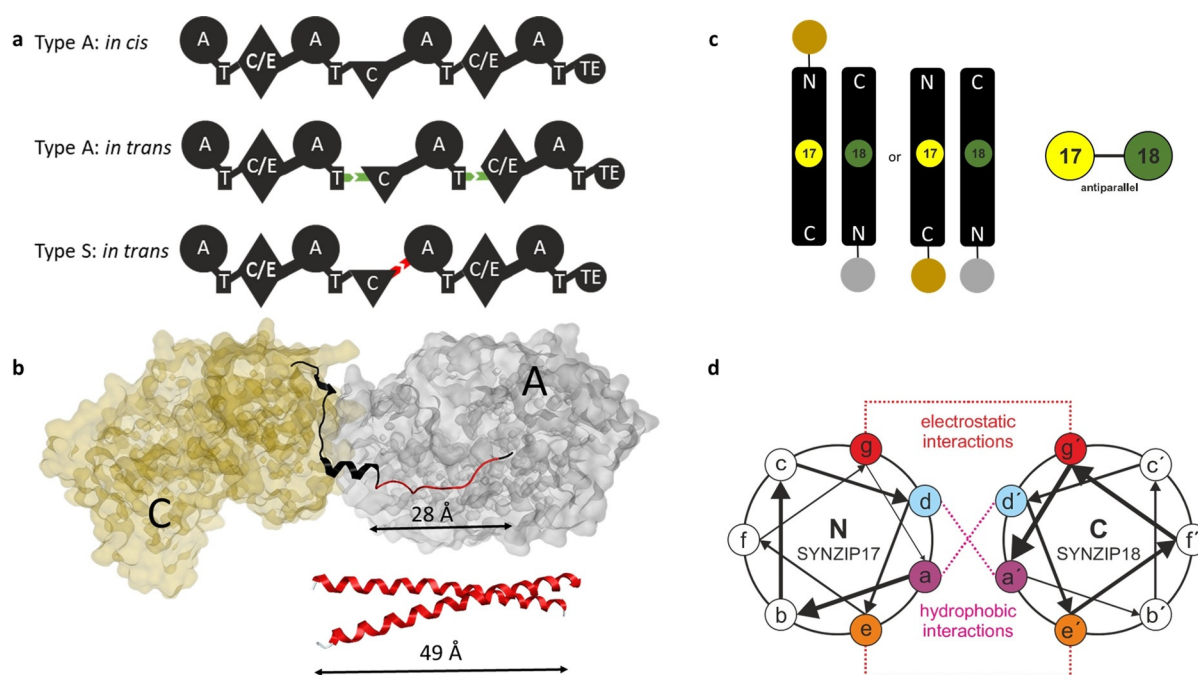
However, cloning and engineering modular enzymatic assembly lines, due to their mere size (up to 1.81 million Da (kolossin-producing synthetase (KolS)<sup>[25]</sup>) and conserved repetitive sequence stretches, naturally comes with a series of difficulties. Up to now, our engineering efforts and success in generating chimeric single protein type A NRPSs relied on transformation associated homologous recombination (TAR) based yeast cloning strategies<sup>[19,26,27]</sup> (Figure 1a). Although TAR cloning even enabled us to generate NRP libraries by applying the recently introduced eXchange Unit Condensation domain concept,<sup>[27]</sup> which applies a fusion site within the linker region of the pseudo-dimeric C domain, compared to *in vitro* cloning methods (e.g. Gibson, HiFi, Hot Fusion cloning) we were facing a much longer and arduous workflow (Figure 1).

In our efforts to simplify the NRPS engineering workflow, we came across synthetic zippers (SZs).<sup>[28]</sup> SZs are computationally calculated and designed synthetic peptides that interact with high affinity ( $K_D < 10$  nM) via a coiled-coil structural motif, enabling the specific association of two proteins.<sup>[29]</sup> Coiled coils generally can fold into two, three or four 20–50 AA long amphipathic  $\alpha$ -helices and oligomerize into homo- or heterodimers. This structural motif is found in many proteins such as in human bZIP transcription factors.<sup>[30]</sup> They are characterized by a repeating seven residue pattern, denoted as (abcdefg)<sub>n</sub>, with predominantly hydrophobic AAs at positions a and d and hydrophilic AAs at positions e and g, which contribute to tight interface formation. Previous work characterized the specificity, orientation, affinity and oligomerization state of 27 SZs, summarized in a specification sheet that is now available as a ready to use synthetic biology tool kit<sup>[28,29]</sup> for almost unlimited combinations within linear or orthogonal networks.<sup>[28]</sup>

Herein, we explored the ability of SZs<sup>[28]</sup> to manipulate collinear type A NRPSs by introducing artificial *in trans* regulation, as it was recently also shown for polyketide synthases.<sup>[31]</sup> 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<sup>[28]</sup> (Figure 2a),



**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. 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) Overview of type A and type S NRPSs. In comparison to natural type A NRPS subunits interacting *in trans* via natural docking domains (DDs, green), the *in trans* interacting NRPS subunits in type S NRPSs are linked via synthetic zippers (SZs, red). b) Top: excised C (gold)—A (grey) di-domain and linker region (ribbon representation) from the SrfA-C termination module (PDB-ID: 2VSQ). As known from the previously published eXchange Unit (XU) concept the conserved WJ-[NATE motif within the C-A linker region is an ideal intersection point to covalently recombine (*in cis*) canonical A-T-C subunits. To meet the native domain distance criteria by the *in trans* connection of eXchange Units via the introduction of SZs in the C-A linker region, a 28 Å spanning stretch of amino acids (highlighted in red) has to be removed (XtpS: XU 2-PQQPVAVIDILSSTERTLLLKTWJ-[NATETVYPES-XU 3). Bottom: Model of 41 AAs comprising SZ17:18 pair showing its size in relation to the C-A linker region. c) Absolute orientations of the non-covalently connected proteins of interest (grey and brown circles) by the application of the antiparallel interacting SZ17:18 pair. The proteins of interests are either positioned on opposite or similar sites of the antiparallel SZ complex. d) Helical wheel diagram of the coiled coil formed by  $\alpha$ -helical SZs highlighting the characteristic interactions between hydrophobic (purple (a), blue (d)) and polar (orange (e), red (g)) amino acids in the 7-residue (heptad, a<sup>(i)</sup>-g<sup>(i)</sup>) repeats.

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 (Figure 1b).

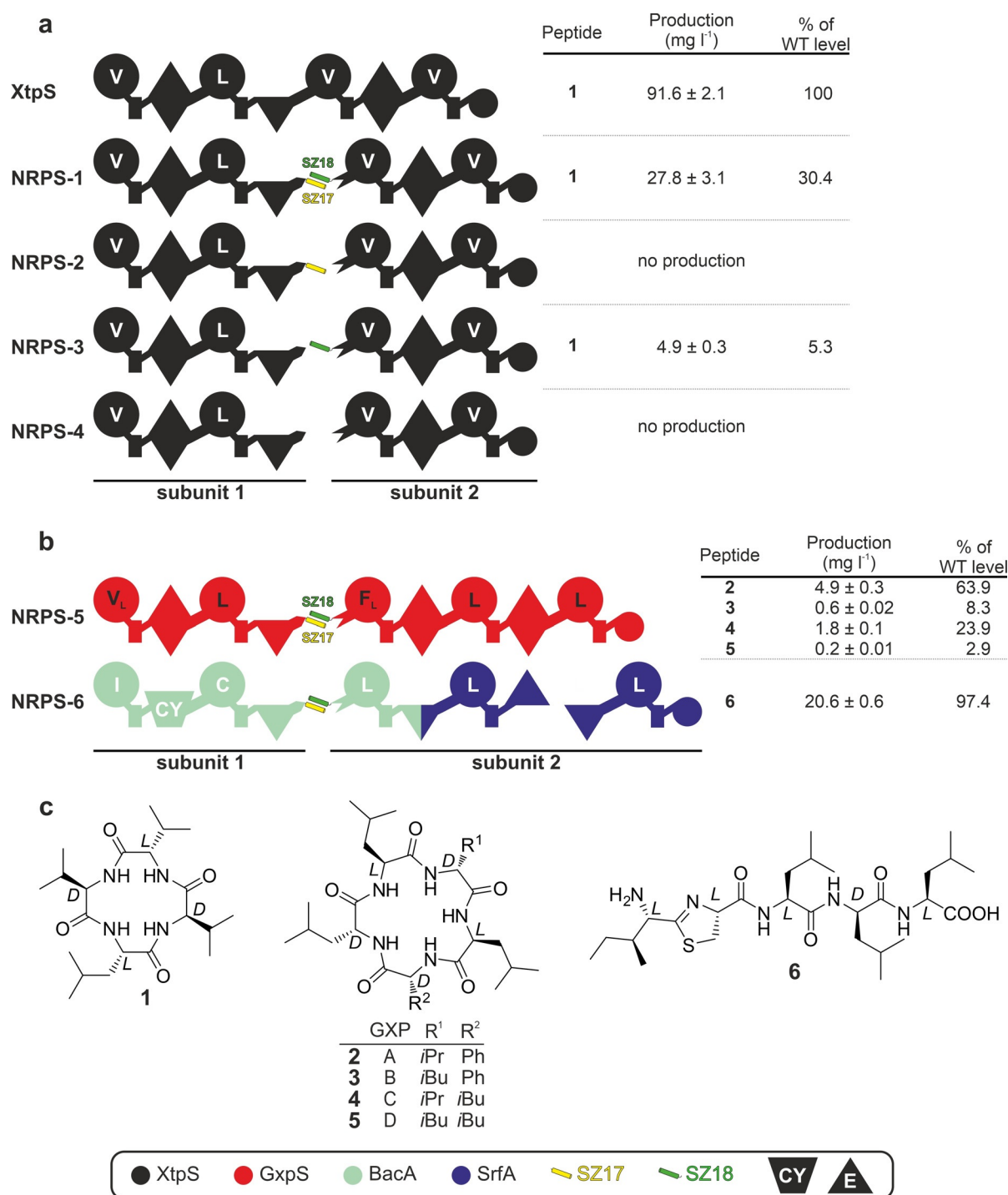
## Results and Discussion

Seeking to overcome present limitations of mega-synthase cloning and bioengineering, we explored possibilities to reduce the complexity of targeted BGCs, via functionally splitting them into separately co-expressible subunits (Figure 1b). Extending from the recently published XU concept<sup>[19]</sup> (Figure 1), we decided to split NRPSs into two subunits between consecutive XUs at the previously defined WJ-[NATE motif of the conformationally flexible C-A linker<sup>[19,32,33]</sup> region (Figure 1b & Figure 2). As already known,<sup>[19,32]</sup> this splicing position bears several advantages.<sup>[32,33]</sup> 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 (Figure 2b). With the aim of keeping the introduced steric hindrance as minimal as possible,<sup>[32,33]</sup> we chose the shortest readily available anti-parallel interacting SZ pair 17 & 18<sup>[28]</sup> (Figure 2c & d), and also removed  $\approx 10$  AAs from the unstructured N-terminus of resulting subunits 2, carrying the modules and domains downstream of the splicing position (Figure 2b).<sup>[27,28]</sup>

### 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; Figure S1) from the Gram-negative entomopathogenic bacterium *Xenorhabdus nematophila* HGB081.<sup>[34]</sup> We split XtpS into two subunits in between XUs 2 and 3. Four artificial two component type S NRPSs (Figure 3a) were constructed and heterologously produced in *E. coli* DH10B::*mtaA*<sup>[35]</sup>—either with SZs fused to both subunits (NRPS-1: subunit 1-SZ17, SZ18-subunit 2); only fused to subunit 1 (NRPS-2: subunit 1-



**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 Figure 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.

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  $\approx 30\%$  ( $28 \text{ mg L}^{-1}$ ) yield compared to WT XtpS (Figure 3a, Figure 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 much lower yields of **1**. Despite lacking SZ17, the C-terminus of XtpS subunit 1 might form a leucine-rich  $\alpha$ -helical structure (cf. Figure 2b & PDB-ID: 2VSQ) that could interact with SZ18 of subunit 2 and mediate an impaired but catalytically active C-A interface.<sup>[19,32,33,36]</sup>

Additionally, SZ17:18 were used to split the GameXPepptide A–D (2–5)-producing NRPS (GxpS,<sup>[37,38]</sup> Figure 3b: NRPS-5, Figure S3) and the recombinant thiazole-peptide (6)-producing NRPS (RtpS,<sup>[27]</sup> Figure 3b: NRPS-6, Figure S4). Whereas GxpS originates from the Gram-negative bacterium *Photorhabdus luminescens* TT01, RtpS was constructed previously<sup>[27]</sup> from building blocks of Gram-positive origin (using NRPSs for the production of bacitracin<sup>[39]</sup> and surfactin<sup>[40]</sup>). Both resulting type S NRPSs (Figure 3b) showed good to very good titres of the desired peptides. NRPS-5 produced **2** (Figure S3) with yields of  $\approx 64\%$  ( $4.9 \text{ mg L}^{-1}$ ) compared to WT GxpS and NRPS-6 produced **6** (Figure S4) at WT RtpS level ( $\approx 20 \text{ mg L}^{-1}$ ).

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 (see supplementary information).

### 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 arise: (Q1) Can type S NRPSs also be efficiently used for NRPS bioengineering purposes to produce natural product like peptide derivatives, as we have shown in previous work for the XU concept,<sup>[19]</sup> and (Q2) how do SZs impact peptide production compared to covalently fused NRPSs?

To answer these questions, we generated and co-expressed four sets of synthetic NRPSs (Figure 4a, NRPS-7 – -14) by recombining building blocks from XtpS,<sup>[34]</sup> GxpS,<sup>[37,38]</sup> RtpS,<sup>[27]</sup> and the szentiamide<sup>[41]</sup>-producing synthetase (SzeS<sup>[42]</sup>), analysed the culture extracts by HPLC-MS/MS and compared the resulting peptide titres. 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,<sup>[19]</sup> -14), constructed according to the XU approach and largely in line with the XUs' specificity rule to prevent potential substrate specificity issues<sup>[19,22]</sup> at the upstream C domains' acceptor site.

In sum, six out of eight recombinant NRPSs were functional (NRPS-7 – -12; Figure 4a). Whereas type S NRPS-7 showed moderately decreased ( $\approx 47\%$ ) 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** (Figure 4d, Figure 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 a result of the XtpS TE domain substrate specificity (cf. Figure 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,

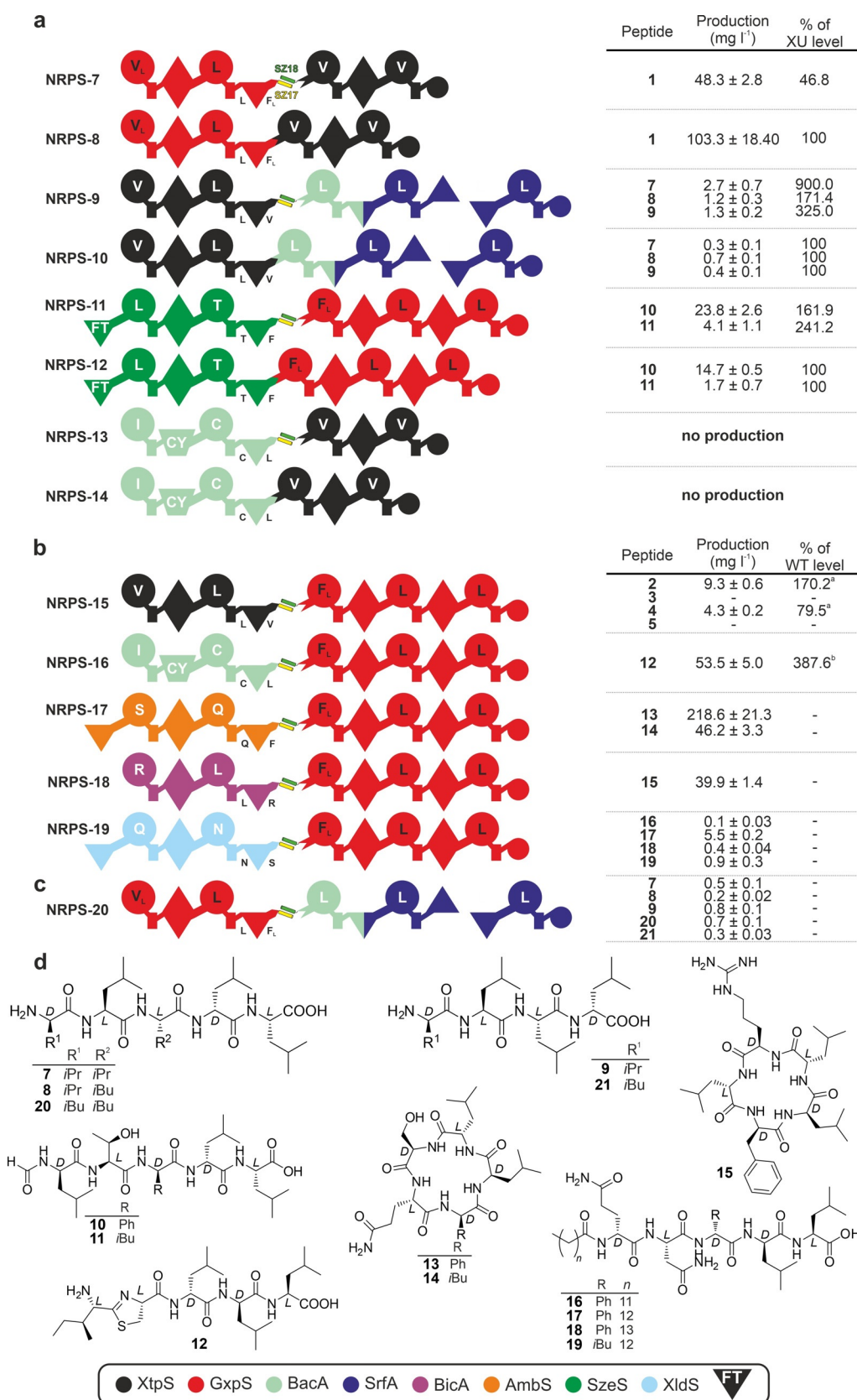
NRPS-9 and -10, are producing peptides **7–9** and are composed of building blocks from Gram-negative (*X. nematophila* HGB081) and -positive (*B. licheniformis* ATCC 10716, *B. subtilis* ATCC 21332) origin. Whereas NRPS-10 exhibits the typical impaired biosynthesis when XUs from Gram-negative and -positive origin are combined,<sup>[19,27]</sup> i.e., synthesising **7** (linear  $\nu\text{LLL}$ ; *D*-AAs in italics and lowercase throughout this work) in yields of  $0.3 \text{ mg L}^{-1}$ , type S NRPS-9 shows a 9-fold increased titre of **7** ( $2.7 \text{ mg L}^{-1}$ ), indicating that observed impairments might be caused at the level of translation, rather than the proteins themselves. 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.

### Reuse of NRPS-Encoding Plasmids Leads to Rapid Generation of Synthetic NRPSs

Type S NRPSs not only open the possibility to convert single protein type A NRPSs into separately expressible subunits (Figure 1b, Figure 2, & Figure 3), but also to the reuse (Figure 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 NRPSs 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 (Figure 1a).

To showcase possible applications, in a first step we created three additional type S subunits 1 (Figure 4b, NRPS-17 – -19) from XUs 1–2 of NRPSs producing ambactin (AmbS,<sup>[35]</sup> *X. miraniensis* DSM 17902), bicornutin (BicA,<sup>[43]</sup> *X. budapestensis* DSM 16342), and xenolindicin (XldS,<sup>[35]</sup> *X. indica* DSM 17382). These subunits 1 (Figure 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 (Figure 4a: NRPS-7, -9, -11, -13; Figure 4b: NRPS-15 – -19; Figure 4c: NRPS-20; Figure S17: NRPS-21 – -28).

In brief, 9 out of 18 type S NRPSs showed catalytic activity (Figure 4a & b, Figure S17), synthesising 21 different linear- (**7–12**, **16–21**, Figure S7, S9, S12, S15, S16), cyclic- (**1–5**, **13–15**, Figure S5, S11, S13, S14), lipo- (**16–19**, Figure S15), formyl- (**10**, **11**, Figure S9), and thiazoline (**12**, Figure S12) containing peptides (**1–21**; Figure 3c & 4c) in yields ranging from  $\approx 0.1$ – $220 \text{ mg L}^{-1}$ . From these results it immediately became apparent that a large number of chimeric NRPSs and associated NRPSs could be generated in no time, and with only a minimum of necessary wet lab work. Moreover, from a detailed analysis



**Figure 4.** Bio-combinatoric potential of bipartite type S NRPS. a) Production yields of type S NRPSs compared to analogous recombinant covalently fused NRPSs (NRPS-7 to NRPS-14). b/c) Further type S NRPSs, each generated by co-expressing two individual subunits (NRPS-15–NRPS-20). For NRPS-15 & -16 production levels were compared to heterologously expressed WT BGCs of GxpS (<sup>a</sup>) and RtpS (<sup>b</sup>), respectively. For additional type S NRPS including non-functional examples see Figure S17. d) The structures of 7–21 produced from NRPS-7 to NRPS-20 expressed in *E. coli*. See Figure 1 and 3 for assignment of the domain symbols; further symbols: FT, formyl-transferase domain. The native C domains' donor and acceptor site substrate specificity is given below selected C domains.

of the functional and non-functional NRPS systems in Figure 4 and Figure S17, the following can be concluded:

All type S NRPSs sharing the same subunit 2 originating from GxpS (NRPS-5) were functional (Figure 4a: NRPS-11, Figure 4b: NRPS-15 – -19, Figure 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 (Figure 4b), produced peptides **15** ( $\approx 40 \text{ mg L}^{-1}$ , Figure S14) and **16–19** ( $0.1\text{--}5.5 \text{ mg L}^{-1}$ , Figure 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.<sup>[35]</sup> 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.<sup>[27]</sup> 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.<sup>[44]</sup> 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.<sup>[23]</sup> 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\**l*IL (**12**;  $\approx 53 \text{ mg L}^{-1}$ ; Figure S12). In its natural NRPS context as well as *in vitro*, the A3 domain of GxpS prefers Phe over Leu.<sup>[27]</sup> 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's specificity. Similar effects of engineered NRPSs, exhibiting chimeric C-A interfaces<sup>[19,44]</sup> or C domains,<sup>[27]</sup> 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 (Figure S17). In light of all co-expression experiments (Figure 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 (Figure 4a) confirms 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.

Reusing type S subunits also allows the functional co-expression of building blocks from Gram-negative and -positive origin. Albeit only 3 (NRPS-9, -16, -20) out of 8 combinations (Figure S17) showed catalytic activity, type S NRPSs 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  $54 \text{ mg L}^{-1}$  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 (**22/23** and **24/25**) at high titre up to  $86 \text{ mg L}^{-1}$  related to the unpaired activity of GxpS subunit 2 (Figure S18). Due to the promiscuous GxpS A3 domain, **22/23** and **24/25** differ from each other at position one, either carrying Phenylalanine or Leucine. In addition, **22** and **24** show a *D-D-L* configuration, whereas **23** and **25** 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.<sup>[45]</sup>

### 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.<sup>[31]</sup> Here we reported the use of high-affinity SZs to split native single protein NRPSs into two individual subunits by introducing them at the previously described fusion point of the XU concept.<sup>[19]</sup> Thus, artificially *in trans* regulated assembly lines were generated not only representing 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. Having SZs at hand, peptide libraries quickly can be constructed with high success rates even when based on NRPSs originating from different strains or phyla. It might also be conceivable to combine different biosynthetic pathways co-occurring in one strain *in situ*, by introducing SZs at the genomic level. 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.

Finally, we believe that SZs will greatly accelerate NRPS research—not only for engineering purposes, but also for the *in vivo* characterization of domain and module specificities (e.g., focusing on C and TE domains). 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). Open access funding enabled and organized by Projekt DEAL.

## Conflict of interest

Goethe University has filed a patent on this NRPS engineering technology. The patent is currently pending.

**Keywords:** artificial docking domains · biosynthesis · engineering · non-ribosomal peptide syntheses · novel natural products

- [1] R. D. Süßmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821; *Angew. Chem.* **2017**, *129*, 3824–3878.
- [2] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490–504.
- [3] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715–738.
- [4] F. Kopp, M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735–749.
- [5] K. A. Bozhüyük, J. Micklefield, B. Wilkinson, *Curr. Opin. Microbiol.* **2019**, *51*, 88–96.
- [6] M. F. Byford, J. E. Baldwin, C.-Y. Shiao, C. J. Schofield, *Chem. Rev.* **1997**, *97*, 2631–2650.
- [7] R. H. Baltz, V. Miao, S. K. Wrigley, *Nat. Prod. Rep.* **2005**, *22*, 717–741.
- [8] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15585–15590.
- [9] 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*, 1340–1351.
- [10] 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*, 4155–4167.
- [11] C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert, *Nat. Commun.* **2018**, *9*, 4366.
- [12] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, D. H. Sherman, *ACS Chem. Biol.* **2009**, *4*, 41–52.
- [13] J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *ACS Chem. Biol.* **2020**, *15*, 982–989.
- [14] C. Kegler, H. B. Bode, *Angew. Chem. Int. Ed.* **2020**, *59*, 13463–13467; *Angew. Chem.* **2020**, *132*, 13565–13569.
- [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*, 913–923.
- [16] X. Cai, S. Nowak, F. Wesche, I. Bischoff, M. Kaiser, R. Fürst, H. B. Bode, *Nat. Chem.* **2017**, *9*, 379–386.
- [17] X. Cai, L. Zhao, H. B. Bode, *Org. Lett.* **2019**, *21*, 2116–2120.
- [18] C. Beck, J. F. G. Garzón, T. Weber, *Biotechnol. Bioprocess Eng.* **2020**, *25*, 886–894.
- [19] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275–281.
- [20] H. D. Mootz, M. A. Marahiel, *Curr. Opin. Biotechnol.* **1999**, *10*, 341–348.
- [21] P. J. Belshaw, C. T. Walsh, T. Stachelhaus, *Science* **1999**, *284*, 486–489.
- [22] U. Linne, M. A. Marahiel, *Biochemistry* **2000**, *39*, 10439–10447.
- [23] M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nat. Commun.* **2020**, *11*, 4554.
- [24] M. Baunach, S. Chowdhury, P. Stallforth, E. Dittmann, *Mol. Biol. Evol.* **2021**, *38*, 2116–2130.
- [25] 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., *Angew. Chem. Int. Ed.* **2015**, *54*, 10352–10355; *Angew. Chem.* **2015**, *127*, 10493–10496.
- [26] R. D. Gietz, R. H. Schiestl, *Nat. Protoc.* **2007**, *2*, 31–34.
- [27] 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.
- [28] K. E. Thompson, C. J. Bashor, W. A. Lim, A. E. Keating, *ACS Synth. Biol.* **2012**, *1*, 118–129.
- [29] A. W. Reinke, R. A. Grant, A. E. Keating, *J. Am. Chem. Soc.* **2010**, *132*, 6025–6031.
- [30] K. J. Lumb, C. M. Carr, P. S. Kim, *Biochemistry* **1994**, *33*, 7361–7367.
- [31] M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla, M. Grninger, *ACS Chem. Biol.* **2019**, *14*, 426–433.
- [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*, 235–238.
- [33] A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science* **2008**, *321*, 659–663.
- [34] C. Kegler, F. I. Nollmann, T. Ahrendt, F. Fleischhacker, E. Bode, H. B. Bode, *ChemBioChem* **2014**, *15*, 826–828.
- [35] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290–1294.
- [36] K. Bloudoff, T. M. Schmeing, *Biochim. Biophys. Acta Proteins Proteomics* **2017**, *1865*, 1587–1604.
- [37] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün, *Chem. Eur. J.* **2012**, *18*, 2342–2348.
- [38] F. I. Nollmann, C. Dauth, G. Mulley, C. Kegler, M. Kaiser, N. R. Waterfield, H. B. Bode, *ChemBioChem* **2015**, *16*, 205–208.
- [39] D. Konz, A. Klens, K. Schörgendorfer, M. A. Marahiel, *Chem. Biol.* **1997**, *4*, 927–937.
- [40] P. Cosmina, F. Rodriguez, F. de Ferra, G. Grandi, M. Perego, G. Venema, D. van Sinderen, *Mol. Microbiol.* **1993**, *8*, 821–831.
- [41] B. Ohlendorf, S. Simon, J. Wiese, J. F. Imhoff, *Nat. Prod. Commun.* **2011**, *6*, 1247–1250.
- [42] 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; *Angew. Chem.* **2019**, *131*, 19133–19139.
- [43] S. W. Fuchs, C. C. Sachs, C. Kegler, F. I. Nollmann, M. Karas, H. B. Bode, *Anal. Chem.* **2012**, *84*, 6948–6955.
- [44] S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süßmuth, E. Dittmann, *Cell Chem. Biol.* **2016**, *23*, 462–471.
- [45] M. Kaniusaitė, R. J. A. Goode, J. Tailhades, R. B. Schittenhelm, M. J. Cryle, *Chem. Sci.* **2020**, *11*, 9443–9458.

Manuscript received: February 25, 2021

Revised manuscript received: March 31, 2021

Accepted manuscript online: May 20, 2021

Version of record online: June 27, 2021