



University of Groningen

Nonribosomal peptide synthetases

Zwahlen, Reto Daniel

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Zwahlen, R. D. (2018). Nonribosomal peptide synthetases: Engineering, characterization and biotechnological potential. [Groningen]: University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER II

Identification and characterization of nonribosomal peptide synthetase modules that activate 4-hydroxyphenylglycine

Reto D. Zwahlen,¹ Remon Boer,³ Ulrike M. Müller,³
Roel A.L. Bovenberg,^{2,3} and Arnold J.M. Driessen^{1,4}

¹Molecular Microbiology, Groningen Biomolecular Sciences and
Biotechnology Institute, University of Groningen, Groningen,
The Netherlands

²Synthetic Biology and Cell Engineering, Groningen
Biomolecular Sciences and Biotechnology Institute, University
of Groningen, Groningen, The Netherlands

³DSM Biotechnology Centre, Delft, The Netherlands

⁴Kluyver Centre for Genomics of Industrial Fermentations,
Julianalaan 67, 2628BC Delft, The Netherlands

submitted to PLOS One

Abstract

4-hydroxyphenylglycine (hpg) is a small non-proteinogenic amino acid, present in secondary metabolites such as teicoplanin, ramoplanin or enduracidin. Due to its physico-chemical properties, hpg occupies a crucial role in the determination of compound structures, thereby establishing therapeutically relevant properties. Compounds containing hpg are predominantly synthesized by modular Nonribosomal peptide synthetases (NRPS), containing adenylation domain bearing modules, which can specifically integrate hpg into a growing peptide. NRPS may require the specific binding of an integral helper protein group, the MbtH homologues. Using pFAM software as a basis for the identification of potential hpg activating adenylation domains, 12 modules and their MbtH helper proteins in a total of 17 domain setups, were selected, expressed in *E. coli* and characterized *in vitro* including their MbtH homologs, and subsequently subjected to expression and substrate specificity characterization experiments. Heterologous expression in *E. coli* was observed with 9 targets, and with all 17, when co-expressed with MbtH or upon introduction of an N-terminal maltose binding protein (MBP) tag, respectively. Pyrophosphate exchange assays revealed, that all domains exhibited activity towards L- or D-4-hpg and underlined the essential role of MbtH homologues in the activation of hpg activating domains. In addition, comparing the adenylation velocities, we demonstrated the crucial role of NRPS starter modules, which showed an up to 5-fold increased activity when compared to any elongation module in this study. Overall, we demonstrate here a medium throughput system, allowing for the identification and characterization of specific NRPS modules. In the context of future compound and NRPS engineering efforts, this may serve as a fundament for the selection of domains and modules with promising specificity and activity, eventually allowing for the creation of novel bioactive compounds.

Introduction

Multi-modular enzymes represent a highly evolved and complex family of large, polycatalytic, biochemical machineries. One of the most versatile representatives of this superfamily are the nonribosomal peptide synthetases (NRPS). The production of antibiotics, immunosuppressants, cytostatics, but also pigments can all be attributed to NRPS. Members of this group of modular enzymes can consist of up to 15 [1] functionally distinct parts or modules. Modules are characterized by their distinct functions in recognition, activation and condensation of a dedicated compound into a growing nonribosomal peptide (NRP). In order to allow for all the aforementioned functions, a module contains a set of functionally distinct parts, or domains, each carrying out a distinct catalytic sub-step. Thus, a minimal module must contain an adenylation (A), thiolation (T) and condensation (C) domain. The adenylation domain, consisting of a A_{core} and A_{sub} part [2] is responsible for the recognition, adenylation and thioesterification of a substrate. Through the hydrolysis of ATP, an AMP-substrate conjugate is formed, which is subsequently transferred to the free thiol group of the 4'-phosphopantetheinyl-moiety (ppant), which is anchored to the downstream T domain [3–4]. This thioesterification is guided by a structural rearrangement of the A_{sub} domain, which itself undergoes a rotation by about 140° [5–6] relative to the A_{core} domain. The ppant arm, linked to the activated substrate, is itself covalently attached to a highly conserved serine residue which is part of the GGXS motif of the T domain [7–8]. To finalize peptide bond formation, the activated T-ppant-substrate moiety associates with the C domain where the peptide bond formation is guided. Upon a nucleophilic attack, a dipeptide is created which remains attached to the downstream ppant arm [9]. An exception to those minimal conditions are standalone modules and initiation or N-terminal modules of multi modular NRPS enzymes. These NRPS modules lack C domains due to the absence of an upstream module. A series of additional domains might be present in NRPS enzymes, but none of them is essential for the peptide chain formation except for the C-terminal thioesterase (Te) domain, which is crucial for product release. In the absence of such a domain, standalone thioesterase enzymes may fulfill this function.

Multiple highly selective domains for epimerization (E), halogenation, oxidation and a multitude of pluripotent C domains have been identified [10–12]. Despite the abundance of many functionally distinct domains, some reactions still require direct or indirect involvement of other factors. These may involve essential steps, such as the attachment of the phosphopantetheinyl moiety at the thiolation domain which is a CoA dependent process carried

out by a phosphopantetheinyl-transferase (*sfp*) [13]. An example of potentially essential chaperones is the highly conserved MbtH protein [14–15]. The precise role of these relatively small proteins, i.e., ~70 amino acids, has remained elusive. However, MbtH proteins have been shown to increase the adenylation activity [16] as well as the levels of NRPS expression [17]. Moreover, high affinity association with A domains seems to play a crucial role in the MbtH activity [18–19]. MbtH homologs are essential for some bacterial NRPS enzymes, but seem entirely absent from eukaryotic organisms such as fungi. Also, the number of MbtH copies and variants varies strongly, even among strains from the same organism [20] while the impact on functionality is very diverse. It ranges from a complete loss of functionality [21] to insignificant variations in product levels [15] as gene deletion studies suggest [22]. Even the replacement of standalone MbtH proteins with a covalently linked, MbtH-like domain has been demonstrated, using RubC [23]. Recently, the first A domain-MbtH structure has become available, suggesting a set of conserved residues which are tightly involved in the MbtH-NRPS association [19].

Many different NRPS products, enzymes, domains and catalytical functions have been identified and assigned. A domains have long been considered to be a bottleneck in NRPS systems, fulfilling an essential gatekeeper function. Therefore, they were in the focus of characterization, prediction and engineering attempts for an extended period of time [24–26]. This has led to the identification of important motifs for adenylation [27], substrate specificity, as well as a better understanding of the domain arrangement through structural analysis [2]. Despite this knowledge and use of an ever increasing reservoir of structures and predictive software, successful site directed engineering efforts are still exceptional [25;28–30]. Experiments targeting the elucidation of catalytic sub steps support the importance of A domains, though also novel targets in substrate specificity emerged, especially C domain characterization. T domains, the smallest and catalytically inactive domain, rely mostly on a core GGXS motif resolving around the highly conserved serine residue, necessary for *sfp* dependent ppant attachment [7]. Due to their very high degree of structural conservation and the availability of crystal structures, the overall structure is well understood. Not only single domains were structurally characterized, but also multi-domain structures as well as whole modules [31–33]. More recent studies attribute an engineering potential to T domains, since it has been shown that synthetic T domains can improve product output *in vivo* and *in vitro* [34]. Nevertheless, conformational changes of the NRPS enzymes during their catalytic cycle are not well understood. Thus, the potential for accurate predictions in the context of engineering efforts are limited.

Despite many studies, C domains remain the most elusive. Early experiments indicate that NRPS enzymes, including C domains, show considerable promiscuity in substrate incorporation [35]. However, the C domain never played an important role in engineering attempts. A turnaround in C domain perception was marked with the discovery of the essential HHxxDG motif [36] involved in peptide bond formation. Further experiments revealed additional conserved sequences related to stereochemical selection [37], indicating important potential bottlenecks for engineering. In terms of sequence conservation and domain size, C domains are exceptionally diverse [38] and are linked to multifunctionality [10]. This has been underlined, as condensation-epimerization [10] domains were discovered. Recently, more novel functions, most notably oxygenase recruitment, has been described [12]. In biochemical studies, NRPS catalytic sub steps have been investigated and likely bottlenecks were identified [39]. However, crucial information about overruling complexities remain widely unknown as no complete NRPS crystal structure is available. Even with the high number of predictive tools and genomic data, a systematic analysis of domains remains to be a delicate task. Given the reservoir of novel amino acids that can be incorporated, the exploitation of A domains for novel NRP production will remain an important challenge.

4-hydroxyphenylglycine (hpg) is a non-proteinogenic amino acid that is present in various NRPs. The 4-hydroxy group on the phenol ring makes hpg especially valuable for covalent cross-linking, side branching and circularization of large NRPs [40–41]. The inclusion of higher order structures can vastly improve relevant properties and spectra of bioactive substances. Thus, the introduction of hpg into scaffold molecules such as antibiotics, fungicides and other relevant compounds has the potential to create novel classes of pharmaceuticals, bypassing contemporary problems, such as the widespread antibiotic resistance. Using fundamental techniques for compound [42], gene and domain determination [43–44], paired with web based software for substrate specificity prediction [45], a selection of putative hpg activating modules were identified and selected. These were produced in *E. coli* in the absence and presences of MbtH proteins, purified and assayed for activity and specificity. These data provide a reference for future exploration of hpg units and the construction of novel chimeric NRPS enzymes that incorporate hpg in the NRP product.

Materials and methods

Selection and determination of hpg activating domains

Putative hpg activating modules were identified using the protein databases UniRef100, NCBI environmental as well as others as a resource. In a first step putative NRPS protein sequences were extracted, based on concurrent presence of Pfam [46] motifs for AMP-binding, ppant-binding, and condensation domains. The hypothetical NRPS domain structures were determined and potential adenylation domains classified as part of a starter or elongation module. To predict the preferred amino acid bound by these adenylation domain sequences, they were analyzed using the NRPSpredictor platform [47] which includes the identification of the “specificity conferring” code by Stachelhaus [27]. The NRPS modules containing adenylation domains with at least 70 % identity to the Stachelhaus code for putative hpg binding were used for final selection.

Cloning, plasmids and culture conditions

Cloning was performed using *E. coli* DH5 α . Selecting with 50 μ g/ml neomycin for pSCI242/243 plasmids (pBR322ori; pBAD, araC; FdT); 100 μ g/ml ampicillin for pMAL-c5x plasmids (pMB1ori; bla; pTac; lacIq; rrnB T2) and 15 μ g/ml chloramphenicol for pACYctac-MbtH plasmids (p15Aori; cat; pTac; lacIq), respectively. *E. coli* BL21 (DE3) strain was used for expression of the adenylation domains. All cultures were grown using 2xPY (15 g/l bacto-tryptone, 10 g/l yeast extract, 10 g/l sodium chloride, pH 7.0) at 37 °C and 200 rpm. Synthetic DNA fragments of domains were ordered at DNA 2.0 and codons optimized for expression in *E. coli*. Fragments were subsequently subcloned using the available NdeI \times SbfI sites (pMAL-c5x, N8108S, New England Biolabs Inc., Ipswich, MA).

Expression and purification of hybrid NRPS and hpg modules

Cultures were grown to an OD₆₀₀ of 0.6 in 2xPY medium, transferred to 18 °C and 200 rpm for 1 h, and subsequently induced using 0.5 mM IPTG and 0.2 % L-arabinose. Harvest followed 18 hours after induction by spinning at 3500 g for 15 minutes. After resuspension in lysis buffer (50 mM HEPES pH 7.0, 300 mM NaCl, 2 mM DTT, complete EDTA free protease

inhibitor (Roche, Basel, CH, REF 04693159001), cells were disrupted using sonication (6 s/15 s; on/off, 60 cycles, 10 μm amplitude) and cell-free lysate obtained by centrifugation at 4 °C, 13000 g, 15 minutes. Enzymes were purified by means of Ni-NTA bead (Qiagen, Hilden, Ger, No. 30210) supported his-tag affinity purification using gravity flow. Wash steps were performed using wash buffer (50 mM HEPES pH 7.0, 300 mM NaCl, 20 mM imidazole) and a one-step elution using elution buffer (50 mM HEPES pH 7.0, 300 mM NaCl, 250 mM imidazole). Imidazole was removed, while simultaneously concentrating the sample with Amicon®-50 spin filters (Amicon® Ultra Ultracel — 50K, Merck Millipore, Billerica, MA, USA, UFC 805024). Protein concentrations were determined using BioRad DC assay kit. All fractions were subsequently analyzed on a 10 % SDS-PAGE. Insoluble expression was determined using the solubilized cell debris in 8 M UREA. Gels were stained using a 0.025 % coomassie stain and images taken using an imaging cabinet (FUJIFILM, Tokyo, Jp, LAS-4000). A domain — MbtH stoichiometry was quantified by means of 2-D densitometric analysis using AIDA Image Analyzer v.4.22 software.

Substrate profiling of putative hpg activating domains

The substrate profile and promiscuity of putative hpg activating domains was evaluated using the *in vitro* pyrophosphate exchange kit EnzCheck (EnzCheck, Thermo Fischer Scientific, Waltham, MA, USA, E6645). Isolated enzyme was tested at a concentration of 1 μM in a 96-well setup and a total reaction volume of 100 μl . L- or D-4-phenylglycine (Pg); L- or D-4-hydroxyphenylglycine (hpg) and L-phenylalanine, serving as negative control, were tested respectively. All substrates were screened at a concentration of 0.1 and 1 mM. The absorption at 360 nm of the coupled indicator substrate MESG (2-amino-6-mercapto-7-methylpurine ribonucleoside) was measured in 5 minutes intervals over a period of 4 h. Activity was compared as a function of (PPi) $\mu\text{M min}^{-1}$ (enzyme) μM^{-1} .

Results

Selection and determination of putative hpg activating domains

Putative hpg activating domains were identified using various databases. NRPS sequences were identified, based on Pfam [46] motifs and obtained

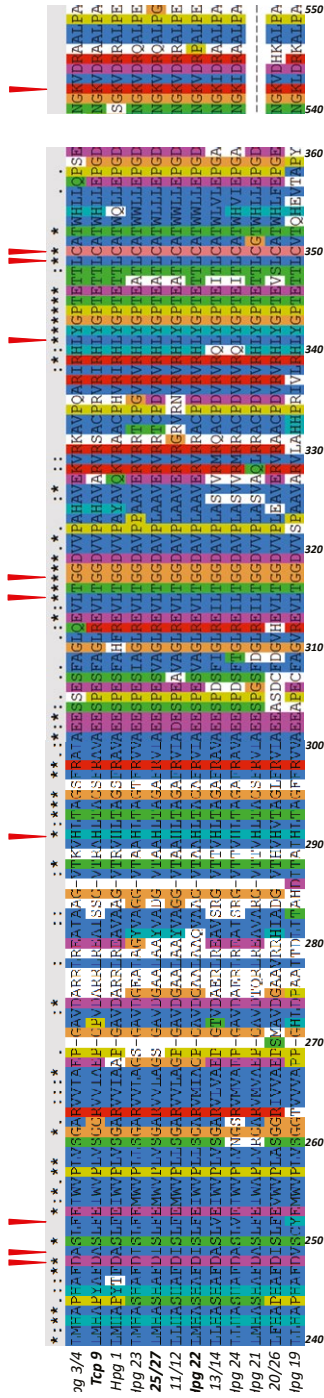
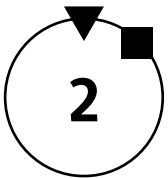


Figure 1 — Stachelhaus code of selected adenylation domains.

The 10 amino acid motif, involved in adenylation specificity is indicated (red). Motifs were predicted using NRPSpredictor [47] and the alignment was made using ClustalX. The motif is almost conserved throughout all domains.

Table 1 — Final selection of target domain constructs.

<i>Construct</i>	<i>Organism</i>	<i>NRPS setup</i>	<i>Initiation</i>	<i>Reference</i>
<i>Tcp9</i>	<i>Actinoplanes teichomyceticus</i> ATCC53649	<u>AT</u> CATE	Y	[51]
<i>Hpg1</i>	<i>Streptomyces toyocaensis</i> NRRL 15009	<u>AT</u> CATE	Y	[52]
<i>Hpg3/4</i>	<i>Nonomuraea</i> sp. ATCC 39727	<u>AT</u> CATE	Y	[53]
<i>Hpg11/12</i>	<i>Amycolatopsis mediterranei</i>	CATE CAT	N	[54]
<i>Hpg13/14</i>	<i>Amycolatopsis mediterranei</i>	CATE CAT	N	[54]
<i>Hpg19</i>	<i>Streptomyces coelicolor</i> A3(2)	CAT CATE CAT <u>CATE</u>	N	[55]
<i>Hpg20/26</i>	<i>Streptomyces lavendulae</i>	<u>CATE</u>	N	[56]
<i>Hpg21</i>	<i>Streptomyces toyocaensis</i> NRRL 15009	CATE CAT	N	[52]
<i>Hpg22</i>	<i>Actinoplanes teichomyceticus</i> ATCC53649	<u>CATE</u> CAT	N	[51]
<i>Hpg23</i>	uncultured	<u>CATE</u> CAT	N	[57]
<i>Hpg24</i>	<i>Amycolatopsis orientalis</i>	CATE CAT	N	[58]
<i>Hpg25/27</i>	uncultured	<u>CATE</u> CAT	N	[57]

Construct code, domain setup of target NRPS and corresponding module are indicated. Furthermore module placement, Initiation or other is shown. Constructs derived from the same module, in different domain setups are indicated with double numbers and adjusted colors, e.g. *Hpg XX/XY*

adenylation domain sequences were subsequently analyzed using the NRP-Spredicator platform [47]. *Hpg* affinity was finally evaluated using NRPSpredicator data, leading to a list of 58 potential candidate modules. Sequences of 43 modules had at least 80 % identity with the predicted Stachelhaus [27] *hpg* core motif (Supplementary figure 1). Upon further analysis of the putative targets, 43 modules could be subdivided into 7 initiation and 36 elongation modules. Additionally, we elucidated that those domains can be assigned to 28 multi-modular NRPS, arranged in 14 secondary metabolite gene clusters. 13 out of 14 can be assigned to a bacterial host, though the origin of two remains elusive, as they were obtained from metagenomic data.

After consideration of the adenylation domain placement within the module as well as the availability and feasibility of source material, 12 module adenylation domains were selected (Figure 1). The adenylation domains were subsequently ordered in different setups, either as single A domain or including their adjacent T, Te or C domains. All genes were codon optimized for *E. coli* by the manufacturer (DNA 2.0). This led to a total of 18 constructs as indicated in Table 1. In addition to the NRPS, their putative corresponding MbtH homologs were identified. MbtH proteins have a chaperoning function, which is potentially essential for adenylation activity and expression. Identification followed taking the conserved MbtH core motif [17], leading to a

selection of 7 MbtH homologs associated with dedicated NRPS gene clusters. Their corresponding sequences were furthermore analyzed for hypothetical NRPS binding motifs (Figure 2A/B).

Heterologous expression profiling

In order to determine the expression properties of the selected putative hpg activating domains, they were subjected to small-scale expression experiments. Constructs in the pSCI242 vector were expressed without the addition of MbtH proteins. Subsequently, the soluble fraction as well as the cell debris or insoluble fraction were analyzed. 9 out of 17 constructs showed soluble expression of their respective domain(s). However, the expression levels varied strongly. Hpg14, 16 and 20 showed very low amounts hardly traceable using coomassie staining but with the occurrence of insoluble protein (data not shown). One variant, Hpg18, allowed for the detection of insoluble expression only. The remaining 6 Hpg variants (Hpg11, 12, 21, 22, 23, 24) did not result in any heterologously-expressed protein.

In order to improve soluble protein expression, a maltose binding protein (MBP) was fused to the N-terminus of the constructs. Expression trials were repeated, using the MBP-Hpg A domain constructs with and without co-expression of MbtH proteins. This led to significantly improved expression behavior for all tested constructs. We could observe expression of all constructs and eliminate insoluble protein almost entirely, in favor of the highly soluble MBP-Hpg variant. The qualitative expression results were independent of MbtH utilization, although, co-expression increased the total amount of soluble MBP-Hpg. The positive effect of MBP-tag introduction is best illustrated comparing Hpg11, 21, 22 and 23, which were previously not expressed at all (data not shown). Another notable observation is the high affinity binding of MbtH variant Tcp13. It did not only increase protein amount, but also co-eluted with all tested domains. To further elucidate the promiscuity and effect of the different MbtH variants, 4 A domain constructs and their 3 respective MbtH variants were chosen for co-expression trials. The selection includes two domains and two MbtH homologs of the teicoplanin biosynthetic cluster, starter module A domain Tcp9 (AT) and elongation module A domain Hpg22 (A) plus the MbtH proteins Tcp13 and Tcp17, and the constructs Hpg25 (A) and Hpg27 (ATE), comprising the same elongation A-domain and their respective MbtH VEG8, derived from a metagenomic sample. Comparing the performance of the different MbtH variants, Tcp13 showed the overall best performance with all tested constructs (Table 2). Focusing on teicoplanin

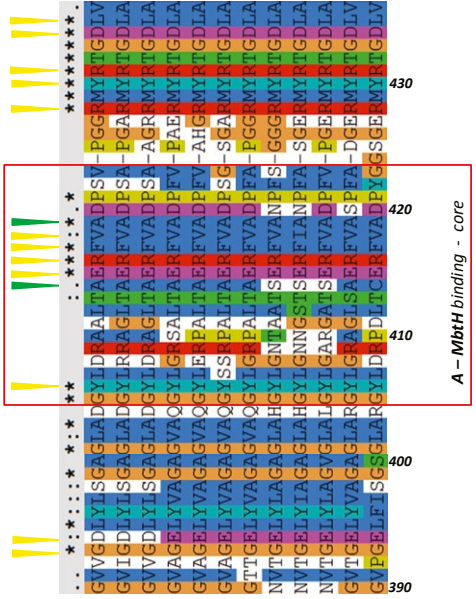
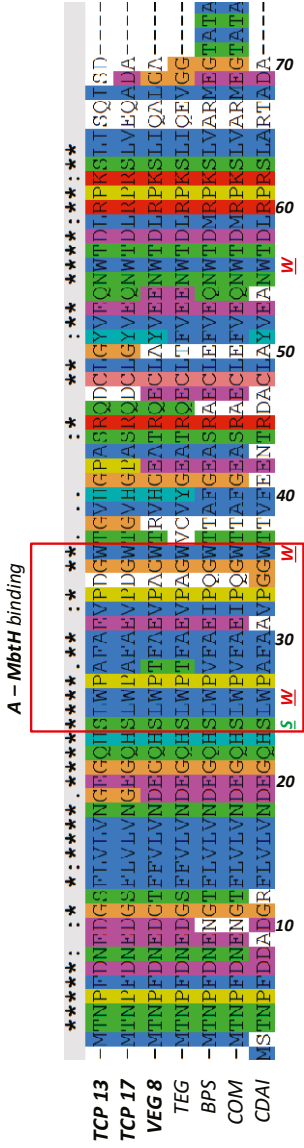



Figure 2 A — Alignment of Mbth variants used in the various experiments.
 The alignments were made in ClustalX [59], in depth analyzed variants are in bold. Highly conserved tryptophan (W) residues are highlighted in red. Additionally, in green, a central serine residue (S) resolving around the hypothetical A-Mbth interaction site (red box) is indicated. Position of the interaction site were determined according to structural studies by [19].

Figure 2 B — Hypothetical binding side on A domains.
 The core binding motif resolves around A414 and A419 (green). Other residues potentially involved in binding are indicated (yellow) [19]. The core motif is widely unaltered among Tcp9, Hpg22 and Hpg25/27.

Table 2 — Pyrophosphate release rates of hpg modules, co-expressed with selected MbtH variants.

Values are expressed as a function of $(\text{PPi}) \mu\text{M min}^{-1} (\text{Enzyme}) \mu\text{M}^{-1}$. The color code is set at red = low (0)/ green = high (3). The high activity of Tcp9 is the most striking feature of this data set. N=2; error of mean overall = 0.056.

		<i>D-hpg</i>		<i>L-hpg</i>		<i>D-pg</i>	<i>L-pg</i>	<i>L-phe</i>
		0.1 mM	1 mM	0.1 mM	1 mM	1 mM	1 mM	1 mM
<i>Tcp9</i>	<i>Tcp13</i>	2.07	5.28	1.18	1.16	0.05	1.32	0.00
	<i>Tcp17</i>	1.71	4.63	1.20	1.34	0.05	1.44	0.00
	<i>Veg8</i>	3.72	8.44	1.23	1.40	0.07	2.32	0.00
<i>Hpg22</i>	<i>Tcp13</i>	0.08	0.66	0.80	1.03	0.00	0.17	0.00
	<i>Tcp17</i>	0.11	0.63	0.95	1.04	0.04	0.23	0.00
	<i>Veg8</i>	0.00	0.86	1.38	1.54	0.00	0.09	0.00
<i>Hpg25</i>	<i>Tcp13</i>	0.14	0.92	0.56	0.59	0.01	0.17	0.00
	<i>Tcp17</i>	0.17	1.03	0.70	0.64	0.02	0.14	0.00
	<i>Veg8</i>	0.18	1.39	0.61	0.61	0.02	0.20	0.00
<i>Hpg27</i>	<i>Tcp13</i>	0.12	0.72	0.54	0.57	0.01	0.15	0.00
	<i>Tcp17</i>	0.11	0.62	0.48	0.52	0.01	0.12	0.00
	<i>Veg8</i>	0.27	1.42	0.84	0.88	0.02	0.27	0.00

min  max

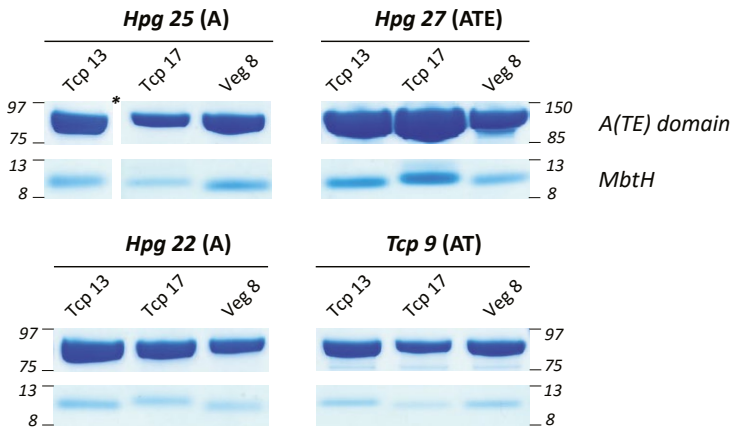


Figure 3 — Hpg22, 25, 27 and Tcp9 co-expression with the MbtH Tcp11, 13 and VEG8.

The constructs Hpg22, 25, 27 and Tcp9 were co-expressed with three MbtH variants. The protein input was adjusted according to culture OD. Differences can be observed between target domain setups (A, AT, ATE) and the corresponding MbtH variants used.

associated domains, VEG8 seems to surpass Tcp17 in terms of its beneficial effects on expression. Interestingly, Hpg25 expression increased equally by co-expression of VEG8, intrinsic to the domain, and Tcp13. The impact of the non-native MbtH is even more drastic for Hpg27, in combination with Tcp17, which exhibits the weakest effects for the other enzymes. Quantitation of the expression shows that despite affecting the levels of soluble protein, no alterations in the 1:2 stoichiometry of A domain to MbtH was evident.

Substrate profiling of hpg domains

To confirm the previously predicted specificity towards hpg, the enzymes were subjected to a pyrophosphate exchange reaction to assess their substrate dependent adenylation potential. Therefore, domains were overexpressed and purified using affinity chromatography targeting the C-terminal his tag. Isolated domains and modules were adjusted to a final concentration of 1 μM and screened for L- or D- hpg and L- or D-pg at 0.1 and 1 mM, respectively. Phenylalanine (1 mM) and substrate free reactions served as negative controls. An initial screen showed activity for 15 out of 16 A domains towards L-hpg and/or D-hpg (Supplementary table 3). To compare the impact of domain setup and MbtH variants, four A domain constructs were selected for an in depth analysis. Tcp9 (AT) and Hpg22 (A), of the teicoplanin cluster, as well as Hpg25 (A) and 27 (ATE) of the VEG cluster, representing the same selection as previously stated for the in depth expression analysis (Figure 3). The enzymes were co-expressed with one of their intrinsic MbtH variants Tcp13, Tcp17 or VEG8 and subjected to *in vitro* analysis. Comparing the specificities of the selected domains, an overall higher adenylation velocity can be observed with reactions containing L-hpg and D-hpg. Furthermore, all enzymes show activity towards L-pg, however, only Tcp9 (AT) + VEG8 shows significant activity towards D-pg. None of the reactions containing L-phenylalanine showed any product formation (Table 2). Moreover, there was no clear overall effect on the L- and D-hpg stereoselectivity. Though a slight D-hpg preference was observable with Hpg22, which is even more pronounced for Tcp9. A 5-fold increased velocity towards D-hpg was measured comparing to the 3 other constructs in the selection. Tcp9 showed overall higher velocities, in a range of 0.07 to 8.44 $\mu\text{M min}^{-1} \mu\text{M}^{-1}$ for D-pg and D-hpg, respectively. The variance within Hpg22, Hpg25 and Hpg27 was generally lower. Rates of Hpg22 at 0.04 and 1.54 $\mu\text{M min}^{-1} \mu\text{M}^{-1}$ for D-pg and D-hpg were observed, respectively. The latter activity represents the highest velocities for the remaining three constructs.

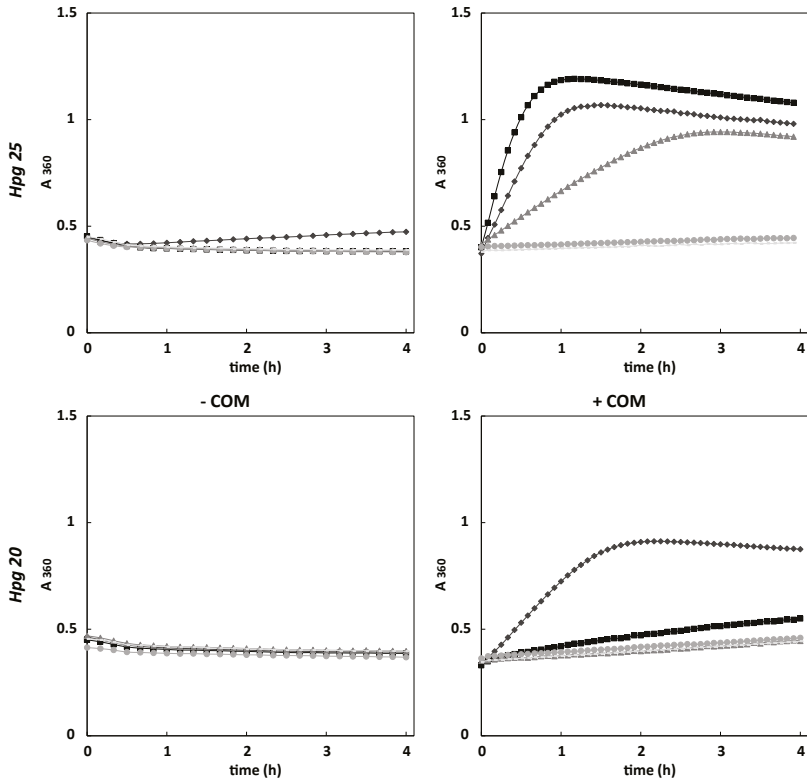


Figure 4 — Effect of Mbth co-expression and co-purification on the adenylation activity of domains Hpg20 and Hpg25.

A slight turnover of L-hpg by Hpg25 is detected without an Mbth, the activity increases dramatically upon the presence of its intrinsic Mbth variant VEG8. The impact on Hpg20 is even more drastic. Activity could only be determined after co-expression with the native Mbth COM.

In addition to the substrate specificity, the performance of the different Mbth homologs was evaluated. VEG8 appears to be the most promiscuous homologue, as it improves the catalytic properties of the A domains in the most significant manner. Tcp13 and 17 exhibit comparable properties, although at slightly lower levels. To assess the effects of omitting Mbth, Hpg20 and Hpg25 were compared with and without COM and Tcp13, respectively (Figure 4). Even though, Hpg25 showed notable activity towards L-hpg, the velocity increased dramatically upon co-expression of VEG8. An even more drastic effect can be observed with Hpg20. Without COM, no substrate

activation was observed. However, upon COM addition, adenylation of the substrates L-hpg and to a lower degree D-hpg were measured.

Discussion

Adenylation domains are essential core components of NRPS enzymes. Their ability to recruit a substrate with a high specificity makes them key elements in the synthesis of complex nonribosomal peptides. The diversity of dedicated substrate recognition is tremendous, as over 500 substrates have been identified [42]. However, due to a significant degree of structural conservation among A domains, specificity prediction software is well developed. Several programs allow for considerably accurate substrate predictions, including the NRPSpredictor [47]. This allowed the selection of domains in this study, which have the potential to adenylate the predicted 4-hydroxyphenylglycine substrate. Moreover, linker regions, domains and modules were accurately pinpointed, leading to a set of fully functional constructs.

In addition to protein domain prediction, corresponding adenylation helper proteins or MbtH homologs, were successfully identified, even from unannotated sources, as for VEG8 and TEG. Because of the high degree of conservation, 7 MbtH sequences were identified, selected and subjected to co-expression and purification studies. Even though the precise stoichiometry of A domains: MbtH is still poorly defined [16], we consistently observed a 1:2 ratio, confirming other studies [16–17]. Furthermore, a typical co-elution pattern emerged. Every tested domain allowed for MbtH binding of native, as well as non-native MbtH variants. Due to the high degree of structural conservation among NRPS modules and MbtH proteins, the question, if non-native MbtH do influence A domain behavior was addressed. It has been shown previously [48] that upon deletion or silencing of MbtH homologs, another genome encoded variant can compensate for the loss. This is in line with the behavior of the tested Tcp13, Tcp17 and VEG8 variants. All three homologs allowed for improved expression and significantly increased amino acid activation rates, when tested *in vitro* with the domains Tcp9, Hpg22, 25 and 27. However, the degree of their influence seems to depend on the homolog used, and importantly, native copies did not always result in the strongest effects. Furthermore, also the domain setup of the A domain seems to play a key role as illustrated by the comparison of Hpg25 and Hpg27, which consist of A only and A + Te, respectively. Even though, this is not representative for all tested domains, it points at an overruling structural and conformational mechanism, contributing to the elucidation of the still poorly defined function of MbtH

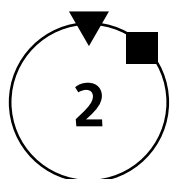
proteins. The differences in performance may be attributed to altered association properties. Focusing on the hypothetical interaction site as described by [19], two positions within the binding region of VEG8 compared to Tcp13 and 17 are altered. Residues A28T and D34A represent the only changes in a stretch of 13 amino acids. Especially the change from aspartic acid to alanine at position 34 seems rather drastic. Taking into account, that the A domain interaction sites are unaffected (Figure 2B), this mutation might play a key role in the observed differences of MbtH performance. However, only a thorough analysis of both complementing docking sites, paired with extensive mutational studies, could lead to a conclusive answer.

Finally, the narrow range substrate profiles of the selected enzymes were determined. Overall, every module, independent of domain setup or module placement, has significant activity towards L- and D- hpg, and in most cases also exhibits L-pg adenylation activity. However, only Hpg25 shows traceable activity without MbtH addition, underlining the essential role of MbtH proteins in the substrate activation process. Depending upon the MbtH homolog used, an effect on activity, non-linear with co-expression performance, can be observed. This suggests that there is at least a dual role of MbtH proteins, i.e., chaperoning the expression/folding and promoting the activity.

Despite MbtH homologs, also module placement within the NRPS seems to influence the kinetic properties *in vitro*, best illustrated with the Initiation module Tcp9 (Table 2). The reason for this may be in the independence of adjacent, upstream domains and modules, allowing for a higher degree of flexibility, thus more rapid conformational changes of the A_{sub} domain. It could further suggest, that due to the elevated adenylation rates, initiation modules could serve as gatekeepers for the NRPS. Product formation cycles may not be initiated without activation of the first module, which is also indicated by the kinetic intermediate step analysis of GrsA [39]. Foremost, in contrast to previous findings [49], we demonstrate the essential role of MbtH homologues in the activation of hpg specific adenylation domains.

Conclusions and Perspectives

We have successfully demonstrated how NRPS modules, activating a non-proteinogenic compound, can be identified, overexpressed and characterized. Through bioinformatics tools, alongside a non-radiolabel, multi-well, pyrophosphate exchange assay, a medium to high throughput experimental setup was established. Due to a high degree of conservation among enzymes of the ANL superfamily, including NRPS adenylation domains, this



setup allows for the identification and characterization of novel domains, activating any of the over 500 known substrates. With respect to the rapid improvements in engineering techniques [50] for multi-modular enzymes, this method provides a fast and straight forward way of increasing the target pool of domains. By using this pool, this could give rise to novel NRPS systems, which are able to efficiently synthesize customized compounds with new application spectra and contribute to ever growing demand of novel bioactive compounds, including the thriving search for new, alternative antibiotics.

Acknowledgements

The Authors would like to give their gratitude to Christian Rausch, for the work done on the development of adenylation domain prediction tools and the identification of potential targets for this study. This work was supported by the BE-Basic Foundation, an international public–private partnership.

References

1. Bode HB, Brachmann AO, Jadhav KB, Seyfarth L, Dauth C, Fuchs SW, et al. Structure elucidation and activity of kolossin A, the D-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase. *Angew Chem Int Ed Engl. Germany*; 2015;54: 10352–10355. doi:10.1002/anie.201502835
2. Von Döhren H, Dieckmann R, Pavla-Vrancic M. The nonribosomal code. *Chem Biol*. 1999;6: 273–279.
3. Ku J, Mirmira RG, Liu L, Santi D V. Expression of a functional non-ribosomal peptide synthetase module in *Escherichia coli* by coexpression with a phosphopantetheinyl transferase. *Chem Biol*. 1997;4: 203–207. doi:10.1016/S1074-5521(97)90289-1
4. Neville C, Murphy A, Kavanagh K, Doyle S. A 4'-phosphopantetheinyl transferase mediates non-ribosomal peptide synthetase activation in *Aspergillus fumigatus*. *Chembiochem*. 2005;6: 679–85. doi:10.1002/cbic.200400147
5. Reger AS, Wu R, Dunaway-Mariano D, Gulick AM. Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry*. 2008;47: 8016–25. doi:10.1021/bi800696y
6. Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT. Crystal structure of DltA. Implications for the reaction mechanism of nonribosomal peptide synthetase adenylation domains. *J Biol Chem*. 2008;283: 32484–91. doi:10.1074/jbc.M800557200
7. Marahiel MA, Stachelhaus T, Mootz HD. Modular peptide synthetases involved in nonribosomal peptide synthesis.

- Chem Rev. American Chemical Society; 1997;97: 2651–2674.
doi:10.1021/cr960029e
8. Weber T, Marahiel MA. Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*. 2001;9: R3–R9.
doi:10.1016/S0969-2126(00)00560-8
 9. Challis GL, Naismith JH. Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol*. 2004;14: 748–756. doi:10.1016/j.sbi.2004.10.005
 10. Balibar CJ, Vaillancourt FH, Walsh CT. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol*. 2005;12: 1189–200.
doi:10.1016/j.chembiol.2005.08.010
 11. Teruya K, Tanaka T, Kawakami T, Akaji K, Aimoto S. Epimerization in peptide thioester condensation. *J Pept Sci*. 2012;18: 669–77. doi:10.1002/psc.2452
 12. Haslinger K, Peschke M, Brieke C, Maximowitsch E, Cryle MJ. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature*. 2015;521: 105–9. doi:10.1038/nature14141
 13. Mofid MR, Finking R, Essen LO, Marahiel MA. Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism. *Biochemistry*. 2004;43: 4128–36.
doi:10.1021/bi036013h
 14. Buchko GW, Kim CY, Terwilliger TC, Myler PJ. Solution structure of Rv2377c-founding member of the MbtH-like protein family. *Tuberculosis*. Elsevier Ltd; 2010;90: 245–251.
doi:10.1016/j.tube.2010.04.002
 15. Boll B, Taubitz T, Heide L. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 2011;286: 36281–36290.
doi:10.1074/jbc.M111.288092
 16. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, et al. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. American Chemical Society; 2010;49: 8815–7.
doi:10.1021/bi1012854
 17. Zhang W, Heemstra JR, Walsh CT, Imker HJ. Activation of the pacidamycin pacl adenylation domain by MbtH-like proteins. *Biochemistry*. 2010;49: 9946–9947. doi:10.1021/bi101539b
 18. Davidsen JM, Bartley DM, Townsend CA. Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein Nocl. *J Am Chem Soc*. 2013;135: 1749–59.
doi:10.1021/ja307710d
 19. Herbst D a., Boll B, Zocher G, Stehle T, Heide L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenyating enzymes. *J Biol Chem*. 2013;288: 1991–2003.
doi:10.1074/jbc.M112.420182
 20. Baltz RH. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J Ind Microbiol Biotechnol*. 2011;38: 1747–1760.
doi:10.1007/s10295-011-1022-8
 21. Tatham E, Sundaram Chavadi S, Mohandas P, Edupuganti UR, Angala SK,

- Chatterjee D, et al. Production of mycobacterial cell wall glycopeptidolipids requires a member of the MbtH-like protein family. *BMC Microbiol.* 2012;12: 118. doi:10.1186/1471-2180-12-118
22. Wolpert M, Gust B, Kammerer B, Heide L. Effects of deletions of mbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiol.* 2007;153: 1413–23. doi:10.1099/mic.0.2006/002998-0
 23. Boll B, Heide L. A domain of RubC1 of rubradirin biosynthesis can functionally replace MbtH-like proteins in tyrosine adenylation. *ChemBioChem.* 2013;14: 43–44. doi:10.1002/cbic.201200633
 24. Uguru GC, Milne C, Borg M, Flett F, Smith CP, Micklefield J. Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. *J Am Chem Soc.* 2004;126: 5032–3. doi:10.1021/ja048778y
 25. Zhang K, Nelson KM, Bhuripanyo K, Grimes KD, Zhao B, Aldrich CC, et al. Engineering the substrate specificity of the DhbE adenylation domain by yeast cell surface display. *Chem Biol.* Elsevier Ltd; 2013;20: 92–101. doi:10.1016/j.chembiol.2012.10.020
 26. Wang M, Zhao H. Characterization and engineering of the adenylation domain of a NRPS-like protein: a potential biocatalyst for aldehyde generation. *ACS Catal.* 2014;4: 1219–1225. doi:10.1021/cs500039v
 27. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* 1999;6: 493–505. doi:10.1016/S1074-5521(99)80082-9
 28. Watanabe K, Oguri H, Oikawa H. Diversification of echinomycin molecular structure by way of chemoenzymatic synthesis and heterologous expression of the engineered echinomycin biosynthetic pathway. *Curr Opin Chem Biol.* 2009;13: 189–196. doi:10.1016/j.cbpa.2009.02.012
 29. Thirlway J, Lewis R, Nunns L, Al Naakeb M, Styles M, Struck AW, et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew Chemie — Int Ed.* 2012;51: 7181–7184. doi:10.1002/anie.201202043
 30. Kries H, Niquille DL, Hilvert D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem Biol.* 2015;22: 640–8. doi:10.1016/j.chembiol.2015.04.015
 31. Tanovic A, Samel SA, Essen L-O, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science.* 2008;321: 659–63. doi:10.1126/science.1159850
 32. Lee TV, Johnson LJ, Johnson RD, Koulman A, Lane GA, Lott JS, et al. Structure of a eukaryotic nonribosomal peptide synthetase adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J Biol Chem.* 2010;285: 2415–27. doi:10.1074/jbc.M109.071324
 33. Mitchell CA, Shi C, Aldrich CC, Gulick AM. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry.* 2012;51: 3252–3263.
 34. Beer R, Herbst K, Ignatiadis N, Kats I, Adlung L, Meyer H, et al. Creating

- functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol Biosyst.* 2014;10: 1709–18. doi:10.1039/c3mb70594c
35. Baldwin JE, Shiau CY, Byford MF, Schofield CJ. Substrate specificity of L-delta-(alpha-aminoadipoyl)-L-cysteiny-D-valine synthetase from *Cephalosporium acremonium*: demonstration of the structure of several unnatural tripeptide products. *Biochem J.* 1994;301, Pt 2: 367–72.
 36. Roongsawang N, Siew PL, Washio K, Takano K, Kanaya S, Morikawa M. Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. *FEMS Microbiol Lett.* 2005;252: 143–151. doi:10.1016/j.femsle.2005.08.041
 37. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol.* 2007;7: 78. doi:10.1186/1471-2148-7-78
 38. Boettger D, Bergmann H, Kuehn B, Shelest E, Hertweck C. Evolutionary imprint of catalytic domains in fungal PKS-NRPS hybrids. *ChemBioChem.* 2012;13: 2363–2373. doi:10.1002/cbic.201200449
 39. Sun X, Li H, Alfermann J, Mootz HD, Yang H. Kinetics profiling of gramicidin S synthetase A, a member of nonribosomal peptide synthetases. *Biochemistry.* 2014;53: 7983–9. doi:10.1021/bi501156m
 40. Hubbard BK, Thomas MG, Walsh CT. Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol.* 2000;7: 931–942. doi:10.1016/S1074-5521(00)00043-0
 41. Hoertz AJ, Hamburger JB, Gooden DM, Bednar MM, McCafferty DG. Studies on the biosynthesis of the lipopeptide antibiotic ramoplanin A2. *Bioorganic Med Chem.* Elsevier Ltd; 2012;20: 859–865. doi:10.1016/j.bmc.2011.11.062
 42. Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, Kucherov G. NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* 2008;36: D326–31. doi:10.1093/nar/gkm792
 43. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39: W29–37. doi:10.1093/nar/gkr367
 44. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. *Nucleic Acids Res.* 2015;43: W30–8. doi:10.1093/nar/gkv397
 45. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* 2011;39: W362–7. doi:10.1093/nar/gkr323
 46. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* England; 2016;44: D279–85. doi:10.1093/nar/gkv1344
 47. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using

- transductive support vector machines (TSVMs). *Nucleic Acids Res. England*; 2005;33: 5799–5808. doi:10.1093/nar/gki885
48. McMahon MD, Rush JS, Thomas MG. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in mycobacterium tuberculosis. *J Bacteriol.* 2012;194: 2809–2818. doi:10.1128/JB.00088-12
49. Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W. The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. *FEMS Microbiol Lett.* 2006;262: 85–92. doi:10.1111/j.1574-6968.2006.00368.x
50. Doekel S, Coëffet-Le Gal M-F, Gu J-Q, Chu M, Baltz RH, Brian P. Nonribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology.* 2008;154: 2872–80. doi:10.1099/mic.0.2008/020685-0
51. Sosio M. Organization of the teicoplanin gene cluster in *Actinoplanes teichomycticus*. *Microbiology.* 2004;150: 95–102. doi:10.1099/mic.0.26507-0
52. Pootoolal J, Thomas MG, Marshall CG, Neu JM, Hubbard BK, Walsh CT, et al. Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc Natl Acad Sci U S A.* 2002;99: 8962–7. doi:10.1073/pnas.102285099
53. Sosio M, Stinchi S, Beltrametti F, Lazzarini A, Donadio S. The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. *Chem Biol.* 2003;10: 541–549. doi:10.1016/S1074-5521(03)00120-0
54. Recktenwald J, Shawky R, Puk O, Pfening F, Keller U, Wohlleben W, et al. Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology.* 2002;148. doi:10.1099/00221287-148-4-1105
55. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature.* 2002;417: 141–7. doi:10.1038/417141a
56. Chiu HT, Hubbard BK, Shah AN, Eide J, Fredenburg RA, Walsh CT, et al. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci U S A.* 2001;98: 8548–53. doi:10.1073/pnas.151246498
57. Banik JJ, Brady SF. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc Natl Acad Sci U S A.* 2008;105: 17273–7. doi:10.1073/pnas.0807564105
58. van Wageningen AA, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, et al. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol.* 1998;5: 155–162. doi:10.1016/S1074-5521(98)90060-6
59. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23: 2947–8. doi:10.1093/bioinformatics/btm404

Supplementary data

Supplementary data Table 1 — First selection of targets.

Stachelhaus code was determined in NRPSpredictor. Targets for final selection are highlighted in grey.

organism	enzyme	module
<i>Actinoplanes</i> ATCC 33076	NRPS	1
<i>Actinoplanes</i> ATCC 33076	NRPS	4
<i>Actinoplanes</i> ATCC 33076	NRPS	5
<i>Actinoplanes</i> ATCC 33076	NRPS	2
<i>Actinoplanes</i> ATCC 33076	NRPS	4
<i>Actinoplanes</i> ATCC 33076	NRPS	8
<i>Chlorobium phaeobacteroides</i> DSM 266	Amino acid adenylation domain	1
<i>Herpetosiphon aurantiacus</i> ATCC 23779	Amino acid adenylation domain	1
<i>Herpetosiphon aurantiacus</i> ATCC 23779	Amino acid adenylation domain	2
<i>Streptomyces griseus</i> subsp. <i>Griseus</i> JCM 4626	NRPS	1
<i>Adineta vaga</i>	NRPS-like protein	3
Uncultured soil bacterium	NRPS	1
Uncultured soil bacterium	NRPS	1
Uncultured soil bacterium	NRPS	1
Uncultured soil bacterium	NRPS	2
Uncultured soil bacterium	NRPS	1
Uncultured soil bacterium	NRPS	1
Uncultured soil bacterium	NRPS	2
<i>Bacillus mycoides</i> Rock 1–4	ATP-dependent leucine adenylation	1
<i>Dickeya dadantii</i> (strain Ech703)	Amino acid adenylation domain	1
<i>Chitinophaga pinensis</i> (strain ATCC 43595)	AMP-dependent synthetase and ligase	1
<i>Streptomyces</i> sp. ACT-1	Amino acid adenylation domain	1
<i>Streptomyces roseosporus</i> NRRL 15998	NRPS	2
<i>Streptomyces lividans</i> TK24	NRPS	5
<i>Segniliparus rotundus</i> (strain DSM 44985)	Amino acid adenylation domain	1
<i>Amycolatopsis orientalis</i>	NRPS	1
<i>Amycolatopsis orientalis</i>	NRPS	2
<i>Streptomyces fungicidicus</i>	NRPS	2
<i>Streptomyces fungicidicus</i>	NRPS	4
<i>Streptomyces fungicidicus</i>	NRPS	8
<i>Streptomyces fungicidicus</i>	NRPS	1
<i>Streptomyces fungicidicus</i>	NRPS	4
<i>Streptomyces fungicidicus</i>	NRPS	5
<i>Shewanella denitrificans</i> (strain DSM 15013)	Amino acid adenylation domain	1
<i>Frankia</i> sp. (strain Ccl3)	Amino acid adenylation domain	3
<i>Nocardia farcinica</i>	NRPS	1
<i>Actinoplanes teichomyceticus</i>	NRPS	1
<i>Actinoplanes teichomyceticus</i>	NRPS	2
<i>Actinoplanes teichomyceticus</i>	NRPS	1
<i>Actinoplanes teichomyceticus</i>	NRPS	1
<i>Actinoplanes teichomyceticus</i>	NRPS	2
<i>Actinoplanes teichomyceticus</i>	NRPS	1
<i>Nonomuraea</i> sp. ATCC 39727	NRPS	1
<i>Nonomuraea</i> sp. ATCC 39727	NRPS	1
<i>Nonomuraea</i> sp. ATCC 39727	NRPS	2
<i>Streptomyces toyocaensis</i>	NRPS	1
<i>Streptomyces toyocaensis</i>	NRPS	1
<i>Streptomyces toyocaensis</i>	NRPS	2
<i>Amycolatopsis mediteranei</i> DSM 5908	NRPS	1
<i>Amycolatopsis mediteranei</i> DSM 5908	NRPS	2
<i>Streptomyces lavendulae</i>	NRPS	1
<i>Streptomyces lavendulae</i>	NRPS	1
<i>Streptomyces lavendulae</i>	NRPS	2
<i>Streptomyces lavendulae</i>	NRPS	1
<i>Streptomyces lavendulae</i>	NRPS	1
<i>Streptomyces lavendulae</i>	NRPS	1
<i>Streptomyces coelicolor</i>	NRPS	6
<i>Streptomyces coelicolor</i>	NRPS	1
<i>Streptomyces roseosporus</i> NRRL 11379	NRPS	2

accession code	Stachelhaus motif	predicted specificity	module code
gi 112084466 gb ABI05682.1 _m1	DAYHLGLLCK	D-hydroxyphenylglycine	
gi 112084466 gb ABI05682.1 _m4	DAYHLGLLCK	D-hydroxyphenylglycine	
gi 112084466 gb ABI05682.1 _m5	DAYHLGLLCK	D-hydroxyphenylglycine	
gi 112084469 gb ABI05683.1 _m2	DAFHLGLLCK	D-hydroxyphenylglycine	
gi 112084469 gb ABI05683.1 _m4	DAYHLGLLCK	D-hydroxyphenylglycine	
gi 112084469 gb ABI05683.1 _m8	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_A1BDX6_m1	DVYYLGGICK	L/D-hydroxyphenylglycine	
lcl UniRef100_A9AUJ0_m1	DIFHFLGIK	L/D-hydroxyphenylglycine	
lcl UniRef100_A9B4Q2_m2	DVYYLGGICK	L/D-hydroxyphenylglycine	
lcl UniRef100_B1W2Q4_m1	DMYHLGLMDK	L/D-hydroxyphenylglycine	
lcl UniRef100_B3G4H6_m3	DIQHLVLLVK	L/D-hydroxyphenylglycine	
lcl UniRef100_B7T1B9_m1	DAFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_B7T1C0_m1	DAFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_B7T1C1_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg25/27
lcl UniRef100_B7T1C1_m2	DAVHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_B7T1D0_m1	DAFHLGLLCK	L/D-hydroxyphenylglycine	
lcl UniRef100_B7T1D2_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg23
lcl UniRef100_B7T1D2_m2	DALHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_C3AKA8_m1	DARHLVLLAK	L/D-hydroxyphenylglycine	
lcl UniRef100_C6CCJ1_m1	DIFHFLGIK	L/D-hydroxyphenylglycine	
lcl UniRef100_C7PT83_m1	DARHLALLVK	L/D-hydroxyphenylglycine	
lcl UniRef100_D1XA32_m1	DMYHLGLMDK	L/D-hydroxyphenylglycine	
lcl UniRef100_D6AU62_m2	DIYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_D6ES40_m5	DVYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_D6ZFQ6_m1	DAFHPGFLAK	L/D-hydroxyphenylglycine	
lcl UniRef100_O52820_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg24
lcl UniRef100_O52820_m2	DAVHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y21_m2	DAYHLGMLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y21_m4	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y21_m8	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y22_m1	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y22_m4	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q06Y22_m5	DAYHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q12IB7_m1	DVRHLALLAK	L/D-hydroxyphenylglycine	
lcl UniRef100_Q2JA75_m3	DAWHLGLMCK	D-hydroxyphenylglycine	
lcl UniRef100_Q5Z1T0_m1	DALHPGHVCK	L/D-hydroxyphenylglycine	
lcl UniRef100_Q6ZZJ4_m1	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q6ZZJ4_m2	DALHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q6ZZJ6_m1	DAFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q70AZ7_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg22
lcl UniRef100_Q70AZ7_m2	DALHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q70AZ9_m1	DAFHLGLLCK	D-hydroxyphenylglycine	Tcp9
lcl UniRef100_Q7WZ66_m1	DAFHLGLLCK	D-hydroxyphenylglycine	Hpg3/4
lcl UniRef100_Q7WZ74_m1	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q7WZ74_m2	DALHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q8KLL3_m1	DAFHLGLLCK	D-hydroxyphenylglycine	Hpg1
lcl UniRef100_Q8KLL5_m1	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q8KLL5_m2	DAFHLGLLCK	D-hydroxyphenylglycine	Hpg21
lcl UniRef100_Q939Z0_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg11
lcl UniRef100_Q939Z0_m2	DAVHLGLLCK	D-hydroxyphenylglycine	Hpg13/14
lcl UniRef100_Q93N86_m1	DAFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q93N87_m1	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q93N87_m2	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q93N88_m1	DIFHLGLLCK	D-hydroxyphenylglycine	Hpg20/26
lcl UniRef100_Q93N89_m1	DIFHLGLLCK	D-hydroxyphenylglycine	
lcl UniRef100_Q924X6_m6	DVYHLGLLCK	D-hydroxyphenylglycine	Hpg19
lcl UniRef100_UPI0001AF6D2A_m1	DALHPGHVCK	L/D-hydroxyphenylglycine	
lcl UniRef100_UPI0001AF7653_m2	DIYHLGLLCK	D-hydroxyphenylglycine	

Supplementary data Table 2 — Overview of all over expression experiments.

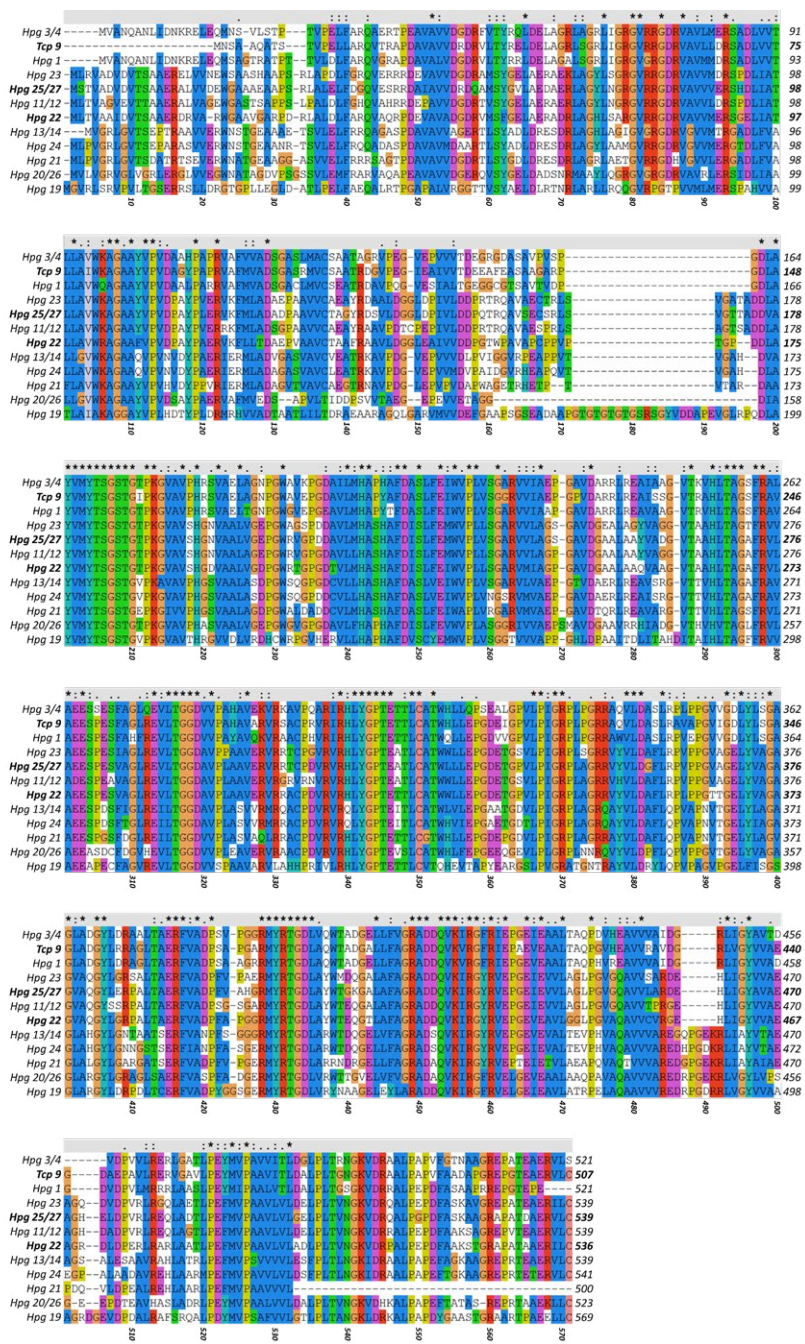
S is soluble, I insoluble expression of the construct; N/A not applicable; *MbtH* indicates the *MbtH* variants which were used for co-expression trials. Underlined *MbtH* enabled activity with the dedicated construct.

Construct	pSCI 242/243		pMAL-c5x		<i>MbtH</i>
	S	I	S	I	
<i>Tcp9</i>	n/a	n/a	Yes	Yes	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg1</i>	n/a	n/a	Yes	Yes	<u>Tcp13</u> ; <u>Veg8</u>
<i>Hpg3</i>	n/a	n/a	Yes	Yes	<u>Tcp13</u> ; <u>Veg8</u>
<i>Hpg11</i>	No	No	Yes	Yes	<u>Tcp13</u> ; <u>Bps</u>
<i>Hpg12</i>	No	No	Yes	Yes	<u>Tcp13</u>
<i>Hpg13</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u>
<i>Hpg14</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u>
<i>Hpg15</i>	Yes	Yes	N/A	N/A	N/A
<i>Hpg16</i>	Yes	Yes	N/A	N/A	N/A
<i>Hpg17</i>	Yes	Yes	N/A	N/A	N/A
<i>Hpg18</i>	No	Yes	N/A	N/A	N/A
<i>Hpg19</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u> ; <u>CDAI</u>
<i>Hpg20</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u> ; <u>Com</u>
<i>Hpg21</i>	No	No	Yes	Yes	<u>Tcp13</u>
<i>Hpg22</i>	No	No	Yes	Yes	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg23</i>	No	No	Yes	Yes	<u>Tcp13</u> ; <u>Teg</u>
<i>Hpg24</i>	No	No	Yes	Yes	<u>Tcp13</u>
<i>Hpg25</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg26</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u> ; <u>Com</u>
<i>Hpg27</i>	Yes	Yes	Yes	Yes	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>

Supplementary table 3 — Overview of the adenylation activity determinations for the different amino acid substrates and the different combinations of adenylation domain constructs with co-purified *MbtH*-like proteins.

*MbtH*s enabling any adenylation activity are underlined. N/A not applicable.

Construct	L-hpg	D-hpg	L-pg	D-pg	L-phe	<i>MbtH</i>
<i>Tcp9</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg1</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Veg8</u>
<i>Hpg3</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Veg8</u>
<i>Hpg11</i>	Yes	No	No	No	No	<u>Tcp13</u> ; <u>Bps</u>
<i>Hpg12</i>	N/A	N/A	N/A	N/A	N/A	N/A
<i>Hpg13</i>	Yes	No	No	No	No	<u>Tcp13</u>
<i>Hpg14</i>	Yes	No	No	No	No	<u>Tcp13</u>
<i>Hpg15</i>	N/A	N/A	N/A	N/A	N/A	N/A
<i>Hpg16</i>	N/A	N/A	N/A	N/A	N/A	N/A
<i>Hpg17</i>	N/A	N/A	N/A	N/A	N/A	N/A
<i>Hpg18</i>	N/A	N/A	N/A	N/A	N/A	N/A
<i>Hpg19</i>	Yes	No	No	No	No	<u>Tcp13</u> ; <u>CDAI</u>
<i>Hpg20</i>	Yes	No	No	No	No	<u>Tcp13</u> ; <u>Com</u>
<i>Hpg21</i>	Yes	Yes	No	No	No	<u>Tcp13</u>
<i>Hpg22</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg23</i>	Yes	No	No	No	No	<u>Tcp13</u> ; <u>Teg</u>
<i>Hpg24</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u>
<i>Hpg25</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>
<i>Hpg26</i>	Yes	No	No	No	No	<u>Tcp13</u> ; <u>Com</u>
<i>Hpg27</i>	Yes	Yes	Yes	No	No	<u>Tcp13</u> ; <u>Tcp17</u> ; <u>Veg8</u>



Supplementary figure 1 — Alignment of all adenylation domains in this study. ClustalX was used to align the sequences of all tested adenylation domains.

