

Type S Non Ribosomal Peptide Synthetases for the rapid generation of tailor-made peptide libraries

Nadya Abboud^{1,2}, Tien Duy Vo², Jonas Watzel², Kenan A. J. Bozhueyuek^{1,2*} and Helge B. Bode^{1,2,3*}

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

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

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

* Corresponding author

Abstract

Bacterial natural products in general, and non-ribosomally synthesized peptides in particular, are structurally diverse and provide us with a broad range of pharmaceutically relevant bioactivities. Yet, traditional natural product research suffers from rediscovering the same scaffolds and has been stigmatized as inefficient, time-, labour- and cost-intensive. Combinatorial chemistry, on the other hand, can produce new molecules in greater numbers, cheaper and in less time than traditional natural product discovery, but also fails to meet current medical needs due to the limited biologically relevant chemical space that can be addressed. Consequently, methods for the high throughput generation of new-to-nature natural products would offer a new approach to identifying novel bioactive chemical entities for the hit to lead phase of drug discovery programmes. As a follow-up to our previously published proof-of-principle study on generating bipartite type S non-ribosomal peptide synthetases (NRPSs), we now envisaged the *de novo* generation of non-ribosomal peptides (NRPs) on an unreached scale. Using synthetic zippers, we split NRPS in up to three subunits and rapidly generated different bi- and tripartite NRPS libraries to produce 49 peptides, peptide derivatives, and *de novo* peptides at good titres up to 145 mgL⁻¹. A further advantage of type S NRPSs not only is the possibility to easily expand the created libraries by re-using previously created type S NRPS, but that functions of individual domains as well as domain-domain interactions can be studied and assigned rapidly.

Introduction

Natural products (NPs) have been used throughout the ages for the treatment of a wide range of medical conditions¹ and still continue to be of particular importance in drug development today²⁻⁴. Especially bacterial NPs derived from modular megasynth(et)ases, such as non ribosomal peptides (NRPs) and polyketides (PKs), made a major contribution to modern pharmacotherapy, inter alia, for tackling infectious diseases and cancer⁵. Nevertheless, although NPs are structurally diverse and bioactive with advantageous properties beyond Lipinski's rule of five⁶, like higher molecular mass and a greater molecular rigidity which can be valuable in tackling protein-protein interactions⁷, they also pose challenges for drug discovery. These challenges are mainly due to technical barriers in the identification, characterisation, isolation, screening and optimisation of natural products, which remain time, labour and cost intensive^{2, 8}. As a result, and due to the lack of adequate solutions, the pharmaceutical industry withdrew from traditional natural product research. With the rapid emergence of antimicrobial resistances (AMRs)^{2-4, 9-11} and recent technological advances addressing the challenges, such as advances in cultivation¹², DNA sequencing^{13, 14, 15}, bioinformatics¹⁶, and synthetic biology¹⁷ interest in natural product research has been reignited⁸.

Despite the progress made, the complexity of NP structures make it difficult to generate synthetic NP derivatives, for example to explore structure-activity relationships and develop hits to leads. Thus, many clinical derivatives have been created by means of semi-synthesis⁴, i.e. azithromycin¹⁸ and cephalosporin¹⁹. Due to technical and chemical limitations, such modifications are often limited to a few synthetically accessible functional groups, leaving the backbone structure untouched. A commonly stressed solution to this problem is bioengineering, as it provides access to a wider range of structural diversity beyond the limitations of synthetic chemistry.^{20, 21} But as rational reprogramming efforts have been met with limited success, progress in the synthetic biology of NPs is of great importance. In this and previous work, we have therefore focused on developing tools enabling the reproducible, rapid, and simple genetic manipulation of biosynthetic gene clusters (BGCs) for the biosynthesis of NRP derivatives and even new-to-nature peptides.²²⁻²⁴

At the heart of our research are BGCs, encoding multi-functional enzymatic protein machines that enable the biosynthesis of peptides independently of the ribosome. These machines, denoted as non-ribosomal peptide synthetases (NRPSs) are in fact assembly lines, and are, *inter alia*, responsible for the synthesis of most antibiotic drug scaffolds in current clinical use, such as penicillin G, vancomycin, and daptomycin.²⁵ In a NRPS assembly line, multiple repeating modules are responsible for selection, activation, programmed functional group modifications, and coupling of an amino acid to the growing peptide chain. An archetypal minimal module consists of three core domains: an adenylation (A) domain, which selects and activates a substrate, a thiolation (T) domain, onto which the activated amino acid is covalently attached to, and a condensation (C) domain, which catalyses peptide bond formation of the bound amino acid to the growing peptide chain. Additionally, several optional *in cis* or *in trans* acting modification domains can be present, introducing structurally complex motives into the peptide chain, for instance epimerisation, methylation, hydroxylation, and glycosylation patterns.²⁵

Most recently, we developed a novel *in trans* acting synthetic type of NRPSs (type S)²² with reduced structural complexity compared to wildtype (WT) *in cis* acting type A NRPSs. Type S NRPSs are characterised by ‘small’ individually expressible chimeric NRPS protein subunits with attached synthetic leucine zippers, referred to as SYNZIPs (SZs)²⁶.²⁷ Type S subunits functionally can be co-expressed and are quickly interchangeable to generate new assembly lines and peptide derivatives, respectively, in no time and with only a minimum of lab work involved. We were able to showcase how type S NRPSs can be created via splitting one protein NRPSs into two individually expressible proteins (subunits) in between eXchange Unit (XU) building blocks (A-T-C tri-domain units)²² by leveraging the established splicing position within the C-A di-domain linker region (W)[NATE]²³ to introduce SZs. Due to the high bio-combinatorial potential of type S NRPSs and the possibility to reuse formerly cloned type S subunits in new combinations, they bear the great chance to accelerate NRPS research and NRP based early drug discovery efforts.

Here we show the potential of type S NRPSs beyond the limitations of the XU concept. We not only sought to demonstrate (I) the possibility to create a bi-partite type S NRP library by using the building blocks of only one single NRPS system, but (II) to introduce SZs within all possible NRPS linker regions. Eventually, (III) to further increase the bio-combinatoric potential of type S NRPSs, we aimed at dividing distinct BGCs into three individually expressible subunits to create tri-partite NRPS libraries.

Results

I. Bi-partite type S NRPS library

To create a bi-partite type S NRPS²² library from only one single starting NRPS, we chose the GameXPeptide synthetase (GxpS) from *Photorhabdus luminescens* TT01²⁸. GxpS, which is responsible for the biosynthesis of four cyclic GameXPeptides (GXP) A-D (**1-4**), was split into all possible subunits in between individual XUs. These subunits differed in the number of XUs ranging from 1-4 XUs, both from the *N*- and *C*-terminus. In detail, we split GxpS between XUs 1 & 2, 2 & 3, 3 & 4, and 4 & 5, resulting in four initiating (subunit-1a, -1b, -1c, and -1d) and four terminating building blocks (subunit 2, -2a, -2b, -2c, and -2d) (Fig. 1a). To functionally create type S NRPS systems, each individual initiating subunit was heterologously co-expressed in *E. coli* DH10B::mtaA²⁹ together with each of the terminating subunits. With this procedure, 16 unique type S NRPSs were generated from a single WT NRPS.

Notably, HPLC-MS analysis revealed that all but one type S NRPSs (Fig. 1c, NRPS-10) showed catalytic activity, producing detectable amounts of overall 34 unique linear and cyclic peptides (**1-34**, Fig. 1) of varying length at titres up to 145 mgL⁻¹ (Fig. 1). Throughout the present work, the resulting peptides (Table S1) and yields were confirmed by HPLC-MS/MS and comparison of retention times with synthetic standards (*c.f.* Supplementary Information S1-35).

An additional strength of this approach lies in the possibility to study and characterise individual domains (*i.e.*, C & TE) as well as domain-domain (*i.e.*, C-A) interface interactions with respect to their substrate specificity or compatibility. An NRPS domain-

domain interface is defined as the region of interdomain contact (e.g., A-T, T-C, C-T), involving (mostly hydrophobic) interface forming amino acid-residues at the respective domains solvent accessible surface. The interdomain interfaces are thought to form flexible and changing domain-domain contacts during the course of the catalytic cycle, which help the NRPS machinery to run and carry out catalytic reactions in an orchestrated manner^{25, 30}. Traditionally, such characterisations were only done *in vitro*.³¹ However, here, the presented GxpS derived NRPS set (NRPS-1 to -16) not only enables interesting conclusions concerning the compatibility of differing C-A interface types, but also to quickly deduce the GxpS_TE-domains' capacity to cyclise peptides differing in length from the wild type (WT) products.

At a glance, the chimeric set of 16 GxpS derived type S NRPSs consists of six truncated (NRPS-1 to -6, Fig. 1b), five elongated (NRPS-7 to -12 Fig. 1c), and five WT-length assembly-lines (NRPS-13 to 16, Fig. 1d). From the subset of truncated NRPSs 1 to 6 we detected linear di- (**6 & 7**), tri- (**8-10**), and tetra- (**11-20**) peptides, showing the expected range of derivatives due to the substrate promiscuity of GxpS_A1 (Val, Leu) and A3 (Phe, Leu). While cyclic dipeptides like diketopiperazines are found in nature³² or as truncated NRPs³³ but mainly are generated by cyclodipeptide synthases using tRNA-activated amino acids³⁴, the generation of tri-peptides is at least also conceivable – although we are not aware of any cyclic tri-peptides generated by NRPSs in nature but chemically synthesized examples exist³⁵. Cyclic tetra-peptides are common and numerous NRPS-based examples exist (e.g. fungisporin³⁶, tentoxin³⁷, azumamides A–E³⁸, xenotetrapeptide³⁹). However, according to our data, the GxpS-TE does not seem to be able to cyclise the synthesised tri- and tetra-peptides, suggesting that a length of five consecutive building blocks is its lower limit for cyclisation (NRPS-13 to -16).

Looking at the upper boundary, NRPS-7 to -9 were able to produce and cyclise hexapeptides (**21-25**) and NRPS-11 even synthesised cyclic hepta-peptides (**27-29**). However, the TE-domain seems to have reached its capacity to efficiently cyclise peptides at a length of 7 consecutive amino acid building blocks. Although we also detected cyclic peptides (**22-27**) from production cultures of the 8-modular NRPS-12,

these cyclic peptides are presumably autocatalytically cyclised shunt products of enzyme-bound hexa- and heptapeptides. NRPS-12 only was capable to hydrolyse the synthesised octapeptides (**30-34**), resulting in linear octapeptide derivatives (Fig. 1 d).

Overall, through this simple and quick experimental procedure, we found that the TE domain of GxpS is quite versatile, accepting a range of peptides from two to at least eight building blocks, but is only able to effectively catalyse cyclisation within a narrow range of five to seven building blocks. In turn, information about TE domains' substrate specificities and preferences, respectively, gained via generating a series of type S NRPSs will help to guide future engineering projects in identifying suitable termination domains. This is of particular interest when it comes to large scale NRPS engineering campaigns, as TE-domains that are less prone to a peptide's amino acid sequence and length are more versatile and can help to overcome TE-domain limitations of artificial NRPS systems. The approach shown is not only superior to *in vitro* characterisation in terms of workload, but also much cheaper (no SNAC peptides need to be synthesised), more robust (no spontaneous or autocatalytic side effects), scalable, can be performed in high throughput, and does not suffer from the *in vitro* bias known for excised NRPS proteins as recently described.⁴⁰

Type S NRPSs can also be used to study C domain specificities and C-A interface compatibilities. Since in the case of GxpS (NRPS-1 to -16), the respective activated and incorporated amino acids are too similar to each other to draw valuable conclusions on C domain specificities – but this was shown previously^{22, 40} – we focused on characterising the compatibility of different C-A interfaces with each other. In general, based on the particular reactions C domains catalyse (condensation or epimerisation and condensation), their donor-T domain bound substrate range (acetyl-, formyl-, D- or L-amino acids), acceptor-T domain bound substrate range (L-amino acids, amines) and from phylogenetic reconstructions, they are categorised into five different groups: ^LC_L, ^DC_L, dualC (C/E), C_{start}, and C_{term}.²⁵ However, in our example the C-A interface depends on whether there is a C or C/E domain directly upstream of the A domain resulting in C-A or C/E-A type interfaces – as evidenced by phylogentic analysis (Fig S36⁴¹).

In brief, depending on whether or not the interface type naturally occurring at a specific site has been altered, we observed major differences in the production titres between NRPSs of similar length and amino acid composition, suggesting that C-A interface types indeed play an important role when it comes to (re-)designing NRPSs. For instance, the four-modular type S NRPSs (NRPS-4 to -6, Fig. 1b) produced tetra-peptides (**11-20**) with yields varying widely, ranging from 4.1 mg/L to 144.5 mg/L. As could be expected, the best producing system, NRPS-6 (144.5 mgL⁻¹), has the same interface type (C/E-A) and an interface most similar to that of the WT C₅-A₅ interface (Identity 89.3 %, supplementary information Table S5). For NRPS-4, in which the interface type changed from C-A to C/E-A, we observed high (63 mg/L) but still significant lower yields than for NRPS-6. Eventually, expression of NRPS-4 resulted in the lowest titre (4.1 mg/L), indicating that changes from C to C/E have a greater impact on production than changes from C/E to C, since here the interface type changed from C/E-A to C-A. These observations further are supported by NRPS-2 and -3 and by NRPS-7 to -9. NRPS-2 and -3 synthesised **8**, **9** and **10**. While NRPS-2 with the same interface type as the WT produced **8** and **9** at titres of 19 mg/L and 18 mg/L, respectively (Fig. 1 b), the switch from C/E-A type to C-A type in NRPS-3 resulted in a sharp drop in production of **10** to 2.3 mg/L. Accordingly, comparing the titres of the hexa-peptide (**21-25**) producing NRPSs (NRPS-7 to -9), the WT-like C-A interface harbouring NRPS-9 showed ~40 to ~20-fold higher titres than NRPS-7 and -8, respectively (Fig. 1 c).

In combination with the most recently published extended gatekeeping function, describing the influence of C domains and the particular formed C-A interface on the catalytic activity and substrate selectivity of A domains⁴⁰, this data helps us to refine the NRPS design principles published previously even further. We assume that altering the C domain type directly upstream of an A domain of interest substantially impairs C-A di-domain contacts, resulting in reduced catalytic activity of the A domain and therefore overall productivity of the respective NRPS protein. In retrospect, this might also explain why some of our previously published recombinant NRPS systems showed reduced

221 production titres - while others showed no impairment or even increased catalytic
222 activity.^{22, 23}



Figure 1. Bi-partite GxpS Library. **a** SZ 17:18 insertion between XUs 1 & 2, 2 & 3, 3 & 4 and 4 & 5 for the generation of four initiating (subunit-1a, -1b, -1c, -1d) and four terminating building blocks (subunit 2, -2a, -2b, -2c, -2d). Generated truncated type S GxpS systems (NRPS-1-6) (**b**), elongated type S GxpS systems (NRPS-7-12) (**c**), and wild type length type S GxpS systems (NRPS-7-12) (**d**) are shown. Corresponding peptide yields (mg/L) as obtained from triplicate experiments. 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.

II. Other SZ insertion sites

To further explore and extend the applicability of SZs for the construction of type S NRPSs beyond the borders of the XU concept²³, we decided to introduce SZ17:18 within the T-C (NRPS-16) and A-T (NRPS-17) linker regions of the cyclic xenotetrapeptide (**5**; cyclo(vLvV)) producing synthetase (XtpS) from *Xenorhabdus nematophila* ATCC 19061³⁹ (Fig. S37) – and compared resulting peptide yields with WT XtpS and NRPS-15 (Fig. 2), which was constructed previously.²² Both, NRPS-16 and -17 synthesised **5** at ~86% and ~280% compared to WT XtpS and NRPS-15, respectively.

While the catalytic activity of NRPS-16 was not surprising, as the introduced SZs are mimicking natural DDs⁴², the observed good activity of NRPS-17 was unexpected. During a catalytic cycle of a module, especially the A-T interaction is considered as highly dynamic. After the adenylation reaction, the A_{sub}-domain must fulfil a torsion of 140° in respect to the A_{core}-domain such that the *holo*-T-domain can meet the distance to the activated amino acid (thiolation reaction).²⁵ Thus, it was assumed that the additional rigidity, inserted by the structured α -helical amino acid stretches of the SYNZIPs, would result in loss of function. The recently gathered structural data of large constructs of the linear gramicidin synthesising NRPS (LgrA)⁴³, 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-T⁴⁴, module skipping⁴⁵, and presumably also SZs.

More dipartite type S NRPSs (NRPS-39 to -44), split in between (T)[C] and within modules (A)[T] as well as within C domains (C_{Dsub})[C_{Asub}] are depicted in Fig. S38 and S39.

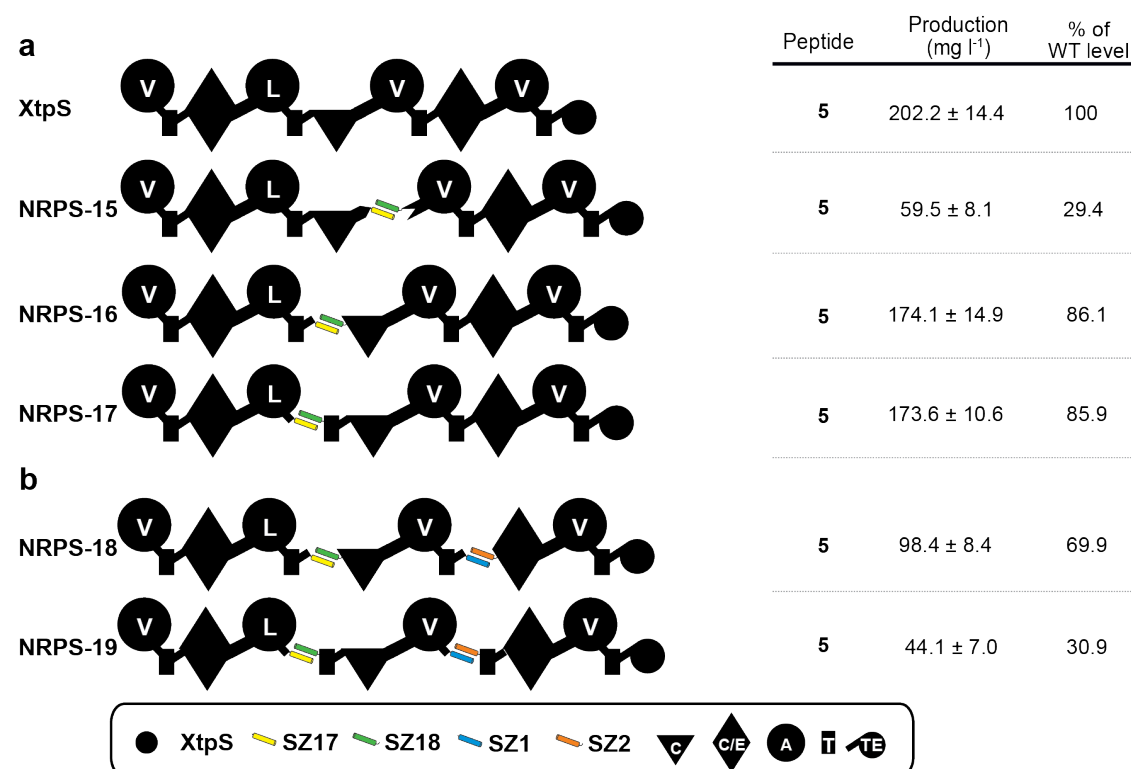


Figure 2. Other splicing positions and Tri-partite type S XtpS. **a** SZ17:18 insertion within the C-A (NRPS-17), T-C (NRPS-18) and A-T (NRPS-19) linker (**a**) XtpS split into three subunits by the insertion of the SZ1:2 pair. SZ17:18 and SZ1:2 pairs were inserted within T-C (NRPS-20) and A-T (NRPS-21) (**b**). Corresponding peptide yields (in mg/L) are determined from triplicate experiments.

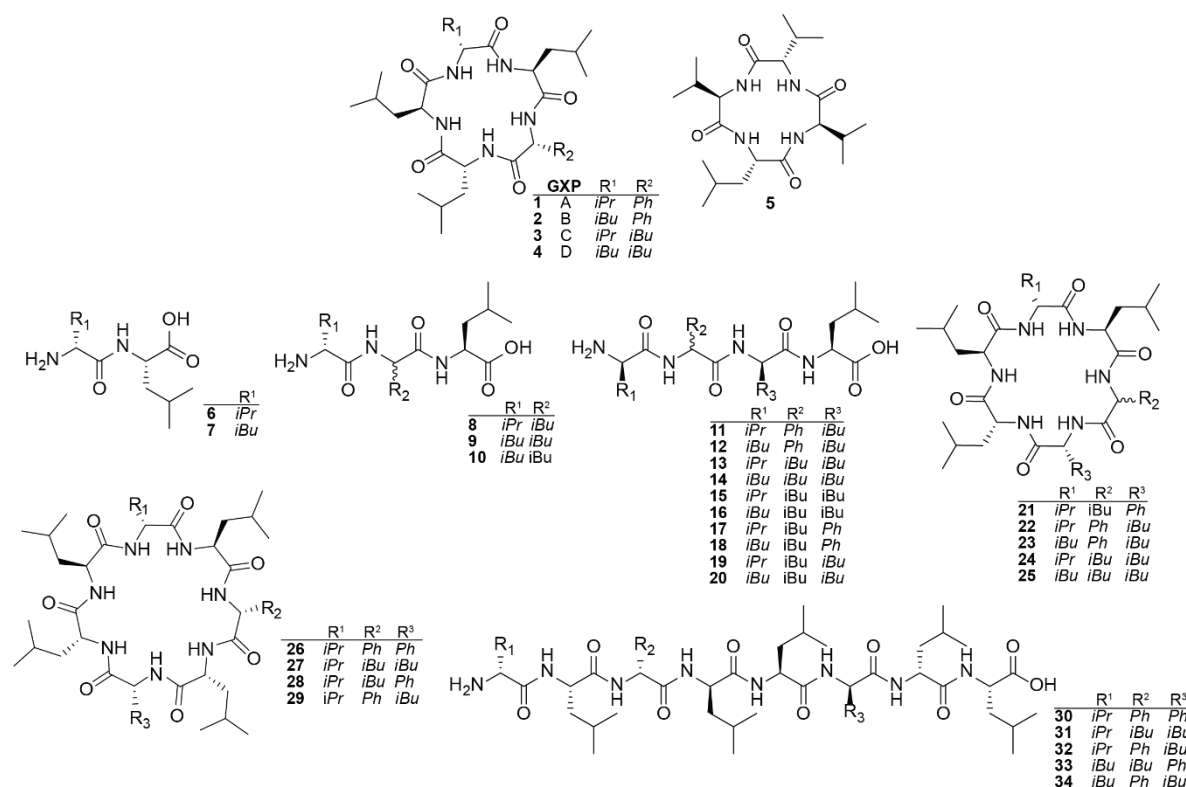


Figure 3. Structures of produced compounds. GameXPeptide A-D (1-4), xenotetrapeptide (5) and GXP derivatives (6-34) are depicted.

III. Tri-partite type S NRPS library

The potential of bi-partite type S NRPSs to generate bio-combinatoric libraries from a small set of NRPS subunits was shown previously²² and above (Fig. 1 & 2a). But, the bio-combinatoric potential could further be increased if it were possible to split NRPS systems into three or more subunits (*c.f.* Supplementary Information S40 and S41).

For a first proof of concept to functionally split NRPSs into three subunits, we again targeted XtpS and inserted, next to SZ17:18, a second SZ pair (SZ1:2) into both, NRPS-16 and -17, to build an orthogonal interaction network (Fig. 2b). The resulting tri-partite type S NRPSs -18 and -19 are split in between modules 2-3 & 3-4 (NRPS-18) and within the A-T linker regions of modules 2 & 3 (NRPS-19), respectively (Fig. 2 b). Both, NRPS-18 and -19, produced **5** with 69.9% and 30.9% compared to WT XtpS but also with

decreased yields compared to their bipartite counterparts (NRPS-16 & -17). In addition to the cumulative effect of inserted impairments, caused by a higher degree of engineering, we assume that SZ1:2 also contributed to the reduced production titre of **5** since the SZ1:2 pair is significantly longer than SZ17:18 and probably disturbs catalytic reactions of the tri-partite type S XtpSs by the inserted additional rigidity. Although NRPS-18 produced **5** at slightly higher titres than NRPS-19, in a next step we decided to use the A-T splicing position (NRPS-17 & -19) for the construction of a small but diverse tri-partite NRPS library with subunits derived from various *Photorhabdus* and *Xenorhabdus* strains, because T-C-A tri-domains as a catalytically active unit to reprogram NRPSs are underrepresented.

Eleven NRPS subunits with attached SZs were extracted from five different BGCs, namely from GxpS, XtpS as well as from the gargantuanin (GarS), xenolindicin (XldS)²⁹, and the szentiamide (SzeS)⁴⁶ producing synthetases. Overall 18 (NRPS-21 to -38) from 45 possible 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. 3, Figs. S24– S40). Despite the method's general simplicity, the overall efficacy or recombination potential of T-C-A units compared to XUs appears to be slightly more restricted²². For example, neither co-expression of all type S subunits to reconstitute SzeS, nor any combination involving the Ser and Thr specifying subunits from XldS and GarS, yielded any detectable peptide, respectively. These results probably indicate an incompatibility of formed chimeric A-T interfaces or substrate incompatibilities at the respective C domains donor site. Yet, in light of previous results concerning C domain specificities^{31, 40, 47-49}, the latter seems to be unlikely. Especially as we were not even able to reconstitute catalytic activity of the tri-partite SzeS, we concluded that the respective subunits might have lost their functionality. 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 or their ability to 'communicate' with downstream subunits – as was already expected for NRPS-19 (Fig. 2b).

Furthermore, for some tri-partite NRPSs (NRPS-32, -34, -35, -37 and -38), we were able to detect peptides (**43-46**, **48**, and **49**) only in very low amounts, which might be explained by the aforementioned impairment within the A-T domain interface and/or the mere length of the chosen SZ1:2 pair. Taking these points into consideration, we assume that productivity can significantly be increased when another fusion site is chosen, another SZ pair is chosen, or, if possible, SZ1:2 is truncated. Nevertheless, the quantity of produced peptides should not distract from the strength of this method: *The ability to generate an enormous variety of new-to-nature NRPSs in an unprecedented short time and with a minimum of lab work involved.* The goal is to produce a large variety of NRPs to generate structural diversity rather than to produce peptides on an industrially relevant scale. This can be focused on in a second step, when the most interesting candidates have been found.

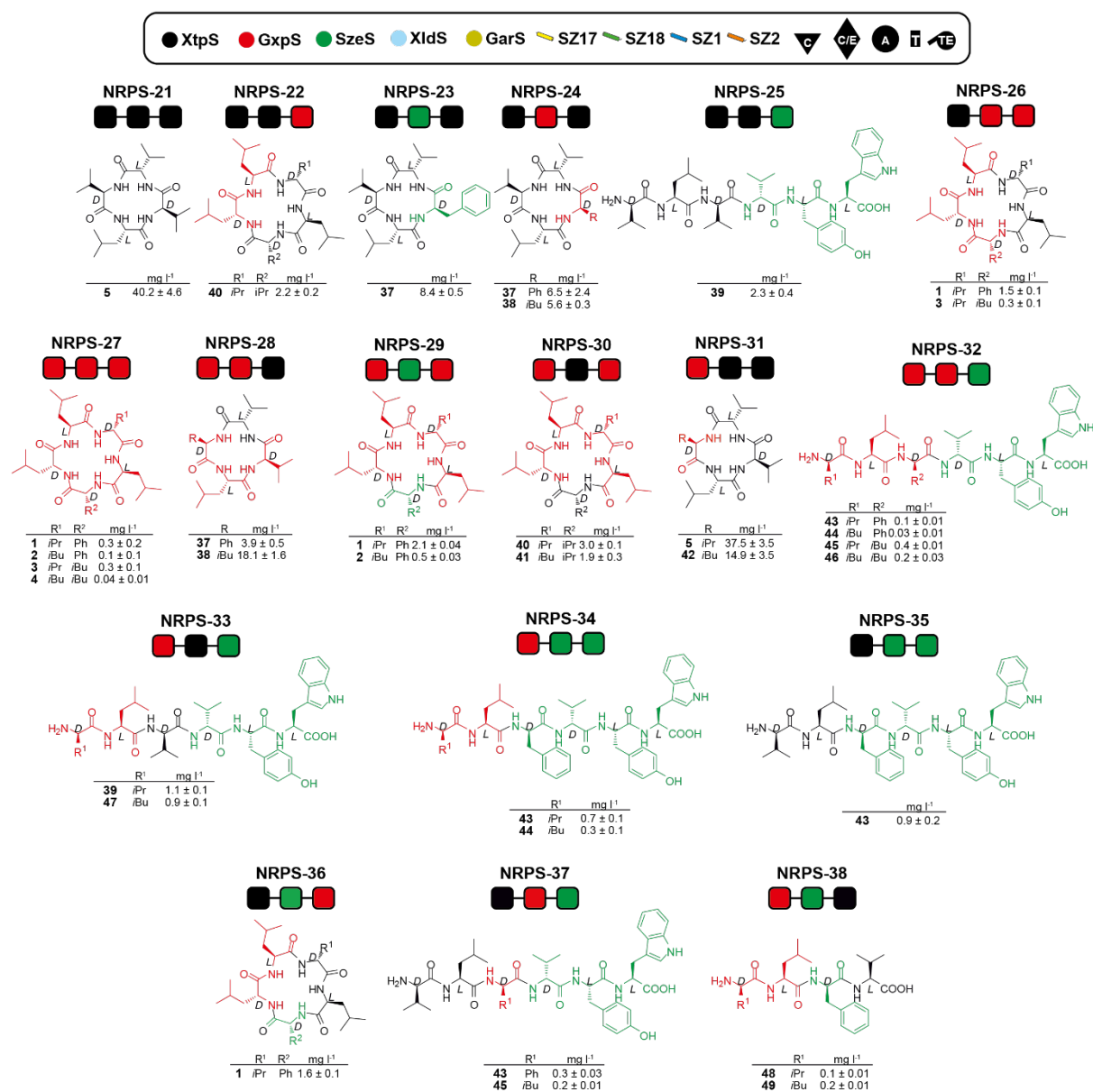


Figure 4. Tri-partite type S NRPS library. The upper section, generated building blocks, depicted in the symbol assignment as introduced previously, are illustrated. Solid lines represent functional combinations. In the lower section, building blocks were simplified and illustrated as boxes representing subunit 1 to 3. From 11 generated and 8 functional building blocks, a total of 18 type S NRPS were confirmed as functional by HPLC-MS. Corresponding peptide yields (mg/L) were obtained from triplicate experiments.

Conclusion

Combinatorial chemistry can generate potential new compounds in a greater number, cheaper and in less time than the traditional NP discovery process^{4, 50}. Yet, a lot of large chemical screening collections have failed in practice, because they do not exhibit enough diversity within the biologically relevant chemical space^{2, 51}. Nevertheless, although ‘NP-derived’ libraries (scaffold is identical to the NP scaffold) and ‘NP-inspired’ libraries⁵² (scaffold is closely related to NP scaffolds) have proven to be enriched in bioactivity over typical collections of synthetic compounds, NP discovery has strongly decreased over the past 20 years^{2, 53-56}. In part because of extraction and supply issues⁵⁷, and the frequent (re-) discovery of natural products with structural precedent in the literature – structurally unique scaffolds represent a decreasing percentage of the total number of isolated NPs. For NRPs, this becomes quite clear from the example of the Norine database⁵⁸. It is dedicated to collect all discovered non-ribosomally synthesised peptides, currently containing ~1800 peptides, of which ~800 peptides are characterised, but only ~175 scaffolds are unique. Interestingly, from the characterised set of peptides already more than 500 different NRP building blocks are known^{59, 60}, spanning a combinatorial chemical space for novel peptides beyond the total number of molecules in the whole universe and comprehension of the human mind.

Consequently, splitting known BGCs into individually expressible subunits and to recombine them simply by attaching SZs and co-expressing a variety of unrelated BGC subunits in a high throughput manner puts us in a position to easily enlarge the known structural diversity, and to outcompete both, traditional natural product research and combinatorial synthetic chemistry approaches. This especially might be true for the

generation of type S NRPSs to discover novel antimicrobials because in this particular case, the high throughput generation and production of antimicrobial peptides can be coupled to bioactivity testing against human pathogens, *i.e.*, via nanoFleming⁶¹, a miniaturized and parallelized high-throughput inhibition assay.

With this study we sought to showcase that, although nature still bears an enormous variety of natural products only waiting to be discovered⁶², methods for the identification and characterisation of new scaffolds from nature are still far from surpassing the throughput of new chemical entities that can be generated from type S NRPSs. Imagine, that you want to create a library of 10^6 NRP variants. According to the current state of the art (pre-type S), this would require the design and construction of 1.000.000 NRPSs, which of course is not feasible. By applying bi-partite type S NRPSs one can reduce the workload to 2000 type S NRPSs building blocks – 1000 building blocks for each subunit. This in turn implies that only 100 building blocks per subunits, or 300 in total, must be created if common one-protein type A NRPSs are split into three individually expressible subunits (100 subunits A x 100 subunits B x 100 subunits C = 10^6 combinations). Even if some type S systems are non-functional or do not synthesise the desired products, either due to the incompatibility of interface types, C and TE domain specificity issues, or capacity limitations of TE domains (Fig. 1), these limitations are no longer a significant problem, as sufficient diversity is still produced. Especially along with our previously developed design principles²²⁻²⁴, we have created a powerful toolset to predictably engineer new-to-nature NRPSs.

The striking advantage of type S NRPSs, however, is that once generated libraries can be enlarged and extended continuously and at any time, since generated subunits are not covalently linked. Moreover, a great variety of type S NRPSs can already be achieved from a small set of BGCs, as exemplified by turning GxpS into 16 artificial NRPSs (Fig. 1). The bi-partite type S GxpS library (Fig. 1), although not intended, led to the discovery of a further puzzle piece complementing our previously established NRPS design principles^{20, 23, 24, 40} – the importance of different C-A interface types, which have never been considered before for the construction of un-impaired high yielding artificial NRPSs.

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).

References

1. Dias, D.A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2**, 303-336 (2012).
2. Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot (Tokyo)* **65**, 385-395 (2012).
3. Moloney, M.G. Natural Products as a Source for Novel Antibiotics. *Trends Pharmacol Sci* **37**, 689-701 (2016).
4. Wright, P.M., Seiple, I.B. & Myers, A.G. The evolving role of chemical synthesis in antibacterial drug discovery. *Angew Chem Int Ed Engl* **53**, 8840-8869 (2014).
5. Atanasov, A.G., Zotchev, S.B., Dirsch, V.M. & Supuran, C.T. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* **20**, 200-216 (2021).
6. Doak, B.C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem Biol* **21**, 1115-1142 (2014).
7. Lawson, A.D.G., MacCoss, M. & Heer, J.P. Importance of Rigidity in Designing Small Molecule Drugs To Tackle Protein-Protein Interactions (PPIs) through Stabilization of Desired Conformers. *J Med Chem* **61**, 4283-4289 (2018).
8. Payne, D.J., Gwynn, M.N., Holmes, D.J. & Pompliano, D.L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40 (2007).
9. Ferri, M., Ranucci, E., Romagnoli, P. & Giaccone, V. Antimicrobial resistance: A global emerging threat to public health systems. *Crit Rev Food Sci Nutr* **57**, 2857-2876 (2017).
10. Ventola, C.L. The antibiotic resistance crisis: part 1: causes and threats. *P t* **40**, 277-283 (2015).
11. Miethke, M. et al. Towards the sustainable discovery and development of new antibiotics. *Nat Rev Chem*, 1-24 (2021).
12. Ling, L.L. et al. A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455-459 (2015).
13. Bennett, S. Solexa Ltd. *Pharmacogenomics* **5**, 433-438 (2004).
14. Shen, R. et al. High-throughput SNP genotyping on universal bead arrays. *Mutat Res* **573**, 70-82 (2005).
15. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat Biotechnol* **34**, 518-524 (2016).
16. Blin, K. et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-w87 (2019).
17. Paddon, C.J. & Keasling, J.D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol* **12**, 355-367 (2014).
18. Jelić, D. & Antolović, R. From Erythromycin to Azithromycin and New Potential Ribosome-Binding Antimicrobials. *Antibiotics (Basel)* **5** (2016).

19. DeMarinis, R.M. et al. Semisynthetic cephalosporins. Synthesis and structure-activity relationships of 7-sulfonylacetamido-3-cephem-4-carboxylic acids. *J Med Chem* **19**, 754-759 (1976).
20. Bozhüyük, K.A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to produce new antibiotics. *Curr Opin Microbiol* **51**, 88-96 (2019).
21. Calcott, M.J. & Ackerley, D.F. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol Lett* **36**, 2407-2416 (2014).
22. Bozhüyük, K.A.J., Watzel, J., Abbood, N. & Bode, H.B. Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases*. *Angew Chem Int Ed Engl* **60**, 17531-17538 (2021).
23. Bozhüyük, K.A.J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat Chem* **10**, 275-281 (2018).
24. 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).
25. Süßmuth, R.D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Ed Engl* **56**, 3770-3821 (2017).
26. 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).
27. 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).
28. Nollmann, F.I. et al. Insect-specific production of new GameXPeptides in photorhabdus luminescens TTO1, widespread natural products in entomopathogenic bacteria. *Chembiochem* **16**, 205-208 (2015).
29. Schimming, O., Fleischhacker, F., Nollmann, F.I. & Bode, H.B. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *Chembiochem* **15**, 1290-1294 (2014).
30. Drake, E.J. et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238 (2016).
31. 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).
32. Bode, E. et al. Promoter Activation in Δ hfq Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angewandte Chemie International Edition* **58**, 18957-18963 (2019).
33. Mootz, H.D. et al. Decreasing the Ring Size of a Cyclic Nonribosomal Peptide Antibiotic by In-Frame Module Deletion in the Biosynthetic Genes. *Journal of the American Chemical Society* **124**, 10980-10981 (2002).
34. Giessen, T.W. & Marahiel, M.A. The tRNA-dependent biosynthesis of modified cyclic dipeptides. *Int J Mol Sci* **15**, 14610-14631 (2014).
35. Jiang, Y. & Zheng, W. Cyclic Tripeptide-based Potent and Selective Human SIRT5 Inhibitors. *Med Chem* **16**, 358-367 (2020).
36. Ali, H. et al. A non-canonical NRPS is involved in the synthesis of fungisporin and related hydrophobic cyclic tetrapeptides in *Penicillium chrysogenum*. *PLoS One* **9**, e98212 (2014).
37. Loiseau, N., Cavelier, F., Noel, J.P. & Gomis, J.M. High yield synthesis of tentoxin, a cyclic tetrapeptide. *J Pept Sci* **8**, 335-346 (2002).
38. Nakao, Y. et al. Azumamides A-E: histone deacetylase inhibitory cyclic tetrapeptides from the marine sponge *Mycale izuensis*. *Angew Chem Int Ed Engl* **45**, 7553-7557 (2006).
39. Kegler, C. et al. Rapid determination of the amino acid configuration of xenotetrapeptide. *Chembiochem* **15**, 826-828 (2014).

- 484 40. Kranz, J., Wenski, S.L., Dichter, A.A., Bode, H.B. & Bozhüyük, K.A.J. Influence of
485 condensation domains on activity and specificity of adenylation domains. *bioRxiv*,
486 2021.2008.2023.457306 (2021).
- 487 41. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D.H. Phylogenetic analysis of
488 condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol*
489 **7**, 78 (2007).
- 490 42. Chiocchini, C., Linne, U. & Stachelhaus, T. In vivo biocombinatorial synthesis of
491 lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic
492 complex. *Chem Biol* **13**, 899-908 (2006).
- 493 43. Reimer, J.M. et al. Structures of a dimodular nonribosomal peptide synthetase reveal
494 conformational flexibility. *Science* **366** (2019).
- 495 44. Keating, T.A., Marshall, C.G. & Walsh, C.T. Reconstitution and characterization of the
496 *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **39**,
497 15522-15530 (2000).
- 498 45. Yan, F. et al. Synthetic biology approaches and combinatorial biosynthesis towards
499 heterologous lipopeptide production. *Chem Sci* **9**, 7510-7519 (2018).
- 500 46. Nollmann, F.I. et al. Synthesis of szentiamide, a depsipeptide from entomopathogenic
501 *Xenorhabdus szentirmai* with activity against *Plasmodium falciparum*. *Beilstein J Org*
502 *Chem* **8**, 528-533 (2012).
- 503 47. Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A. & Essen, L.O. Structural
504 and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide
505 synthetase. *Structure* **15**, 781-792 (2007).
- 506 48. Izoré, T. et al. Structures of a non-ribosomal peptide synthetase condensation domain
507 suggest the basis of substrate selectivity. *Nat Commun* **12**, 2511 (2021).
- 508 49. Dekimpe, S. & Masschelein, J. Beyond peptide bond formation: the versatile role of
509 condensation domains in natural product biosynthesis. *Natural Product Reports* (2021).
- 510 50. Lefevre, F. et al. Drugs from hidden bugs: their discovery via untapped resources. *Res*
511 *Microbiol* **159**, 153-161 (2008).
- 512 51. Harvey, A.L., Edrada-Ebel, R. & Quinn, R.J. The re-emergence of natural products for
513 drug discovery in the genomics era. *Nat Rev Drug Discov* **14**, 111-129 (2015).
- 514 52. Laraia, L. & Waldmann, H. Natural product inspired compound collections: evolutionary
515 principle, chemical synthesis, phenotypic screening, and target identification. *Drug Discov*
516 *Today Technol* **23**, 75-82 (2017).
- 517 53. Demain, A.L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J Antibiot*
518 *(Tokyo)* **62**, 5-16 (2009).
- 519 54. Wetzel, S., Bon, R.S., Kumar, K. & Waldmann, H. Biology-oriented synthesis. *Angew*
520 *Chem Int Ed Engl* **50**, 10800-10826 (2011).
- 521 55. Balamurugan, R., Dekker, F.J. & Waldmann, H. Design of compound libraries based on
522 natural product scaffolds and protein structure similarity clustering (PSSC). *Mol Biosyst* **1**,
523 36-45 (2005).
- 524 56. Cremosnik, G.S., Liu, J. & Waldmann, H. Guided by evolution: from biology oriented
525 synthesis to pseudo natural products. *Natural Product Reports* **37**, 1497-1510 (2020).
- 526 57. Amirkia, V. & Heinrich, M. Natural products and drug discovery: a survey of stakeholders
527 in industry and academia. *Front Pharmacol* **6**, 237 (2015).
- 528 58. Flissi, A. et al. Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res*
529 **48**, D465-d469 (2020).
- 530 59. Walsh, C.T., O'Brien, R.V. & Khosla, C. Nonproteinogenic amino acid building blocks for
531 nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl* **52**, 7098-
532 7124 (2013).
- 533 60. Blin, K. et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster
534 boundary identification. *Nucleic Acids Res* **45**, W36-w41 (2017).

- 535 61. Schmitt, S. et al. Analysis of modular bioengineered antimicrobial lanthipeptides at
536 nanoliter scale. *Nat Chem Biol* **15**, 437-443 (2019).
 - 537 62. Pye, C.R., Bertin, M.J., Lokey, R.S., Gerwick, W.H. & Linington, R.G. Retrospective
538 analysis of natural products provides insights for future discovery trends. *Proc Natl Acad*
539 *Sci U S A* **114**, 5601-5606 (2017).
- 540