Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M. Sc. Anna Degen

born in Püttlingen, Germany

Oral examination date

29.03.2019

Exploring the use of a blue pigment-producing NRPS as a tagging method to easily detect engineered NRPs

Referees:

Prof. Dr. Benedikt Brors

Prof. Dr. Barbara Di Ventura

Acknowledgement

The first thank you goes to **Barbara DiVentura**, my boss,
who was an enthusiastic group leader I came across.
She offered me a challenging synbio project,
which I accepted and enjoyed to dissect.
We investigated NRPSs, these fascinating enzymes,
to re-engineer them, which was difficult at times.
That's why we developed a readout showing blue color,
and for many months, that's what we looked for.
When we finally found it, we thought it was great,
but we also raised more questions to investigate.
"How do the modules communicate with each other?"
became the next question we aimed to bother.
That's why we went *in vitro* and what we found next,
was that they all behave differently, depending on their context.
Now you are a professor of "molecular and cellular engineering"
I wish the best of luck to you and the lab you are steering.

Thank you, **Benedikt Brors**, that you agree,
to take over as my first referee.
Thank you as well for connection to **Aubry Miller**, it was a pleasure,
to visit his lab and record an important last minute measure.
"Thank you!", **Matthias Mayer** and **Stefan Pfeffer**, I want to say,
for joing my defense as examiners without any delay.

The next thanks go out to the members of the committee,
who once a year gave advice regarding the thesis to me.
Thank you to my first referee for three years, **Roland Eils**,
who is now director of the BIH in Berlin, away many miles.
On a great ski retreat after my talk he was stating
"that's why I switched to mathematics, biology was too frustrating".
Thank you, **Jörg Hoheisel**, for scientific and personal support and the decision
to temporarily host me in your DKFZ devision!
Thank you, **Max Cryle**, your expertise on NRPS
helped me immensely to progress.
Also thank you for keeping the contact and support
even after you have moved abroad.

Thank you, **Tiia Kittilä**, for the help with the assay of PP_i release,
so I could perform it and analyze the data at ease.
Thank you **Richard Wombacher**, **Tobias Timmermann**, **Heiko Rudy** who are
at the IPMB in Heidelberg and provided MS measurements and NMR.
Thank you, **Henning D. Mootz** and **Florian Meyerthaler**, for the collaboration
that hopefully soon results in a joint publication.

A big thank you to all former and current members of the DiVentura Group,
most of whom left for Freiburg in 2017 at a swoop.
In particular **Edo**, **Eethan**, **Enoch**, **Daniel**, **Dominik**, **Linda** and **Pierre**,
we had a great (lab-) life, while you were there!
I am happy that we do still share good times,

for example around Christmas with personal rimes.
To Pierre: “let’s still try it, it makes so much sense,
let’s set a date for our back-to-back defense!”
Dominik Niopek, thank you for the accommodation
in your lab, so I did not have to again change location.
I wish you continued success in the coming year
So you can smoothly continue your scientific career!
Mareike and **Sabine**, thank you for the nice atmosphere,
also the **Conrad Lab** for the shared happy hours with cake or beer.
Chiara, David and Farzaneh and Stefan for sharing the office
and digesting science, stories and lunch over coffice!
Wherever I go, I will miss you much,
so please, everyone, do stay in touch!

Auch an **Corinna Sprengart** und **Manuela Schäfer** geht ein großes Dankeschön,
So kompetent und freundlich, da freut jedes Wiedersehn.
Mein Dank gilt auch **Christiane Graffy**, die das BioQuant tapfer zusammenhält
Und **Herrn Marinkovic**, der viel repariert, sodass es nicht zusammenfällt.

Und jetzt mal ab von der Arbeit, will ich diejenigen grüßen,
die mir seit jeher eher die Freizeit versüßen.
Dazu gehören Freunde und Familie, vor allem aus dem schönen Saarland,
danke für die Ermunterung damals und seit ich nach Heidelberg verschwand.
Thank you as well to **Cheryl** and **John**, my hostparents from America,
it is always fun to skype and meet again, near or far.
Danke auch an die Freunde, die ich in Heidelberg gewonnen,
seit wir vor zehn Jahren unseren Bachelor hier begonnen.
Trotzdem man sich im Studium plagte,
könnte wahr sein, was mal jemand sagte:
„irgendwann werden wir zugeben,
das war die beste Zeit in unserem Leben“.
So erfreut zum Beispiel das Beantworten von Fragen
beim Quiz im Mohr jede Woche an Montagen.
Manche bleiben hier, andere zieht es in die Welt,
ich hoffe es klappt dass man sich immer wieder gesellt.

Mein tiefster Dank gilt meiner Mutter, **Christina Degen**,
du hast mir immer den Rückhalt gegeben,
kleinere und größere Schwierigkeiten zu meistern
und mich für viele Sachen zu begeistern.
Ich bin so froh, denn ich kann immer auf dich bauen,
es gibt nichts Besseres als dieses Urvertrauen.
Auch meinem Vater, **Norbert Degen**, möchte ich danken,
für die Unterstützung jeglicher Art und das liebgemeinte Zanken.

Schließlich danke ich dir, lieber **Daniel**, du warst ein Freund und wirst mein Mann,
ich genieße die Zeit mit dir und freue mich, dass ich das auch in Zukunft kann.
Ob auf 12m^2 in Cambridge, 30 Stunden in einem Zug aus den 70er Jahr‘n,
mit dem Sportauto auf einem Pfad für Landmaschinen oder auf einem Vulkan,
wir haben viel erlebt und es immer geschafft, aus Feuerland zurück ins kleine Haus,
so soll es bleiben! Wohin wir auch gehen, mit dir wird das Beste draus.

Abstract

Nonribosomal peptide synthetases (NRPSs) are modular mega-enzymes found in bacteria and fungi that produce nonribosomal peptides (NRPs) in an assembly line fashion. Each module is in charge of adding a specific amino acid (AA) to the growing peptide chain. Three basic domains constitute one NRPS module: the adenylation (A), peptidyl carrier protein (PCP) and condensation (C) domains. The A domain recognizes and activates the AA. An external enzyme, the PPtase, attaches a phosphopantetheine (PPant) arm to the PCP domain which then picks up the activated AA and delivers it to the C domain. The C domain recognizes the growing peptide chain (donor) as well as the new AA (acceptor) and fuses the two together. A special feature of NRPSs is their ability to recognize and incorporate not only proteinogenic AAs, but also other building blocks like fatty acids (FAs) or non-proteinogenic AAs. All building blocks can be further modified through the action of additional domains: epimerization (E), methylation (M) and oxidation (Ox) domains, among others. In this manner a great variety of different NRPs can be synthesized, many of which are bioactive and exhibit anti-microbial or anti-cancer properties. Thus, it is highly desirable to understand how NRPS domains and modules function and find ways to genetically re-engineer them for custom NRP production.

Since the discovery of NRPSs, many efforts have already been made to engineer these enzymes in order to create custom NRPs, but general design rules yet remain elusive. The successful attempts to re-create functional NRPS for the production of novel NRPs include: (i) mutations of the A domain, (ii) subdomain modifications and (iii) rearrangements on the module level. Yet, many engineered NRPSs exhibit only slow reaction rates and low product yields. In some cases, the desired NRP products cannot be detected at all, possibly due to additional control mechanisms that have not been taken into account during the engineering process, such as substrate specificity of C domains. Hence there is still a great need to identify the general rules for successful NRPS engineering in order to exploit the ever-growing molecular toolbox of newly discovered NRPSs for recombinant production of novel bioactive compounds.

In this work I present my attempts to develop an approach to easily monitor the outcome of NRPS manipulation using a pigment-producing synthetase as a genetic tag. To this end, I first investigated two homologous synthetases, IndC and BpsA, which produce the blue pigment indigoidine, and mutants thereof and revised the proposed biosynthesis mechanism. I then created a series of fusion constructs between modules coming from different NRPSs and IndC/BpsA to test if indigoidine-tagged peptides could be produced. I identified a promising construct for which point mutations in the upstream module resulted in weaker or null pigment production. However, the expected indigoidine-tagged AA was not detectable, which could be due to the fact that indigoidine production inevitably leads to the separation of the donor AA. These results raised further questions as to whether in a native NRPS, the same modifications lead to congruent effects in neighboring modules. I addressed this question using a fragment of a non-engineered NRPS to monitor the activity of the native and modified versions in an *in vitro* assay, which I present in the last part of the results. Surprisingly, the effects of the same set of modifications on neighboring modules did not only differ between the engineered NRP-pigment synthetase and the native NRPS, but also between different modules within the native NRPS. These results hint at highly individual behavior of NRPS modules, depending on the context they are in.

Zusammenfassung

Nichtribosomale Peptidsynthetasen (NRPSs) sind modulare Megaenzyme, die in Bakterien und Pilzen vorkommen und nichtribosomale Peptide (NRPs) in Fließband-Art produzieren. Jedes Modul ist dafür verantwortlich, der wachsenden Peptidkette eine spezifische Aminosäure (AA) hinzuzufügen. Drei grundlegende Domänen bilden ein NRPS-Modul: die Adenylierungs- (A-), Peptidyl-Carrier-Protein- (PCP-) und Kondensations- (C-)Domänen. Die A-Domäne erkennt und aktiviert die AA. Ein externes Enzym, die PPtase, bindet einen Phosphopantethein-Arm (PPant) an die PCP-Domäne, der dann die aktivierte AA aufnimmt und an die C-Domäne weiterleitet. Die C-Domäne erkennt die wachsende Peptidkette (Donor) sowie die neuen AA (Akzeptor) und verbindet die beiden miteinander. Das Besondere an NRPSs ist ihre Fähigkeit, nicht nur proteinogene AAs zu erkennen und einzubauen, sondern auch andere Bausteine wie Fettsäuren (FA) oder nicht-proteinogene AAs. Alle Bausteine können durch zusätzliche Domänen weiter modifiziert werden: dazu gehören die Epimerisierungs- (E-), Methylierungs- (M-) und Oxidations- (Ox-)Domänen. Auf diese Weise kann eine große Vielfalt verschiedener NRPs synthetisiert werden, von denen viele bioaktiv sind und antimikrobielle oder tumorunterdrückende Eigenschaften haben. Daher ist es äußerst wünschenswert, die Funktionsweise von NRPS Domänen und Modulen zu verstehen und Wege zu finden, um sie genetisch für die benutzerdefinierte NRP-Produktion neu zu konstruieren.

Seit der Entdeckung von NRPSs wurden bereits viele Anstrengungen unternommen, um diese Enzyme zu rekonstruieren, um benutzerdefinierte NRPs herzustellen, aber allgemeine Regeln dafür sind immer noch schwer aufzustellen. Experimentelle Ansätze funktionelle NRPSs für die Herstellung neuartiger NRPs zu rekonstruieren umfassen: (i) Mutationen der A-Domäne, (ii) Modifikationen von Subdomänen und (iii) Neuanordnungen ganzer Module. Viele neu angeordnete NRPSs weisen jedoch niedrige Reaktionsgeschwindigkeiten und geringe Produktausbeuten auf. In einigen Fällen konnte das gewünschte Produkt überhaupt nicht nachgewiesen werden, was möglicherweise auf zusätzliche Kontrollmechanismen, z.B. durch die C-domäne, zurückzuführen ist und bei der Neukonstruktion der NRPSs nicht berücksichtigt wurde. Es besteht daher nach wie vor ein großer Bedarf, die allgemeinen Regeln für ein erfolgreiches NRPS-Engineering zu ermitteln, um die ständig wachsende molekulare Toolbox neu entdeckter NRPS für die rekombinante Produktion neuartiger bioaktiver Verbindungen zu nutzen.

In dieser Arbeit zeige ich Ansätze zur Entwicklung einer Methode zur einfachen Evaluierung des Effekts von NRPS Manipulationen auf deren Aktivität anhand der Bildung eines farbigen Produkts durch eine genetisch an das modifizierte NRPS angehängten Pigmentsynthetase. Zu diesem Zweck untersuchte ich zwei homologe Synthetasen, IndC und BpsA, die das blaue Pigment Indigoidin produzieren. Die Analyse verschiedener Mutanten dieser Synthetasen führten zu der Überarbeitung des bisher anerkannten Biosynthesemechanismus. Anschließend erstellte ich eine Reihe von Fusionsenzyme aus Modulen verschiedener NRPSs und IndC/BpsA, um zu testen, ob mit Indigoidin markierte Peptide hergestellt werden. Eines der Fusionsenzyme war vielversprechend, da Punktmutationen im Upstream-Modul zu einer abgeschwächten oder gar keiner Pigmentproduktion führten. Jedoch ließ sich die erwartete mit Indigoidin markierte AA nicht nachweisen, was darauf zurückzuführen sein könnte, dass die Indigoidinproduktion zwangsläufig zur Abspaltung der Donor-AA führt. Diese Ergebnisse warfen weitere Fragen auf, z.B. ob in einem natürlichen NRPS die gleichen Mutationen zu den gleichen Effekten in

benachbarten Modulen führen. Um diese Frage zu beantworten, habe ich drei natürlich zusammenhängende Module eines NRPS *in vitro* untersucht. Dabei habe ich herausgefunden, dass sich die Auswirkungen der gleichen Modifikationen auf benachbarte Module nicht nur zwischen dem konstruierten NRP-Pigment Fusionsenzym und dem natürlich zusammenhängenden NRPS, sondern auch zwischen verschiedenen Modulen innerhalb des natürlichen NRPS, unterscheiden. Diese Ergebnisse deuten auf ein sehr individuelles Verhalten von NRPS-Modulen hin, je nachdem, in welchem Kontext sie sich befinden.

Contents

1. Introduction

1.1 Preface

“Antimicrobial resistance has become one of the biggest threats to global health and endangers other major priorities, such as human development.”

The United Nations (UN)[1] as well the World Health Organization (WHO)[2] have recognize this fact and issued a global action plan[3]. The five major objectives of this plan are: (i) to raise and increase awareness of antimicrobial resistance through more effective communication, education and training; (ii) to accumulate more knowledge through research; (iii) to decrease the number of infections through prevention measures; (iv) to limit the use of antibiotics to sensible cases in human and animal health and (v) to increase investment in research towards new medicines, diagnostic tools, vaccines and other interventions.

It takes a global effort to advance each of these objectives and step by step realize this action plan. I decided to take a little step towards the second objective (ii) during my PhD project and investigated non-ribosomal peptide synthetases (NRPSs), the assembly lines of many antibiotics and other bioactive compounds. They naturally occur in many bacteria and fungi and their products are among our “most important antibacterial, antifungal, antiviral, immunosuppressant, and anticancer drugs”[4]. Much research has been done towards finding, understanding and manipulating these mega-enzymes, the advances of which I will introduce in the first part of this thesis. At the same time, many aspects remain elusive. In this work I focus on the development of an NRPS tagging procedure for a visual readout of NRPS engineering suitable for high-throughput screening. In addition, I investigated the intramolecular cooperation within a native but truncated NRPS, in its original form and upon different local disruptions. My results contribute to a better understanding of NRPSs while at the same time raising more questions about these fascinating molecular machines.

With my PhD thesis, I take a tiny step towards finding out how NRPSs work in more detail. Of course, it is a tiny step, a little dent in what is already known about NRPSs and a tiny dent in the humongous circle of human knowledge, much like Matt Might nicely put it in his “illustrated guide to a Ph.D”[5]. Nonetheless, every tiny dent contributes to enlarging this circle of knowledge and may one day lead to, in our case, new medicines.

1.2 Non-ribosomal peptide synthetases and their natural products

Non-ribosomal peptide synthetases (NRPSs) produce, as their name implies, peptides. They are not the ribosome; hence they synthesize non-ribosomal peptides (NRPs). In contrast to the ribosome, which is the main peptide and protein producer across all forms of life, NRPSs synthesize their peptides independently from a DNA sequence-derived blueprint. This approach has the advantage that NRPSs are not limited to the 22 L-amino acids encoded by DNA but can choose from a tremendous pool of building blocks to form their peptides. Additionally, they cannot only form linear peptides, but can make them circular, branched or intertwined resulting in a vast variety of shapes that the peptides can adopt. With this variety of constituents and shapes, NRPs often have useful chemical properties and/or interfere with biological processes, such as cell-wall biosynthesis or enzyme activity. In addition, they are less vulnerable to degradation, e.g. by peptidases, than linear peptides. For this reason, many bacteria and fungi

afford to produce the large NRPSs to have an evolutionary advantage through the production of an NRP with a certain, helpful characteristic.

Of course, being enzymes, NRPSs and their products are not completely independent from the ribosome. As for every other enzyme, the ribosome builds NRPSs according to the genetic code. Thus, to ultimately change the sequence of an NRP, one needs to change the DNA that encodes its synthetase. What evolution has mastered resulting in the large diversity of NRPs that we find in nature today is much more complicated in the laboratory setting. While the genetic code has been deciphered and we exactly know the effect that changing the DNA sequence will have on the amino acid sequence, the non-ribosomal code has only been partially cracked.

It is known by now, how the molecular arrangement of NRPSs is linked to the sequence of their final product. Their modular setup and the contribution of each domain within these modules have largely been uncovered. However, despite these great advances, it remains difficult to predict which effect the change of individual amino acids or the rearrangement of whole domains within NRPSs have on the sequence of the NRP they produce. Thus, it remains difficult to rationally engineer NRPSs to produce custom peptides with changed bioactivity.

Alternative natural product producers

NRPSs are not the only bioactive natural product producers: Polyketide Synthetases (PKSs) equally contribute to natural bioactive compound production, following a similar pattern of modular substrate elongation. The elongation reaction is similar to fatty acid synthetases with the basic building blocks being actyl- and malonyl-CoA. NRPSs and PKSs can form complexes with each other. There are plenty of very comprehensive review articles featuring PKSs[6–9]. Another alternative route to bioactive natural peptides is represented by the ribosomally synthesized and post-translationally modified peptides (RiPPs)[10]. However, in this work I am covering and using primarily NRPSs, which I will introduce in more detail in the following sections, starting with basic, linear NRPSs.

1.2.1 *Linear NRPSs*

A basic, linear NRPS consists of an initiation module, followed by one or more elongation modules and is completed by a termination module. Each module is in charge of adding a specific building block to the growing natural product[11], the non-ribosomal peptide, NRP. Thus, the size and sequence of an NRP is dictated by the number and order of modules within the NRPS that produces it. Linear NRPSs can range in size from two[12] to 22[13] modules. These can be distributed across several distinct enzymes, which cooperate via communication domains (section 1.2.9). The building blocks they incorporate include the standard proteinogenic L-amino acids, but also fatty acids (FAs)[14–16] and non-proteinogenic AAs[17,18]. Through internal and external auxiliary domains (section 1.2.7), the structural and chemical variety of NRPs is further increased resulting in up to 500 different monomers found in natural products[19]. This great variety has evolved over millions of years to produce bioactive NRPs providing its host organism protection against e.g. other microbes[4,20]. However, in modern medicine, many NRPs have also been discovered to be anti-fungal, anti-viral, anti-cancer and immunosuppressant[4,20–24]. A particularly comprehensive overview of different NRPs, their structures, their host organism(s), their function and mode of action is provided by Süssmuth and Mainz[25].

Besides linear NRPSs, more complex versions exist, such as iterative and non-linear NRPSs (section 1.2.10)[23]. In the following sections, I will give an example of a well-studied linear NRPS and then introduce the general biosynthetic cycle of NRP formation before describing the obligatory and auxiliary domains and their roles in more detail.

1.2.2 *A well-studied example of a linear NRPS: the tyrocidine synthetase*

Here I want to highlight one example of a linear NRPS: the tyrocidine synthetase (**Figure 1A**), which produces the antimicrobial cyclic NRP tyrocidine (**Figure 1B**) in many Bacillus species[26].

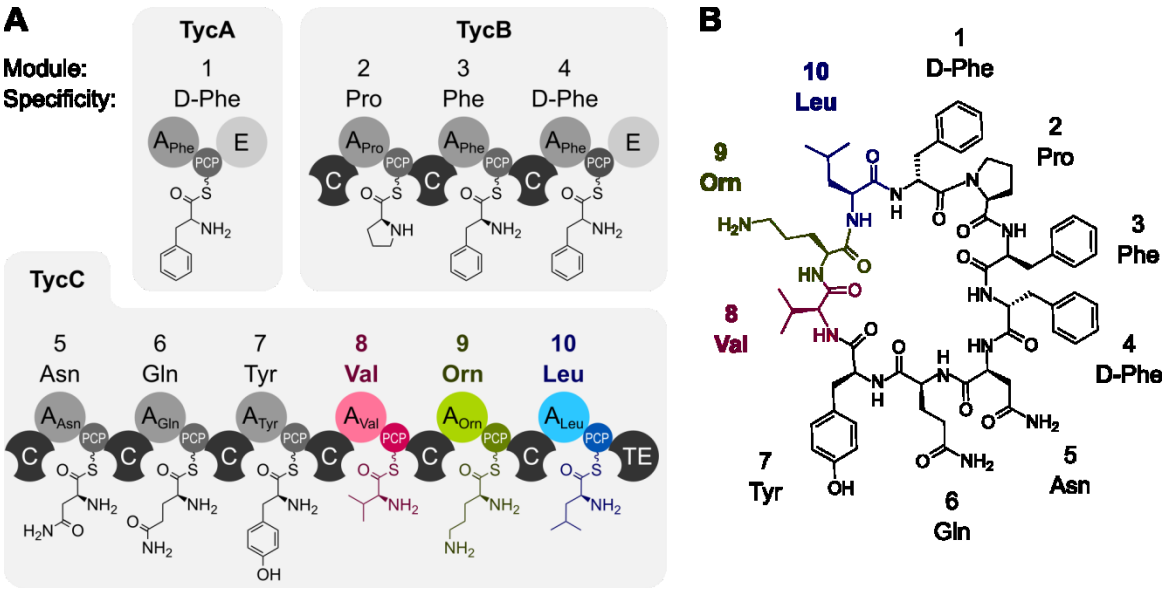

Figure 1 | The tyrocidine synthetase produces the antimicrobial cyclic peptide tyrocidine. (A) Schematic depiction of the three enzymes that make up the tyrocidine synthetase. Numbers refer to the modules. Modules used in this study are number 5 and 6, shown in gray and modules 8, 9 and 10, shown in color. Amino acids are depicted attached to the PPant arms. **(B)** Chemical structure of tyrocidine.

In 1963, the tyrocidine synthetase was the first NRPS to be discovered. Tatum and coworkers found that the cellular production of the cyclic peptide tyrocidine[26,27] was not affected after the addition of a ribosome inhibitor and thus an alternative synthesis pathway was inevitable[28]. Purification and *in vitro* study of the polyenzyme responsible for tyrocidine synthesis was reported in 1970[29,30]. Further analysis contributed to the concept of a modular setup of the NRPS with the number of modules matching the number of residues of tyrocidine[11]. In 1997, the complete DNA sequence of the tyrocidine biosynthesis gene cluster of *Brevibacillus Brevis* (ATCC 8185) and detailed biochemical analysis of each of its components was reported by Mootz and Marahiel[31] and extended by others[32–36]. In my thesis, I used parts of the tyrocidine synthetase for engineering an AA-pigment synthetase, namely modules 5 and 6 specific for asparagine and glutamine, and for the analysis of native NRPS modules *in vitro,* namely modules 8, 9 and 10 specific for valine, ornithine and leucine, highlighted in different colors (**Figure 1**).

1.2.3 *Biosynthetic cycle of NRP formation*

The very basic scheme of a linear NRPS is shown in **Figure 2**, including the initiation module, elongation module(s) and the termination module.

Each NRPS module is composed of multiple domains that cooperate to incorporate a specific building block into the growing peptide chain. A basic initiation module consists of an adenylation (**A**) domain, that specifically recognizes its substrate amino acid (here called X) and activates it via adenylation (**Figure 2**, 1.) of the carboxyl group while simultaneously releasing a pyrophosphate (PP_i). The second crucial domain within an initiation module is the peptide carrier protein (**PCP**) domain, also termed thiolation (T) domain. This domain requires a posttranslational modification to be functional. Specifically, a 4'-phosphopantetheine "arm" (PPant) is attached to a conserved serine within the PCP domain (**Figure 2**, 0.) by an external 4'-phosphopantetheinyl transferase (PPtase). This PPant arm serves as the point of attachment for the activated substrate via thiolation (**Figure 2**, 2.), which is shuttled to the elongation domain. In addition to the A and PCP domain, the elongation module also contains a condensation (**C**) domain, which facilitates peptide bond formation (**Figure 2**, 3.) between the upstream substrate ("donor", here X) and the downstream substrate ("acceptor", here Y), forming the intermediate product "XY". This cycle of activation, thiolation and condensation is repeated N times, adding one building block per elongation module. The last building block is added by the termination module producing the peptide "$XY_{[N]}Z$", before it reaches the thioesterase (TE) domain. The TE domain cleaves the thioester bond between the intermediate substrate and the upstream PCP domain and temporarily binds it at a conserved serine residue via an acyl-O-TE intermediate. Release of the final product is facilitated through a nucleophilic attack either by water or an intramolecular nucleophile resulting in either a linear or a cyclic NRP, respectively (**Figure 2**, 4.). In the following sections each basic and several additional domains will be described in more detail.

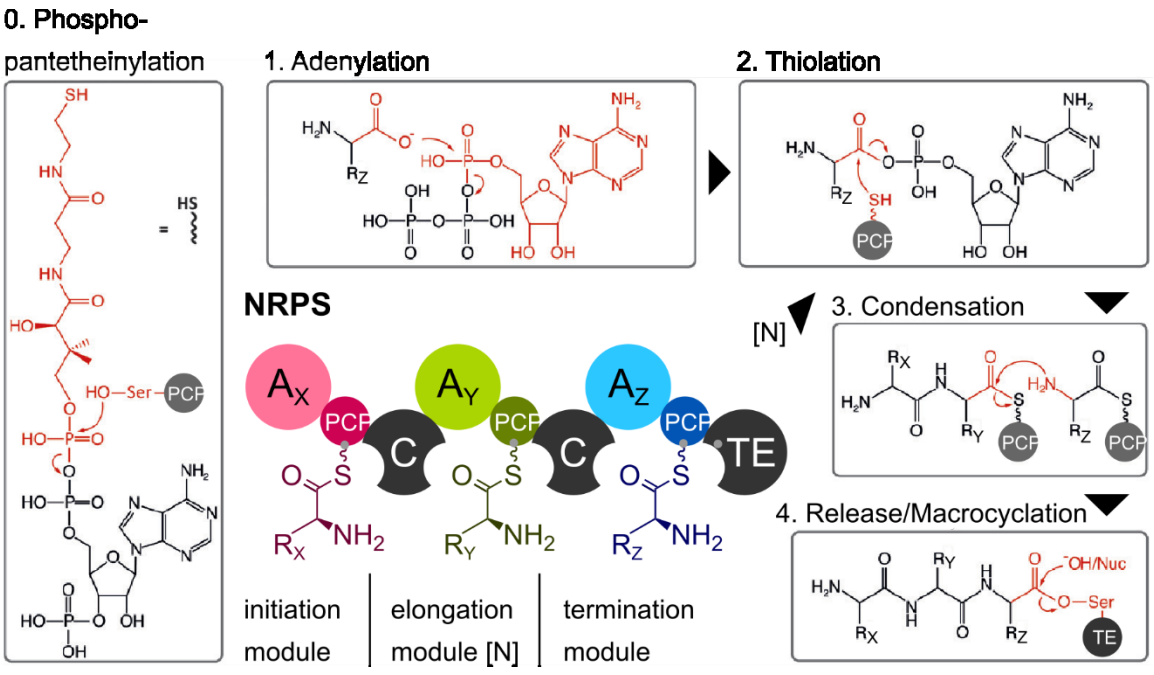

Figure 2 | The biosynthetic cycle of NRP formation. A standard NRPS consists of an initiation, one or more elongation and a final termination module. Each module again consists of various domains, each playing a crucial role in the NRP formation process. $A_{X, Y, Z}$ - Adenylation domains specific for amino acids with sides chains X, Y and Z; PCP - Peptide Carrier Protein (= Thiolation domain - T); C - Condensation domain; TE - Thioesterase domain. Conserved serine residues within PCP and TE domains are highlighted as little grey dots. This figure was modified from Süssmuth and Mainz[23] with permission of the publisher.

1.2.4 *Adenylation (A) domains*

A domains recognize and activate their substrates via adenylation

A domains, the adenylation domains of NRPSs, belong to a large family of adenylate forming enzymes: the ANL (**A**cyl-CoA synthetases, **N**RPS adenylation domains, and **L**uciferase enzymes) superfamily[37]. Other members of this family include acyl- or aryl CoA synthetases and luciferase oxidoreductases. On average, NRPS A domains are comprised of about 500 residues summed up to a size of about 55 kDa. Like all members of the ANL family, NRPS A domains catalyze a two-step adenylation reaction[37]: In the first step adenosine triphosphate (ATP) and the substrate are bound and adenosine monophosphate (AMP) is attached to the carboxyl group of the substrate resulting in the activated substrate-adenylate and release of pyrophosphate (PP_i). In the second step a nucleophile, here the thiol group of the PPant arm, attacks the reactive intermediate to release the final product and AMP from the A domain.

The first correlations between peptide elongation and substrate activation were made in the gramicidin S[38,39] and the tyrocidine synthetase[29,30] in 1969 and 1970. Verification that indeed A domains of an NRPS specifically recognize and activate their substrates could only be provided 25 years later, once individual A domains were purified and analyzed *in vitro* using an ATP/PP_i exchange assay[40–42]. With the elucidation of the structure of the Phe activating first A domain of the gramicidin S synthetase in 1997[43] and other A domains[44–46], also in conjunction with their cognate PCP domain[47,48], further insights into the substrate recognition and adenylation process were gained. It became clear that the A domain is divided into two subdomains, a large, N-terminal core domain A_{core} of about 400 residues and a small, C-terminal subdomain A_{sub} of about 100 residues. The two domains are linked via a hinge loop[49,50]. The active site, where the amino acid and Mg·ATP are joined together, is located in the A_{core} close to the interface between the two subunits.

Furthermore, ten conserved signature sequences (A1 - 10) could be identified and assigned specific roles in substrate binding, structural arrangement or catalysis[51–53] (**Table 1**). The A_{core} domain contains regions A1 - A7 while A8 - A10 are located on the A_{sub} domain. An extended post A10 motif (LPxP) has been recognized to play a role in correct positioning of a lysine within the active site of A10 and in A domain PCP interaction[54].

Table 1 | A domain signature sequences and their roles in substrate binding and activation. Φ represents an aromatic amino acid residue, while x stands for any residue. Table modified from Labby *et al.*[55] with permission from the publisher.

core	consensus sequence	role
A1	L(T/S)YxEL	N-terminus of domain, caps an α-helix[53] [structural]
A2	LKAGxAYL(V/L)P(L/I)D	properly aligns Gly78[53] [structural]
A3	LAYxxYTSG(S/T)TGxPKG	acts as a loop and positions the β,γ-phosphates correctly[53] [substrate binding]
A4	ΦDxS	aromatic residue terminates an α-helix that forms side of acyl-binding pocket[53] [substrate binding]
A5	NxYGPTE	invariant glutamic acid coordinates Mg^{2+} ion; adenine ring of ATP is stacked against aromatic residue[53] [structural and substrate binding]

A6	GELxIGx(V/L)ARGYL	stabilizes distorted β-sheets in the N-terminal domain[53] [structural]
A7	Y(R/K)TGDL	aspartic acid residue is 100% conserved and hydrogen bonds with ATP through the ribose hydroxyls[53] [substrate binding and catalytic]
A8	GRx[D/K]xxxKxxGxxxELxxxE	arginine stabilizes the ribose through its hydroxyls; a hinge is formed at aspartic acid residue; in the thioester-forming conformation, the glycine forms part of the PPant tunnel[53] [structural and substrate binding]
A9	(L/V)PxΦM(L/V/I)P	stabilizes thioester-forming conformation by properly positioning residues to interact with PCP domain[56] [catalytic]
A10	NGK(V/L)DR	in the adenylate-forming conformation, lysine is within the active site[53] [catalytic]

A domain specificity is regulated by amino acids lining the active site pocket

With the structures of substrate-bound A domains, the active site pocket and the specificity conferring residues could be identified[57,58] and refined[59]. This enables the prediction of substrate specificity from the DNA sequence of an unknown A domain[60–62].

A domain activity is accompanied by conformational rearrangements

For single modules, the A domain was reported to adopt three different conformations: (i) an open (O) conformation, where no ligand is bound, (ii) an adenylation conformation, where the small C terminal A_{sub} domain is turned by ~48° to complex Mg^{2+}, ATP and the amino acid substrate to perform the adenylation reaction and finally (iii) a thiolation (T) state, in which the A_{sub} domain is rotated by ~140° around the large N terminal A_{core} to tether the activated amino acid to the PPant arm of the PCP domain[45,49,63,64]. This A domain alternation mechanism is expected to contribute most to the conformational rearrangements that are necessary for the growing peptide chain to travel along the enzyme.

Some A domains require MbtH-like proteins to support their fold and function

MbtH-like proteins (MLPs) are small proteins of about 70 amino acids named after their first discovery in the mycobactin biosynthesis gene cluster[65] in *M. tuberculosis*. They are required by some A domains for activity[66,67] or increase solubility and product formation[68–71]. From their consensus sequence ($NXEXQXSXWPX_{[5]}PXGWX_{[13]}LX_{[7]}WTDXRP$)[72] the tryptophan residues have been identified to contribute critically to the interaction with the A domain[73,74]. Several crystal structures capturing A domains and MLPs in complex[73,75,76] suggest that MLPs interact mainly with the A_{core} domain but do not have direct contact with the substrate or the small A_{sub} domain. A gel filtration experiment showed that in complex, the A domain has a slightly lower apparent molecular weight than without the MLP thus indicating a more rigid conformation of the MLP bound A domain[76]. Note that for some NRPSs, MLPs are completely dispensable[77].

Promiscuous A domains lead to a variety of different peptides formed by the same NRPS

While the A domain normally determines the building block that is incorporated via its specificity conferring residues as described above, there are examples of promiscuous A domains, which add to the diversity of NRPs produced by one NRPS.

The nostopeptolide A synthetase features an A domain that accepts three different branched hydrophobic amino acids (Ile, Leu and Val)[78] and is thus able to synthesize three versions of nostopeptolide A. Another example is the anabaenopeptin synthetase, where the initial A domain is even capable of activating two structurally different amino acids: arginine and tyrosine. The crystal structure revealed that this promiscuity is possible due to arginine adopting the same conformation as the tyrosine in the active site pocket of the A domain [79]. Despite the fact that we know how A domains recognize their substrates and thereby control the NRP sequence, there are exceptions and influences by neighboring domains which cannot be predicted yet. *In vitro*, A domain specificity was shown to be dependent on what type and how much of the upstream C domain was present in the analyzed construct [80,81].

Assays for analyzing A domain activity

During the adenylation reaction catalyzed by the A domains, ATP is consumed while PP_i and AMP are released. The change in each reactant over time can be used as a measure to study A domain activity *in vitro.* In **Figure 3**, different assays are summarized according to the reactant they detect.

ATP/PP_i exchange assays (**Figure 3A**) measure the reverse adenylation reaction through the incorporation of [^{32}P]-PP_i into ATP[82–84]. They exhibit a high sensitivity but require radiolabeled ATP. A recently presented assay uses [^{18}O] labeled ATP and thus is non-radioactive[85]. The hydroxylamine-trapping assay (**Figure 3B**) detects the amino acid-AMP intermediate, from which AMP is cleaved yielding a signal at 540 nm. This assay is useful for A domains that bind PP_i tightly[86]. The commercial kit AMP-Glo™ (Promega) detects AMP (**Figure 3C**) by first depleting ATP from the reaction and then turning the AMP back to ADP and ATP followed by a luciferase assay. Colorimetric phosphate assays (**Figure 3D**) detect PP_i release by cleaving PP_i into two monophosphates P_i that react with either Malachite Green or MesG to form a colorful product that can be quantified[87–90]. Another type of PP_i release assay is coupled to the decrease of NADH via an enzymatic cascade taking PP_i as a starting substrate[91] (**Figure 3E**) and has been recently adapted to the study of NRPSs[70]. In this assay, A domain activity is ultimately measured via quantification of the decrease in absorption at 340 nm, the absorption maximum of NADH[92].

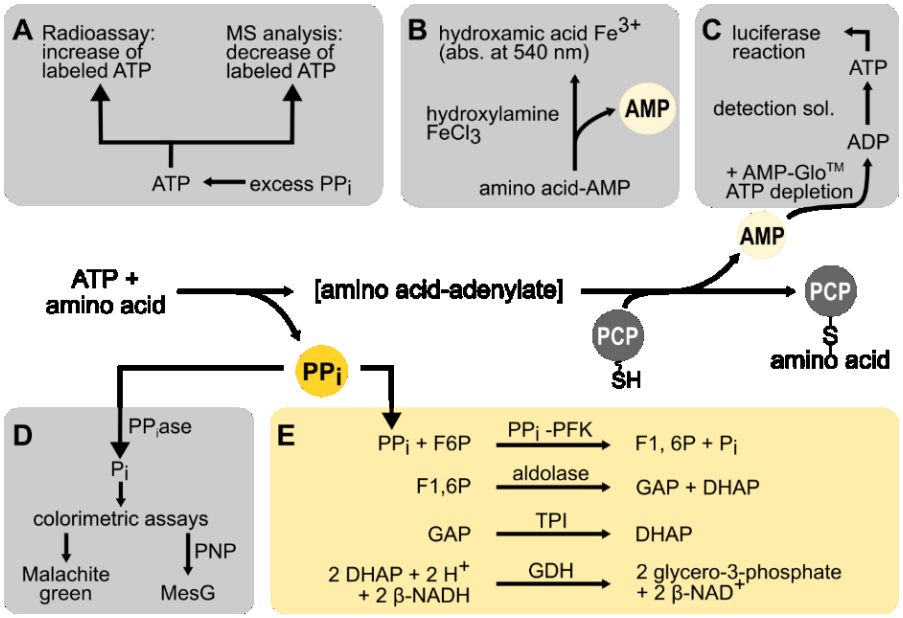

Figure 3 | Overview of A domain activity assays. The two step adenylation reaction catalyzed by A domains uses ATP and produces PP_i and AMP. **(A)** ATP–PP_i exchange assays; **(B)** hydroxylamine-trapping assay **(C)** AMP-Glo™ (Promega) **(D)** Colorimetric phosphate assays (PP_iase: pyrophosphatase; PNP: purine nucleoside phosphorylase: MesG: 2-amino-6-mercapto-7-methylpurine ribonucleoside). **(E)** NADH coupled PP_i release assay (F6P: d-fructose-6- phosphate; F1,6P: fructose-1,6-diphosphate; PP_i-PFK: phosphofructokinase (pyrophosphate dependent); GAP: glyceraldehyde-3-phosphate; DHAP: dihydroxyacetone phosphate; TPI: triosephosphate isomerase; GDH: glycererophosphate dehydrogenase). Figure modified from Kittilä *et al.*[71] with permission from the publisher.

In this work, I use the coupled PP_i release assay described last for the *in vitro* investigation of NRPS A domains, because it does not require radioactivity, it is suitable to study A domains also within a larger NRPS complex[70] and can be performed in a photospectrometer to detect one distinct round of activation or in a plate reader for higher throughput measurement of repeated activation cycles.

1.2.5 *Peptidyl carrier protein domains and their activation through PPtases*

PCP domains, also called thiolation (T) domains, are small domains of less than 100 amino acids or about 10 kDa arranged in a 4-5 helix bundle structure as identified in X-ray[93–95] and NMR[96–99] studies. They are located downstream of the A domains in an NRPS to take on the adenylated substrate and present it to adjacent C (and other auxiliary) domains. To do so, PCP domains need a posttranslational modification: a 4'-phosphopantetheine (PPant) moiety with a length of 18 Å is transferred from a Coenzyme A (CoASH) precursor onto a conserved serine ([I/L]GG[D/H]S[L/I])[100,101] by an external PPtase[102] (**Figure 2**, 0.). The thiol (-SH) group at the terminus of the PPant arm serves as the nucleophile to attack the activated carboxy group of the adenylated substrate and thereby tethers it covalently to the PCP domain via a thioester bond (**Figure 2**, 2.).

PPtases are crucial for NRPS activity[103] and can block efficient substrate elongation when accepting and transferring e.g. acetyl-CoA. Mis-primed PCP domains can be liberated again by TE type II domains, present in many NRPS encoding gene clusters (section 0). This laxness of PPtases towards their substrate can also be exploited to bypass A domain selection by loading amino acid-thioesters directly onto the PCP domain[104–106] or evolve a PPtase to introduce non-native compounds[107]. PPtases are promiscuous with regards to which NRPS they attach the PPant arm to[108]. Sfp, a PPtase from the *surfactin Srf* gene cluster in *Bacillus subtilis*[105], is a well-studied promiscuous PPtase, often used to phosphopantetheinylate NRPS *in vitro* and *in vivo*[109–111].

The PPant arm is used by PCP domains to shuttle the intermediate products between adjacent domains

As described before, NRPSs undergo large conformational changes during their biosynthetic cycles. One might expect the peptide carrier protein domain with its PPant arm to be involved in these movements to a great extent, however, in contrast to previous findings in an NMR study[96], PCP domains seem to be quite rigid. According to several X-ray structures and structural alignments, they do not change their conformation, neither upon phosphopantetheinylation, nor dependent on their neighboring domains[94,95,112]. These findings support the notion that A domains are the major driving force of the conformational changes in NRPSs, which then re-orient the PCP domains to fulfill their task. Though, an additional layer of regulation and elasticity might be imposed by the PCP domain loading state as proposed recently[64,113,114].

In general, PCP domains need to shuttle intermediate substrates between adjacent domains and must thus engage in transient contacts with them. Hydrophobic and ionic interactions between the A domain and its PCP domain could be attributed to helix 2 of the PCP domain and helix 11 of the A_{core}[50]. Further, the loop connecting helix 1 and 2 of the PCP domain (at the end of which, the conserved serine is located) interacts with the last structural motif of the A_{sub} domain via a network of charged residues[98]. In addition, the A-PCP linker region has been discovered to contain a conserved motif (LPxP) shared among ~70% of linkers, where a mutation of the Leu in the enterobactin synthetase greatly reduced product formation[54].

In the downstream direction, the PCP interacts with and discriminates between C or TE domains. The exchange of an elongation with a termination PCP domain prevented product formation[115]. Point mutations in the C-terminal part of the exchanged PCP domain could rescue product formation in this engineered NRPS hinting at specific interaction of these residues with the downstream domain.

The importance of the linker regions between the PCP and its adjacent domains in both directions has also been shown in an earlier publication from the Di Ventura lab[116]. In this study, the native PCP domain of an indigoidine synthetase was replaced by external or artificial PCP domain. Successful product formation of the engineered NRPS largely depended on how much of the native linker region was included in the exchanged part.

Assays to monitor successful substrate transfer to PCP domains

A frequently used method is a thiolation assay[117], in which a radiolabeled substrate is presented to the NRPS module(s) of interest. Upon thiolation the enzyme gets precipitated and excess substrate is removed by washing. The radioactivity of the remaining enzyme is measured by a

scintillation counter indicating the amount of radioactively labeled substrate covalently bound to the PCP domain. A more direct way of assessing substrates bound to the PCP domain is mass spectrometry of attached substrates[118–122]. The NRPS fraction is subjected to LC-MS either as a whole or after enzymatic digest. The fractions containing the PPant arm can be traced back to previous fractionation states and the mass of the substrate attached can be assessed.

1.2.6 *Condensation domains*

Condensation (C) domains have a size of about 450 amino acids or 50 kDa and have been first recognized in 1995 due to their conserved HHxxxDG motif[123]. They form a V shaped[124] pseudo-dimer reminiscent of the chloramphenicol acetyl transferase[125]. The two legs of the V can be separated into an N-terminal and a C-terminal lobe, forming a tunnel for the connection of the donor and acceptor substrates. The conserved motif is located within the N-terminal lobe, close to the center of the tunnel.

C domains form a peptide bond between the donor and the acceptor substrates

Condensation domains catalyze the peptide bond (amide bond) formation between donor (upstream) and acceptor (downstream) amino acids[32] presented to them through the PPant arm. Formation of the peptide bond is facilitated by a nucleophilic attack of the acceptor amino group on the donor carbon engaged in the thioester bond (**Figure 2**, 3.), which is resolved in the process.

Although, the conserved HHxxxDG motif is crucial for catalysis its exact nature is ambiguous: Mutations of the second histidine, for example, can lead to (almost) complete obstruction[32,126] or to an only two-fold reduction of peptide bond formation[127], which points towards a catalytic role of the second His. The same mutation can also render an NRPS module (here part of TycB of the tyrocidine synthetase) insoluble, hinting at a structural function[35]. Mutation of aspartate (D) within the catalytic motif inactivates the enzyme[35,128].

Overall, the catalytic process of peptide bond formation by C domains is thought to be driven by stabilization of catalytic intermediates rather than acid-base catalysis. However, more crystal structures with one or both substrates bound are needed to fully recapitulate the catalytic mechanism[23].

The C domain acts as a checkpoint for correct directionality and specificity of the reaction partners

The condensation reaction is the one that ultimately drives product elongation. Consequently, C domains control proper directionality of this elongation process[129]. A recent crystal structure of the first condensation domain of a calcium-dependent antibiotic (CDA) synthetase can adopt two different states, a more closed and a more open state[130]. It is hypothesized that this minor conformational change aides the proper directionality of the reaction, e.g. opening only after the desired acceptor amino acid has entered the tunnel[130]. Interestingly, this acceptor amino acid bound conformation is also adopted by the C domain during donor substrate activation by the upstream A domain. Hence it is likely that during the condensation reaction, the proximate donor substrate is already prepared for thiolation to improve the overall enzyme efficiency[50].

C domains act as an additional gatekeeper ensuring that the correct substrates are condensed. This feature is more pronounced at the acceptor than at the donor side[33,129,131], though the

donor side was also shown to be size- and stereo-selective with regards to the immediate donor residue and other upstream building blocks[132,133]. Thus, likely there is a substrate recognition step as well, but in contrast to A domains, no specificity conferring residues could be identified in a large phylogenetic analysis of C domains[134]. However, and this is also in contrast to A domains, this phylogenetic analysis could identify individual residues within the conserved sequences[51] from which you can distinguish if a C domain is a starter or an elongation C domain. Sequence analysis can furthermore predict whether elongation C domains catalyze a condensation reaction between two L-amino acids ($^{L}C_{L}$) or between a D- and an L-amino acid ($^{D}C_{L}$)[134]. How L-amino acids can be epimerized prior to condensation within the NPRS will be discussed in the next section.

Starter C domains have been first identified in Daptomycin[135] and a CDA[136] synthetases as part of the initiation modules and are responsible for condensing a fatty acid onto the acceptor amino acid[51]. Functional evidence of this type of condensation reaction was produced by the examination of the surfactin lipoinitiation module[15]. The starter C domain indeed catalyzed the amide bond formation between 3-hydroxymyristic acid-S-CoA, activated by an external fatty acyl CoA ligase, with the PPant tethered substrate activated by the initiation module A domain, glutamate (Glu)[15].

Alternative roles of (evolved) C domains

Condensation domains have evolved over time to form structurally homologous domains with the ability to accomplish other relevant functions within NRPSs[137], such as epimerization[138], cyclization[139], b-lactam formation[140], or recruitment[141] of auxiliary enzymes. Mainly in fungal NRPS complexes, C domains can also replace the TE domain and promote macrocyclic release of the final product[142,143]. This is also the only example to date of any type of C domain crystal structure represented with the donor substrate oriented in a catalytically competent state[143].

1.2.7 *Auxiliary domains in* cis *and* trans

Auxiliary (or tailoring) domains expand the basic NRPS domains and add an additional opportunity for NRP diversification in cis, as part of the NRPS, or in trans, as a separate enzyme. Accordingly, these additional modifications are either introduced along with NRP elongation, prior to or following the production process. Here are few examples of auxiliary domains[23] and the reactions they perform to modify their substrates, some of which will be described in more detail in the paragraphs below: cyclization (Cy), epimerization (E), formulation (F), ketoreduction (KR), methyl transfer (M), oxidation (Ox), and reduction (R). This list is by no means complete and new tailoring domains are still being identified. For a very detailed update and overview on auxiliary domains, please see[23].

Larger auxiliary domains such as E, F and Cy domains act as a discrete domains located in the beginning or rationally along the NRPS (assembly) line. Smaller auxiliary domains, like M, Ox and R domains are often embedded within A domains, preferably between core sequences A2 and A3 or between A8 and A9[55] (**Table 1**).

Epimerization domains

Epimerization (E) domains epimerize L-amino acids to D-amino acids in trans[144,145] or more commonly in cis[51,146]. Therefore, they are usually embedded in the NRPS complex, downstream of the PCP domain, on which the PPant tethered substrate is then presented for epimerization.

E domains display a C domain-like structure, however, there are critical differences between the two: (i) the C domain binding site for acceptor substrate is blocked in E domains, e.g. by eleven additional residues in TycA[138]; (ii) the floor loop between the N- and C-lobe is expanded by at least 5 residues in E domains, which enable specific interactions with the PCP domain[138]; (iii) the linker region between PCP and E domain of the gramicidin synthetase GrsA PCP-E di-domain was found to align in a structured interaction along the E domain surface leading to correct positioning of the rigid PCP domain and thus of the PPant bound substrate towards the active site[147]. In contrast to PCP-C domain linkers, the connection between PCP and E domains plays a prominent role in enzymatic recognition and binding.

Methylation domains

Methyltransferases or methylation (M) domains use S-adenosylmethinonine (SAM) as a co-substrate to perform backbone N-methylation of NRPs[148]. Other attachment sites like O-, S-, and C-methylation as well as N,N-dimethylation have also been reported[23,149,150]. The methyltransferase domains have a size of about 45 kDa and are usually embedded within the A domain core sequences A2-A3 or A8-A9[55]. They can also be located between two domains[151] or act in *trans* as a standalone enzyme[152]. Bioinformatic comparison of M domains has identified conserved signature motifs from which the type of methylation can be predicted to a certain degree[153]. Very recently, the structure of an M domain embedded in TioS A_{Val}(A8-9) domain in complex with its MbtH-like protein partner TioT was solved[154]. This is the first available structure of an interrupted A domain and could pave the way to better understand how integration of auxiliary into A domains can yield a functional collaboration between the two.

Oxidation domains

Likewise, Oxidation (Ox) domains are generally embedded within A domains of the substrate they oxidize. The position between A8 and A9 core motifs seems to be an evolutionary favorable insertion point of auxiliary domains[55], even though Ox domains can also locate at other positions such as downstream of the PCP domain as observed for MtaC of the myxothiazol PK/NRP synthetase [155].

Three conserved signature motifs for Ox domains have been identified[55] (**Table 2**). Sequences Ox1 and Ox2 allow the classification of NRPS integral Ox domains to a larger group of FMN-dependent oxidoreductases found as distinct enzymes. The contribution of each conserved motif on the oxidation reaction remains unknown.

Table 2 | Conserved motifs of NRPS integrated FMN dependent Ox domains.

core	consensus sequence
Ox1	KYxYxSxGxxY(P/G)VQ
Ox2	GxxxG(L/V)xxGxYYY(H/D)P
Ox3	IxxxYG

Examples of PKS/NPRS complexes harboring Ox domains are EpoB and MtaD of the epothilone and myxothiazol biosynthesis pathways, respectively[155,156]. The Ox domain-integrated A domains accept L-cysteine, cyclize it to form a thiazoline and subsequently perform the oxidation to create a thiazol moiety[157]. Another example of an Ox domain integrated in a glutamine specific A domain can be found in the indigoidine synthetase, which will be discussed in section 1.2.11.

1.2.8 *Thioesterase domains*

Thioesterase (TE) domains (~280 aa) release the final peptide product formed by the upstream modules[158,159] most commonly in form of a cyclic or a linear product. They comprise a high sequence variety, contain very few conserved sequence stretches[160] and are able to facilitate a vast number of reactions. Unlike for e.g. A domains, the substrate they accept and reaction they catalyze cannot be predicted from conserved motif sequences or phylogenic analysis[161]. More detailed information about different types of TE domains are covered in comprehensive review articles[162,163]. In addition, reductase or condensation domains can also account for product release in some cases[162]. In the following I will focus on the aspects of TE domains that are key for this work.

TE domains feature a catalytic pocket and a flexible lid-like structure

As described above, thioesterase domains can differ largely in their protein sequence, the confirmation they adopt and in the reaction they catalyze. However, some features are common to most of them, as established by crystal structures of different TE domains[164–169]: (i) an active site pocket harboring the catalytic triad Ser-His-Asp, where the serine is embedded in a GXSXG motif. (ii) an alpha/beta fold, where alpha helices and beta sheets are alternating in a rather conserved order and finally (iii) a flexible lid region, which consists of one or more alpha-helices and is located between beta sheets 6 and 7[170].

Thioesterase domains release the final NRP as a linear or a cyclic product

The general release mechanism of TE domains occurs in two successive steps, very similar to other hydrolases[171,172]. In short, the intermediate peptide chain is transferred from the last PCP domain PPant arm onto the active site serine of the TE domain from where it is released through and intra-, or intermolecular nucleophilic attack.

In the first step, the intermediate substrate is loaded onto the active site serine. For this purpose, the His-Asp dyad draws a proton from the alcohol of the serine side chain making it more nucleophilic. The intermediate NRP is presented to the TE domain linked to the PPant arm of the PCP domain. The linking thioester bond is then attacked by the negatively charged serine side chain forming the acyl-O-TE intermediate and releasing the PPant thiolate.

In the second step, the substrate is offloaded from the TE domain as the final product. To that end, an intramolecular or external nucleophile approaches the scene. It is activated by either the active site histidine after deprotonation by the leaving thiolate as proposed by Townsend[173,174] or by another external cue. The nucleophile attacks the carbonyl of the acyl-O-TE intermediate, releasing the final NRP and a seryl alkoxide, which requires immediate reprotonation. If the nucleophile was a water molecule, hydrolysis of a linear peptide product occurs. If the nucleophile was the N-terminus of the NRP, a (macro-) cyclic product is released.

The latter case is the most common scenario, since cyclic peptides can adopt specific, entropically favored conformations needed for the interaction with target structures. In addition, cyclic peptides are less susceptible to degradation by peptidases thus making them more stable[175]. In addition, there is a range of exceptional NRPSs, which also require more exceptional product release strategies. Some of these will be shortly discussed in sections 1.2.10.

Conformational rearrangements of the TE domain during product release

The core region of thioesterase domains does not undergo any larger rearrangements to facilitate product release. In their native conformation, they provide an oxyanion hole through hydrogen bonding from two backbone amide NH groups, which stabilizes the acyl-O-TE intermediate during formation and breakdown[169] without any further conformational switch required. However, the aforementioned lid region has been shown to exist in two different states, an open and a closed one[169]. In an NMR study it could be additionally confirmed that this lid is moving from one state to the other during substrate processing[176] suggesting that it does have a role during this process. Potentially, it excludes intrusive molecules like water from the active site pocket. Yet, the active site pockets of different TE domains can have different volumes and in addition differently sized lids can limit this volume without stringent implications for substrate specificity[163].

Variations from the catalytic triad sequence motif can be found in natural NRPSs

For the catalytic triad there are examples of natural deviations from the Ser-His-Asp sequence, which have been long known for other alpha/beta hydrolases[172,177]. For example, instead of the serine there is a cysteine serving the same purpose in few NRPSs[163]. The histidine has also been found replaced by proline in the mycolactone TE[178]. Finally, aspartate can also be substituted by glutamate or serine[163].

Thioesterase type II domains regulate a variety of reactions in trans *including regeneration of misprimed PCP domains*

In addition to the thioesterase type I described above, in many NRPS gene clusters a second type of thioesterase can be found, termed type II (in short TEII)[158]. They mostly act in *trans* and remove mis-primed PPant arms of NRPSs thus regenerating the enzyme for another futile cycle of NRP production[167,179]. In rare cases, they can also participate in substrate selection or (intermediate) product release[180]. In this way, they can improve the NRP yield from heterologously expressed NRPSs[180,181].

1.2.9 *Communication domains*

Communication (Com) domains mediate the intermolecular communication and directionality between two adjacent NRPS modules which are expressed as separate proteins to ensure proper peptide propagation along the enzymatic complex[182]. They are short regions located at the C-terminal end of the donor module and the N-terminus of the acceptor module. Structural analysis revealed that their contact is facilitated by anti-parallel helix-hand interactions[183,184]. Communication domains can be found in the tyrocidine, plipastatin, surfactin, gramicidin S and lichenysin Synthetases[182]. Note that there are also NRPS complexes, e.g. the andrimid PKS/NRPS synthetase, that lack communication or other docking domains, suggesting that alternative interaction mechanisms might exist[185].

1.2.10 *Iterative and non-linear NRPSs*

Besides linear NRPSs, other types of NRP producing enzymes have been identified: iterative and non-linear NRPSs. Iterative NRPSs produce their NRPs by iteratively adding up the basic monomer or peptide sequence that is defined by the specificity of their modules[23]. In this way the NRPS machinery can be much smaller and still produce a longer NRP. Cyclo-oligomerization catalyzed by iterative NRPS has been shown for the bacterial gramicidin S[186] and enterobactin synthetase, as well as the fungal beauvericin and enniatin synthetases[187]. This oligomerization

can be achieved by temporarily storing the intermediate components on the terminal PCP domain and the active site serine of the TE domain before oligomerization.

For non-linear NRPSs, the sequence of the NRP is not reflected by the module specificity or arrangement of their synthetase[188]. A recent study suggests that they might be more common than initially expected[189]. An example of an NRP produced by non-linear NRPS is fungisporin and related cyclic tetrapeptides in *Penicillium chrysogenum*[190].

1.2.11 *Special cases: Pigment synthetases with focus on indigoidine and dodecylindigoidine synthetases*

A large variety of natural pigments produced in bacteria, fungi and archaea[191–194]have been discovered and many of them originate from NRPSs and PKSs. Some compoundsare currently applied as antifungal, antibacterial and antitumor agents or as natural dyes for food or clothing[193]. A recent review article focusses on natural blue pigments[195], including indigoidine and a derivative, N,N-dodecylindigoidine.

These two blue pigments are produced by extraordinary NRPSs of the possibly iterative kind, which will be described in greater detail below.

Indigoidine and its proposed synthesis mechanism

The blue pigment indigoidine was noted to be produced in *Pseudomonas indigofera*[196] in 1890, isolated from the Spree river in Germany. Many chemical properties of indigoidine were also described then already, e.g. its insolubility in water and many organic solvents, its decomposition in sodium hydroxide (NaOH), the transformation into a yellow-brown substance in concentrated hydrochloric acid (HCl).

The chemical structure of indigoidine (5,5'-diamino-4,4'-dihydroxy-3,3'-diazadiphenoquinone-(2,2')) was solved in 1965[197] by the Heidelberg based Nobel-Prize Laureate Richard Kuhn and a chemical synthesis pathway was presented[198]. Other solvents for indigoidine were identified to be dimethylsulfoxide (DMSO), dimethylformamide (DMF) and pyridine. The molar extinction coefficient in DMF was determined to be ε (Indigoidine in DMF) = 23 442 M^{-1} cm^{-1} [198]. During and following the chemical synthesis process, other derivatives of indigoidine were also identified, e.g. the colorless but fluorescent reduced leuco-indigoidine[197,199].

The biosynthetic gene cluster harboring the indigoidine synthetase gene was first discovered in blue pigment producing[200] plant pathogen *Erwinia chrysanthemi*[201] (**Figure 4A**). Homologues indigoidine synthetase genes have been found across the bacterial kingdom, within the genus of *Streptomyces* (e.g. *lavendulae* ATCC 11924[202]), *Pseudomonas* (e.g. *flourescens*[203] and indigofera[204]), *Photorhabdus* (e.g. *luminescens*[205], where the expression was silenced, but could be triggered by promotor exchange), *Microbispora* and *Dickeya* strains and others. The two indigoidine synthetases I will investigate further and use in this work are IndC from *Photorhabdus luminescens*[205] and BpsA from *Streptomyces lavendulae* ATCC 11924[199,202].

The indigoidine synthetase is a single module NRPS of about 1280 amino acids, which consists of an A domain (accession number cd05930) specific for L-glutamine, followed by a PCP domain (CL0314) and a final TE (alpha/beta hydrolase, cl21494) domain. The A domain is interrupted by a mcbC type nitro-reductase domain (cd02142), which binds the cofactor FMN to reduce it to

$FMNH_2$ and concomitantly oxidize the glutamine substrate (**Figure 4A**). The $FMNH_2$ is autocatalytically recovered via oxidation with molecular oxygen[206].

The indigoidine synthetase is schematically depicted in **Figure 4B** and converts two L-glutamine molecules into the blue pigment indigoidine. The proposed synthesis mechanism is shown in **Figure 4C**. Glutamine is recognized and activated by the A domain and thiolated to the PCP domain. Either before or after thiolation, the oxidation domain introduces a double bond in the glutamine side chain, whereupon the amino group of this side chain executes a nucleophilic attack on the thioester bond. Thus, the oxidized cyclic glutamine is proposed to be released from the enzyme and undergo a tautomerization, where the double bond shifts within the cyclic molecule to the neighboring C-C bond. The second isomer is suggested to non-enzymatically dimerize into the blue pigment indigoidine[55,201].

It remains unknown at which step exactly the oxidation reaction occurs. Furthermore, in this proposed mechanism, the function of the TE domain is unclear and seems neglectable, which is unlikely given that evolution put it there.

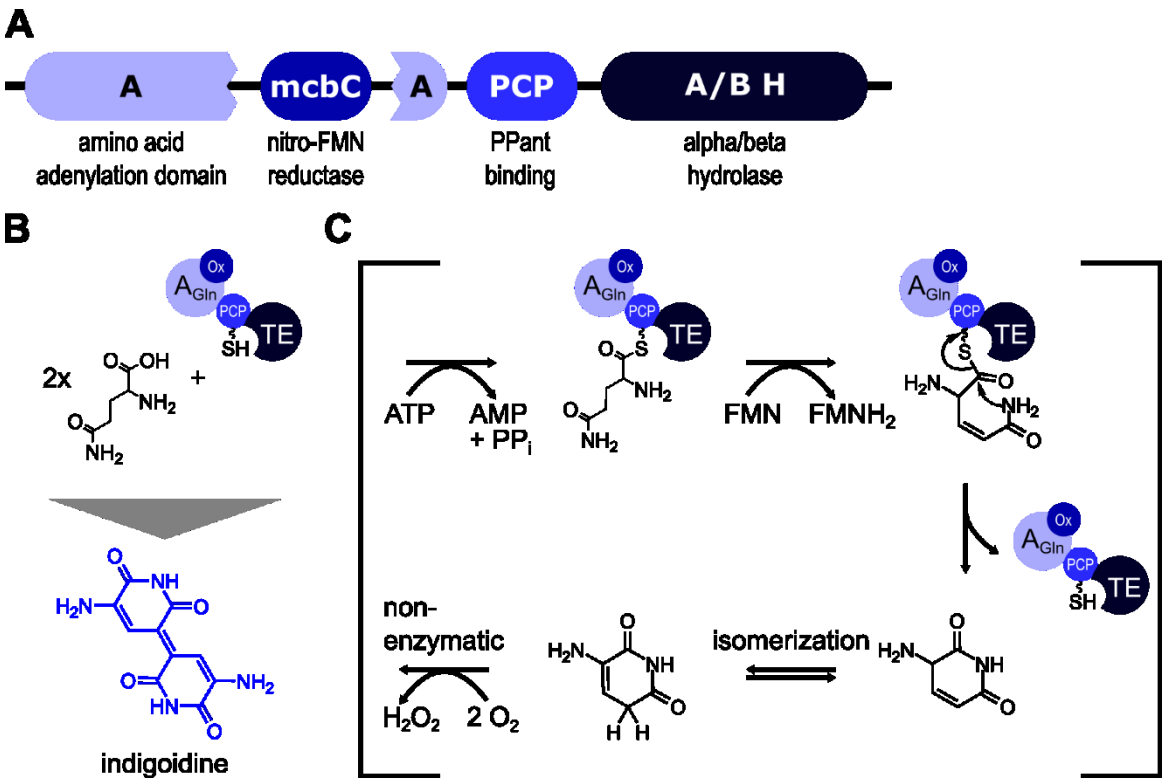

Figure 4 | Indigoidine and its proposed synthesis mechanism by the indigoidine synthetase. (A) Domain representation of the indigoidine synthetase. The A domain is interrupted by a mcbC like nitro-FMN-reductase and followed by a PCP and TE domain (alpha/beta hydrolase). **(B)** Schematic depiction of the indigoidine synthetase. The A domain is specific for glutamine and ultimately converts two L-glutamine molecules into one indigoidine, that has an intense blue color. **(C)** Proposed mechanism of indigoidine production. Upon thiolation, the glutamine is suggested to be oxidized and released from the enzyme via an internal nucleophilic attack. The oxidized and circularized intermediate undergoes isomerization and is hypothesized to non-enzymatically dimerize to form indigoidine.

The natural role of indigoidine is not finally resolved

The role of indigoidine in the natural context is ambiguous. Its presence has been investigated in various organisms and contexts, some of which will be described below to give an overview of possible roles that indigoidine plays in nature.

In *Streptomyces lavendulae*, indigoidine production is controlled via the SARP (Streptomyces antibiotic regulatory protein) regulatory cascade including IM-2 and its receptor FarA[207,208], which is also responsible for the production of antibiotics D-cycloserine and different nucleoside antibiotics. In *Roseobacter Phaeobacter sp. Strain Y4I*, indigoidine production was linked to the inhibition of *Vibrio fischeri* on agar plates[209]. Another *Roseaobacter* strain (*Leisingera sp. JC1*) resides in the reproductive system of the Hawaiian bobtail squid (*Euprymna scolopes*) and is partly deposited into the jelly coat of their eggs, to protect them from *Vibrio fischeri* colonization[210]. On the other hand, indigoidine synthetase null mutant *Roseobacter* strains were less motile and faster to colonize an artificial surface[209], which reveals the pleiotropic effects of indigoidine. Indigoidine exerts pleiotropic effects on oxidative stress as well. Indigoidine synthetase null mutant *Roseobacter* were more resistant to hydrogen peroxide[209] while for *Erwinia* the presence of indigoidine lead to an increased resistance toward oxidative stress[201]. For *Vogesella sp. strain EB* indigoidine might be a precursor for an indigoidine-derived pigment termed cryo-indigoidine, which prevents freezing of the bacterium in low temperatures and increases survival in the cold, iron-oxidizing environment of Andean Patagonia[211].

Overall, we can summarize that indigoidine seems to have a mild antimicrobial effect on other bacterial entities and is involved in surface colonization and oxidative stress response, however with strains-specific outcomes.

Indigoidine as a reporter molecule

Indigoidine has a dark blue color and is synthesized by a single, post-translationally modified enzyme from a widely available, non-absorbing, non-toxic substrate (L-glutamine). This combination makes indigoidine an ideal reporter molecule, which has been exploited in various applications.

It has been used as an alternative reporter molecule for blue/white screening of successful cloning in bacteria[212]. In mammalian cells, indigoidine could also be produced and was subsequently reduced to the fluorescent leuco-indigoidine serving as a selection marker in FACS[199]. Since a PPtase has to add a PPant arm to the PCP domain of the indigoidine synthetase for proper function, indigoidine formation has been used to validate characteristics of known PPtases *in vivo* and *in vitro*[108]. When placed upstream of an indigoidine synthetase gene, the strength of engineered promotors and other regulatory sequences could be investigated[213]. The same principle was applied to explore naturally occurring metabolite-sensitive repressors in bacteria[214]. Once purified, an indigoidine synthetase was employed to measure glutamine concentration *in vitro* from different biological samples[215].

In the examples described above, the indigoidine synthetase remained unaltered to produce the blue pigment as a reporter molecule for testing PPtases, regulatory sequences and glutamine concentrations. In contrast, indigoidine synthetase has also been used to monitor the success of domain and in particular PCP engineering of the single module NRPS itself, by us[116] and others[216], as illustrated in more detail in section 1.2.5.

N,N-dodecylindigoidine and its proposed synthetase in Shewanella violacea DSS12

Another exceptional but less well studied pigment is N,N-dodecylindigoidine. It consists of an indigoidine core with two dodecanoic (=lauric) acid side chains attached to the N-terminal amino group via a peptide bond (**Figure 5B**). It has a purple color with an absorption maximum at 636 nm in chloroform[217].

The source of this pigment is *Shewanella violacea* DSS12, which is a marine bacterium that was found in the Ryukyu Trench near Japan at a depth of 5 110 m[218]. Cultivated on Marine Agar plates it produces a violet pigment in the form of long crystals, which turned out to be N,N-dodecylindigoidine[217]. The genome of *Shewanella violacea* has been sequenced completely[219] and searched for potential genes related to its pigmented phenotype using a Bidirectional Best Hits (BBH) approach in comparison to the pigment producing *Rheinheimera baltica* DSMZ 14885[220,221]. Among others, a putative N,N- dodecylindigoidine synthetase (WP 013053246.1), hereafter referred to as "dIndS", as well as an Sfp type PPtase (WP_013050488.1) have been identified[221].

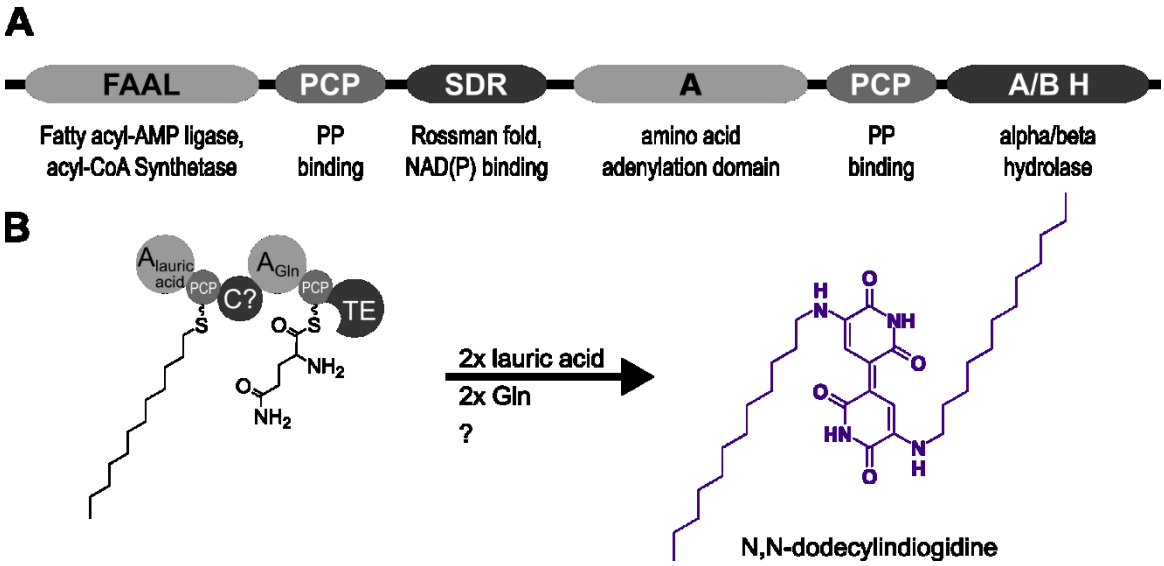

Figure 5 | Predicted conserved domains of the putative N,N-dodecylindigoidine synthetase "dIndS". (A) Conserved domains of the protein WP 013053246.1 in *Shewanella violacea* according to the Conserved Domain Database. FAAL = fatty acyl AMP ligase, PCP = phosphopantetheine attachment site, SDR = short-chain dehydrogenases/reductases, Rossman fold for NAD(P) binding, A = adenylation domain of NRPSs and A/B H = alpha/beta hydrolases. **(B)** Possible domain arrangement and product formation.

According to the Conserved Domain Database[222], dIndS consists of the following domains (**Figure 5A**): FAAL = fatty acyl AMP ligase (accession number cd05931), PCP = phosphopantetheine attachment site (pfam00550), SDR = short-chain dehydrogenases/reductases with a Rossman fold for NAD(P) binding (cl25409), A = adenylation domain of NRPSs (cd05930) and A/B H = alpha/beta hydrolases (cl21494). Deduced from the potential product formation process of indigoidine, I proposed a similar process for N,N-dodecylindigoidine (**Figure 5B**). The domain order and function suggest that the FAAL recognizes and activates dodecanoic acid, which is then transferred to the PPant arm attached to the first PCP domain. The second A domain likely has a specificity for glutamine, which is transferred to the PPant arm of the second PCP domain.

The two A domains are connected via a NAD(P) binding domain, which could condense the fatty acid and the amino acid to a secondary amine by reductive amination. How glutamine is oxidized and cyclized to ultimately form the indigoidine core remains unknown. None of the other Bidirectional Best Hits matched an external oxidation domain[221]. The exact role of the final TE domain (alpha/beta hydrolase) also remained unclear, similar to the indigoidine synthetase TE domain, which is of the same type.

1.2.12 *Interactions and interfaces between NRPS domains*

A number of crystal structures and cryo-electron micrographs of multi-domain NRPSs trapped by mechanistic inhibitors, have become available to study the conformation in certain states of the NRP synthetic cycle[53,47,104,50]. Yet, the complete picture remains elusive: Do complex NRPSs adopt a conformational pattern along which the growing peptide chain travels in an orderly manner, like suggested by Marahiel[223]? Or do they exist in an disorganized state and find their reaction partners if present rather by chance - a hypothesis based on work by the Gulick lab[104,224]?

Either way, the synthesis process must be very efficient to avoid mis-initiation or premature product release, both ultimately leading to truncated NRPs futile to the NRPS host organism. How the modules communicate with each other in the larger context to ensure this efficiency to date remains a secret.

1.3 Ways of expanding and harnessing the toolbox of NRPSs to find and engineer novel drugs based on natural products

Now that I have summarized what we know about NRPSs, their setup, their mode of function and the great diversity and importance of their products, I want to focus on how to expand and use that knowledge for the discovery and development of novel antibiotics or other drugs. In this second part of the introduction, I want to present current approaches to find novel natural products and their synthetases. These findings contribute to expanding the natural toolbox of known NRPSs to be used for engineering novel, custom NRPs. However, there are still some problems to be overcome before we can fully exploit that toolbox. I will introduce the problems and show successful examples of NRPS engineering.

1.3.1 *The search of novel antibiotics in (un-)cultivatable organisms*

The great advances of high throughput sequencing, (mass) spectrometry, bioinformatical analysis and integrated approaches enable large screens of potential NRP producer organisms[225], of which an estimated 99% of which are uncultivatable. Many of them are marine organisms[226–228] or live in the soil[229,230] or on plants[231,232]. Recently, several new (potential) natural product producers have been identified via genome mining[233–235] or improved mass spectrometry approaches[120,236]. In a particularly successful approach, Ling *et al.* first obtained single colonies of uncultivatable bacteria from soil using a new device named iChip, then identified colonies with antimicrobial potential via co-culture and finally discovered teixobactin[229], a novel antibiotic that “kills pathogens without detectable resistance”[229].

New information about NRPS or NRP discoveries can be added to comprehensive databases that have already been established for NRPSs and PKSs, such as Norine[237], Clustermine360[160] and antiSMASH[238,239] database.

1.3.2 *Aspects of NRP production by fermentation*

To date, most marketed natural products are fermented in their original producer organism and purified further to be formulated into an administrable product[23]. Chemical total- or semisynthesis have been reported for a variety of natural products, however, often the fermentation process is less expensive, better scalable and possibly more environmentally friendly since no or less hazardous chemicals are employed in the production process.

Besides the original producer, also heterologous hosts have been successfully established to express complete NRPS gene clusters. In this way, the whole production pathway of e.g. anti-bacterial[240] or anti-tumor[241,242] natural products was reconstituted in *E. coli*. Expression of orphan NRPSs under the control of an inducible promotor has led to the identification of natural products silenced in their native producer strain[205,243]. Other hosts like yeast and *Aspergillus* strains have been engineered to process fungal NRPSs[244,245].

For heterologous expression of NRPS gene clusters, a couple of problems also have to be taken into account: The sheer size of these gene clusters can cause difficulties in handling and cloning. Though, novel DNA recombination strategies like Gibson assembly[246] and recombination in yeast[247,248] have provided remedy for this problem. Another aspect is the difference in codon use by many of the original producer strains. Codon optimization might be necessary prior to successful expression in a different host strain. Since (commercial) gene synthesis is advancing rapidly to be more accurate, more efficient and less expensive, producing complete codon optimized genes for heterologous expression will be more commonly used.

1.3.3 *NRPS engineering has the potential to generate novel drugs*

What we know about the mode of action of NRPSs and the pharmaceutical impact of their products has become evident so far. How to use this knowledge to engineer artificial systems to generate custom NRPs and thus potentially novel drugs will be described in the coming sections.

Different routes can be taken to make NRPSs synthesize different products. One route is precursor-directed biosynthesis, in which alternative substrates are presented to the enzymes while the supply of the original substrate is limited eventually forcing the NRPS to accept and incorporate the alternative precursor. This approach can be paired with mutasynthesis that takes a mutated strain incapable of producing the original product without external supplement of at least one of the precursors which is then exchanged by the alternative substrate[249,250]. In my thesis, I focus on a different route, namely design engineering and combinatorial biosynthesis, where NRPSs are engineered and recombined based on previous knowledge to achieve the predefined outcome. This concept was first introduced by Marahiel and co-workers in 1995[251]. In some cases, design engineering is complemented by precursor directed or mutasynthesis to improve the results.

The strategies of design NRPS engineering reach from the introduction of point mutations and subdomain modifications to whole module alterations or even the reconstitution of new-to-nature NRPSs from scratch. The more complex the modifications become, the higher their potential to produce truly custom peptides. On the other hand, the expected and unexpected problems also increase with higher complexity. I will now summarize the successful examples of NRPS engineering and the problems that arise even in successful instances. These

problems might be the reason why despite a lot of effort, engineered NRPS still fall short of the expectations that many scientists had upon their discovery.

1.3.4 *Examples of successful engineering approaches of NRPSs and their limitations*

Successful is a relative term in NRPS engineering. In a perfect world scenario, "successful" would mean that the engineered version of the enzyme is just as active as the wild type version, also *in vivo*, and that the substrate and reaction specificity is retained, just towards the altered substrate. This is certainly not the case for most engineered NRPS. In some cases, a comparison with the original NRPS *in vitro* and *in vivo* is hard to assess, so the effect of engineering cannot be expressed in numbers. Thus, I will describe examples, which are considered successful if an alternative product of the engineered NRPS was detected. When the levels of modified product or enzyme activity are only a fraction of those of the wild type, reasons for this shortcoming will be discussed. Usually these shortcomings provide an insight to the limitations still encountered in NRPS engineering and demonstrate what to consider for the following engineering approaches.

1.3.5 *Point mutations*

As the most prominent substrate selector in NRPSs, A domains are logical targets of small-scale manipulations like point mutations to alter their specificity towards a desired substrate. In addition, the introduction of point mutations does presumably not perturb the native intra-, and intermodular interfaces of the NRPS.

Following the proposal of the specificity conferring code of A domain substrate binding pockets[57], the second module of the surfactin synthetase (SrfB A_{Asp}) was mutated to change substrate specificity from L-Asp to L-Asn[252]. According to the code, three residues should have been changed, but the biochemical analysis revealed, that a single point mutation was enough (H322E). When the original A_{Asp} domain of the *surfactin* gene cluster was replaced by the modified A_{Asn} domain, the modified surfactin [Asn(5)] was detected in *Bacillus subtilis,* however, at lower levels and only along with the original surfactin. In addition, the specificity change recognized *in vitro* was accompanied by a drastic drop in catalytic activity[252]. These results suggest, that engineering A domain specificity is possible via the introduction of point mutations. However, they also underline the great impact of the C domain acceptor side, which in this case allows the unnatural substrate Asn to pass through, but at the same time uses A domain mis-initiation of Asp to form the original product, overruling A domain specificity. A similar problem arose when mutating the seventh A domain specific for Asp (A^7_{Asp}) of a calcium dependent antibiotic (CDA) synthetase to match the specificity conferring code of Asn, so the converse case of the first example[253]. While *in vitro* the change in specificity was demonstrated, replacing the original by the mutated A domain in *Streptomyces* mainly led to the production of a hexapeptidyl intermediate. This premature product release is likely caused by the upstream C domain at the acceptor side, which recognizes the modified substrate Asn only weakly stalling the production process and releasing the hexapeptide by unintended hydrolyzation[253].

Other approaches of successfully altering the A domain specificity via point mutations have been described. For example, alternative andrimids were produced *in vivo* after site-directed mutagenesis followed by a mutasynthesis approach switching the substrate specificity of an A domain from L-Val to L-Ile or L-Leu[254]. While the most functional versions yielded near wild

type levels of the alternative products, other versions (specificity altered to L-Ala and L-Phe) only reached up to 1.3% of the production level of the unmodified enzyme[254]. In another example, yeast cell surface display was employed to alter the substrate specificity of DhbE. *In vitro* a switch in substrate preference of 200-fold was reached. However, all top selected variants that were reintroduced into the native gene cluster, were unable to load the alternative substrate onto its partner aryl carrier protein (ArCP-) domain because of a substrate specificity independent mutation in the active site (H234W)[255]. A directed evolution approach via successive saturation mutagenesis (SM) of the eight non-conserved residues lining the substrate binding pocket of, in this case, TycA yielded a mutant where the switch in substrate preference from L-Phe to L-Ala was 10^5 in total[256]. An L-Phe accepting A domain in the gramicidin synthetase that was mutated to accept non-natural aromatic substrates modified with azide and alkyne groups for subsequent chemical modifications via click reactions[257]. A single point mutation was sufficient to achieve a 10^5-fold specificity switch towards the non-native substrate, without a great loss in catalytic efficiency. The alternative substrates were incorporated *in vitro* and *in vivo* into a diketopiperazine, even in the presence of the native substrate phenylalanine[257].

The most comprehensive and eventually also one of the most successful A domain engineering experiments combined several methods to make TycA accept (S)-β-Phe instead of the native L-Phe[258]. TycA was subjected to random mutagenesis of several residues of the active side in combination with rational shortening of a loop connecting β-sheets 13 and 14 based on the structure of VinN, an A domain naturally incorporating β-Phe. The resulting library of about 10^6 variants was expressed and displayed on the cell surface of yeast. Successful variants of TycA that could activate and thiolate a clickable β-Phe were detected via an immunofluorescence labelling strategy and were identified in a FACS screen. When the most efficient variants were complemented with the downstream modules TycB and C, β-Phe-containing peptides were produced at a high level *in vitro* (~1 mmol) and yielded high titers in *E. coli* (~100 mg l^{-1}).

Engineering efforts of A domain specificity can also reduce the promiscuity of an NRPS: The fusaricin synhetase, for example, naturally produces a mixture of peptides by incorporating L-Tyr, L-Val, L-Ile, L-allo-Ile, or L-Phe at the third position. The peptide featuring L-Phe at this position is the most favorable for it has the highest antimicrobial activity. Thus to shift the equilibrium, the A^3 domain was successfully mutated to match more stringent L-Phe incorporating A domains[259].

In conclusion, NRPS engineering via point mutations has the largest effect when mutating the specificity conferring code of the A domain, which is why these residues represent the most common target. An apparent advantage of this kind of NRPS engineering is the evasion of interfering with the native inter- and intramolecular cooperation of the assembly line. However, with the increasing number of studies using this approach, it became evident, that the A domain is not the only domain that determines which building block is incorporated. Instead, the acceptor side of C domains also governs this process to different extents. Moreover, the successful substrate specificity switch of an A domain usually comprised the exchange of standard amino acids of similar size and/or polarity. Thus, mutating the A domain binding pocket alone is not sufficient in many cases and further modifications e.g. through mutasynthesis, are helpful to ensure the altered A domain could also be re-integrated into the native cluster *in vivo* to produce the modified product.

1.3.6 *Subdomain modifications*

Subdomain modifications are also targeted at A domains of NRPSs, for the same reasons and anticipated advantages as point mutations. However, for this type of modification, it is less obvious, what part of the domain to change, hence, there are fewer examples.

In two different studies, the successful swap of the substrate binding pocket between A domains within the hormaomycin[260] and the gramicidin[261] synthetases were reported. In both studies, less conserved regions within the A domain were identified and used as integration sites for the heterologous substrate binding pocket. In the gramicidin synthetase, the subdomain ($\neq A_{sub}$) of the initiation module $GrsA_{Phe}$ was exchanged with nine different subdomains encoding various specificities. All chimeras could be purified, however only the integration of a Val specific binding pocket lead to adenylation activity and the formation of a Val-Pro diketopiperazine (cyclic dipeptide) when the downstream module GrsB was supplemented[261]. Note that also for this type of A domain modification, the reaction rate for the chimera was 300-fold slower than for native GrsA and the demonstration of *in vivo* activity is missing.

Another A subdomain modification is the addition of a heterologous auxiliary domain into the naturally rather conserved site between A8 and A9 motifs. Two different methylation M domains, from KtzH(M_H) and TioS(M_{3S}), were inserted into the naturally uninterrupted A domain of Ecm6 from a *Streptomyces* strain[262]. In both versions, the A domain retained its activity *in vitro* and in addition the site-specific methylation patterns of the integrated M domains could be observed on the native substrate. From the engineering point of view, this was a great achievement. However, it remains questionable, whether the methylated substrate can be accepted and processed in the native context of the other Ecm modules. A previous attempt to exchange a complete uninterrupted A domain by an A(M) domain in a bimodular actinomycin NRPS caused a drastic decrease in the catalytic efficiency of the upstream C domain[263].

As for the point mutations, the success of A subdomain modifications largely depends on how well the modified domain can still interact with - and if the altered substrate can be processed by - the surrounding domains. If these two prerequisites are met, also for unknown reasons, even fusion sites within otherwise conserved domains can yield a functional chimeric enzyme.

1.3.7 *Whole domain and module rearrangements*

When rearranging whole domains and modules, an important aspect is finding the right fusion site between the non-cognate neighbors. Two different strategies are often employed: The first strategy is to find non-conserved regions for the fusion. The reasoning behind this strategy is that in un-conserved regions, changes have been tolerated during evolution, so an engineered fusion is unlikely to perturb an important function. The same strategy was also employed in the subdomain exchange described before[261]. The second strategy is based on the opposite assumption: a fusion site is searched in highly conserved regions, if both future neighbor domains share the same conserved sequence. Thus, the conserved sequence is supposed to maintain its role in the fused enzyme. The advantage of the second strategy is that fusion sites in conserved sequences can be located precisely within many NRPSs and therefore permit to fuse different NRPSs in the same position.

For whole module rearrangements alongside the fusion sites between the individual domains, another aspect to consider is the composition of a module. Naturally, a C-A-PCP module

structure is assumed, however, a A-PCP-C module has also been successfully reengined[264,265]. Even a $C_{C\text{-}term.}$-A-PCP-$C_{N\text{-}term.}$ module was reported very recently to serve as a general exchange module[266]. The anticipated advantage of exchanging whole modules is the functional cooperation within the native module to e.g. circumvent the problem of the C domain acceptor side gatekeeper function, if the C-A domains naturally belong together. On the downside, the native linker or interface between two adjacent domains is interrupted at the fusion site, which could lead to additional complications. Here I will summarize the engineering approaches starting with the exchange of individual domains, followed by di-domains and conclude with whole module rearrangements.

A very early exchange of A domains of tyrocidine and bacitracin synthetases to produce custom dipeptides *in vitro* resulted in a product yield similar to the native excised dipeptide synthetase *in vitro*[267]. This result nurtured the hope that through A domain exchange any custom NRPS could be produced. However, this hope was soon damped. In a different study, the pyoverdine Pvd synthetase producing a fluorescent compound was used as a model NRPS. Substitution of the native PvdD A domain specific for threonine with different synonymous A domains produced high levels of wild type pyoverdine *in vivo*[216,268]. When substituting with non-synonymous A domains, however, the modified Pvd synthetase failed to produce any modified product, instead trace amounts of the original pyoverdine were detected[216,268]. This effect was attributed to the C domain again, condensing the native substrate which has been falsely activated by the substituted A domain, as described before[252].

PCP domains have been the target of domain exchanges as well. The Heidelberg team of the iGEM (international Genetically Engineered Machine) competition in 2013, supervised by Barbara DiVentura and Roland Eils, substituted the PCP domain of the indigoidine synthetase IndC (section 1.2.11) with PCP domains of other NRPSs and several synthetic PCP domains that have been rationally engineered from a multiple sequence alignment[116]. They found that most substituted PCP domains did not lead to the production of indigoidine, not even the PCP domain of the homologous BpsA could restore pigment production. However, one of the synthetic PCP domains successfully replaced the native one and was further investigated with regards to the substitution borders. The highest titer of indigoidine production was reached when the insertion was on the left side directly downstream of the A domain (A-PCP linker provided by the inserted PCP domain) and on the right side at the end of the PCP domain (native PCP-TE linker provided by IndC)[116]. These results suggest that also the linker regions play an important role for NRPS engineering. Another PCP domain exchange specifically investigated the switch between elongation and termination module PCP domains[115]. They found that substituting one for the other kind was not successful. A random mutagenesis of the unfunctional inserted PCP domain generated functional variants. Rescue of substrate production was particularly attributed to mutations at positions +4 and +24 from the active site serine. And indeed, in a multiple sequence alignment, these two positions showed a discrimination between PCPs interacting with downstream C domains in contrast to PCP domains passing on the substrate to TE domains[115]. A di-domain exchange study replacing the A, A-PCP, or PCP-E domains of TycA (**Figure 1A**) with variants from the bacitracin synthetase showed that engineered versions could epimerize the alternative substrates Trp, Ile, and Val *in vitro* at reduced efficiency[269]. The success of these di-domain substitutions also largely depended on the PCP domain substituted, which needed to be naturally followed by an E domain to yield an epimerized alternative product.

Thus, the PCP domain does not only discriminate between downstream C and TE domains, but also E domains[269].

Instead of exchanging (part of) modules of a large NPRS enzyme, communication domain relocation was also successfully used to change the order of NRP synthesis and thus the sequence of the peptide product. In this way, three modules of the surfactin, bacitracin and surfactin synthetases could be reconfigured with matching communication domains and indeed they produced the expected tri-peptides[183], also *in vivo*[270]. Introduction of point mutations into the Com domains of the plipastatin synthetase changed their affinity for each other and thus the sequence of substrate incorporation. This resulted in the production of a variety of differently sized and arranged permutations of plipastatin[271]. While the modification or translocation of Com domains is a promising approach to NRPS engineering, it does not circumvent the problems such as interfering C domain specificity.

Deletion of modules or translocation of the TE domain within an NRPS was shown to produce the expected cyclic NRP of decreased ring size for the surfactin synthetase[272,273] and shorter cyclic as well as linear products of the plipastatin synthetase[274]. However, deletion of single domains or whole modules of plipastatin synthetase causes a module skipping process and thereby to an unexpected decrease of the product ring size by two building blocks instead of only the one that was deleted[275], a phenomenon that was also observed in a natural NRPS[276]. A successful case of module addition to the balhimycin synthetase to turn the native heptamer into an octamer was accompanied by the loss of a P450 tailoring function and thus also provides the disadvantage of this engineering approach[277].

A very comprehensive set of studies with the aim to re-engineer the calcium dependent antibiotic (CDA) daptomycin synthetase[135] to acquire improved variants included C-A di-domain, C-A-PCP module substitutions as well as deletion of auxiliary enzymes. Overall, about 120 modified daptomycin synthetases were generated, 42 of which were produced *in vivo* at levels between 3-100% of the wild type and thus sufficient for purification and *in vitro* characterization of the antimicrobial efficiency[278–281]. In this way, one of the few examples of a modified antibiotic with improved characteristics actually produced by an engineered NRPS could be presented[279].

Whole module exchanges of the C-A-PCP unit were successful in the lichenysin synthetase[282,283]. The first, Gln specific module was exchanged for a Glu specific one to produce a modified lipocyclopeptide in *Bacillus* close to wild type levels[282]. Through the recombination of C-A-PCP units of the enniatin, beauvericin, and PF1022 fungal NRPSs, chimeric synthetases were created which produced the expected peptides in *Aspergillus*[284]. The latter example indicates that fungal NRPS modules might be less problematic to reengineer than the bacterial ones, which have been used in most studies to date.

Choosing a different unit for the whole module exchange is another strategy that was introduced as a “XU” concept and used A-PCP-C units for the rearrangement. The basis for a new exchange unit was the discovery of a conserved motif within the C-A domain linker (WNATE) chosen as the fusion site, which lies within a flexible loop that is not otherwise responsible for interacting with its neighboring domains[265]. Based on this system, an exchange unit can only be followed by an alternative unit, where the A domain specificity matches the specificity of the natural downstream module, to ensure proper functioning of the C domain at the acceptor side[265]. Following these rules of using the conserved fusion sites and matching specificities, Bode

and co-workers successfully combined up to five modules of seven different NRPSs via homologous recombination in yeast and expressed them in *E. coli*. Several GameXPeptide[285] (a cyclopentapeptide) derivatives as well as completely novel peptides were produced by these recombinant NRPSs in the one to two digit mg/l range[265]. An even more independent type of exchange unit is represented by $C_{C\text{-}term.}$-A-PCP-$C_{N\text{-}term.}$, where the fusion site is located in the flexible hinge loop between the N- and C-terminal lobe of the C domain[266]. This second-generation exchange unit ($XU_{2.0}$) was reported to allow the free recombination of modules, quasi-independent from their native up- and downstream partner, because the C domain acceptor and donor side are always presented with their cognate substrate. In this way, different $XU_{2.0}$ could be combined in yeast[247] to give rise to a NRPS library for GameXPeptide derivative production[266]. Finally, it was tested whether and how an elongation A domain can be turned into an initiation one for *in vivo* NRP production and found that in order to achieve this conversion, either the complete or at least the C-terminal lobe of the native upstream C domain needs to be present. When only the C-A domain linker was left at the N-terminus, no product was detected after expression in *E. coli*[266]. While *in vitro* neither the use of an elongation A domain as an initiation domain[264] nor the combination of (C-A-PCP) modules of non-matching specificities[34] has caused the inactivity of the rearranged NRPS, this is the first time these achievements were reported in *in vivo.*

1.3.8 *Open questions and problems of NRPS re-engineering*

Nature has been very successful in exchanging NRPSs between species, reengineering them to produce novel and more potent bioactive molecules and even inserting external domains to add to the diversity of their products. We are not yet able to take advantage of the same strategies in the laboratory environment. The original hope in NRPS engineering was to elucidate general design rules to reliably modify these natural assembly lines to produce custom peptides and test them for enhanced bioactivity. If this is ever possible, as in evolution, only time can tell.

Research up until now has identified diverse characteristics of NRPSs which can turn into pitfalls when trying to modify them as described in the previous sections. Among them are C domain specificity at donor and acceptor side, boundaries between the rearranged domains, turning elongation into initiation A domains, a discrepancy between results observed *in vitro* and *in vivo*, unforeseeable effects of neighboring domains, their interfaces and interdomain linkers. Some of these issues can be prevented by smart design engineering. Others can be resolved by directed evolution[286], a process for which the 2018 Nobel Prize in chemistry was awarded to Frances Arnold, George Smith and Gregory Winter[287]. Following either strategy, a great number of variants need to be screened in order to find the active ones.

Structural guidance for domain and module rearrangements has been provided by crystal structures of individual domains as well as multi-domain constructs trapped in a specific state by point mutations[168] or mechanism-based inhibitors[48,50,76,104]. Each structure represents a staged snapshot of the synthesis process, but the overall movie has yet to be shot. One step in this direction was recently taken by comparing the crystal structure and negative stain electron microscopy (EM) images of a A-PCP-C tri-domain construct in complex with an MbtH-like protein[76]. The EM images supported the idea that domains within an NRPS module adopt many conformations and do not seem to exist in a set of well-defined states[224]. NMR studies provided insights into conformational changes in solution, but are limited to relatively small proteins and

therefore only individual PCP and TE domains or PCP-TE di-domains have been analyzed[96,165,288]. Thus, it is still unclear how individual modules are structurally and functionally embedded within the context of full-length or at least multi-modular truncated NRPSs.

To sum up, a successful NRPS engineering approach according to the current state of the art includes a series of complementary considerations and experiments: (i) consider functional and structural knowledge for design engineering where possible; (ii) compensate unforeseen errors in the design using directed evolution; (iii) screen a sufficiently large number of engineered variants for activity at best *in vivo*. Using a combination of these approaches, maybe it is possible to come one step closer to generating NRPs with enhanced bioactivity via NRPS engineering.

2. Aims of the study

The easiest readout for any kind of assay is color. Color can be detected by the bare human eye even quantitatively to a certain degree. Using relatively basic equipment, the amount of pigment causing the coloration of the cells or the medium can be precisely quantified. In this work I explore the possibility to append a unimodular NRPS that produces a blue pigment (**Figure 6**) at the end of any NRPS, be it engineered or natural. The idea is to use the color to easily monitor the production of the peptide *in vivo*, which would represent a break-through in projects aimed at engineering large libraries of novel NRPSs. The underlying assumption is that the pigment would be produced only if the upstream modules work together and the engineered peptide is made. The peptide would be tagged with the pigment. I, therefore, further assume that the pigment would not lose, but possibly change, its color once fused to other AAs.

To achieve this goal, I set myself the following tasks:

Analyze the blue pigment indigoidine and its synthetase

As a pigment I chose indigoidine, which is produced by a standalone unimodular NRPS from two glutamine molecules. I aim to investigate two indigoidine synthetases, namely IndC and BpsA, *in vivo* and *in vitro*. Further I want to test whether I can assess the effects of modifications on the indigoidine synthetase via pigment quantification. In addition, I want to explore a putative synthetase of dodecylindigoidine (a naturally tagged-version of indigoidine) to see whether this enzyme could also serve our purpose.

Construct fusions of single NRPS modules to the indigoidine synthetases IndC and BpsA to test if the tagging works

In the second part, I aim to put the idea of tagging to the test. I will construct fusions between a module incorporating a certain AA and IndC/BpsA. To increase the chances of success, I selected a module from the tyrocidine synthetase incorporating asparagine that is naturally followed by a glutamine-incorporating module. This part of the project will help me tackle open questions in NRPS engineering, such as (i) where to best fuse the two non-cognate NRPSs, (ii) how to turn elongation into initiation modules and vice versa and (iii) how mutations directed to a specific module affect this module and its neighboring modules.

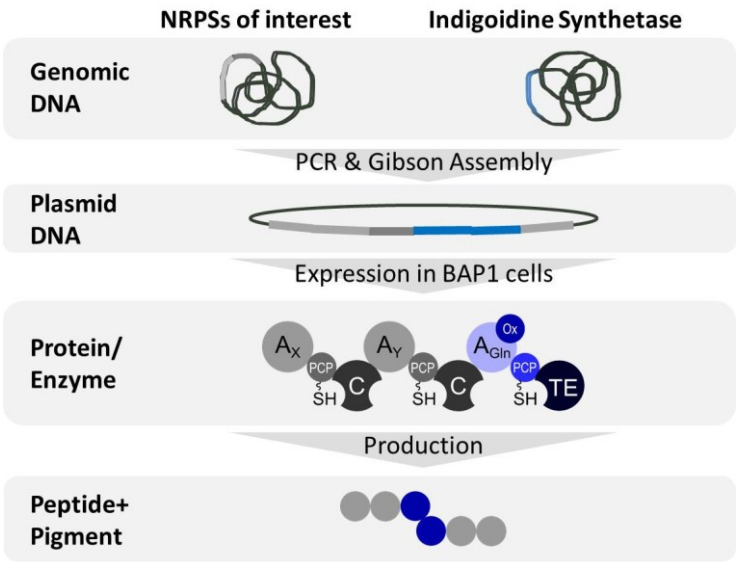

Figure 6 | Concept of tagging NRPSs with an indigoidine synthetase to monitor NRPS manipulations using an easy, visual readout.

Analyze inter- and intramodular communication

Finally, beyond the tagging approach, I wish to address the question of how different modules interact with each other using constructs consisting of either one, two or three modules extracted from the same natural NRPS, in this case the tyrocidine synthetase. I aim at introducing point mutations on the A and PCP domains on a certain module and study their effects when the module is alone or connected with other modules. In this way I hope to gain further insights into the inter-, and intra-modular communication between NRPS modules which could be very important for future engineering approaches.

3. Materials and Methods

3.1 Materials

3.1.1 *Chemicals*

Unless otherwise indicated, chemicals were purchased from Sigma-Aldrich (St. Louis, Missouri, USA) or Carl Roth (Karlsruhe, Germany). Kits for plasmid purification, DNA gel-extraction and PCR clean-up as well as the corresponding vacuum manifold were purchased from Qiagen (Hilden, Germany). For bacterial genomic DNA extraction, GenElute Bacterial Genomic DNA Kit (Sigma Aldrich) was used. Custom peptides were produced by PSL Peptide Specialty Laboratories (Heidelberg, Germany). Dansyl-L-glutamine was ordered from Santa-Cruz Biotechnology (Dallas, Texas, USA). Short, single-stranded custom DNA oligos (primers) were ordered from Sigma-Aldrich. Long, double stranded custom DNA oligos were ordered as gBlocks® from IDT Integrated DNA Technologies (Coralville, Iowa, USA). As DNA vectors for replication and protein expression in *E. coli* I used pTrc99A[289], pSB1C3[290], pCDFDuet™ and pET-28(a)+ (both Merck, Darmstadt, Germany). For more details on the expression vectors, please see **Table 3**. Ni-NTA beads from Bio-Rad (Hercules, California, USA) were used for immobilized metal affinity chromatography in Poly-Prep® chromatography columns (also Bio-Rad).

Table 3 | Vectors and their characteristics.

Vector:	Resistant to:	Other features:
pTrc99A	ampicillin	Hybrid trp/lac promoter; IPTG inducible; pBR322 origin of replication
pCDFDuet	spectinomycin	Two multiple cloning sites (MCS) each with an upstream T7 promoter, lac operator, and ribosome binding site (rbs); CDF origin of replication; lacI gene; IPTG inducible
pET-28(a)+	kanamycin	T7 promoter; IPTG inducible; pBR322 origin of replication
pSB1C3	chloramphenicol	pUC19-derived pMB1 origin of replication; IPTG inducible; T7 promoter
pMiniHimar-Rb1	kanamycin	Transposase under Plac promotor; R6K origin of replication; inverted repeats flanking “transposon”

3.1.2 *Labware*

Graduated Pipette Tips TipOne® in sizes of 10 µl, 200 µl and 1000 µl were ordered as re-fill packs from Starlab (Hamburg, Germany). Serological Pipettes Costar® Stripettes® in sizes 5 ml, 10 ml and 25 ml were purchased from Corning® (New York, USA). Single use polypropylene 14 ml tubes were bought from Greiner Bio-One (Frickenhausen, Austria). Reusable glass 14 ml tubes were provided by Schott Duran (DWK Life Sciences, Wertheim, Germany). Before each usage, glassware was cleaned and autoclaved. Glass bottles and flasks were provided by Schott Duran

as well. Micro reaction tubes with a safe lock lid in the volume of 1.5 ml and 2 ml were bought from Eppendorf (Hamburg, Germany). Micro reaction tubes with a volume of 0.2 µl were purchased as individual tubes and as 8-well tube strips from Starlab. BRAND® 1.5 ml PP screw cap tubes with an attached lid were purchased from Sigma-Aldrich. 15 ml and 50 ml reaction tubes were purchased from Greiner Bio-One (Cellstar®) or Corning (Falcon®). Clear 96 well plates (PlateOne, flat bottom) were bought from Greiner Bio-One as well.

3.1.3 *Equipment*

A set of mechanical pipettes (Tacta by Sartorius, Göttingen, Germany) to transfer small amounts of liquids were used in combination with graduated pipette tips for volumes between 0.5 µl to 1 ml. An electric pipette boy (pipetboy comfort by Fisher Scientific) was used in combination with serological pipette tips to transfer volumes of 5 to 25 ml. Thermo Cyclers for DNA amplification were used from Bioyzym (LifeTouch Thermal Cycler) and Eppendorf (Mastercycler Nexus Gradient GSX1 Thermal Cycler). Separated DNA fragments in an agarose gel stained with GelRed were visualized in a UV chamber with attached CCD camera (both Intas Science Imaging Instruments, Göttingen, Germany). Tabletop centrifuges for 24 samples up to a volume of 2 ml were used from Beckman Coulter (Brea, USA), Microfuge 22R and 20R. For larger samples, the Allegra 25R and Avanti J-26XP also from Beckman Coulter were used. The tabletop thermo shaker PHMT was fabricated by Grant Instruments (Cambridge, UK). DNA and protein concentrations were measured on a NanoDrop™ 2000c (Thermo Fisher Scientific). The pH of solutions was measured with a 50+ series bench pH meter by Chromservis (Prague, Czech Republic). Bacterial cells were lysed using a Sonopuls HD 2070 with ultrasound transducer UW 2070 with a MS 73 micro tip by BANDELIN electronic (Berlin, Germany). To remove water from samples to complete dryness, they were frozen in liquid nitrogen and dried in an Alpha 1-2 LD plus lyophilizer by Christ (Osterode am Harz, Germany) attached to a chemistry hybrid pump by Vacuumbrand (Wertheim, Germany). Small molecule MS was performed on a micrOTOF II (Bruker: Billerica, Massachusetts, USA) by Heiko Rudy, IPMB. LC-MS of small molecules were performed on an Agilent 6100 Series (single Quadrupole, electrospray atmospheric pressure ionization) instrument with a Kinetex 2.6 µm C18 column (Phenomenex, 50 x 2.1 mm). NMR spectra were recorded on a Varian Mercury Plus 500 MHz spectrometer and a Varian Mercury Plus 300 MHz spectrometer by Tobias Timmermann at the IPMB.

3.1.4 *Enzymes*

Restriction endonucleases, CutSmart buffer and the Gibson Assembly Master Mix were purchased from New England Biolabs NEB (Ipswich, Massachusetts, USA). The high fidelity DNA polymerase Phusion Flash High-Fidelity was used as the 2X PCR Master Mix from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Lower fidelity Taq polymerase OneTaq® was used as the 2X Master Mix in standard buffer including loading dye from NEB. Random mutagenesis was reached using an error-prone Taq DNA polymerase and PCR conditions that favor the introduction of point mutations, all provided in the error-prone PCR reaction kit by Jena Bioscience (Jena, Germany). *In vitro* protein expression was tested with the PURExpress® Protein Synthesis Kit by NEB.

Enzymes (and substrates) for the online pyrophosphate assay were all purchased from Sigma-Aldrich. The reduced NADH powder was ordered from GERBU Biotechnik (Heidelberg, Germany).

3.1.5 *Bacterial strains*

One Shot™ TOP10 Chemically Competent *Escherichia coli* were purchased from Thermo Fisher Scientific and used to prepare chemically competent *E. coli* in house. Lei Fang from the Pfeifer Lab (Prof. Blaine Pfeifer, Department of Chemical and Biological Engineering, University of Buffalo, USA) kindly provided the iGem team Heidelberg 2013 with the *E. coli* BAP1 strain[109] engineered from *E. coli* BL21(DE3) with a stably integrated PPtase (Sfp from *Bacillus subtilis*[105]). BAP1 cells were also made chemically competent in house. We used TOP10 cells for plasmid amplification and BAP1 cells for protein expression.

Shewanella violacea DSS12 was provided by Institute Pasteur, Paris, France. *E. coli* EC100 pir+ is a cloning strain for plasmids with an R6K ori and was purchased from Epicentre Technologies (Madison, Wisconsin, USA). Both were used to experimentally validate the putative N,N-dodecylindigoidine synthetase dIndS.

Photorhabdus laumondii luminescens (DSM 15139) and *Streptomyces lavendulae lavendulae* (DSM 40708) were ordered from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Leibzig, Germany) as glycerol stocks. *Bacillus amyloliquefaciens* (DSM 7) genomic DNA was also ordered from DSMZ. *Brevibacillus parabrevis* (ATCC 8185) was kindly provided to the iGem team Heidelberg 2013 by Prof. Mohamed A. Mahariel (University of Marburg, Marburg, Germany). *Bacillus subtilis 168* was generously provided by Dr. Ilka Bischofs-Pfeifer, MPI for Terrestrial Microbiology, Marburg, Germany. Prof. Dr. Erhard Bremer, Laboratory for Microbiology, Faculty of Biology, University of Marburg kindly provided us with *Bacillus licheniformis* DSM 13. Prof. Fussenegger sent us the expression plasmids for *bpsA* codon optimized for bacterial cells pMM6419. These strains and their genomic DNA served as templates to extract NRPS genes of interest.

3.1.6 *Buffers, media and antibiotics*

For DNA separation, TEA buffer (AppliChem) was bought as 10x stock solution. To prepare running buffer and 1 % agarose gels (LE agarose, Biozym, Hessisch Oldendorf, Germany), it was diluted to 0.5 % in deionized water. DNA was visualized using 1 % GelRed® (Biotum, Fremont, CA, USA) in the 6x DNA loading dye (TriTrack, Thermo Fisher Scientific) and in the DNA ladder (GeneRuler Ladder Mix, also Thermo Fisher Scientific).

For protein separation, SDS-PAGE was performed using NuPAGE (Thermo Fisher Scientific) gels and 20x buffers. Depending on protein size, 10 % Bis-Tris, 4-12 % Bis-Tris or 3-8 % TA gels were utilized (1 mm thick) in combination with MES, MOPS or TA buffer diluted to 1x in deionized water.

LB-Miller (lysogeny broth according to Miller) medium was prepared by dissolving 5 g/l Bacto™ yeast extract (Bacto Biosciences, Franklin Lakes, New Jersey, USA), 10 g/l Bacto™ Tryptone (Bacto Biosciences) and 10 g/l sodium chloride in ultrapure water (GenPure, TKA Wasseraufbereitungssysteme, Niederelbert, Germany). For LB agar plates, 12 g/l bacteriology grade agar (AppliChem, Darmstadt, Germany) was added. Before use, the medium was autoclaved at 121 °C for 20 minutes. Once cooled to below 40 °C, the desired antibiotic can be added to the medium. For agar plates, the medium was poured into petri dishes (Greiner Bio-One) and allowed to harden at room temperature. Once the antibiotic is added, the medium is stored at 4 °C.

Antibiotics are used in the following concentrations (**Table 4**):

Table 4 | Antibiotics concentrations in stock and working solutions.

Antibiotic	Stock solution [mg/ml]	Dilution factor	Working concentration [µg/ml]
Ampicillin (AppliChem)	100	1:1000	100
Chloramphenicol (AppliChem)	30	1:1000	30
Kanamycin (AppliChem)	50	1:1000	50
Spectinomycin (AppliChem)	100	1:1000	100

Supplemented Difco Marine Broth 2216 (MB), pH 7.6, contains 5 g/l peptone, 5 g/l yeast extract (e.g. BD BactoI Yeast Extract, 212750), 2 g/l casamino acids and 100 ml 10x MB salt concentrate (1 g/L ferric citrate, 194.5 g/l sodium chloride, 59 g/l magnesium chloride, 61.6 g/l magnesium sulfate hexahydrate, 36 g/l calcium chloride hexahydrate, 5.5 g/l potassium chloride, 1.6 g/l sodium bicarbonate, 0.8 g/l potassium bromide, 0.571 g/l strontium chloride hexahydrate, 0.22 g/l boric acid, 0.04 g/l sodium silicate, 0.024 g/l sodiumfluoride, 0.016 g/l ammonium nitrate and 0.100 g/l disodium phosphate 46ehydrate). For supplemented MB agar plates, 12 g/l bacteriology grade agar was added.

M9 minimal medium was modified from the original recipe by Miller[291]. Our optimized version for NRPS production and pigment absorption measurements contained 1x M9 salts (M9-Minimal salts 5x, powder, SERVA, Heidelberg, Germany), 1 mM $MgSO_4$, 0.5 % casamino acids (OmniPur®, Merck, Darmstadt, Germany) and 2 mg/ml Glucose. Antibiotics were added according to **Table 4**. The final M9 minimal medium was filtered prior to use.

3.1.7 *Informatics tools*

DNA sequences were handled with ApE, A plasmid Editor[292], for gene and domain annotation, plasmid map construction, cloning strategies, sequencing result alignments and DNA sequence translation. Primer design for site directed mutagenesis was assisted by PrimerX[293] according to the parameters described by Edelheit *et al*[294]. Protein sequences were aligned using Clustal Omega[295] and protein parameters were estimated based on their sequence with ExPASy ProtParam[296,297]. Protein structures were downloaded from PDB[298] or predicted in RaptorX[299] and visualized in UCSF Chimera[300]. NRPS protein sequences were accessed from online databases Norine[237], ClusterMine360[160] and in general from NCBI[301]. NRPS A domain specificity prediction was predicted with NRPSsp[61].

The plate reader was programed, and data was recorded via the Magellan™ software (also provided by TECAN). Data analysis was performed in Excel (Microsoft, Redmond, Washington, USA). Graphics were created in InkScape[302] and PowerPoint (Microsoft). Protein Alignments were made pretty in JalView[303]. NMR data was analyzed in the ACD/Labs (Toronto, Ontario, Canada) NMR Processor Academic Edition 12.0. The bibliography was created in Zotero[304] and this Thesis was drafted in Word (Microsoft).

3.2 Methods

3.2.1 *Prepare chemically competent E. coli*

To prepare chemically competent *E. coli,* the desired strain was plated on an LB agar plate (without antibiotic if not otherwise indicated) and grown over night at 37 °C. Over the following day, the plate was stored at 4 °C before an individual colony was picked, transferred into 50 ml LB medium in a 200 ml glass flask and incubated overnight at 37 °C, shaking at 180 rpm. In the morning, 400 ml LB medium were inoculated with 4 ml overnight culture and grown at 37 °C and 180 rpm to an OD600 of 0.5-0.6. In the meantime, 200 ml 100 mM $CaCl_2$ and 20 ml 10 % glycerol in 100 mM $CaCl_2$ were chilled on ice. About 200 1.5 ml micro reaction tubes were also precooled on large metal racks on ice and covered with aluminum foil to prevent contamination. Once the bacterial culture has reached the desired OD600, cells were pelleted in 50 ml reaction tubes for 20 minutes at 4100 rpm at 4 °C. The supernatant was discarded and a washing step followed. On that account pellets were re-suspended in 5 ml ice cold 100 mM $CaCl_2$ solution. After combining two pellet re-suspensions each per reaction tube, the volume was filled up to 50 ml with 100 mM $CaCl_2$ solution and incubated for 30 minutes on ice. Cells were then pelleted again at 3500 rpm for 15 minutes at 4 °C and the supernatant discarded. All pellets were then combined in 10 ml 10 % glycerol in 100 mM $CaCl_2$ and pipetted up and down until homogeneous. Another 10 ml 10 % glycerol in 100 mM $CaCl_2$ was added to the cell re-suspension up to a total volume of 20 ml. Finally, 100 µl cell aliquots were dispensed in the pre-cooled micro reaction tubes with a multistep pipette (Multipette® plus, Eppendorf), frozen in liquid nitrogen and stored at -80 °C.

3.2.2 *DNA amplification via PCR and purification*

DNA fragments for molecular cloning were amplified using the Phusion Flash high fidelity DNA polymerase in a 2X reaction mixture. Primers (listed in section 6.1), template, ultrapure water and up to 3 % DMSO were combined according to the manufacturer's instructions prior to adding the Master Mix in a 1:1 volume ratio. Samples were then placed in a thermal cycler running the following temperature cycles (Table 5).

Table 5 | Temperature cycles for DNA amplification using high fidelity Phusion Flash DNA polymerase.

	Temperature [°C]	Time [s]
1 - Initial Denaturing	98	10
2 - Denaturing	98	1
3 - Primer Annealing	49 - 72, primer dependent	5
4 - Elongation	72	15/kb
5 - Final Elongation	72	20/kb

Steps 2 to 4 were repeated 30 times.

Amplified DNA fragments were separated on a 1 % agarose gel stained with GelRed for 30 minutes at 135 Volts in 0.5 % TAE buffer. A DNA ladder was run in parallel on each gel to estimate the DNA fragment size and bands were visualized in a UV chamber. DNA fragments of the expected size were excised from the gel using a scalpel and purified using a gel-extraction kit according to the manufacturer's instructions. Purified DNA fragments were stored at -20 °C in ultrapure water.

3.2.3 *Cloning via Gibson Assembly*

Gibson Assembly is an isothermal reaction joining DNA fragments that contain an overlapping sequence of minimum 16 bp with their adjacent piece of DNA[246]. We used a commercial 2X Gibson Assembly Master Mix to join purified DNA fragments. The necessary overlap was created by performing a PCR with primers that have the desired overlap as a non-complementary attachment at their 5` end. Vector backbones were obtained either via PCR or by classical restriction enzyme digest. The reaction itself was performed according to the manufacturer's recommendations, with small adjustments. In short, all DNA fragments including the backbone were added in a micro reaction tube at a concentration of 0.01 pmol (the smaller the fragment, the higher the applied molar concentration) to a volume of 5 µl. To the DNA mix, the 2X Gibson Assembly Master Mix was added in a 1:1 volume ratio and the reaction mix was incubated at 50 °C for 15 minutes to 1 hour, depending on the amount of DNA fragments to be joined. Finally, the reaction mix was diluted 1:4 in ultrapure water and transformed into chemically competent TOP10 cells.

3.2.4 *Transformation*

First, the desired plasmid was transformed into chemically competent bacterial cells. For that, 0.5-4 µl purified plasmid DNA was combined with 25-100 µl thawed competent cells on ice and incubated for 5-30 minutes. The amount of DNA and cells as well as the initial incubation step on ice depend on the state of the DNA: For a retransformation of purified DNA plasmid, less DNA/cells and shorter incubation times were chosen while for a transformation of a DNA mix after Gibson assembly more DNA/cells were combined and a longer incubation time was allowed. Cell-DNA suspension was then heat shocked for 30 seconds at 42 °C, followed by incubation on ice for 2 min. For Ampicillin resistant plasmids, transformed cells were immediately plated out on a pre-heated LB agar plate containing ampicillin and incubated o/n at 37 °C. For plasmids featuring a different resistance gene, 900 µl LB medium without antibiotic was added to the transformed cells and they were grown at 37 °C shaking at 180-200 rpm for one hour prior to being plated on the pre heated LB agar plate plus matching antibiotic and incubated at 37 C° o/n. Colonies on LB agar plates can be stored for up to 5 days at 4 °C. For long term storage, a glycerol stock was prepared and stored at -80 °C.

3.2.5 *Colony PCR and plasmid extraction from positive clones*

To check whether an *E. coli* colony contained the desired plasmid, we performed a colony PCR. Individual colonies were resuspended in 50 µl ultrapure water and 1 µl of this resuspension was used as a template for PCR. Primers were chosen to bind on two different fragments of the Gibson Assembly and amplify a sequence no longer than 4 000 bp. 0.5 µl of 10 µM forward and reverse primer were added to the template as well as 3 µl ultrapure water. As a DNA polymerase, we used OneTaq polymerase as a 2X Master Mix, which was added to the reaction

mix in a 1:1 ratio and run in a thermal cycler according to the manufacturer's instructions (**Table 6**). Steps 2 to 4 were repeated 30 times. Expected fragment size was checked on a 1 % agarose gel.

Table 6 | Temperature cycles for DNA amplification using Taq polymerase.

	Temperature [°C]	Time [s]
1 - Initial Denaturing	95	300
2 - Denaturing	95	15
3 - Primer Annealing	49 - 68, primer dependent	15
4 - Elongation	68	60/kb
5 - Final Elongation	68	60/kb

Positive clones were picked and used to inoculate a 4 ml overnight culture in LB medium plus the appropriate antibiotic. The desired plasmid was then extracted from the bacterial cells using the QIAprep Spin Miniprep Kit (Qiagen). DNA sequence was verified via Sanger sequencing at GATC (Konstanz, Germany). Purified DNA plasmids were stored at -20 °C in ultrapure water. A list of plasmids which were used or generated in this work is provided in section 6.2.

3.2.6 *Cultivation of Shewanella violacea and whole genome extraction*

Untargeted gene disruption of *Shewanella violacea* was performed by Konrad Herbst as described before[305]. Conjugation of *S. violacea* with the disruptive pMiniHimar-RB1 plasmid was performed at 8, 12 and 18°C and plated out with a dilution factor of 1:10, 1:10 and 1:100 respectively as described in the internship protocol of Konrad Herbst[221]. *S. violacea* with disrupted genomes were incubated on kanamycin containing supplemented marine broth (MB) agar plates for 2 weeks at 8°C. Subsequent storage of these plates at 4 °C for several weeks lead to darker pigmentation of some of the colonies. To identify the disrupted genes, four pigmented and four non-pigmented colonies were picked and grown in supplemented MB medium plus kanamycin to an OD of minimum 0.6 at 180 rpm and 8 °C. The cells were harvested and whole genome extraction was performed using the GenElute Bacterial Genomic DNA Kit (Sigma Aldrich) according to the manufacturer's instructions. To identify the disrupted genes, two approaches were chosen: Firstly, the extracted genomic DNA was digested with BamHI, religated and transformed into electro-competent *E. coli* EC100 pir+ to recover DNA fragments which have the transposon construct integrated. Secondly, I was specifically looking for an integration event within the putative dIndS gene, so I performed a PCR across the putative gene using three primer pairs spanning whole sequence. In case transposon integration happened, the PCR product should be longer (~4.5 kbp instead of 2 kbp).

3.2.7 *Site-directed mutagenesis*

For site-directed mutagenesis, I modified a protocol developed by Edelheit *et al.*[294] Forward and reverse primers containing the desired mutation were designed according to the conditions proposed in their publication. Two separate PCR reactions were set up in a total of 10 µl each,

both containing the template plasmid, one with the forward primer and one with the reverse primer. DNA plasmids containing the mutation were thus amplified linearly as a single stranded DNA fragment by the Phusion Flash polymerase master mix. After the PCR, both reactions were combined, denatured and slowly re-annealed to favor the annealing of forward and reverse strands containing the mutation. The reaction was then digested over night with DpnI according to the manufacturer's recommendations to eliminate the template plasmids. Finally, the reaction mix was transformed into TOP10 cells and mutations were verified by Sanger sequencing.

3.2.8 *Random mutagenesis via error-prone PCR*

To perform directed evolution, I chose error-prone PCR to randomly introduce mutations solely into a desired DNA sequence. I used the error-prone PCR kit by Jena Bioscience and followed the manufacturer's instructions. Since the maximum optimal length for random mutagenesis with the polymerase provided in this kit is 4 000 bp (oral communication from Jena Biosciences), longer PCRs were split up into two reactions using overlapping primers for Gibson Assembly. Mutated DNA fragments were then purified like described before and inserted into a non-mutated backbone via Gibson Assembly.

3.2.9 *Bacterial glycerol stock*

To prepare a glycerol stock, an individual colony of the bacterial cells transformed with the desired plasmid(s) was used to inoculate 4 µl LB medium plus suitable antibiotic. The bacterial culture was grown o/n at 37 °C shaking at 180-200 rpm. After 16-18 hours, 800 µl o/n culture were combined with 200 µl 100 % Glycerol, transferred by a trimmed 1000 µl pipette tip, into a 1.5 ml tube with screw cap. The glycerol stock was then stored at -80 °C.

3.2.10 *Small scale protein expression and absorption measurement*

Small scale protein expression was mainly performed to screen new fusion NRPSs for their ability to produce the blue pigment indigoidine. I found the colorless M9 minimal medium to be best suited for protein expression as well as detecting the produced pigment via absorption measurement afterwards. Thus, a starter culture consisting of 4 ml LB medium plus antibiotic was inoculated with a single colony of plasmid transformed into the expression cells BAP1 and grown o/n at 37 °C shaking at 180-200 rpm.

After 16 to 18 hours, 3 ml M9 minimal medium plus suitable antibiotic were inoculated with 6 µl of the overnight culture and grown at 37 °C and 200 rpm to an OD600 of 0.8-1.0. Cultures were then cooled down on ice for 5 minutes before inducing protein expression by adding 0.5 mM IPTG. Further, proteins were expressed at 18 °C and 200 rpm for 18-24 hours. The absorption of 200 µl overnight culture in a clear 96 well plate was then measured in the plate reader from 300 to 800 nm.

3.2.11 *Calculation of indigoidine production from absorption of the overnight culture*

In a smallscale overnight expression culture, (engineered) NRPSs are expressed, post-translationally modified with a PPant arm on a specific serine within each T domain and thus potentially active in the expression strain. Since the project aimed to easily identify functional engineered NRPSs via a blue pigment tag, the readout of indigoidine synthesis should be easy as well. Therefore, I measured the absorption of the overnight expression culture and calculated

the relative amount of blue pigment according to Myers *et al*[306]. The absorption maximum of indigoidine lies at 600 nm, which is the same absorption at which you typically measure bacterial cell density. Thus, to assess the amount of pigment independently from the cell density, I measured the absorption of 200 µl overnight expression culture in a clear 96 well plate in the TECAN plate reader at both, 600 and 800 nm. The ratio δ between OD600 and OD800 of a negative control sample (e.g. no blue pigment expression or an un-induced sample) was calculated and used to infer the relative amount of indigoidine in the samples of interest as follows:

$$\text{relative pigment production of construct X} = \text{OD600}_{\text{X}} - (\delta * \text{OD800}_{\text{X}})$$

Each construct was expressed in triplicate, induced from the same overnight culture, and relative pigment production was calculated. For IndC or BpsA wild type the OD800 was elevated due to the high absorption at 600 nm and thus the OD800 of the respective negative control was assumed for relative pigment amount calculation. Finally, the relative pigment production of all samples was normalized to the wt indigoidine synthetase. In order to take small changes in the procedure into account, like cell density of the overnight culture or duration of expression, one negative and one positive control were expressed in parallel for each batch of small-scale protein expression. In that way, distinct calculation of pigment production per experiment could be performed.

3.2.12 *Calculation of indigoidine concentration from absorption values*

The molecular extinction coefficient of indigoidine[195] in DMF is ε = 23400 M^{-1} cm^{-1} and can be used to calculate the approximate indigoidine concentration from an absorption measurement.

$$c\ (indigoidine) = \frac{OD_{600}}{(\varepsilon\ *\ d)}$$

3.2.13 *Large scale protein expression*

Large scale expression cultures were set up to purify the NRPSs of interest. In this case, an overnight culture was set up from an individual colony like previously described and 800 µl overnight culture was used to inoculate a 400 ml LB medium in a 1 l flask. The expression culture was grown at 37 °C shaking at 180-200 rpm for about 3 h to an OD600 of 0.8-1. Once this OD was reached, the incubator was cooled to 18 °C for about 30 minutes. Once this temperature was reached, protein expression was induced with 0.5 mM IPTG. Proteins were expressed for 18 to 20 hours before being harvested in 400 ml centrifuge tubes at 4000 g and 4 °C for 30 minutes. Supernatant was discarded and the cell pellet was frozen (either plain or re-suspended in buffer) at -80 °C.

3.2.14 *His-tag based affinity protein purification*

Proteins were all equipped with a hexa-histidine tag (His6-tag) and purified via immobilized metal affinity chromatography using Ni-NTA beads. The cell pellet of a 400 ml overnight expression culture of the desired protein was resuspended in 12 ml lysis buffer (50 mM TRIS or HEPES, pH at 4 °C adjusted to 7.5, 5 mM imidazole, 500 mM NaCl and 12.5 % glycerol) and frozen at -80 °C for a minimum of 12 hours. Resuspended pellets were then thawed at room temperature (about 25 °C) shaking lightly until the solution was almost free of ice crystals. All further steps were performed on ice or at 4 °C. Thawed samples were lysed via sonication (100 % power, 3 min, 30 s intervals, 6/10s pulses) and cleared for 35 minutes at 15 000 rpm and

4 °C. The supernatant containing the desired protein was added to 1.5 ml to 2 ml of Ni-NTA beads in 50 % water (so 0.75 to 1 ml of pure beads) and incubated for 2 h gently shaking at 4 °C for the proteins to bind to the beads. The bead slurry was added to chromatography columns and the supernatant was eluted via gravity flow. The beads were washed twice with 3-4 column volumes of lysis buffer. Finally, the protein was eluted in two steps with 1.5 column volumes elution buffer (50 mM TRIS or HEPES, pH at 4 °C adjusted to 7.5, 250 mM imidazole, 500 mM NaCl and 12.5 % glycerol). Different fractions of the eluate were collected manually and 0.75 µl were spotted onto a nitrocellulose membrane (Thermo Fisher Scientific). The membrane was then stained for 5 minutes in Ponceau S solution to identify the fractions with the highest protein concentration. Those fractions were pooled and added onto Amicon® Ultra 0.5 ml spin columns with a 100, 50 or 10 kDa MWCO, depending on protein size (Merck). Purified proteins were then concentrated and buffer exchanged to storage buffer (50 mM sodium phosphate buffer, 20 % glycerol, pH 7.25 at 4 °C) five to ten times at 14 000 rpm for 7 minutes at 4 °C. In this way, smaller protein contaminations were also reduced. Proteins were then either stored at 4 °C for immediate use in an *in vitro* assay or glycerol content was increased to a total of 40 % and proteins were stored at -20 °C. Protein amount was measured via absorption at 280 nm and molar concentrations were calculated through division by the molecular weight of the respective protein.

3.2.15 ***Protein separation via SDS-PAGE***

Size analysis of proteins of interest was performed via SDS-PAGE. For a quick expression check, 400 µl of overnight expression culture were pelleted at 13 000 g for 2 minutes. The supernatant was discarded and the pellet was resuspended in 30 µl 1x FastBreak™ Cell Lysis Reagent (Promega, Madison, Wisconsin, USA) to lyse the cells. 10 µl 4x NuPAGE™ LDS Sample Buffer (Thermo Fisher Scientific) were added to each sample before denaturing the proteins for 10 minutes at 95 °C. Before loading into the gel's wells, samples were spun down at 19 000 g for 10 minutes to pellet the DNA and only load the protein containing solution. For size comparison, one lane with Precision Plus Protein™ Dual Color Standards (Bio-Rad) was run in parallel to all samples. For 15 well gels, 10 µl sample was loaded per pocket, for 10 well gels, 20 µl sample was loaded into each pocket. The type of gel was determined by the predicted protein sizes to be separated. Gels were run in the matching 1x buffer in the electrophoresis chamber at a constant 180 Volts for one hour. The plastic support around the gel was then removed and the gel stained in a Coomassie staining solution (0.1% Coomassie Brilliant Blue G-250 (Thermo Fisher Scientific), 50 % methanol and 10 % glacial acetic acid) for 30 minutes to 1 hour, shaking lightly. Protein bands became visible after destaining the gel overnight in deionized water. Gel bands were documented with a CCD camera (iPhone 5, Apple, Cupertino, California, USA) against a white background.

For SDS-PAGE of purified proteins, the procedure was very similar. A small sample was diluted in 50 mM TRIS buffer, loading dye was added to 1x final concentration before denaturing the proteins at 95 °C. The DNA pelleting step could be omitted and the remaining steps were performed as described above.

All proteins in this work, from small scale expression to purified proteins were checked for expression and the correct size on an SDS-PAGE. Only in cases where it added particular information did I show these gels.

3.2.16 *In vitro NRP formation by purified enzymes*

After successful purification, BpsA and its TycC:BpsA fusion constructs were used to form their products *in vitro*. The basic components needed include the reaction buffer (TRIS, pH 7.5-7.9 or sodium phosphate buffer, pH 8), ATP (between 0.5 – 5 mM), $MgCl_2$ (between 2 – 20 mM), substrate (between 0.1 – 1 mM) and enzyme (between 0.1 – 3 µM). For the formation of indigoidine, a higher pH of around 8 is beneficial, possibly to aid the autocatalytic re-oxidation of $FMNH_2$ to FMN.

3.2.17 *Coupled online Pyrophosphate release assay and k_{cat} calculation*

The coupled online pyrophosphate release assay was performed to measure A domain activity of purified NRPSs under different conditions and with different substrates[70]. In short, successful substrate activation by the NRPS A domain leads to the release of one pyrophosphate PP_i, which is the first substrate in a cascade of enzymatic reactions ultimately leading to the oxidation of 2 β-NADH to 2 NAD^+. Since NADH absorbs strongly at 340 nm while NAD^+ does not, amino acid activation by NRPS A domains can indirectly be assessed by measuring the decrease of absorption at 340 nm in the online pyrophosphate assay.

Enzymes and substrates for the online PP_i release assay were purchased from Sigma-Aldrich followed by dilution or resuspension in 50 mM TRIS-HCl (**Table 7**). All amino acids were dissolved in water.

Table 7 | Components for the online pyrophosphate assay, their amount, volume of TRIS, pH 7.5, needed to prepare the stock solution, stock concentration and final concentration.

Component	**amount**	**V [ml]**	**C_{stock}**	**C_{final}**
D-Fructose 6-phosphate disodium salt hydrate	100 mg	1.1	300 mM	3 mM
Fructose-6-phosphate Kinase, Pyrophosphate-dependent from Propionibacterium freudenreichii (shermanii)	10 UN	0.5	20 U/ml	0.1 U/ml
Aldolase from rabbit muscle	100.29 U	1	100 U/ml	1 U/ml
Triosephosphate Isomerase from rabbit muscle	33 651 U/ml	3.7	2 500 U/ml	5 U/ml
α-Glycerophosphate Dehydrogenase from rabbit muscle	200 UN	0.29	500 U/ml	5 U/ml
ATP (pH ~6.5)	5 g		0.1 M	0.5 mM
$MgCl_2$			100 mM	2 mM
Amino acid(s): N, Q, V, O, K, L, Dipeptides: NQ, AQ, GQ, dansylQ (in100 % DMSO) Tripeptide: dFNQ			0.1 M	1 mM
ß-NADH (Gerbu Biotechnik)	1 g		0.1 mM	0.8 mM

The online PP_i release assay itself was performed in a total volume of 100 µl in 50 mM TRIS-HCl buffer, pH 7.5 at 30 °C in a 96 well plate. Per NRPS to be analyzed, a master mix was created by adding the buffer first, followed by other components except the NADH, amino acids and enzyme. The amino acids were dispensed in the individual wells of the 96 well plate in a volume of 5 µl (to be 1 mM in the total reaction volume). Finally, NADH and the enzyme of interest (to final concentrations between 0.5 µM and 3 µM) were added. The master mix was mixed by inverting the tube and 95 µl were dispensed in each well of the plate. After combining the master mix with the substrate, the reaction is started and the 96 well plate is placed in the plate reader, where the absorption at 340 nm is measured over time. Each enzyme-substrate combination and the respective negative control (without amino acid) were measured in triplicate. k_{cat} values were calculated like described by Kittilä *et al*[70] in Excel. The decrease in absorption at 340 nm was fitted in the linear range (ΔOD_{340} min^{-1}) and divided by the arithmetic product of the molar extinction coefficient of NADH[92] (ε_{340}(NADH) = 6 220 M^{-1} cm^{-1}), the path length d (100 µl per well in a flat bottom 96 well plate ~0.2 cm), the enzyme concentration c_{enzyme} (in µM) and the factor 2, due to two NADH being converted per PP_i release.

$$k_{cat} = \left(\frac{\Delta OD_{340}\ min^{-1}}{\varepsilon_{340}(NADH) * d * c_{enzyme} * 2} \right)$$

For each enzyme-substrate combination, the mean k_{cat} value of three measurements was calculated and background subtracted by the mean value of the negative control. The error was determined by subtracting the negative control standard deviation SD from the sample SD, including error propagation ($\text{error}\ (k_{cat}) = \sqrt{SD^2_{\text{neg. ctrl}} + SD^2_{\text{sample}}}$).

3.2.18 *Thiolation assay using radiolabelled amino acids*

(Experiments were performed by Florian Meyerthaler, Mootz lab, Institute for Biochemistry, WWU Münster)

Thioester formation of radiolabelled leucine was investigated according to a modified protocol7. Enzymes were mixed at 0.5 µM in a total volume of 100 µL TRIS HCl (pH 7.5) together with 10 mM MgCl2, 2 mM TCEP, 2 mM L-Val and/or L-Orn, 10 µM L-Leu (980 pmol of non-labelled and 20 pmol of [3H]-L-Leu (Hartmann Analytics, Braunschweig, Germany)). The reaction was started by adding 5 mM ATP and quenched after 1 min with ice-cold 800 µL TCA (10 % trichloroacetic acid solution). Proteins were coprecipitated with 15 µL ice-cold BSA (25 mg-1ml-1 solution) and pelleted for 30 min at 13 000 rpm. The pellets were washed twice with 800 µL ice-cold TCA and resuspended in FA (10 % formic acid solution). After addition to 3 mL scintillator liquid, the amount of radiolabelled leucine covalently attached to the PPant arm of the respective enzyme was measured in a scintillation counter (Beckman Coulter LS 6500 Liquid Scintillation Counter).

3.2.19 *Indigoidine extraction from* **E. coli**

Indigoidine extraction from an *E. coli* overnight culture co-expressing an indigoidine synthetase, IndC or BpsA, and a PPtase (Sfp) was performed according to a modified protocol described by Yu *et al*[307]. The fermentation broth was centrifuged for 30 minutes at 4 000 g resulting in a pellet consisting of indigoidine and bacterial cells. The pellet was lyophilized in vacuo. The dried pellet was resuspended in pure DMSO causing the indigoidine to dissolve, in contrast to the cell debris. In another round of centrifugation for 30 minutes at 4 000 g, the solution was cleared of the

non-dissolvable components and the indigoidine containing supernatant was passed through a 0.45 µM Durapore® PVDF membrane spin filter (Merck).

3.2.20 *Identification of indigoidine via MS and NMR spectroscopy*

Indigoidine, dissolved in 100 % DMSO either extracted from bacterial expression culture or pelleted after an *in vitro* reaction, was diluted with 100 % methanol in a 1:1 ratio before applying the sample on a micrOTOF-Q II mass spectrometer. Mass to charge ratios (m/z) were scanned between 50 to 1 200 in positive or negative ion mode. The nebulizer was set to 0.4 bar, the dry heater at 200 °C and the dry gas at 4 l/min.

For ^{1}H NMR measurement, the pelleted indigoidine was dissolved in deuterated DMSO-[D6] (VWR chemicals, Radnor, Pennsylvania, USA) and subjected to a frequency of 300 or 500 MHz at 25 °C by Tobias Timmermann at the IPMB. The resulting spectrum was analyzed in the NMR Processor Academic Edition 12.0 by ACD/Labs.

3.2.21 *Identification of substrate intermediate on PPant arm of NRPS mutants*

In order to elucidate the mechanistic details of indigoidine formation by the respective synthetase or one of its fusion constructs with TycC, we sought to identify intermediate substrates stuck on the PPant arm or the active site Serine of the TE domain via mass spectrometry (Thomas Ruppert, ZMBH, Core facility for Mass Spectrometry and Proteomics). We followed a modified protocol described by Henderson *et al*[122]. In detail, we purified the protein of interest as described in section 3.2.13 followed by a buffer exchange of 1 µM protein into 100 µl of 50 mM TEAB and 6 M Urea, pH 7.3. At 25 °C, the protein was first digested with 1 % Lys-C for 4 h. The solution was then diluted 1:5 in 50 mM TEAB (pH 7.3) and digested with 1 % Trypsin at 25 °C overnight. A last digestion step using 1 % Glu-C was performed after another 1:5 dilution with 50 mM TEAB (pH 7.3) for 4 h at 25 °C. The resulting peptides were subjected to an LC-MS and peptide identification was performed by Thomas Ruppert.

3.2.22 *BpsA and TycC5-6:BpsA product formation followed by liquid-chromatography coupled mass spectrometry (LC-MS)*

(Experiments were performed with Aubry Miller, Cancer Drug Development Group Leader, DKFZ, Heidelberg)

The formation of the respective products was analyzed by LC-MS. Enzymes (3 µM) were mixed in sodium phosphate buffer (pH 8, final volume of 50 µL) with 20 mM $MgCl_2$, 1 mM L-Asn and 1 mM L-Gln and incubated at 25 °C. With the addition of 5 mM ATP the reaction was started and after 16 h stopped by removing the enzyme in a 10 kDa MWCO spin filter. The samples were pelleted (14 000 rpm, 10 min). For enzymes producing a blue pigment, the supernatant was discarded and the pigment was redissolved in 100 % DMSO to load on the LC-MS. For enzymes expected to form a peptide, the supernatant was injected onto the LC-MS. 10 µL were subjected to the LC-ES-API-MS (Agilent 6100 series, Single Quadrupole, electronspray atmospheric pressure ionization). The MS was used in negative mode together with a Kinetex 2.6 µm C18 column (30 °C). For indigoidine, the gradient ran from 0 to 6 min from 99 % A and 1 % B to 10 % A and 90 % B at a flow rate of 0.6 min^{-1}. For the expected Asn-Gln dipeptide, the solvent composition ran for 1 min at 99 % A and 1 % B followed by a wash of 100 % B within 3 min at a flow rate of 0.6 min^{-1}. Solvent A contained H_2O with 0.01 % FA (formic acid) and B 100 % acetonitrile with 0.01 % FA. Extracted ion traces of the small molecule masses were recorded.

3.2.23 *TycC 8-, 9-, 10 product formation followed by liquid-chromatography coupled mass spectrometry*

(Experiments were performed by Florian Meyerthaler, Mootz lab, Institute for Biochemistry, WWU Münster)

The formation of the respective peptide products was analyzed by LC/MS. Enzymes (3 µM) were mixed in TRIS-HCl buffer (pH 7.5, final volume of 20 µL) with 10 mM $MgCl_2$, 2 mM TCEP, 100 µM L-Val, 100 µM L-Orn and 100 µM L-Leu and incubated at 37 °C. With the addition of 5 mM ATP the reaction was started and after 1 h quenched with 10 µL of formic acid. The samples were pelleted (13 000 rpm, 5 min), the supernatant mixed with 20 µL acetonitrile and 10 µL were subjected to the LC-ESI-MS (Agilent 6130B Single Quadrupole). The MS was used in positive mode together with a NUCLEODUR® HILIC column (27 °C, gradient from 0 to 3 min of 95% B, 3 to 23 min of 95% to 50% B, from 23 to 25 min of 50% B and from 25 to 28 min of 50% to 95% B). Solvent A contained H2O, 0.15% FA and 10 mM ammonium formate and B H20, 95% acetonitrile, 0.15% FA and 10 mM ammonium formate. Extracted ion traces of the peptide masses were recorded.

4. Results

In this chapter I will show the experimental results I have obtained. I will start with the investigation of two homologous indigoidine synthetases to shed more light on how they produce the blue pigment with special emphasis on how modifications influence this process *in vivo* and *in vitro*. I will focus on how I used the pigment synthetase(s) to tag a module extracted from the tyrocidine synthetase and fine-tune this system to link upstream NRPS modifications to the level of pigment production in *E. coli* and pyrophosphate release *in vitro*. In the third part of the results, I will present further NRP-indigoidine fusion synthetase engineering approaches, also using modules of other NRPSs to be tagged and testing a dodecylindigoidine synthetase for its use as a tag. Finally, I will present results obtained by studying constructs of increasing size and complexity consisting of one, two or three modules extracted from the tyrocidine synthetase to which point mutations have been applied on A and PCP domains.

4.1 Analysis of two homologous indigoidine synthetases

The two homologous indigoidine synthetases I investigated are IndC from *Photorhabdus luminescens* and BpsA from *Streptomyces lavendulae* ATCC 11924. They share 49.1% sequence identity and 64.5% similarity on the protein level. Yet, their domain composition is the same (**Figure 4A**). One major difference is that IndC features 23 cysteines, while BpsA only has 6 cysteines in total. I amplified the *indC* gene directly from the genomic DNA of *Photorhabdus luminescens* and inserted it into the bacterial expression vectors pCDFDuet and pTrc99a. The DNA sequence of BpsA was provided as a codon optimized version for expression in bacterial cells (with 25.6% base pairs changed in comparison to the genomic DNA) by the Fussenegger lab. I changed the backbone for BpsA expression to pTrc99a.

4.1.1 *IndC and BpsA produce the blue pigment indigoidine in* E. coli *and* in vitro

The expression of IndC in BAP1 cells, which have the promiscuous PPtase *Sfp* stably integrated in their genome, leads to the formation of the functional holo-enzyme that produces the blue pigment indigoidine (**Figure 7B, C**). I verified that the pigment is indigoidine by measuring its mass via ESI-MS with the help of Heiko Rudy (IPMB, Heidelberg) (**Figure S 1**) and its composition via ^{1}H NMR with the help of Tobias Timmermann (IPMB, Heidelberg) (**Figure S 2**). Both were as expected for indigoidine[202].

For *in vitro* analysis of the indigoidine synthetases, I optimized the purification procedure based on existing protocols[108,202]. In contrast to BpsA, IndC has not been purified before. Since it harbors almost four times as many cysteine residues as BpsA, the addition of 5 mM dithiothreitol (DTT) as a reducing agent was necessary to prevent intramolecular disulfide bond formation leading to a locked conformation of the enzyme and thus inactivation. I purified both indigoidine synthetases in their holo-version from BAP1 cells. Both formed the blue pigment *in vitro* upon addition of the substrate glutamine, ATP and $MgCl_2$ (**Figure 7D**). The correct composition of indigoidine was verified via ^{1}H NMR (**Figure S 3**). Since BpsA has been more frequently used for *in vitro* studies, does not require reducing agent for purification and was reported to withstand storage at -20 °C[215], this homologue was used for further *in vitro* studies. For analysis of pigment production *in vivo*, IndC is equally suited, since the cytoplasm of *E. coli* naturally is a reducing environment[308,309] and is thus able to keep IndC in the active state.

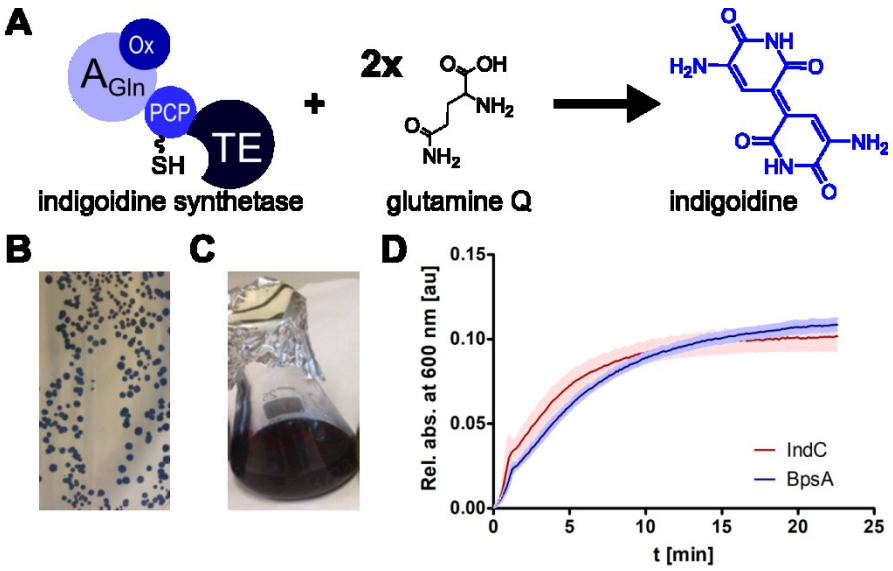

Figure 7 | Indigoidine produced in *E. coli* and *in vitro*. (A) Indigoidine synthetases convert two glutamine molecules into the blue pigment indigoidine. **(B,C)** Expression of IndC in *E. coli* BAP1 cells leads to blue pigment production on LB-agar plates and in liquid culture, respectively. **(D)** Indigoidine production followed *in vitro* via the increase in absorption at 600 nm over time. Enzymes were added at 1 µM, Q at 100 µM in 50 mM TRIS, pH 8, with 20 mM $MgCl_2$ and 5 mM ATP. For the purification of IndC, 5 mM DTT was added to each buffer.

4.1.2 *Effects of mutations to the indigoidine synthetase can be evaluated by pigment production levels in* E. coli

With the aim in mind to use the indigoidine synthetase as a C-terminal tag for an NRPS to monitor the outcome of its re-engineering, I examined whether I could monitor the effects of mutations to the pigment synthetase itself via indigoidine relative quantification upon expression of the mutated enzyme in BAP1 cells.

In order to evoke a change in pigment production and, at the same time, gain insights into the indigoidine production mechanism, I chose to mutate key residues of IndC and BpsA. Some of the residues are crucial to all NRPSs; others are conserved only among indigoidine synthetases and are thus interesting for my purposes (**Table 8**). I introduced all mutations via an established protocol[294] that I adapted using overlapping forward and reverse primers that carry the desired mutation. I expressed the mutated constructs in BAP1 cells overnight and quantified pigment production relative to a negative control and normalized to the wild type level (**Figure 8**).

All mutations within the A domain core motifs completely prevented pigment production, which is what I expected given the crucial roles of the mutated residues (**Table 1**). The mutation of the conserved serine of the PCP domain, which cannot be PPantylated anymore, also lead to a complete abrogation of indigoidine formation in both homologues. For the oxidation domain, some of the point mutants did not lead to color formation (N571S in IndC), while others did not have a significant impact (Y603A/F in BpsA) despite being part of the conserved Ox1 sequence. Contrary to what was expected, the mutation of the conserved lysine from the Ox1 motif did not completely abrogate indigoidine formation in IndC (K605E), even less so when the positively charged lysine (K) was mutated to the neutral alanine (A) instead of to the negatively charged

glutamic acid (E). Only in combination with another point mutation within the same motif (Y615A) did it inhibit indigoidine production in IndC. In BpsA, however, K600E did have the destructive effect as previously reported[202].

Table 8 | Overview of the mutations introduced into IndC and BpsA. Amino acids (AA) are represented by their one letter code. Mutations are described in the format $AA^{wt}positionAA^{mut}$.

domain	consensus sequence	Mutation in IndC	Mutation in BpsA
A	A3: LAYxxY**TS**G(S/T)TGxPKG	T179A/S180A	
A	A7: Y(R/K)TG**D**L	D414A	D410L
Ox	conserved[222]	N571S	
Ox	Ox1: **K**YxYxSxGxxY(P/G)VQ	K605E/A	K600E
Ox	Ox1: **K**YxYxSxGxx**Y**(P/G)VQ	K6005-Y615	
Ox	Ox1: KYx**Y**xSxGxxY(P/G)VQ		Y603A/F
A	A10: NG**K**(V/L)DR	K922A	
PCP	[I/L]GG[D/H]**S**L	S980A	S974A
TE	GX**S**XG	S1108A/C	S1103A/C
TE	conserved	A1135D A1135D/P1136A/G1137E	A1131D/G

In the TE domain, mutation of the conserved serine to either alanine or cysteine again eliminates pigment production. This observation was unexpected since this serine is not involved in the suggested synthesis mechanism of indigoidine (**Figure 4C**), despite being part of the catalytic triad found in many NRPSs. In addition, it does not serve a structural role, so I deduce that in fact it might be involved in the synthesis, possibly even in the tranditional role of transiently storing the intermediate substrate before its release (as described in section 0). The alanine which was mutated within the TE domain is conserved among all indigoidine synthetases. Other NRPSs and the putative dIndS feature an aspartate in this position, as part of the catalytic triad S-H-D. In IndC, mutating this alanine to aspartate abrogated indigoidine formation. Additionally, replacing the following two residues P and G by A and E to match this part of dIndS had the same effect (**Figure 8**). In BpsA, the same A1131 to D mutation decreased indigoidine production by 81% but did not prevent it completely. When I replaced the neutral alanine by the neutral glycine instead of the negatively charged aspartate, the decrease in pigment formation compared to the wt was even less (**Figure 8**). This indicates that a small, neutral amino acid might be required in this position for indigoidine production, while the residue found most often in this position, aspartate, partially or even completely prevented product formation. For BpsA, an SDS-PAGE of an overnight expression sample of the wt and all mutants is shown in **Figure S 4**. Note that the wild type BpsA and all mutants that still produced close-to-wt indigoidine levels exhibited a lower enzyme expression level. This is possibly due to a mild antimicrobial effect of indigoidine[209] leading to slower cell growth of the expressing cells and thus to a lower enzyme level.

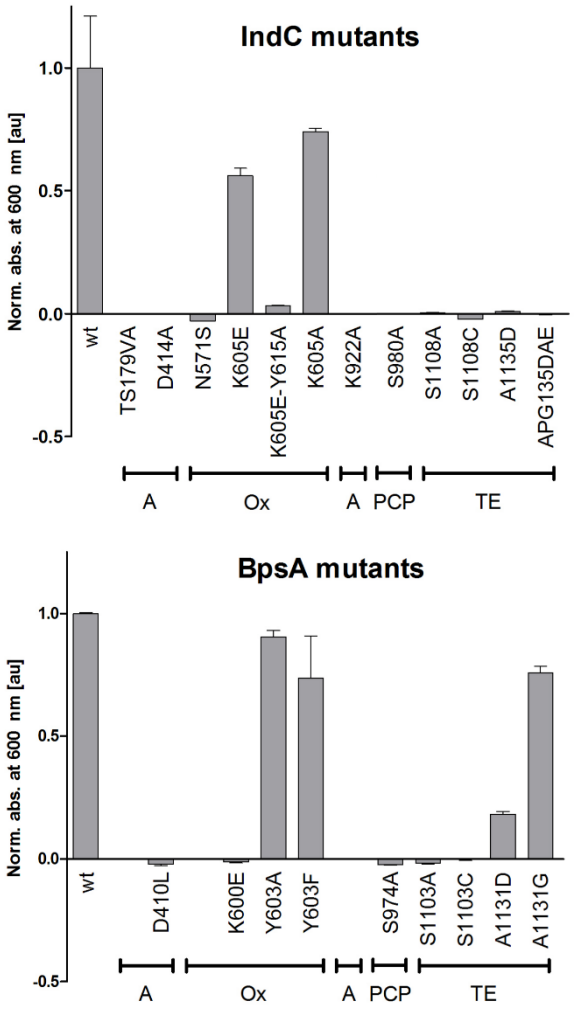

Figure 8 | Effects of point mutations in IndC and BpsA on pigment production. Bar graphs show the background-subtracted absorption at 600 nm of the indicated mutants after overnight expression in BAP1 cells normalized to the wild type synthetase. n = 3 +SD.

The slightly negative values of the absorption at 600 nm can be attributed to the formula I used to quantify the pigment[306] (section 3.2.11) where the ratio δ between OD_{600} and OD_{800} is calculated for the negative control (here, two empty plasmids), multiplied with the OD_{800} of the sample of interest and subtracted from OD_{600}, where the potentially produced pigment would absorb. Thus, the negative control, which does not express any proteins, can proliferate more quickly than the samples expressing the large enzymes. The overall absorption of the negative control as well as the ratio δ are higher than for un-pigmented sample cells of a lower density and thus the overall calculated absorption at the pigment maximum can be below zero.

In conclusion, it is possible to investigate the effect of point mutations on indigoidine synthetases through quantification of the pigment produced by the respective variant *in vivo.* Analysis of corresponding mutations to IndC and BpsA revealed that even two homologue NRPSs do not necessarily react in the same way to the same perturbation.

4.1.3 *Analysis of BpsA mutants in vitro reveals which mutants are still capable of continuous substrate activation*

Having observed the differences in pigment production of two indigoidine synthetases *in vivo* based on absorption measurements, I further investigated BpsA and its mutants *in vitro.* k_{cat} values for glutamine activation were calculated from the online PP_i release assay and summarized in **Figure 9**.

In **Figure 9A**, the averaged raw data of the absorption of 1 µM BpsA wt converting 0 (grey), 0.1 (red), 0.5 (green), or 1 mM glutamine (blue) from triplicate measurements at 340 and 600 nm is shown. The absorption at 340 nm (circles) illustrates the decrease in NADH, which is coupled to the release of PP_i via an enzymatic cascade (**Figure 3E**). This measurement serves as the basis to fit the steady PP_i release per minute in the linear range, subtract the background signal (Q = 0, grey) from the samples and then calculate the k_{cat} values (**Figure 9B**). Absorption at 600 nm (triangles) illustrates the pigment formation, which can be measured in parallel (**Figure 9A**). From the molecular extinction coefficient of indigoidine (though given in DMF) I calculated the production rate, which was 6.05 indigoidine min^{-1} for 1 mM glutamine converted by 1 µM BpsA. Thus, according to product formation, 12.10 glutamine molecules were activated per minute, which is about three (2.97) times less than the activation rate according to the PP_i release assay (35.98 min^{-1}).

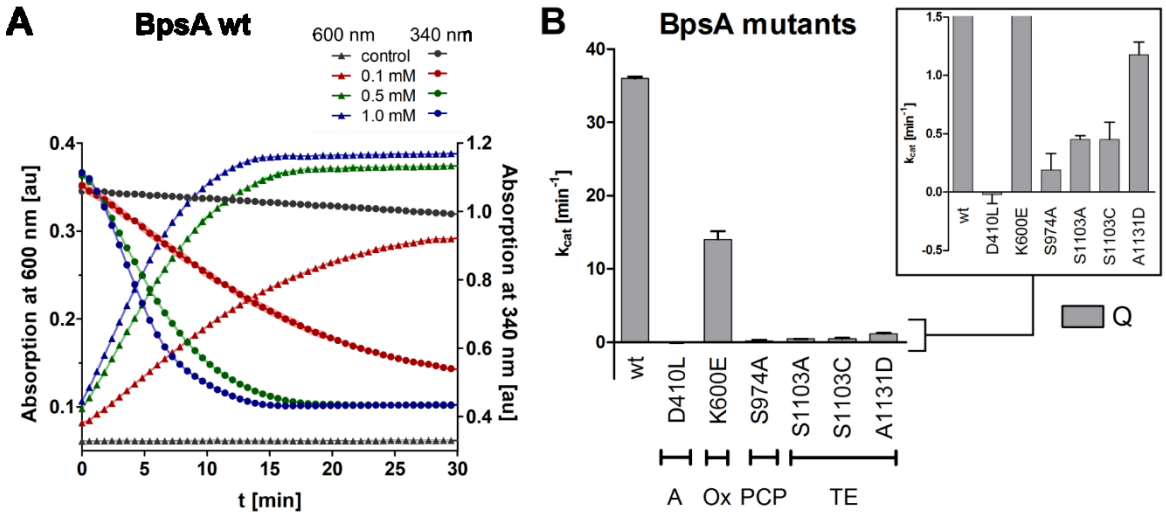

Figure 9 | A domain activity of BpsA wt and mutants in presence of glutamine. (A) Absorption at 340 nm (circles) and 600 nm (triangles) of 0 mM (grey), 0.1 mM (red), 0.5 mM (green) and 1 mM (blue) glutamine and 1 µM BpsA wt over time. Curves represent the mean of three replicates in 100 µl, ±SD in matching transparent color. The absorption at 340 nm was taken as the basis to calculate the k_{cat} for BpsA wt. **(B)** The bar graph shows the k_{cat} values of the indicated BpsA variants after background (no Q) subtraction calculated from the online PP_i release assay. Enzymes were added at 1 µM, glutamine (Q) at 1 mM. The inlay shows a zoom into the lower value bars. n = 3 +SD.

As forecasted from the *in vivo* absorption measurement (**Figure 8**), BpsA wt had the highest activation rate when fed with the substrate glutamine, which amounted to 35.98 min^{-1}. The mutation of the ATP binding residue in the A7 motif (D410 to leucine (L)) indeed prevented ATP

binding and set the k_{cat} value to zero (**Figure 9B**, inlay). The oxidation domain mutant, which *in vivo* showed no pigment formation, is *in vitro* still able to activate glutamine at the rate of 14.02 min^{-1}, corresponding to about 38% of the wt. This finding suggests that activation comes before oxidation during indigoidine production. If thiolation also occurs before the oxidation step would need to be assessed in a separate thiolation assay. PCP and TE domain mutations led to a drastic decrease in substrate activation. For other NRPSs it was shown that they can also activate their substrates without the presence of the PCP domain, but for BpsA, this domain seems to have an influence also on the substrate activation rate. The same is true for the TE domain. While it is not assigned any responsibility in the proposed indigoidine production process, a mutation of the conserved serine in the TE domain prohibited pigment formation *in vivo* and repeated substrate activation *in vitro*, highlighting it as essential. The A1131 to D mutation in the TE domain did not completely abrogate pigment formation *in vivo* (**Figure 8**), which is supported by the *in vitro* k_{cat} value of 1.17 min^{-1} and a retained but strongly affected product formation ability leading to a maximum absorption at 600 nm of 0.012±0.001 au in 100 µl (not shown) in comparison to 0.281±0.002 au for wt BpsA (**Figure 9A**, 1 mM, background subtracted).

Overall, the online PP_i release assay can provide additional information about the ability of modified NRPSs to repeatedly activate their substrate. Thus, for engineered or mutated indigoidine synthetases this assay adds to the understanding of the functionality gained from the visible pigment production.

4.1.4 *Activation of dipeptide substrates by BpsA and its mutants*

Since I want to create an NRP-indigoidine fusion synthetase possibly producing indigoidine-tagged amino acid(s), I explored the option of creating such a molecule by feeding a di- or tripeptide with glutamine (Q) at its C-terminus (**Figure 10A**) to BpsA.

In comparison to the native substrate Q, dipeptides glycine-glutamine (GQ), alanine-glutamine (AQ) and the fluorescently labelled dansyl-glutamine (dansylQ) were activated by wt BpsA *in vitro* at significantly lower rates (**Figure 10B**). For GQ and AQ, even a slight increase in absorption at 600 nm was detected indicating that a pigment was formed. The amount of pigment was too low to be purifed for further analysis. Increasing the incubation time and the reaction volume did not improve the yield. Thus, I was unable to identify the pigment that was produced by BpsA from the dipeptide substrates. In an early attempt to produce indigoidine-tagged amino acids, NQ was fed to IndC and incubated for 72 hours at 25 °C. The blue pigment produced was identified via MS as pure indigoidine (m/z $[M-H]^-$ 247.0) without asparagine attached. In the supernatant, the substrate dipeptide NQ (m/z $[M-H]^-$ 259.0) as well as the individual fragments, Q (m/z $[M-H]^-$ 145.1) and N (m/z $[M-H]^-$ 131.0), were detected. Thus, I could not yet clarify whether the substrate dipeptide NQ decomposed on its own over time and IndC converted the Q fragment to indigoidine, or the NQ fragmentation was enzymatically induced by IndC, actively splitting N from Q to produce pure indigoidine. BpsA did not activate the dipeptide NQ (or only to a very low degree within the large standard deviation) or the tripeptide dFNQ (D-phenylalanine-NQ, **Figure 10B**), again uncovering a difference between the two indigoidine synthetases.

Dipeptides AQ and dansylQ were then presented to one BpsA mutant of the Ox domain (K600E) and two of the TE domain (S1103A and A1131D) and their activation rate was assessed from the

PP_i release assay (**Figure 10C**) and compared to the activation of Q. The TE domain mutants activated the dipeptides at similar rates as the native substrate Q. The Ox domain mutant activity towards the dipeptides however was greatly reduced, from 12.9 min^{-1} to almost zero for AQ (**Figure 10C**). This indicates that the Ox domain might be involved in modifying a XQ (where X stands for any amino acid) dipeptide to be accepted by BpsA, e.g. through detachment of the upstream amino acid from Q.

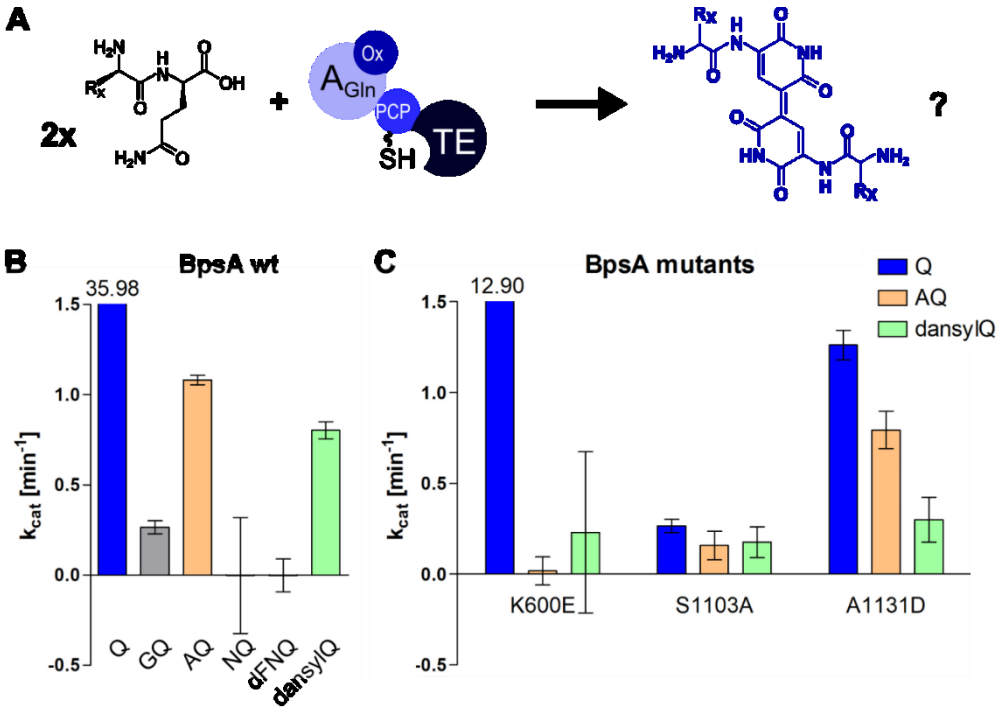

Figure 10 | Activation of different dipeptides by BpsA wt and mutants. (A) Scheme of an example dipeptide XQ serving as a substrate for BpsA to produce indigoidine-tagged X, where X stands for any amino acid. **(B)** Bar graphs show the activation rate of BpsA wt and **(C)** BpsA mutants towards the indicated di- and tripeptides in comparison to the native substrate Q (Figure 9B) as calculated from the online PP_i release assay. Substrates were added at 1 mM, BpsA and its mutants at 1 µM. n = 3 ±SD.

Taken together, these results suggest that indigoidine synthetases cannot form indigoidine-tagged amino acids from dipeptide substrates, because the N-terminal amino acid is split off either non-enzymatically or through an enzymatic process likely involving the Ox domain.

4.2 Development of an asparagine-indigoidine fusion synthetase to monitor the effects of upstream module manipulation with a visual readout

In the following part of the thesis, I will describe how I created a novel fusion enzyme between the asparagine incorporating module of TycC and an indigoidine synthetase (IndC or BpsA). Further, I evaluated and improved the fusion enzyme(s) to find the best version to investigate the effects of upstream module modifications, such as point mutations or different N-terminal start sites, on the downstream pigment synthesis through absorption measurements. In order to be able to link the upstream module modifications to pigment production levels, there is one major prerequisite that needs to be fulfilled: the pigment needs to be produced exclusively if the upstream module is functional. Ultimately, with these experiments I intend to test if an indigoidine synthetase can be genetically fused to any multimodular engineered or native NRPS to monitor the functionality of the upstream NRPS modules via pigment formation.

In the following sections I will describe the different fusion sites and other engineering aspects I followed while creating the asparagine-indigoidine fusion synthetase. Therefore, I will give a quick overview of how I annotated these fusion constructs, because there is no general annotation agreed upon in the field of NRPS research yet. Residues are represented by the single letter amino acid code. I numbered the AAs in the unimodular enzymes, namely IndC, BpsA and dIndS, starting from the first methionine (M1). For the tyrocidine synthetase and other larger NRPSs, the numbering starts at the position where the beginning of the domain is predicted to be (by Pfam[310]) and is, therefore, domain-dependent. Residues before or after the domain border are annotated by "+" for downstream or "-" for upstream. The name of the NRPS that is used in a fusion construct is written as normal text. The individual domains of the NRPS, which are part of the fusion enzyme, are given in parenthesis and connected with hyphens. The last residue of a domain, the so-called fusion site, is separated by an underscore "_". Protein fusion sites are represented by a colon ":". The asparagine-indigoidine fusion synthetases without any fusion site specifications will be called TycC(A_{Asn}-PCP-C-A_{Gln}):IndC/BpsA or TycC5-6:IndC/BpsA.

4.2.1 *Expression of different TycC5-6:IndC fusions leads to fusion-site-dependent pigment production*

In the search of a good starting point for a prototype NRP-indigoidine fusion synthetase, I turned to the well-studied tyrocidine synthetase (**Figure 1**). I tried to circumvent the complications arising when two modules, which do not naturally belong together, are fused, e.g. the additional proofreading activity of the C domain. Thus, I decided to use the modules TycC5, specific for asparagine (N), and TycC6, specific for glutamine (Q) (**Figure 1**). The indigoidine synthetase does not feature an N-terminal C domain. For this reason, in the fusion synthetase, the C domain is contributed by TycC6, naturally condensing N and Q and thus possibly able to cooperate with the indigoidine synthetase, also processing glutamine.

For an initial screen, I generated ten different TycC5-6:IndC constructs fused together at different residues within the A_{Gln} of TycC6 and IndC, where the N-terminal part is still contributed by TycC6 and the C-terminal part by the indigoidine synthetase (**Figure 11A**). I amplified the desired gene fragments from the genomic DNA of *Brevibacillus Brevis* and *Photorhabdus luminescens* and assembled them without scars using Gibson Cloning[311] into the

pCDFDuet(MCS2) expression vector. As an N-terminal start site for all constructs, I chose the "Q-167", 167 AA upstream of the TycC5 A_{Asn} domain. This start site was originally suggested by the Heidelberg iGem team in 2013[312]. For most fusion constructs, I generated the PCP domain mutant of module TycC5 (PCP_{Asn} S28A), mutating the conserved serine to an alanine as described before for IndC and BpsA (**Table 8**). This mutant serves as a control, to try to ensure from the beginning that I select for fusions, which only produce a pigment if the upstream module is working. I expressed the different fusion constructs and their respective PCP_{Asn} mutants in *E. coli* BAP1 cells overnight, calculated the background (= IndC PCP mutant)-subtracted absorption of each variant at 600 nm according to Myers *et al.*[306] and normalized all values to IndC wt pigment levels (**Figure 11B**).

In general, the more of IndC is left intact within the TycC5-6:IndC fusion, i.e. the less of the A_{Gln} domain of TycC6 is included, the more pigment is produced. For the iGem fusion, where IndC is left complete and added directly after the TycC6 condensation domain, the pigment production level is just as high as for wt IndC. While pigment production of the NRP-indigoidine synthetase is necessary to use it as readout for NRPS engineering, it also needs to be ensured that no pigment is produced, in case the upstream module does not work. However, for the first three fusion sites, Asn module inactivation through the mutation of the PCP domain does not result in a drastic decrease of pigment production. This leads me to the conclusion that if the indigoidine synthetase, which is naturally a standalone NRPS, is left intact, it can produce indigoidine much faster than any potential asparagine-indigoidine fusion, which is why the upstream PCP domain mutation does not have much of an effect on pigment production. I confirmed this assumption by purifying and analyzing the produced pigment of the TycC(A_{Gln}_T-161):(M1) and TycC(A_{Gln}_V-37):(N3)IndC fusion enzymes via ESI-MS. Instead of indigoidine-tagged asparagine (calculated exact mass of $C_{18}H_{20}N_8O_8$: 476.14), I identified pure indigoidine (calculated exact mass for $C_{10}H_8N_4O_4$: 248.05, found in ESI^-: $[M-H]^-$ 247.0) for both.

Fusion sites closer to the middle of the two A_{Gln} domains, which were partially inspired by the successful A subdomain swap by Kries *et al.*[261] (section 1.3.6), did not produce any pigment when used as a TycC5-6:IndC fusion site. Thus, I also did not create the corresponding PCP_{Asn} mutants for comparison. Moving the fusion site further towards the A domain C-terminus, incorporating more of the TycC A_{Gln} and less of the IndC A_{Gln}, namely into the A7 core motif ("D372:L415") and upstream of the large A_{core} domain, between A7 and A8 ("R395:G438"), did lead to a pigment production level above the background (**Figure 11B**, inlay). These levels amounted to 2-3 % of those obtained by wild type IndC. However, for these two fusion constructs, the inactivation of the upstream Asn-incorporating module through PCP domain mutation resulted in a substantial decrease in pigment formation. This characteristic makes these two TycC5-6:IndC fusions the best candidates for the development of an artificial NRP-indigoidine synthetase suitable for high-throughput screening of upstream NRPS engineering approaches monitored via absorption at 600 nm. Fusion sites even further downstream, on the A_{sub} domain, post A8 core motif and within the A_{sub}-Ox domain linker did not produce any pigment suggesting that these fusion sites lead to inactive enzymes (**Figure 11B**, inlay).

Overall, I identified several TycC5-6:IndC fusion constructs which can produce a blue pigment, two of which have the desired characteristic of showing less pigment formation upon upstream PCP domain mutation.

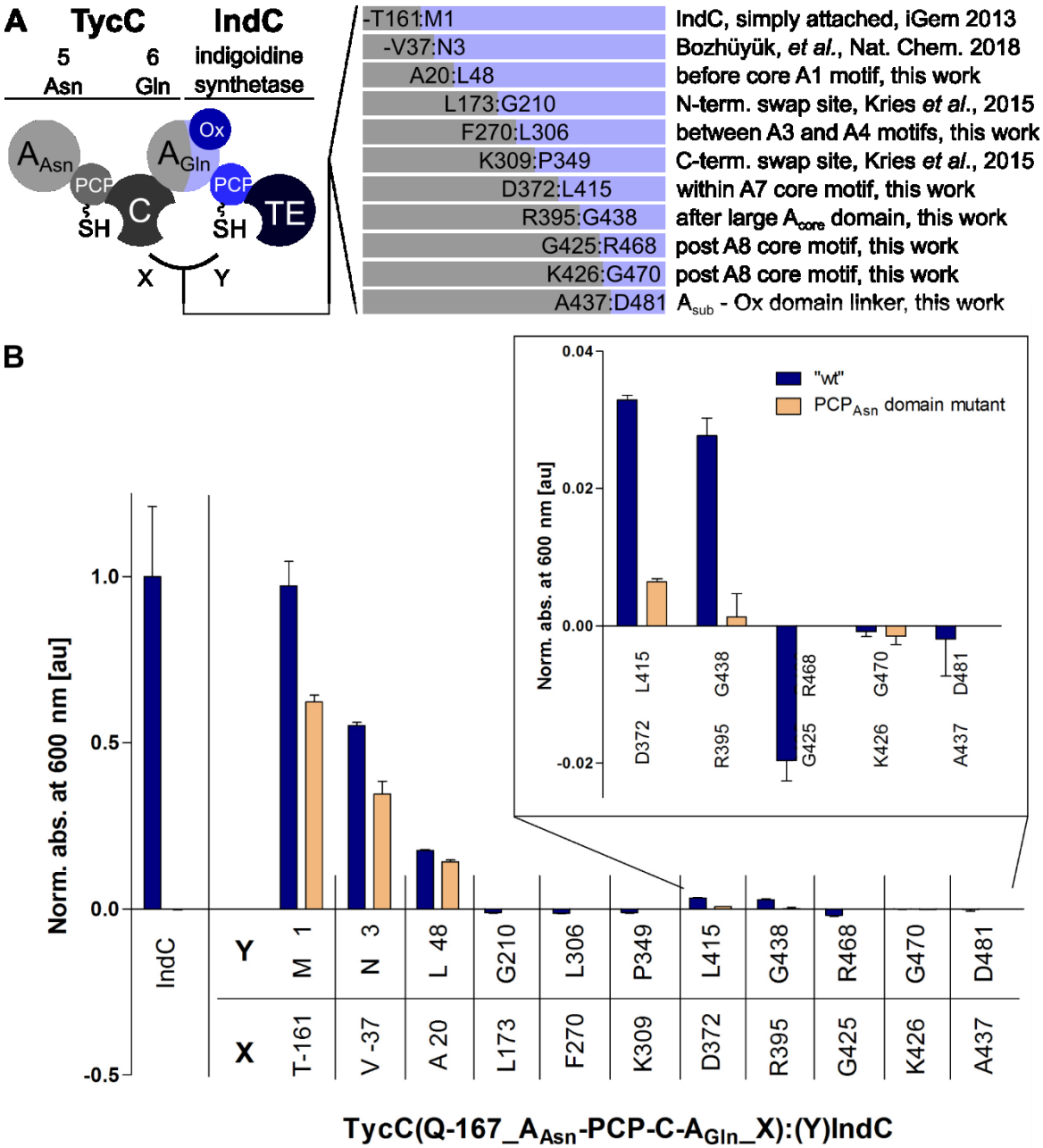

Figure 11 | Pigment production of TycC5-6:IndC constructs fused together at different sites within the A_{Gln} domain and their respective PCP_{Asn} mutants. (A) Scheme of a TycC5-6:IndC synthetase including the position of the fusion site within the A_{Gln} domain, where X indicates the C-terminal position of TycC A_{Gln} and Y the N-terminal position of IndC A_{Gln}, and explanations why they were chosen. **(B)** Absorption at 600 nm measured after overnight expression of the indicated fusion constructs and their PCP_{Asn} mutants from pCDFDuet(MCS2). The background was subtracted, and values are normalized to the absorption of IndC wt. The inlay shows the absorption of the last five fusion constructs enlarged. The start site of all constructs is the same (TycC(Q-167_A_{Asn}). n = 3 +SD.

4.2.2 *The N-terminal start site can have (minor) effects on pigment production*

Another critical point in NRPS engineering, next to the fusion site of two recombined modules, is the new artificial start site. Since there is no apparent difference in the sequence of initiation and elongation A domains, it is hard to estimate how to turn the second into the first[266]. I

addressed this question using the two most promising TycC5-6:IndC constructs described above (section 4.2.1, fusion sites "D372:L415" and "R395:G438", in pCDFDuet(MCS2)): I exchanged the original start site ((Q-167)TycC5) with other start sites and measured the absorption at 600 nm after overnight expression in BAP1 cells, relative to the negative control (IndC PCP mutant, S980A) (**Figure 12**).

Including the native upstream C domain (M1) and artificially adding the first 28 amino acids from the tyrocidine initiation module (TycA(M1-S28):(I-5_A_{Asn})TycC) increased pigment production about two fold for the "D372:L415" fusion site (**Figure 12A**), but only slightly for the "R395:G438" fusion construct (**Figure 12B**), in comparison to the "Q-167" start site, which is located directly downstream from the TycC5 C domain, including the complete linker region. Start site "I-30" mimics the length of the TycA N-terminus and reached a similar pigment production level for the "R395:G438" fusion (**Figure 12B**). I did not test it for the "D372:L415" fusion. Conversely, I tested start sites "Y-18" and "MLA:Y-18", where residues L and A were added upstream of "Y-18", both resembling the IndC N-terminus, solely for fusion site "D372:L415". Both do not differ much in pigment production level from the "Q-167" start. Neither does the fusion construct starting at "I-5", without the TycA part added (**Figure 12A**).

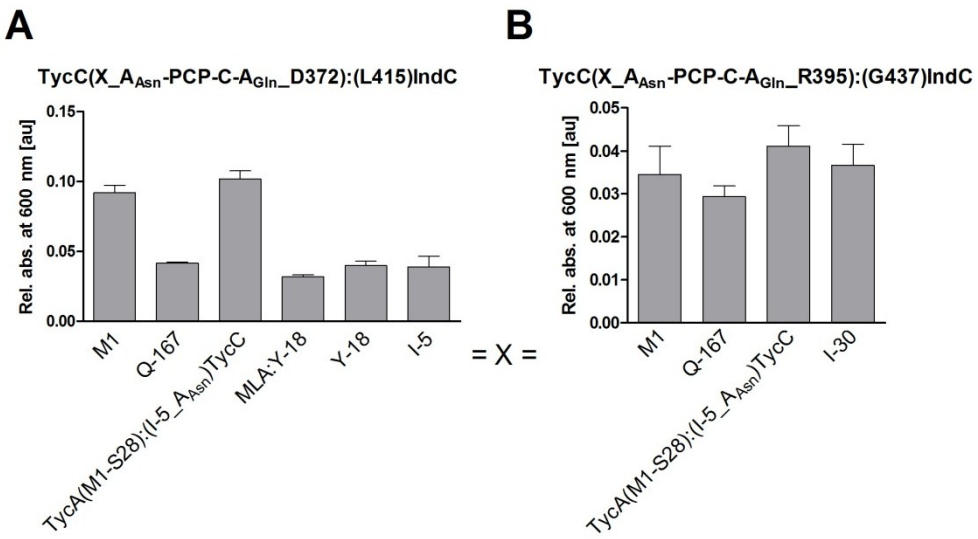

Figure 12 | Different N-terminal start sites can influence pigment production of TycC5-6:IndC fusion constructs in *E. coli*. TycC5-6:IndC fused together at **(A)** D372:L415 and **(B)** R395:G438, featuring different artificial N-terminal start sites, were expressed in E. coli BAP1 from pCDFDuet(MCS2). Bar graphs show their relative absorption at 600 nm in comparison to the background (empty vector). n = 3 +SD.

With this screen I showed that one can indeed evaluate the influence of the N-terminal start site of an upstream module on the downstream module pigment production of an artificial dimodular TycC5-6:IndC fusion construct. It seems that keeping a native start site, either by leaving the upstream C domain in place or by adding the sequence of the native initiation module TycA, can improve pigment production. However, this was only observed for one of the two fusion constructs questioning whether it is a general concept. What this screen cannot tell, is if the TycC5 A domain is indeed functional *in vivo* with the tested start sites, or wether it rather

provides a stable N-terminus for the indigoidine producting module, since I am lacking the proof of product formation or substrate activation. For this reason, I recreated a sole indigoidine synthetase by fusing only the TycC6 A_{Gln} part to IndC and testing different A domain start sites (section 4.2.3).

4.2.3 *A TycC6(A_{Gln}):IndC fusion can recreate an indigoidine synthetase*

For the reconstitution of an indigoidine synthetase consisting of the N-terminal part of the TycC module 6 A_{Gln} domain and the C-terminal part of IndC, I used the same internal fusion site that was successful for the TycC5-6:IndC hybrid enzyme: TycC6(A_{Gln}_D372):(L415)IndC. Different start sites for this fusion construct were then created and the chimeras were expressed in BAP1 cells overnight. Absorption at 600 nm for the different constructs were calculated relative to the background set by the IndC PCP domain mutant S980A (**Figure 13**). The hybrid TycC6:IndC did not yield much pigment when it started at "Q-167", which worked well for the TycC5-6:IndC fusion construct. Start sites "S-42", "A-7" and "V1", which is the first residue of the A_{Gln} domain according to Pfam, did not drastically increase pigment production (**Figure 13**). Adding the IndC N-terminus (M1-M27) or setting the start to the corresponding amino acid "L-19" of the native C-A_{Gln} linker did result in an increase of more than 0.1 au relative to the other start sites. Note that the TycC6:IndC hybrid indigoidine synthetase starting at "Q-167" does not produce much pigment, while shortening the N-terminus to start at "L-19" (corresponding to TycC5(A_{Asn}_Y-18), **Figure 12A**) leads to a remarkable increase in pigment production. Both observations are in contrast to the TycC5-6:IndC fusion construct and the effects on pigment production of the corresponding start sites (**Figure 12**), again indicating that discoveries about one NRPS module are not generalizable to others, not even within the same NRPS. Adding a native N-terminus, this time from IndC, also increases pigment production. To find out whether this N-terminus indeed serves a role in IndC, I removed the first 20 amino acids (start site "A21", corresponding to "A-7" of TycC A_{Gln}) or the first 26 amino acids (start site "M27", corresponding to "V1" of TycC A_{Gln}) of IndC and compared the pigment production levels to those of IndC wt (**Figure S 5**). Both truncations did not lead to a decrease in pigment production in comparison to the wt IndC (**Figure S 5**), indicating that this sequence does not play a crucial role, neither for IndC stability nor activity.

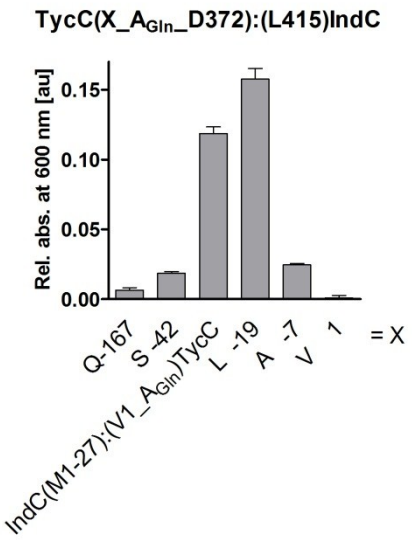

Figure 13 | Pigment production of a reconstituted indigoidine synthetase TycC6:IndC largely depends on the N-terminal start site. Absorption at 600 nm of an overnight expression culture of the indicated hybrid enzymes relative to the negative control (IndC PCP mutant, S980A) from pCDFDuet-MCS2. X indicates the tested start site. n = 3 +SD.

These TycC6:IndC hybrid enzymes are a reconstruction of a native indigoidine synthtase. Thus, it is clear that if the TycC A_{Gln} module could be successfully turned into an initiation module, this fusion enzyme produces the blue pigment. Therefore, all start sites of constructs for which indigoidine production was observed, are suitable to turn this natural elongation A domain into a starter A domain *in vivo*. In contrast, for the TycC5-6:IndC start sites (section 4.2.2), I could not draw the same conclusion, since from the *in vivo* pigment production it is not clear whether the TycC5 A_{Asn} domain successfully acted as an initiation A domain. We cannot support the finding of the Bode group that a natural elongation A domain needs part of or the whole native C domain upstream to function as an initiation A domain *in vivo*[266].

4.2.4 *Changing the expression backbone and indigoidine synthetase used in the fusion improves its properties*

The best candidate for a fusion enzyme so far is TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G438)IndC, since it is able to produce a detectable amount of pigment, which is abolished upon upstream PCP_{Asn} domain mutation. However, this construct also suffers from the very low pigment production of about 3% of the IndC wild type (**Figure 11B**, inlay). Thus, I aimed to improve these aspects of the most promising hybrid enzyme using various approaches, some of which will be presented later (section 4.3). Here I show the improvement obtained when changing expression backbone (from pCDFDuet(MCS2) to pTrc99) and indigoidine synthetase (from IndC to BpsA). TyccC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395) was fused to (G438)IndC or to (G434)BpsA and expressed from either pCDFDuet(MCS2) or pTrc99 in BAP1 cells. Absorption at 600 nm of the overnight expression culture was measured and calculated in relation to the negative control (**Figure 14**). The original TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G438)IndC expressed from pTrc99 produced more pigment than when expressed from pCDFDuet. However, pigment production of the PCP_{Asn} mutant also increased when expressed from pTrc99, though it is still only about half of the wt production level. Using BpsA in the hybrid enzyme resulted in an even higher pigment level, while at the same time decreasing the absorption at 600 nm for the PCP_{Asn} mutant. This makes TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA the perfect candidate for further investigations.

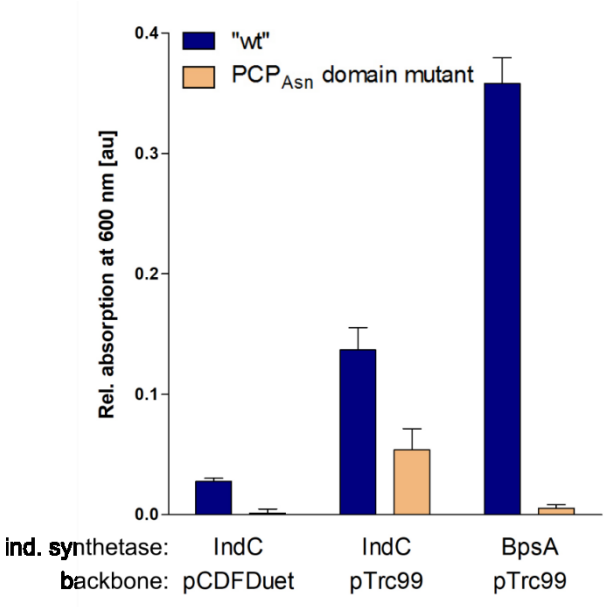

Figure 14 | Influence of the indigoidine synthetase homologue and the expression backbone on pigment production of TycC5-6:indigoidine synthetase fusion constructs and their PCP_{Asn} mutants. Bar graphs show the absorption at 600 nm of the indicated hybrid enzyme in the indicated vector relative to the negative control (PCP_{Gln} domain mutant of the respective indigoidine synthetase in the respective backbone). n = 3 +SD.

4.2.5 *Mutational analysis of TycC5-6:BpsA fusion reveals different effects of Asn-incorporating module inactivation on pigment production in vivo and in vitro*

I further investigated the most promising fusion construct for asparagine-indigoidine tagging, TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA, by analyzing the effects of different point mutations in the different parts of the chimeric enzyme on pigment production *in vivo* (**Figure 15A, B**).

To inactivate the Asn incorporating module 5 of TycC, I mutated the aspartate of the core motif A7 to alanine (A_{Asn} D370A) and the conserved serine of the PCP domain to alanine (PCP_{Asn} S28A), as described before (**Figure 14**). Surprisingly, the A_{Asn} domain mutant produced similar levels of pigment as the "wt", in contrast to the PCP domain mutant that produced very little indigoidine (**Figure 15B**). From the analysis of the corresponding A domain mutations in IndC and BpsA, I expect the A domain activity of this mutant to be abrogated; however this does not seem to influence the downstream pigment producing module. I confirmed this observation by mutating other residues within the core motifs A3, A7 and A10 of the TycC5-6:IndC hybrid enzyme (**Figure S 6**). To ensure this difference of TycC5 inactivation via A domain versus PCP domain mutation is not due to an unintended additional mutation, I sequenced the complete insert of the three variants and did not detect any difference in the sequence other than the intended point mutations. Further I checked the overnight expression cultures via SDS-PAGE for large differences in protein expression level or mutant solubility but did not detect any anomaly (**Figure 15C**). Consequently, I conclude that this distinct effect of TycC5 inactivation on downstream pigment production is indeed accounted for by the intrinsic behavior of the TycC5-6:BpsA/IndC fusion enzyme.

Mutation of the second histidine of the conserved C domain motif HHxxxDG to an alanine ($C_{Asn-Gln}$-H129A) leads only to a minor decrease in pigment production which might indicate that it is not crucial in the condensation reaction between Asn and Gln or that condensation reaction does not occur. Opposed to the corresponding mutation in TycB (section 1.2.9), this mutant is not insoluble, indicating that in TycC5-6 C domain, the second His of this conserved motif does not play a structural role. To inactivate the Gln-incorporating module via A domain mutation, the core motif A7 was chosen again which is located on the TycC6 part of the fusion construct. D372 to A mutation of the hybrid A_{Gln} leads to a complete abrogation of pigment formation revealing that the fusion between TycC6 and BpsA is functional and cooperates with the downstream PCP and TE domains towards product formation, which is inhibited upon A7 motif mutation. Almost all mutations of the BpsA part of the hybrid enzyme also resulted in loss of pigment production as expected from the analysis of the corresponding mutations in wild type BpsA (**Figure 8**). Only the TE-A1131D mutation behaved differently in that it also completely lacked pigment production, while in wild type BpsA it still produced indigoidine to a lower degree.

Overall, it seems that despite the observation that the N-terminus and PCP domain mutation of TycC5 do influence pigment formation of the TycC5-6:BpsA fusion construct, it is not clear whether this tagging approach works. Most data collected so far rather suggests, that the TycC6:BpsA part is able to produce indigoigine *per se* and without the input of the upstream TycC5 module.

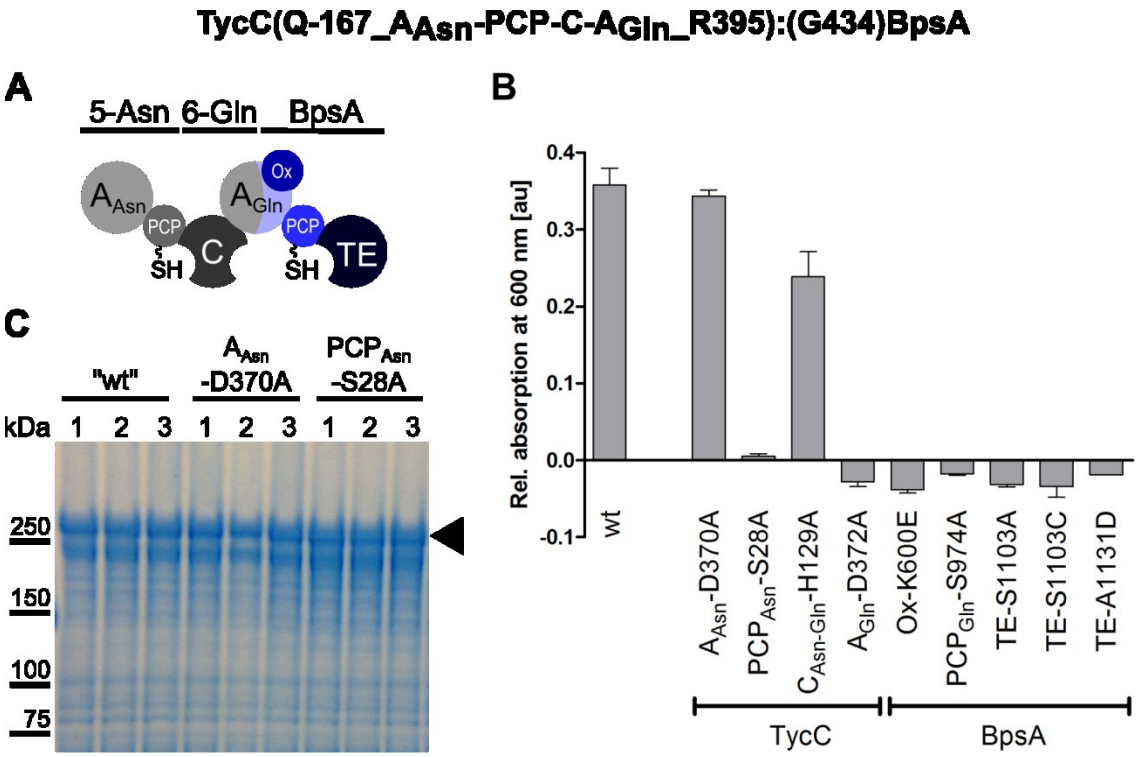

Figure 15 | Influence of point mutations on pigment production by a TycC5-6:BpsA fusion construct. (A) Scheme of the TycC5-6:BpsA hybrid enzyme to conceptually localize the mutations analyzed in (B). **(B)** The bar graph shows absorption at 600 nm of indicated mutants of the "wild type" TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA after overnight expression in BAP1 cells from the pTrc99 vector relative to the negative control (BpsA PCP mutant, S974A). n = 3 +SD. **(C)** Overnight expression of BpsA "wt", A_{Asn}-D370A and PCP_{Asn}-S28A mutants (in triplicate) visualized via SDS-PAGE run on a 3-8% TRIS-acetate gel stained with Coomassie. The position of the hybrid enzymes is indicated by the arrowhead. They run at an expected size of above 250 kDa (calculated size is 274 kDa).

In vitro analysis of TycC5-6:BpsA mutants shows a discrepancy between in vitro and in vivo behavior

To analyse the TycC5-6:BpsA wild type and mutant fusion proteins *in vitro*, I expressed and purified them from *E. coli* BAP1 cells, which should result in the holo-version of the enzymes with the PPant arm attached to the PCP domain due to the co-expressed PPtase Sfp. I then used them in an online PP_i release assay and calculated the k_{cat} values upon feeding no substrate (negative control), only asparagine (N), only glutamine (Q) or a combination of the two (N & Q). The results are presented in **Figure 16**.

In theory, if the TycC5-6:BpsA fusion construct produced indigoidine-tagged asparagine, similar to the structure presented in **Figure 10A**, one would expect similar k_{cat} for N and Q and about twice their value if both substrates were added. However, from this PP_i release assay, I

calculated a very low k_{cat} of the "wt" hybrid enzyme when adding N as the only substrate (~ 0.12 min^{-1}) and a relatively high activation rate for Q alone (~1.4 min^{-1}), which was only slightly increased by adding both substrates. The product formation rate was assessed by measuring the increase in absorption at 600 nm in the linear range.

The "wt" TycC5-6:BpsA analyzed *in vitro* formed no pigment from Asn alone, but Gln was converted into indigoidine at a rate of about 0.63 min^{-1} (corresponding to the conversion of 1.27 Gln min^{-1}), which is very close to the released PP_i (**Figure 16**). For the A_{Asn}-D370A mutant, which produced close to wild type pigment levels *in vivo*, the activation of N was abrogated as expected (no significant difference in comparison to the background, p=0.42 in a two-sample student's T-test), whereas the activation of Q, also in the presence of N, was boosted in comparison to "wt" TycC5-6:BpsA (**Figure 16**). Pigment formation rates of 1.78 and 1.55 min^{-1}, respectively, confirmed this observation.

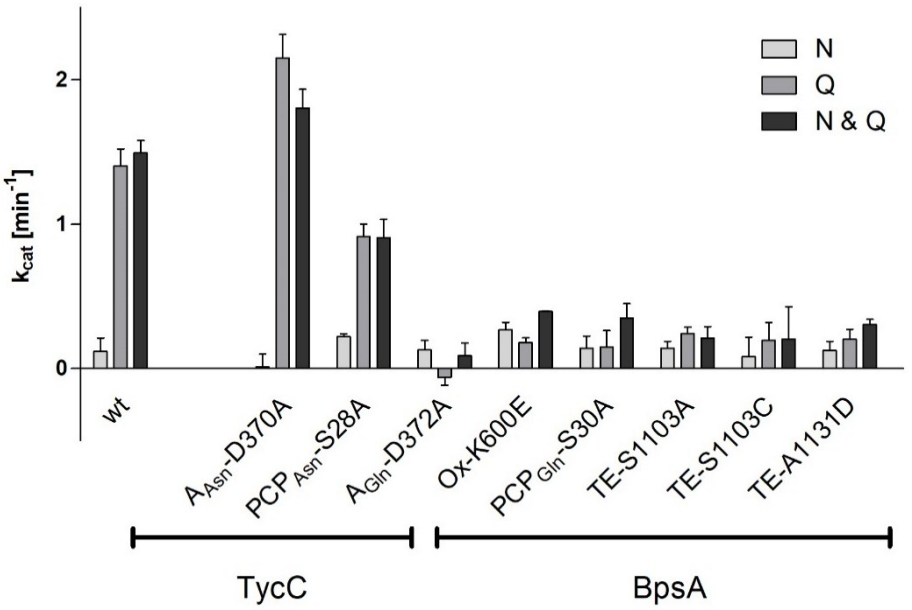

Figure 16 | *In vitro* activity of A domains in TycC5-6:BpsA and its mutants. The bar graph shows the k_{cat} values calculated from the PP_i release assay for the indicated TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA variants upon adding asparagine (N), glutamine (Q) or a combination thereof (N & Q). Reactions where the substrate was omitted were run in parallel for each variant and served as negative controls, which were subtracted from their respective sample values. Enzymes were added at 1 µM, except for A_{Asn}-D370A and TE-S1103A, which were added at 2.5 µM, and substrates at 1 mM in a total volume of 100 µl. n = 3 +SD.

In contrast to the *in vivo* expression of the PCP_{Asn}-S28A mutant, where no pigment production was observed, *in vitro* this mutant produced blue color at a rate of 0.23 and 0.21 min^{-1} (corresponding to 0.46 and 0.41 min^{-1} glutamine conversion) upon Q and N & Q addition, respectively. The same discrepancy was observed for the TE-A1131D mutant of TycC5-6:BpsA, even though the pigment production was minimal. The A_{Gln}-D372A mutant indeed failed to activate Gln and did not produce any pigment. So did the PCP_{Gln}-S974A mutant and the two TE-S1103A/C mutants. In contrast to the effect of the Ox domain mutation in BpsA, where the

glutamine activation decreased by 38% in comparison to the wt, in TycC5-6:BpsA the mutation Ox-K600E led to an almost 800% decrease in glutamine activation relative to the "wt", hinting again at a potential role of the oxidation domain in dipeptide formation or splitting of the same, as discussed above (section 4.1.4). The activation levels of Asn remained comparably low for all TycC5-6:BpsA hybrid enzymes, that should theoretically activate Asn at the native level (**Figure 16**), when no mutations interfere with the TycC5 module. Even when the native upstream C domain was added to the construct, Asn activation by the TycC5 did not increase (section 4.2.6, **Figure 17**, D).

4.2.6 *In vivo and in vitro analyses of TycC:BpsA fusions with different start sites emphasize how intermolecular communication plays a role in pigment production*

For the most promising TycC:BpsA fusion site (TycC(A_{Gln}_R395):(G434)BpsA), I reconstituted an indigoidine synthetase from TycC6:BpsA and analyzed the effect of the N-terminal start site on pigment production *in vivo* (**Figure 17A**) and *in vitro* (**Figure 17B**). Three different start sites of TycC6 A domain were tested for their effect on pigment formation *in vivo* relative to the negative control (BpsA PCP mutant, S974A). The relative absorption at 600 nm of the overnight cultures is shown in **Figure 17A**. Start sites "L-19" and "IndC(M1-M27):(V1_A_{Gln})" were also tested for the TycC6:IndC fusion and led to a relative absorption of 0.16 and 0.12 au, respectively (compare **Figure 13**), demonstrating again the different behavior of BpsA and IndC, also in association with TycC modules. *In vitro* analysis of the same TycC6:BpsA constructs revealed that in contrast to the *in vivo* pigment production, the conversion rate of glutamine is highest for start site "L-19" (**Figure 17B**).

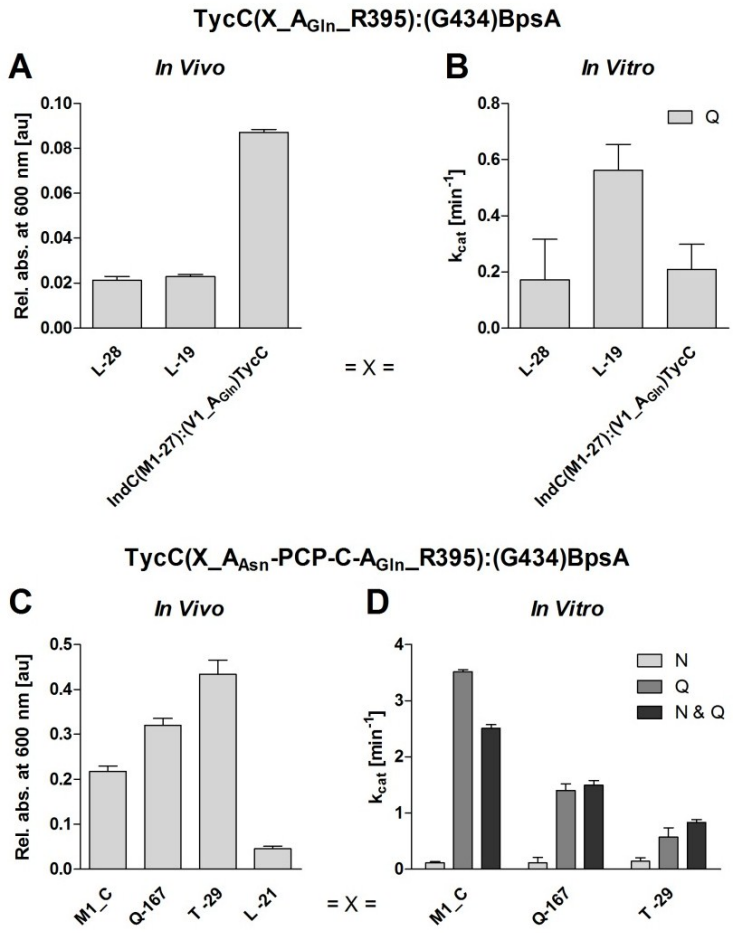

Figure 17 | Comparison of *in vivo* and *in vitro* activity of TycC6:BpsA and TycC5-6:BpsA fusions with different start sites. Relative absorption at 600 nm of **(A)** TycC6:BpsA and **(C)** TycC5-6:BpsA constructs featuring the indicated N-terminal start sites (=X) after overnight expression in BAP1 cells from pTrc99 relative to the negative control (BpsA PCP mutant, S974A). The same variants of **(B)** TycC6:BpsA and **(D)** TycC5-6:BpsA were purified from the BAP1 cells and the activation of their respective substrates, N, or Q, or both, were measured in the *in vitro* PP_i release assay to calculate the k_{cat} values. Enzymes were added at 1 µM, substrates at 100 µM, each. For all bar graphs, the mean of three replicates +standard deviation is shown.

Further, I analyzed the effect of the N-terminal start site on pigment production of TycC5-6:BpsA. Including the entire upstream C domain ("M1_C") or the C-A domain linker ("Q-167") did not have a particularly positive effect on pigment production *in vivo* (**Figure 17C**). The two start sites "T-29" and "L-21", corresponding to "L-28" and "L-19" of TycC6, had converse effects: For TycC6 both led to a similar pigment production level *in vivo*. For TycC5, "T-29" had a favorable effect on pigment production whereas "L-21" decreased pigment production to a value below 0.1 au. In comparison, the *in vitro* results of the PP_i release assay follow an opposite trend with regards to the activation rates (k_{cat}) of glutamine (Q). This rate is highest for start site "M1-C" and decreases for the other start sites, where "L-21" was not tested (**Figure 17D**). Originally, I hoped to increase the activation of asparagine (N) by leaving the native TycC5 C domain in front of the A domain (A_{Asn}), however, the activation rate of N for this start site remains very low, just as for "Q-167" and "T-29". Thus, the activation rate of N & Q combined can likely be attributed to the activation of Q alone (**Figure 17D**). It would be interesting to investigate TycC6:BpsA with the native C domain up front to see if the pigment production would be increased.

4.2.7 *TycC5-6:BpsA produces pure indigoidine in vitro*

I analyzed the products formed *in vitro* by BpsA as a control and TycC5-6:BpsA (TycC(Q-167_A_{Gln}-PCP-C-A_{Asn}_R395):(G434)BpsA), to evaluate which pigment molecule is formed by the hybrid enzyme. On an LC-MS I recorded the LC chromatogram of the products at 600 nm (**Figure 18A, C**) and the ion traces of indigoidine (m/z $[M-H]^-$ 247.0) and indigoidine-tagged asparagine (m/z $[M-H]^-$ 475.0) measured in negative mode MS (**Figure 18B** and **D**).

For both enzymes, one peak with an absorption at 600 nm elutes from the column at 2.9 minutes. This peak coincides with the ion trace of indigoidine, while the ion trace for indigoidine-tagged asparagine does not rise above background level. The identity of the ion trace peaks for indigoidine after the main peak could be the tail of indigoidine still eluting from the column, which is also visible in the chromatogram at 600 nm. These data provide evidence that the TycC5-6:BpsA fusion construct produces pure indigoidine as the only colored product and does not tag indigoidine with asparagine.

An LC-MS measurement of the *in vitro* reaction of the TycC5-6:BpsA oxidation domain mutant (K600E) failed to provide evidence that this mutant produces the dipeptide Asn-Gln (m/z $[M-H]^-$ 259.0), which was a hypothesis I originally had.

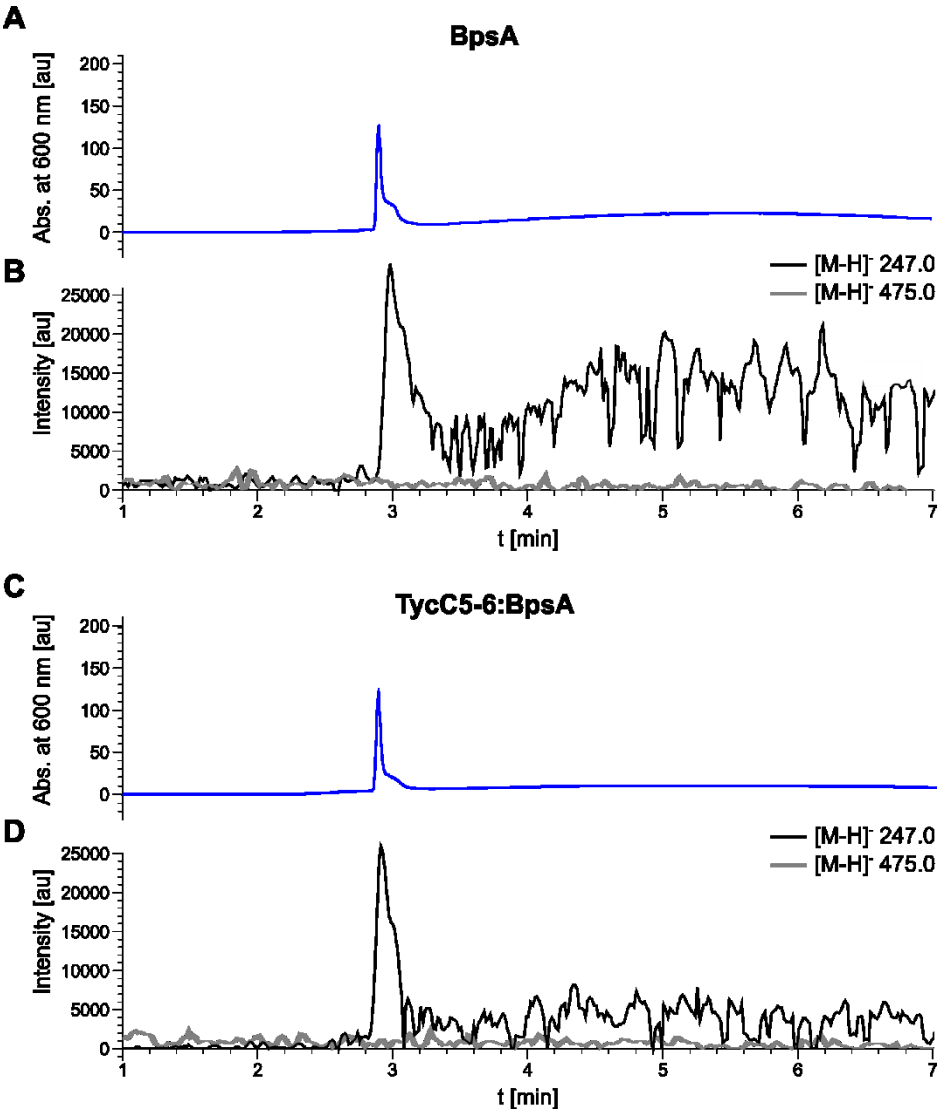

Figure 18 | BpsA and TycC5-6:BpsA both produce pure indigoidine *in vitro*. **(A)** The chromatogram of the BpsA product measured at an absorption wavelength of 600 nm. **(B)** Extracted ion traces of indigoidine (m/z $[M-H]^-$ 247.0, black line) and indigoidine-tagged asparagine (m/z $[M-H]^-$ 475.0, gray line) measured via ES-API MS in negative mode. **(C)** The chromatogram of the TycC5-6:BpsA product measured at an absorption wavelength of 600 nm. **(D)** Extracted ion traces of indigoidine (m/z $[M-H]^-$ 247.0, black line) and indigoidine-tagged asparagine (m/z $[M-H]^-$ 475.0, gray line) measured via ES-API MS in negative mode. Enzymes were added at 3 µM in sodium phosphate buffer, pH 8, with 1mM Gln for BpsA and 1 mM Asn and Gln for TycC5-6:BpsA, 20 mM $MgCl_2$ and 5mM ATP. The pelleted pigment was dissolved in 100 % DMSO before subjection to the LC-MS.

To sum up, despite my efforts to generate a fusion enzyme able to produce only indigoidine-tagged asparagine, the data rather suggest that the TycC6:BpsA part of the fusion works independently to produce pure indidoigine.

4.3 Further aspects and engineering approaches towards an improved NRP-indigoidine fusion synthetase

In this section I will summarize other aspects of engineering NRPSs, indigoidine synthetases and in particular the fusions thereof, always with the aim in mind to create or improve an NRP-indigoidine fusion synthetase to easily monitor the success of NRPS engineering.

4.3.1 *Engineering the post A10 A domain motif of IndC to match an interrupted termination module does not increase pigment production significantly*

(Experiments in this section 4.3.1 were partly executed together with Luise Nottmeyer as part of an internship during her Bachelor studies)

In a recent study, the post A10 (pA10) sequences of different A domains were compared[54]. A distinct motif ("LPxP") was identified and proposed to expand the A10 motif, since it seems to play a role in interacting with the A domain to stabilize the catalytic lysine of the preceding A10 motif. Furthermore, it was reported that this motif is mainly found in the A domains of multimodular NPRS and not in standalone A domains. Thus, I aimed to test whether engineering this post A10 motif could make IndC behave more like part of a multimodular NRPS, all the while decreasing its standalone activity. I aligned different A domain sequences from *Photorhabdus luminescence* and *Brevibacillus Brevis*, along with one interrupted A domain of an elongation module (TgaC, a hybrid PKS/NRPS from the thuggacins synthetase of *Sorangium cellulosum,* elongation A domain interrupted by an Ox domain) and one interrupted A domain of a termination module (PchF of the pyochelin synthetase from *Pseudomonas aeruginosa,* interrupted by a methylation domain) and compared their expanded post A10 motifs to IndC. All sequences from elongation A domains of NRPS found in *Photorhabdus luminescence* as well as the three sequences included from TycC featured the predicted "LPxP" motif (**Figure S 7**). While TgaC also featured the proposed "LPxP" motif, IndC and PchF both shared the conserved leucine, but not the proline residues of that motif, which for PchF is not in line with the findings of Miller *et al*.[54](**Table 9**).

Table 9 | The extended post A10 motifs of IndC, TgaC and PchF. The pA10 motif is underlined.

protein	extended pA10 motif
IndC	YQSLSES
TgaC	RGALPDP
PchF	RKALTGF

I then engineered the IndC expanded post A10 motif to match either the interrupted elongation A domain TgaC (pA10-1: "RKALPEP") or the interrupted termination A domain PchF (pA10-2: "RKALTGF"). Luise and I replaced the native pA10 motif of IndC and TycC5-6:IndC with either one of those two, expressed them in BAP1 cells and assessed pigment production (**Figure 19A**). The absorption at 600 nm of both "wild types" was set to one and the pA10 engineered versions normalized accordingly to have a better comparison of the effects. In absolute numbers, the TycC5-6:IndC fusion produces less pigment than IndC. Changing the pA10 motif of IndC to that of TgaC (pA10-1) led to a decrease in pigment formation of the native standalone IndC and also of

the fusion construct. Changing the pA10 motif of IndC to match that of PchF (pA10-2) led to a decrease in pigment formation by IndC but to a slight increase in pigment formation of TycC5-6:IndC (**Figure 19A**). Hoping that this would be due to the desired effect of having turned IndC into an elongation module, we implemented the changes to the post pA10 motif also in the TycC5-6:IndC A_{Asn} domain and PCP_{Asn} domain mutants and compared their pigment production levels (**Figure 19B**). For the "wt" as well as the mutants of TycC5-6:IndC, the pA10-2 motif led to a slight increase in pigment production, significantly for the PCP_{Asn} domain mutants ($p<0.05$ in a two-sample student's T-test). Consequently, the desired effect of a high increase in pigment production for the "wt" TycC5-6:IndC and a decrease for the mutants of the Asn-incorporating module was not achieved.

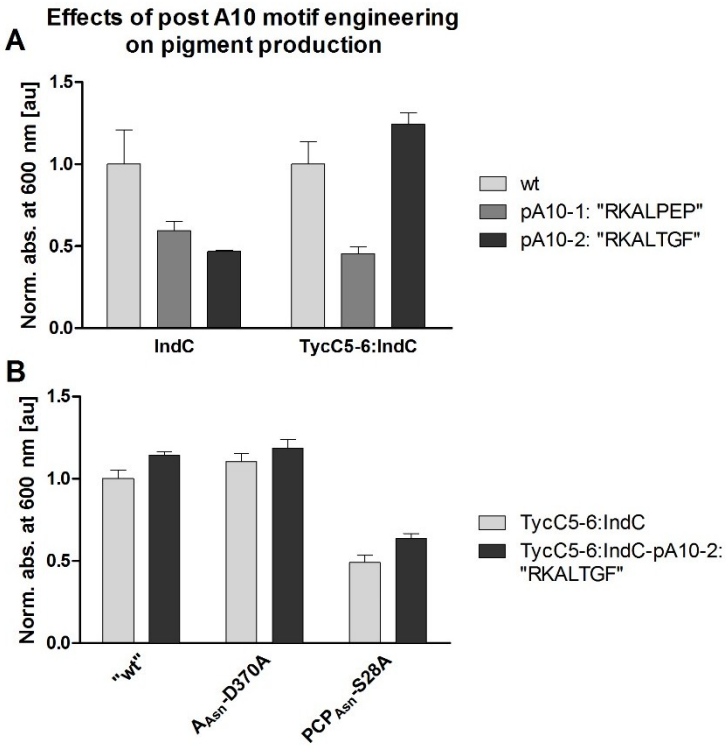

Figure 19 | Engineering the post A10 motif had an effect on pigment production of IndC and TycC5-6:IndC in *E. coli*. (A) In wild type IndC and TycC5-6:IndC, the effect of post A10 engineering was assessed by expressing the indicated variants in BAP1 cells from the pTrc99 vector, measuring the absorption at 600 nm relative to the negative control (IndC PCP mutant, S980A). Values of the wild types were set to one and pA10 variants normalized to the wt for better comparison. **(B)** For pA10 version "2", the effects on pigment production of the wt TycC5-6:IndC and the A_{Asn} domain mutant (D370A) and the PCP_{Asn} mutant (S28A) are depicted. n = 3 +SD.

4.3.2 *Random mutagenesis of TycC5-6:IndC fusions yields a small number of mutants, all of which show reduced pigment production*

(Experiments and data analysis of this section were partly executed by Luise Nottmeyer as part of an internship during her Bachelor studies) (**Figure 20C**)

In a different approach to improve the pigment production levels of TycC5-6:IndC hybrid enzymes, I turned towards directed evolution. I performed random mutagenesis of complete or partial TycC5-6:IndC fusion constructs and cloned them into a non-mutated expression vector. The idea was to transform the mutated variants directly into the expression cells, *E. coli* BAP1, to enable high throughput screening of the different clones. Even though the yield of purified DNA after random mutagenesis PCR was comparable to that for the wild type counterpart, the transformation into BAP1 cells did not yield many colonies under different conditions such as different amounts of Gibson assembly mix or the type of transformation (heat shock versus electroporation). Thus, the conventional route via first cloning into competent *E. coli* TOP10 cells

was chosen. However, the yield of positive colonies (checked via colony PCR for the presence of an insert) was not much higher, despite the optimization of the conditions.

The DNA was purified from the positive colonies and transformed into BAP1 cells, which were then used to express the randomly mutated clones overnight to assess pigment production from the absorption measurement at 600 nm relative to the negative control (**Figure 20A, B**). I partially sequenced four random samples and counted about five to fourteen point mutations per 1 kbp, which corresponds well with the mutation rate between 0.6% and 2% stated by the manufacturer of the mutagenesis kit.

Changing the expression vector from pCDFDuet to pTrc99 to be able to achieve an expression in the cloning strain (TOP10) was not successful since the endogenous PPtase of *E. coli* did not sufficiently activate the hybrid enzyme and even for the "wt" TycC5-6:IndC no pigment production was observed (not shown). However, expression of the same TycC5-6:IndC fusion construct from pTrc99 instead of pCDFDuet in BAP1 cells led to an increase of pigment production for the "wt" (**Figure 20C**, compare **Figure 14**).

Unfortunately, all randomly mutated variants did not produce more pigment than the respective "wt". On the contrary, for almost all of them, the pigment production was largely decreased (**Figure 20A-C**). This is not surprising, since most DNA point mutations do not change the protein sequence at all (e.g. at the last position of a triplet). If the DNA point mutation does change the protein sequence, the likelihood that it has a loss-of-function effect is higher than a gain-of-function effect. Thus, the strategy of introducing random mutations and screening a vast number of mutant enzymes needs to be revised and improved, e.g. by using a robotic system that can handly a huge number of clones, to be able to find a higher pigment producer in several rounds of directed evolution.

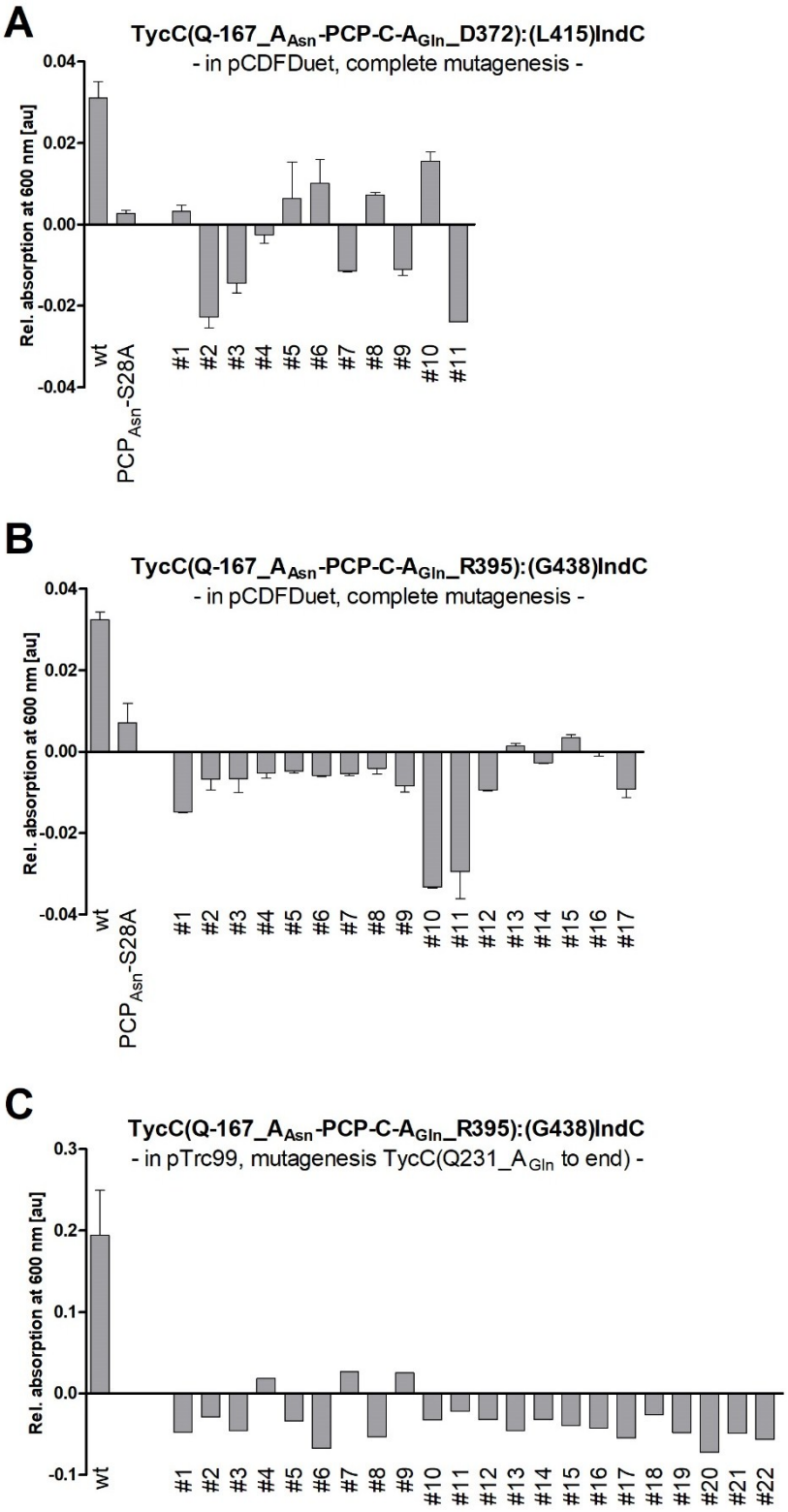

Figure 20 | Random mutagenesis of TycC5-6:IndC fusion constructs did not yield a mutant with higher pigment production. (A-C) The absorption at 600 nm of overnight expression cultures of the indicated constructs, their site-directed (PCP_{Asn}) and randomly mutated variants (numbered consecutively) from the indicated vectors in BAP1 cells are shown relative to the negative control (IndC PCP mutant, S980A). (A,B) n = 3 +SD, (C) n_{wt} = 3 +SD, $n_{mutants}$ = 1

4.3.3 *Mutations to the Ox domain cannot be rescued through external supplementation of the wild type Ox domain*

In a co-culture study from five decades ago, it was found that the only mildly colored leuco-indigoidine, produced by one bacterial strain, could be converted to indigoidine by external oxidation[313]. Since my previous experiments pointed towards a role of the oxidation domain in preventing the formation of indigoidine-tagged peptides and amino acids, I aimed to supplement the oxidation domain externally, though within the expression cells, to first allow the formation of the di-peptide intermediate followed by the oxidation step to produce the pigment. To this end, I expressed the oxidation domain mutants (K600E) of BpsA and TycC5-6:BpsA (TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA) alone and concomitantly with the BpsA oxidation domain (residues 569-695) from a separate vector in BAP1 cells to assess the blue pigment production (**Figure 21**).

External supplementation of the BpsA oxidation domain to BpsA and TycC5-6:BpsA oxidation domain mutants (K600E) did not lead to pigment production (**Figure 21**). A possible reason for this could be that the oxidation domain is still present in the mutant and therefore the external oxidation domain cannot take the place, also not conformationally, of oxidizing glutamine.

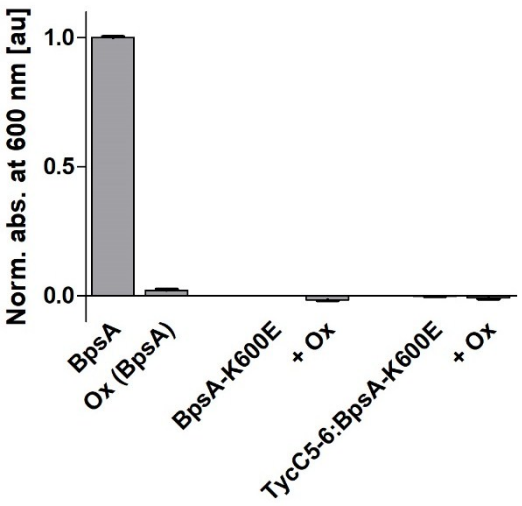

Figure 21 | Co-expression of the excised oxidation domain with the Ox domain mutants of BpsA and and TycC5-6:BpsA does not lead to pigment formation. Bar graphs show the absorption at 600 nm of the indicated constructs after overnight expression, without or with (+ Ox) the BpsA oxidation domain, in BAP1 cells from pTrc99 (NRPS constructs) and pCDFDuet (Ox domain) relative to the negative control (BpsA Ox mutant, K600E) and normalized to the absorption of BpsA wt. n = 3 +SD

4.3.4 *Reconstruction of an indigoidine synthetase from A_{Gln} of TycC6 by inserting the oxidation domain and appending the IndC TE domain did not lead to a functional enzyme*

I tested whether I could turn the glutamine activating module TycC6 into an indigoidine synthetase by inserting the oxidation domain of IndC into it (testing two different sites, I. and II.) and appending the TE domain of IndC to its C-terminus, which is apparently also necessary to form indigoidine. Different start sites were tested for the different oxidation domain insertion sites (1, 2, 3). Exact start and fusion sites are listed in **Table 10**.

Table 10 | Start and fusion sites of TycC6(A_{Gln}):IndC(Ox):TycC6(A_{Gln}-PCP):IndC(TE) constructs created to test indigoidine production in BAP1 cells.

	start	TycC A_{Gln}	IndC Ox	TycC A_{Gln}-PCP	IndC TE
I. 1	Q-167	V1 – D373	L415 - K862	V477_A_{Gln} – PCP_A61	W1015 - end
I. 2	(M1-M27)IndC				
I. 3	-				
II. 1	Q-167	V1 – D373	L415 - I804	K426_A_{Gln} – PCP_A61	W1015 - end
II. 2	(M1-M27)IndC				
II. 3	A-7				

A scheme of the constructs is shown in **Figure 22A**. These constructs were expressed in BAP1 cells and their pigment production was assessed via measuring the relative absorption at 600 nm (**Figure 22B**). None of the constructs could yield a functional indigoidine synthetase.

These data together with the data on the TycC5-6:IndC fusion indicate that keeping the Ox-A_{Gln}-PCP-TE interface of IndC intact is important for its function. I created a similar construct to I. 2 (**Table 10**) before, without re-inserting the C-terminal A_{Gln} and PCP of TycC6, which was able to produce blue pigment (**Figure 13**).

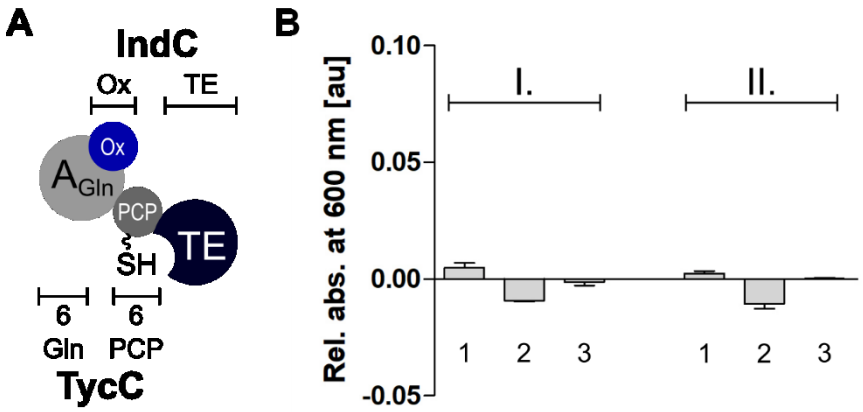

Figure 22 | Inserting the IndC oxidation domain into TycC6 and adding the IndC TE domain does not reconstitute and indigoidine synthetase. **(A)** Schematic depiction of the reconstituted indigoidine synthetase. Parts of TycC are shown in gray, parts of IndC are shown in blue. **(B)** Bar graphs show the absorption at 600 nm of the indicated constructs (refer to **Table 10**) after overnight expression in BAP1 cells from pCDFDuet (MCS2) relative to the negative control (IndC PCP mutant, S980A). n = 3 +SD

4.3.5 *Rescuing removal of the TE domain by its external supplementation can be achieved for BpsA but not for the TycC5-6:BpsA fusion*

Another idea I had was to remove the TE domain of the TycC5-6:BpsA fusion construct to enable a slow (see section 4.4.1, **Figure 28**) formation of the Asn-Gln_{Ox} dipeptide which could then be released by the externally supplied BpsA TE domain to form indigoidine-tagged asparagine or at least prevent standalone activity of the BpsA part, which is impaired by TE domain mutations or deletion.

To this end I created a truncated BpsA construct consinsting of the wt BpsA up to residue Q1014 (BpsA(Q1014) ΔTE) and separately cloned the BpsA TE domain starting at residue E1015 (TE_E1015-A1285) for co-expression in BAP1 cells. The question I intended to answer with these experiments is whether the PCP and the TE domains of BpsA are able to interact *in vivo* even when not on the same molecule. In addition, I created the TE deletion construct TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434)BpsA(Q1014) for the fusion enzyme and also co-expressed it with the BpsA TE domain in BAP1 cells. A schematic depiction of the created constructs is shown in **Figure 23A**. The pigment formation for the TE domain deletion constructs alone or with (+) the external BpsA TE domain was assessed by measuring the relative absorption at 600 nm (**Figure 23B**).

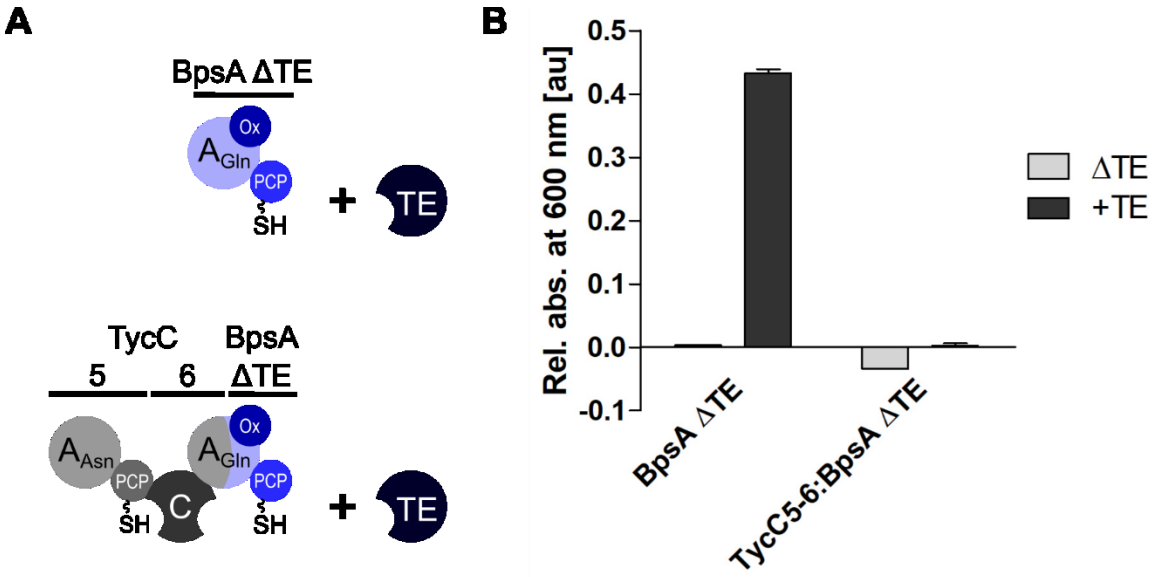

Figure 23 | Rescue of TE domain deletion of BpsA by co-expression of the BpsA TE domain was successful, but not applicable to TycC5-6:BpsA ΔTE. (A) Schematic depiction of the TE domain deletion constructs. Parts of TycC are shown in gray, parts of IndC are shown in blue. **(B)** Bar graphs show the absorption at 600 nm of the indicated constructs after overnight expression in BAP1 cells from pTrc99 (ΔTE constructs) with (+) or without the BpsA TE domain (from pCDFDuet) relative to the negative control (BpsA PCP mutant, S974A). n = 3 +SD

While the indigoidine formation of BpsA could be rescued by supplementing the TE domain *in vivo,* the TycC5-6:BpsA ΔTE construct did not produce any pigment after co-expression with the TE domain. One can speculate that the larger TycC5-6:BpsA ΔTE fusion construct might take on a different conformation, when the TE domain is missing, leading to a sterical hindrance of external TE domain binding. This hypothesis could be examined via analysis of the ΔTE constructs in electron microscopy, alone and in combination with the TE domain. To conclude, this approach does not serve our purpose, but could be used to test different TE domains for their ability to form indigoidine, as also recently proposed by Alistair Brown[314] from the Ackerley group.

4.3.6 *Other NRP-indigoidine synthetase fusions do not produce any pigment*

Of course, the tyrocidine synthetase is not the only NRPS incorporating a glutamine thus being suitable for a fusion to an indigoidine synthetase. I identified three other NRPSs, the lychenisin synthetase LchA from *Bacillus licheniformis* DSM13, the Plipastatin synthetase PpsX from *Bacillus*

subtilis strain 168 and the Iturin synthetase ItuX from *Bacillus amyloliquefaciens* DSM7, which incorporate a glutamine through one of their modules. The LchAA glutamine-incorporating module is an initiation module featuring an upstream starter C domain that incorporates a fatty acid. The PpsD and ItuB glutamine incorporating modules are elongation modules.

For the LchAA:IndC fusion (LchAA(X_A_{Gln}_T368):(T424)IndC) I tested three different start sites (=X), one including the upstream starter C domain ("M1_C") and two starting with the A_{Gln} domain ("L-163" and "M-27"). All these LchAA:IndC constructs did not produce any pigment (**Figure 24A**). In the plipastatin synthetase, PpsD, the glutamine-incorporating module, is naturally preceeded by a proline-incorporating module. I created two different fusions, one reconstituting the indigoidine synthetase (ppsD(L-60_A_{Gln}_R393):(G437)IndC) and one which should tag proline with indigoidine (PpsD(-174V_A_{Pro}-PCP-C-A_{Gln}_R393):(G438)IndC). Both constructs were expressed in BAP1 cells, alone and along with the *Bacillus subtilis* MbtH-like protein YbdZ (**Figure 24B**). In neither case did the PpsD:IndC fusion produce enough pigment to be visible. In the iturin synthetase, the glutamine-specific module features an asparagine-specific module upstream, with the specialty, that an additional epimerization domain (E) turns L-Asn into D-Asn. However, also for this ItuB(N-175_A_{Asn}-PCP-E_{Asn}-C-A_{Gln}_H406):(G434)BpsA fusion, no pigment production could be observed, despite the co-expression of YbdZ.

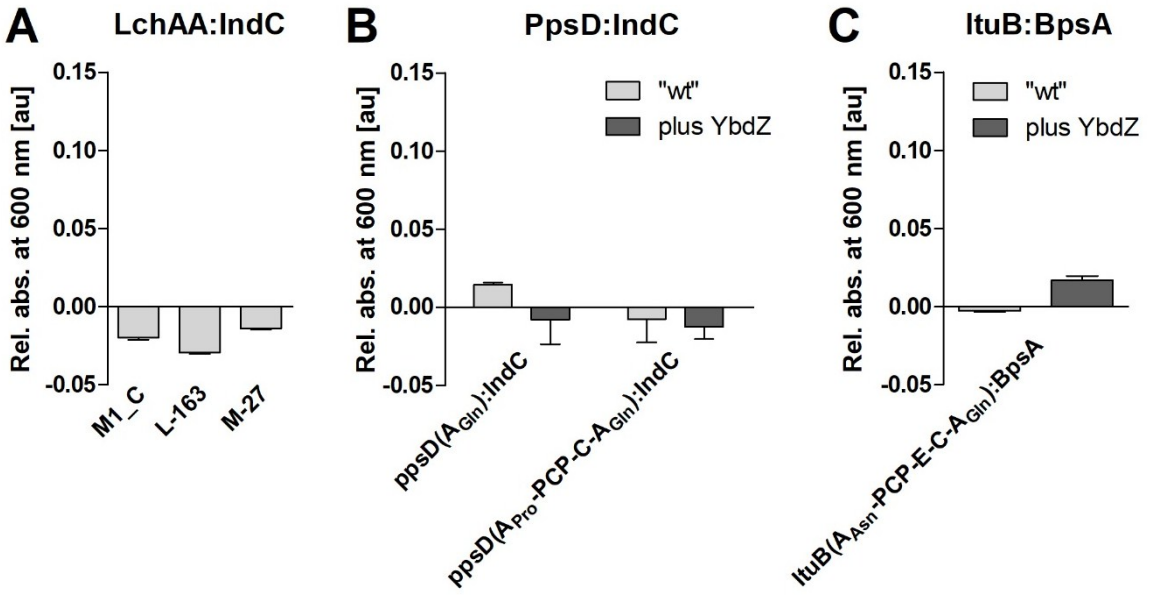

Figure 24 | Fusion constructs of other glutamine incorporating NRPS modules to an indigoidine synthetase did not produce any pigment. (A-C) Bar graphs show the absorption at 600 nm of the indicated constructs after overnight expression in BAP1 cells from (A) pCDFDuet and (B, C) pTrc99 without ("wt") or with (plus YbdZ, from pCDFDuet) relative to the negative control (IndC or BpsA PCP mutants). n = 3 +SD

For the LchAA:IndC fusion including the starter C domain, an external enzyme not present in *E. coli* might be needed to synthesize or prepare the fatty acid for its fusion to glutamine by the starter C domain. However, the fusion site chosen for LchAA:IndC ("T368:T424") also did not lead to pigment production for TycC:IndC, indicating that a different fusion site needs to be tested to draw a conclusion whether this LchAA A_{Gln} is also capable of cooperating with IndC. For the PpsD-indigoidine and the ItuB-indigoidine synthetase hybrids, the corresponding fusion site

to the one that led to pigment production in TycC:IndC/BpsA fusions, were chosen. Yet, neither one of them produced any pigment. This could be due to the absence of an MLP protein which PpsD and ItuB require for their function; IndC and BpsA do not. The co-expression with the MLP YbdZ, however, also did not cause pigment formation, which could be caused by the interference of the MbtH-like protein with the indigoidine synthetase part. However, from the crystal structure of a MLP-bound A domain, it became clear, that the MLP binds at the N-terminal A_{core} domain and not at the C-terminal A_{sub} domain[75,154], which is where the fusion to the indigoidine synthetase is. Theoretically, the oxidation domain of the IndC or BpsA could still sterically interfere with the proper MLP binding and thus prevent proper functioning.

4.3.7 *Attempted experimental validation of the putative dodecylindigoidine synthesis gene cluster of Shewanella violacea DSS12*

(These experiments were performed in part by Konrad Herbst during an internship[221]).

Given the structural similarity between indigoidine and N,N-dodecylindigoidine, we assumed that the latter is also produced by a NRPS-like enzyme in *Shewanella violacea* DSS12. A putative gene (SVI_3984) encoding the violet pigment synthetase (WP 013053246.1) was identified via homology BLAST using a Bidirectional Best Hits (BBH) approach in comparison to the pigment producing *Rheinheimera baltica* DSMZ 14885[221]. In order to validate the putative N,N-dodecylindigoidine synthetase dIndS, we chose an untargeted gene disruption approach as described before[305]. An overview of this approach is depicted in **Figure 25A**. In short, the pMiniHimar-Rb1 plasmid harbors a transposase under the control of the Plac promoter (both shown in gray). This plasmid was conjugated via an *E. coli* host into *Shewanella violacea.* The transposase randomly integrated the inverted repeat (IR, shown in yellow)-flanked R6K origin of replication and the kanamycin resistance (*kan*^{r}) gene (shown in red) into the *Shewanella* genome. The resulting *S. violacea* cells were grown on kanamycin supplemented MB agar plates to select for successful integration. Colonies were checked for pigmentation (**Figure 25B**). Four pigment producing ("1-4 +", circled in purple) and four non-pigment producing ("1-4 -" circled in yellow) *S. violacea* colonies were picked, grown and their whole genome was extracted. At this point, two different means of mapping the integration site were used (**Figure 25A**): a) The genomic DNA was digested, re-ligated and transformed into *E. coli* EC100 pir+, where only re-ligated plasmids harboring the R6K origin of replication and the *kan*^{r} gene could survive and replicate on LB agar plates supplemented with kanamycin. The plasmids would then have been purified from the individual colonies and sequenced using a primer binding to *kan*^{r} facing outwards. b) The genomic DNA was used to perform a PCR covering the complete putative dIndS gene to see whether the *kan*^{r} and the R6K ori were inserted, leading to a longer PCR product.

BamHI whole genome digest seems to have been successful (**Figure S 8**). After re-ligation with T4 DNA ligase, the resulting circularized genomic DNA fragments were transformed into *E. coli* EC100 pir+ via electroporation. However, after several attempts, no colonies were obtained (not shown). The positive control plasmid containing the R6K origin of replication and the kanamycin resistance gene did yield plenty of colonies *E. coli* EC100 pir+, confirming the competence of the bacteria and the correct implementation of the transformation protocol.

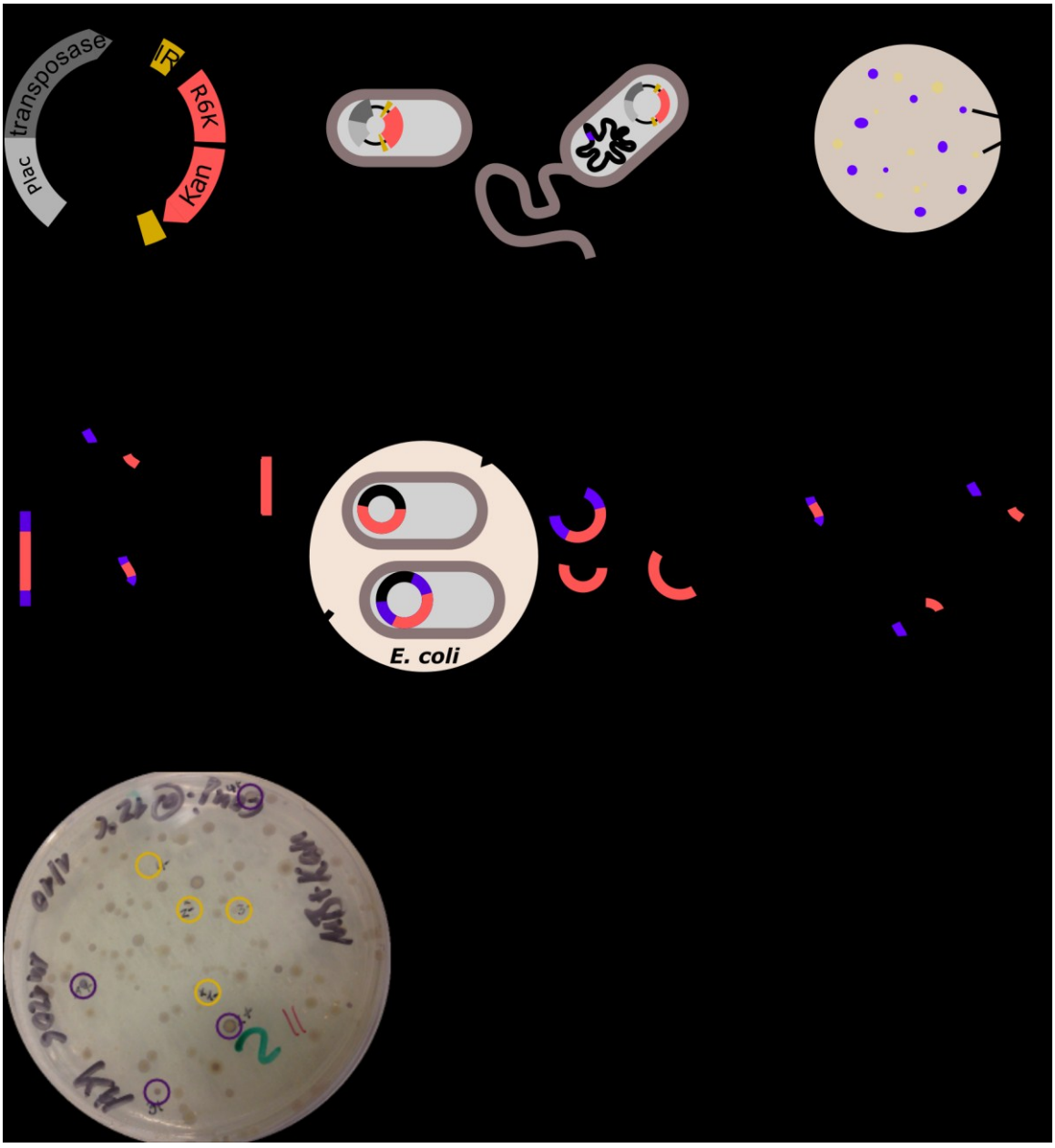

Figure 25 | Experimental approach to validate the putative violet pigment synthetase in Shewanella violacea. (A) Scheme of untargeted gene disruption to identify the gene location responsible for violet pigment production. **(B)** *S. violacea* colonies after conjugation with pMiniHimar-RB1 at 12 °C, diluted 1:10 on Marine Broth 2216 supplemented agar. Four pigment producing (circled in purple) and four non-pigment producing colonies (circled in yellow) were grown in liquid medium and the genomic DNA was extracted for a) digest and relegation or b) a PCR across the putative dIndS.

Thus, we decided for a more targeted analysis of the putative gene whose disruption should abrogate pigment production in *S. violacea*. We designed three DNA primer pairs with which the complete gene could be amplified via PCR in circa 2 kbp steps. In case of an insertion of the IR-flanked R6K/kan^{R} cassette, the PCR product should be larger (about 4.5 kbp). Successful amplification of 2 kbp fragments for both, the pigment producing and non-producing colonies (not shown), indicates that, in none of those colonies, the putative *dIndS* gene was targeted by the transposon.

Given the number of non-pigmented colonies on the MB agar plate and the fact that both pigmented and non-pigmented clones did not show an interruption of the putative dIndS gene, we conclude that there are likely many factors influencing pigment production. Thus, our untargeted gene disruption approach did not lead to the experimental validation of the putative dodecylindigoidine synthesis gene. However, from the bioinformatical predictions, we are confident that WP 013053246.1 is indeed responsible for N,N-dodecylindigoidine production.

4.3.8 *Neither wild type nor engineered dIndS lead to pigment production in* E. coli

Since we were confident about the predicted putative dIndS in *S. violacea,* we continued investigating it. In addition, we applied an engineering approach replacing the lauric acid A domain of dIndS by three different TycC A domains (specific for Val, Orn and Leu) with the aim to produce indigoidine-tagged amino acids (**Figure 26A**). To this end, we amplified the putative pigment synthetase from the *Shewanella* genome and cloned it seamlessly into the second MCS of pCDFDuet. We then determined possible insertion sites for the TycC A domains to replace the lauric acid A domain of dIndS via multiple protein sequence alignment of Clustal Omega. We chose relatively conserved residues, five amino acids upstream and three amino acids downstream of the A domain border according to Pfam (**Figure 26B**). We amplified the dIndS and the backbone, leaving out only the sequence of the lauric acid A domain. Via Gibson assembly we inserted each of the TycC A domains specific for either Val, Orn or Leu amplified with primers featuring the overlapping dIndS DNA sequence of the insertion site. In addition, we amplified the putative *Shewanella* PPtase (WP_013050488.1) and inserted it into the pTrc99 backbone via Gibson assembly.

We expressed the wt dIndS and TycC $A_{Val/Orn/Leu}$-dIndS fusion constructs in *E. coli* BAP1 cells (**Figure 26C**) and co-expressed them with the putative PPtase in M9 minimal medium. Unfortunately, neither the wt nor the engineered versions resulted in pigment production (**Figure 26C**).

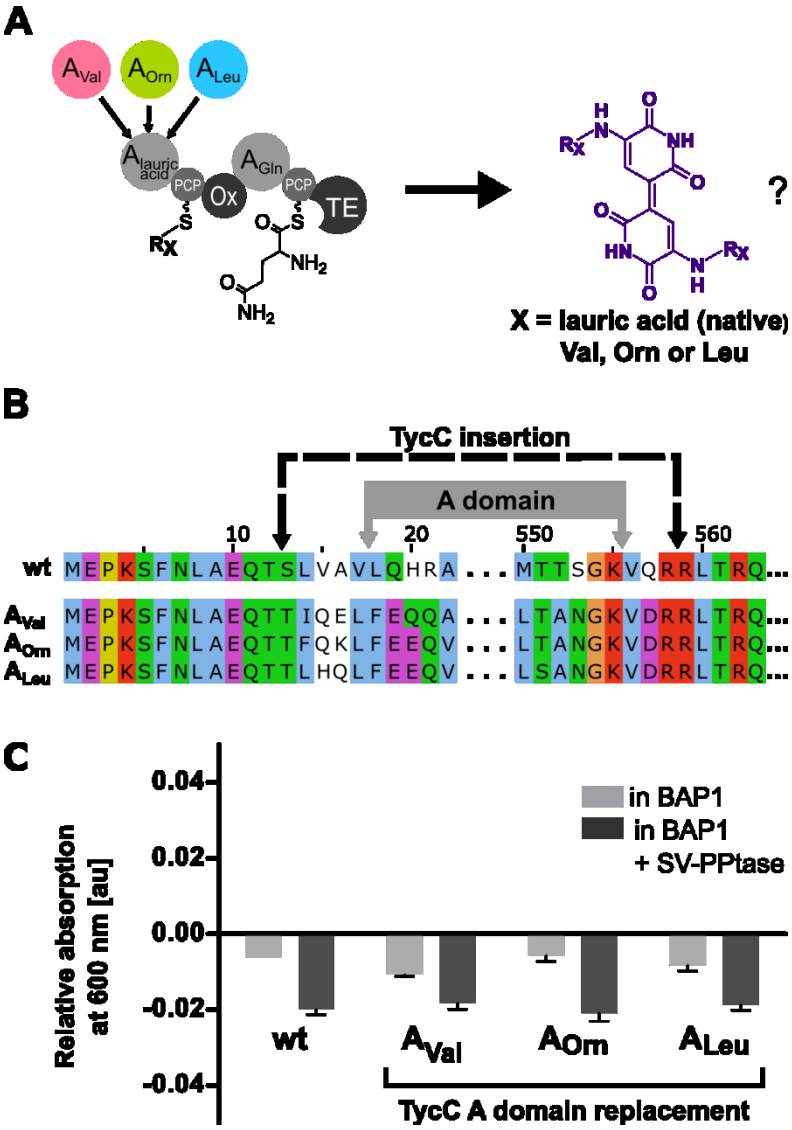

Figure 26 | Neither wild type nor engineered dIndS produce a violet pigment in *E. coli*. **(A)** Engineering approach to create an NRPS that can tag an amino acid with indigoidine. In our engineered versions of dIndS, the wt A domain specific for lauric acid was replaced by TycC A domains (Val, Orn or Leu) to produce N,N-valine-, ornithine-, or leucine-indigoidine. **(B)** Protein Alignment of the insertion sites of the TycC A domains into the wt dIndS. The wt N-terminus of dIndS was kept up to T12, where the respective TycC A domain was inserted up to R559. Thus, the complete lauric acid A domain was replaced. **(C)** Expression of wt and engineered dIndS in E. coli did not lead to pigment production. Neither when expressed in BAP1 alone nor in together with the putative S. violacea PPtase could we detect any pigment, i.e. absorption around 600 nm.

Based on the hypothesis that the putative dIndS we have identified from *S. violacea* is indeed the NRPS responsible for dodecylindigoidine production, I assumed that the TE domain of dIndS is capable or even responsible for the release of an indigodine-tagged lauric acid. I performed a protein alignment of the PCP-TE domains of IndC, BpsA, TycC10+TE and dIndS using Clustal Omega[315] to find suitable fusion sites and obvious differences between the analyzed TE domains. One obvious difference that I found was that dIndS featured a complete catalytic triad, including aspartate (D), which was replaced by alanine (A) in the indigoidine synthetases. Mutations of this residue in IndC, BpsA and the TycC5-6:BpsA fusion are described elsewhere (sections 4.1.2 and 4.2.5). Another obvious difference was located in the PCP-TE linker, where dIndS features an elongated alpha helix in comparison to the other PCP-TE linkers.

I created three different fusion constructs, replacing the IndC PCP-TE or TE domain(s) with the respective dIndS version: (i) IndC(A_{Gln}-Ox_L896):(S1536_A_{A9-A10}-PCP-TE)dIndS, which sets the fusion site right after the IndC Ox domain and before the A9 motif; (ii) IndC(A_{Gln}-Ox-A-PCP_A1013):(A1651_TE)dIndS replacing exclusively the TE domain while still keeping the PCP-TE linker from dIndS, as suggested by Beer *at al.* before[116] and (iii) TycA(M1-S28):TycC(I-5_A_{Asn}-PCP-C-A_{Gln}_D373):(L415)IndC(L896):(S1536)dIndS, where the same PCP-TE domains were replaced as in (i), just added onto the TycC5-6:IndC fusion shown in **Figure 12A**. I expressed these fusion constructs in BAP1 cells and measured their absorption at 600 nm relative to the empty vector control. No pigment production was detected for any of the constructs (**Figure 27**).

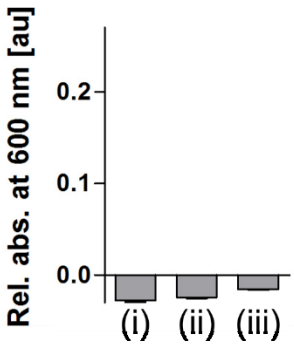

Figure 27 | Exchange of the PCP-TE (i) or TE domain of IndC (ii) and TycC5-6:IndC (iii) fusion construct with the equivalent of dIndS does not lead to pigment production *in vivo*. The indicated constructs (please refer to the main text for details) were expressed in BAP1 cells from pCDFDuet (MCS1) over night. Their absorption at 600 nm was measured relative to the negative control (IndC PCP mutant S980A). n = 3 +SD.

4.4 Context-dependent activity of A domains in the tyrocidine synthetase

(Parts of this section were submitted to Scientific Reports as a research article: A. Degen, F. Meyerthaler, H.D. Mootz and B. Di Ventura "Context-dependent activity of A domains in the tyrocidine synthetase")

During the validation of the asparagine-indigoidine fusion synthetase, I discovered that, in the synthetic dimodular fusion construct TycC5-6:BpsA, inactivation of the first module via A domain mutation did not disable the second module from activating its substrate and turning it into the blue pigment indigoidine *in vivo. In vitro*, this A domain mutation of the first module even seemed to boost the activity of the A domain of the second module in comparison to the wt fusion construct. In contrast, inactivation of the first module by PCP domain mutation completely prohibited pigment production *in vivo*, while *in vitro* activity of the second A domain was still measurable though at a reduced rate along with blue color formation (section 4.2.5, **Figure 16**). To follow up on these discoveries, I decided to investigate the effects of module inactivation on neighboring modules in a (more) native environment.

Thus, I first analyzed the TycC5-6:BpsA dimodular system further, including and omitting the BpsA TE domain and comparing it to the wt excised TycC5-6 dimodule, naturally without TE domain and after adding the BpsA TE domain. To investigate a more natural, multimodular NRPS system, I focused on the last three modules of the tyrocidine synthetase, naturally followed by a TE domain, to analyze the context-dependent activity of A domains. I constructed a library consisting of individual A-PCP di-domains, namely TycC8, 9 or 10, two dimodular constructs (TycC8-9 and TycC9-10, each plus and minus TE domain) and one trimodular construct (TycC8-9-10+TE). All constructs were either kept wt or mutated within the A or PCP domain, at the same amino acids mutated in the fusion constructs before. Upon expression in BAP1 cells, the PPantylated enzymes[56,316] were purified and analyzed *in vitro* using a PP_i release assay[70,91] (**Figure 3E**).

4.4.1 *Activity of A domains embedded in the dimodular constructs TycC5-6 with or without fusion to BpsA*

In order to comprehend if the TycC5-6:BpsA fusion construct behaves differently than a native dimodular construct consisting of TycC5-6 upon upstream domain I created different variants of of TycC5-6, with or without BpsA fused at its C-terminus (**Table 11**). From the best TycC5-6:BpsA fusion construct, I removed the TE domain (TycC5-6:BpsAΔTE) but kept the BpsA oxidation and C-terminal A_{Gln} domains, to see if the substrate activation rate changes without the TE domain. Further I generated the native TycC5-6 dimodule (TycC5-6ΔTE), which naturally does not have any TE domain and no other part of BpsA. Finally, I added just the BpsA TE domain to the C-terminus of TycC5-6 (TycC5-6:TE (BpsA)), namely right after the TycC6 PCP domain. I analyzed these constructs using the online PP_i release assay. Instead of inactivating the domains with point mutations, here I control the individual domain activity by selectively adding the substrates Asn (N) and Gln (Q) individually or simultaneously.

Table 11 | Overview of potential Asn-Gln producing NRPS fusion constructs analyzed using the online PP_i release assay.

name	construct
TycC5-6:BpsA	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434_Ox-A-PCP-TE)BpsA
TycC5-6:BpsAΔTE	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G434_Ox-A-PCP_E1004)BpsA
TycC5-6ΔTE (native)	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}-PCP_A58)
TycC5-6:TE (BpsA)	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}-PCP_A58):(K1005_TE)BpsA

The k_{cat} values for all constructs in the presence of N, Q or both were calculated (**Figure 28**). For TycC5-6:BpsA these conversion rates were presented in **Figure 16** already and show that Q is activated at a much higher rate than N. Here I analyzed the same construct without the TE domain. Deleting the TE domain results in lower activity for all substrate combinations, most severely for Q alone and the combination of N & Q. This effect is very similar to the effect of the various TE domain mutations presented before (**Figure 16**) and argues that indeed the TE domain is greatly involved in product release and is thus a driver of standalone activity of BpsA in the fusion construct.

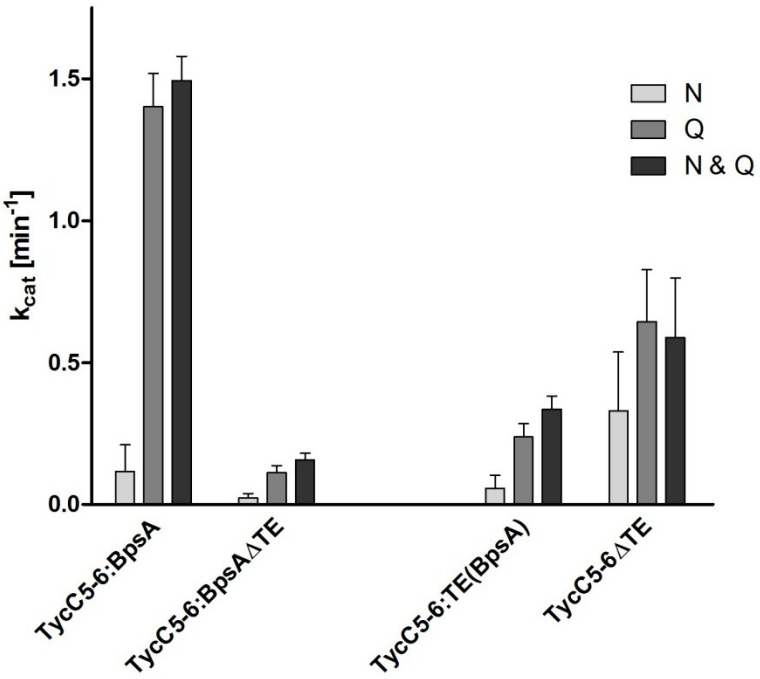

Figure 28 | Impact of the BpsA TE domain on the TycC5-6:BpsA fusion in comparison to the native TycC5-6 construct. The graph shows k_{cat} values for the indicated constructs in the presence of the indicated amino acids, after background subtraction (value w/o substrate). Amino acids were added at 1 mM, enzymes at 0.5 µM. Data represent the mean (+ standard deviation, SD) of three technical replicates.

This hypothesis is supported by the conversion rates of TycC5-6:TE (BpsA), where the TE domain of BpsA is fused to the native TycC5-6 dimodule. For this construct, the conversion of Q is still higher than that of N, however not to the same degree as for the TycC5-6:BpsA construct. This

indicates that the BpsA TE domain preferably releases the oxidized (and possibly cyclized) glutamine to form the pigment. Unmodified glutamine is not the preferred substrate resulting in a rather low activation rate of Q even in the presence of the BpsA TE domain. This assumption is further supported by the native TycC5-6ΔTE dimodular construct, where the conversion rates of all substrates are similar, even without any TE domain, indicating that all substrates can be released by unintended hydrolysis. Note that, irrespective of the presence of a TE domain, each module can activate its substrate independently from the neighboring module.

Because of these results, I decided to switch to a native system where the last module is naturally followed by a TE domain to compare it to the artificial TycC5-6:BpsA fusion construct and the effects that perturbations have on each of the components of the system. Most conveniently, I switched to the last three modules of TycC, ending in a TE domain, which has previously been reported to also release unnatural substrates as a linear product[34].

4.4.2 *Analysis of A-PCP di-domains of TycC8, 9, or 10*

First, I chose two different start sites consistent for each of the last three modules of TycC. The first one (1) is located about 20 bp upstream of the A domain start according to Pfam and was previously identified to work well for TycC5 and 6 in the indigoidine synthetase fusion *in vivo* and *in vitro* (**Figure 12** and **Figure 17**). The second one (2) is located about 30 bp upstream of the A domain start and was previously established by Mootz and colleagues[34]. For both start sites, the A-PCP di-domains TycC8, 9 and 10 were analyzed *in vitro* using the online PP_i release assay to calculate the conversion rate of their native substrates (**Figure 29**).

These experiments show that for the di-domain constructs, the start site though only differing in about 10 bp in length seems to have a large impact on substrate activation. Based on these results, I chose to apply start site (1) for all constructs in the upcoming experiments. The values are within a range of about 1 min^{-1}, making them more comparable, especially with regards to evaluating the effects of domain inactivation.

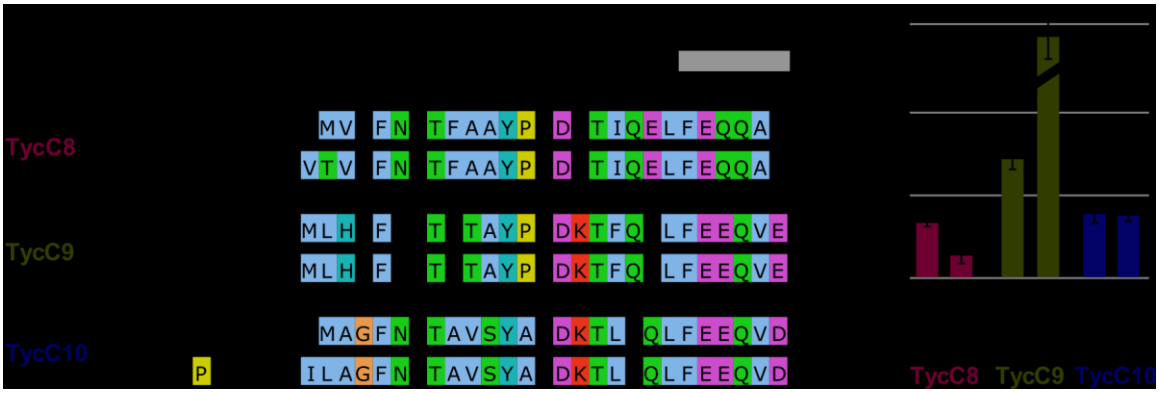

Figure 29 | Impact of the start site on activity of A domains within A-PCP di-domain constructs. (A) Sequence alignment of the indicated constructs, where (1) and (2) indicate two different start sites. Amino acids are color-coded according to the Clustal coloring scheme. Numbers above the sequence indicate the amino acid position relative to the predicted start site of the A domain which is set to zero. **(B)** Bar graphs showing k_{cat} values for the indicated constructs in the presence of the indicated amino acids after subtraction of the background (=no substrate) value. Amino acids were added at 1 mM, enzymes at 0.5 µM. Data represent the mean (± standard deviation, SD) of three independent experiments.

These effects of domain inactivation, I validated using the minimal component of an NRPS system, namely the A-PCP di-domains depicted in the alignment in **Figure 30A**. To inactivate the A domain, we introduced the same point mutation as described before (sections 4.1.2 and 4.2.5), namely the D to A conversion of the aspartic acid of core motif A7 (Y[R/K]TGDL)[51]. These mutants I abbreviate with TycC8^{A*}, TycC9^{A*} and TycC10^{A*}. To inactivate the PCP domain, I also revert to the previously described S to A mutation of the PPant attachment site within the PCP domain [I/L]GG[D/H]SL[100] (**Figure 30B** and **Figure S 10A, D, G**).

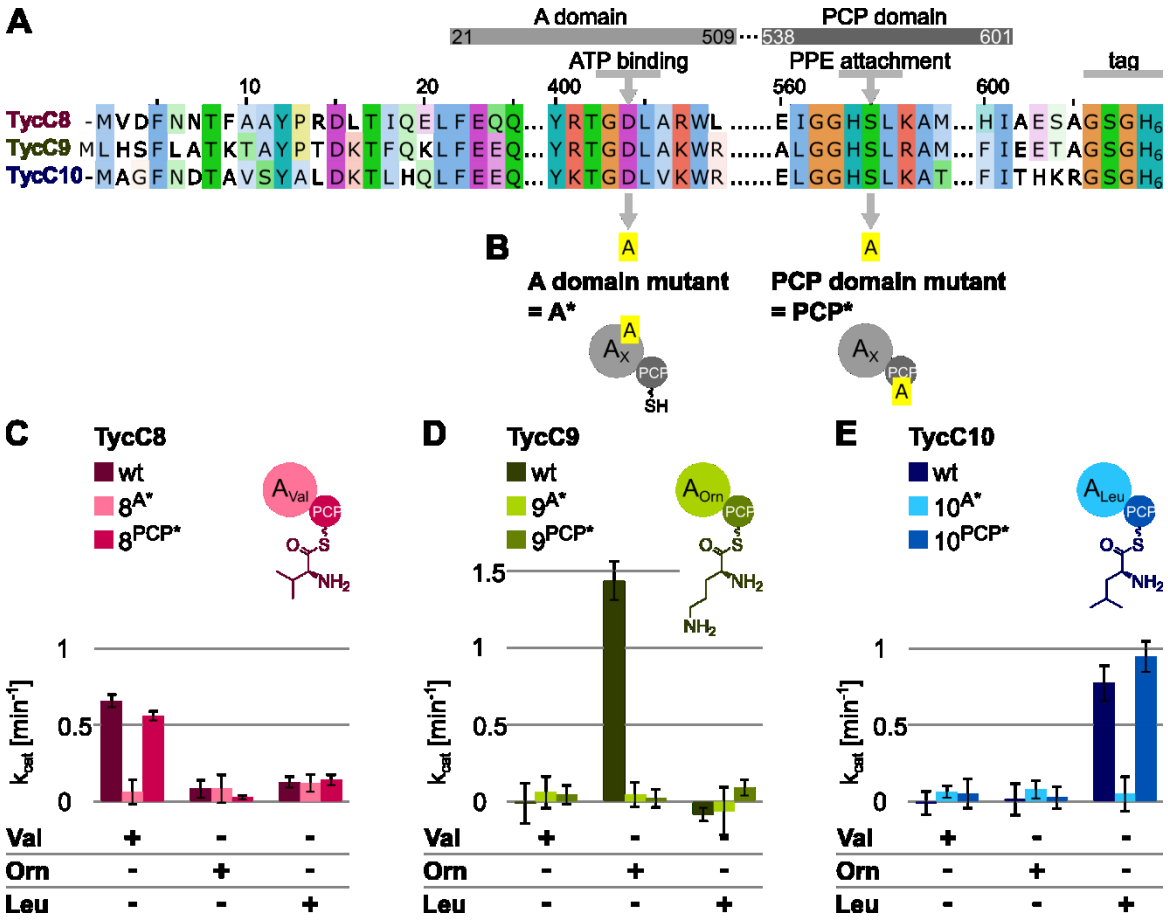

Figure 30 | Activity of A domains within "wild type" or mutated TycC 8, 9 and 10 A-PCP di-domain constructs. (A) Protein sequence alignment of the indicated constructs. Amino acids are color-coded following the Clustal scheme based on their category and degree of conservation. The amino acids of all constructs are numbered consecutively beginning with 1 at the start of construct TycC9. Predicted boundaries for A and PCP domains are shown above the sequences. Gray arrows indicate a residue critical for ATP binding and the serine needed for PPant arm attachment. "Tag" indicates a GSG linker plus the 6x His tag. **(B)** Position and identity of mutations used to inactivate the A and PCP domains. **(C-E)** Upper panel: schematic depiction of construct with amino acid attached to the PPant arm. Colors as in Figure 1, A. Lower panel: bar graph showing kcat values for the constructs in the upper panel in the presence of the indicated amino acids, calculated using the online PP_i release assay after subtraction of the background (=no substrate) value. Amino acids were added at 1 mM, enzymes at 0.5 µM. Data represent the mean (± standard deviation, SD) of three independent experiments.

For all three constructs, the wt A domains showed activity only towards the expected substrates (**Figure 30C-E**), even when additional amino acids were present (**Figure S 10C**, **F** and **I**). As

expected, A domain inactivation lead to an almost complete stall of activity. With the PCP domain mutation, two of the three A domains could still continuously activate their substrates, namely TycC8^{PCP*}and TycC10^{PCP*} (**Figure 30C**, **E**). This observation was reported in other studies before, where either the PCP domain was mutated in the same position or the A domain activity was observed in complete isolation[31]. However, in opposition to these observations, the PCP domain mutant of TycC9^{PCP*} strongly affected the adenylation activity (**Figure 30D**). In search of an explanation we came across an early publication by Mootz and colleagues[34] describing that the ornithine thioester can undergo intramolecular cyclization thereby releasing itself from the PPant arm, which is prohibited in the absence of that very PPant arm in the PCP mutant. This auto release mechanism could also explain why the TycC9 wt conversion of ornithine can be faster than the activity of its two neighbors towards their substrates (**Figure 29A**). To test this assumption, I replaced Orn with Lys, which was also shown to be accepted by TycC9 *in vitro*[34], but cannot undergo this self-cyclization. I fed Orn as the sole substrate to TycC9 (start sites 1 and 2) and TycC9^{PCP*} (start site 1) and measured the k_{cat} value (**Figure S 3J**). For "wt" and PCP mutant with the start site (1), I could not finally resolve this hypothesis. Both were activating Lys only to a very low degree, which makes it difficult to draw a conclusion even though both k_{cat} values are very similar, which is what you would expect if the substrate cannot release itself from the PCP domain. For start site (2), Lys was activated at a higher speed, though not as much as Orn (**Figure 30D**), but I did not have the PCP mutant available for comparison (**Figure S 3J**).

4.4.3 *Activity of A domains embedded in the dimodular constructs TycC8-9 and TycC9-10*

The construction and analysis of the di-domain constructs served to find the best start site and validate the effects of the point mutations as well as the reliability of the online PP_i release assay. The main objective however is to investigate the effects of module inactivation, through either lack of substrate or the introduction of point mutations, on surrounding modules. With that objective, I generated two dimodular constructs, TycC8-9 and TycC9-10 with start site (1), for analysis (**Figure 31A**, **B**). To shortly resume the investigation of the effect that natural and artificially placed TE domains can have on upstream A domains (section 4.4.1), I also added the TycC TE domain to Tyc8-9 (**Figure 31A**) and removed it from TycC9-10 (**Figure 31B**). The four constructs were expressed in and purified from BAP1 cells (**Figure S 9A**, **B**), then analyzed *in vitro* adding the amino acid substrates in various constellations using the online PP_i release assay (**Figure 31C**, **D**).

The two elongation modules, TycC8-9, with an without TE domain, both activated their respective substrates, valine and ornithine, regardless of the absence of the neighboring module's substrate (**Figure 31C**) or presence of another amino acid (**Figure S 9C**). This finding supports the notion, that A domains in elongation modules can continuously activate their substrates without waiting for the upstream A domain to do so first. Comparing the A domain activity of the "wt" TycC8-9ΔTE with the version where the TycC TE domain was added (TycC8-9+TE = TycC(V-19_A_{Val}-PCP-C-A_{Orn}-PCP_A+4):(-F16)TE(TycC)), I realized that the artificially added TE domain did not cause a change in A domain substrate activation pattern. This is an observation I have also made for another dimodular construct TycC5-6:TE(BpsA) (**Figure 28**). Thus, I conclude that an artificial TE domain that naturally releases a different intermediate substrate does not aid substrate release and thus does not accelerate substrate activation rounds of non-native upstream modules.

For TycC9-10 in contrast, where the module 10 is naturally followed by the TE domain, its presence does make a difference (**Figure 31D**). In the presence of both substrates, ornithine and leucine, TycC9-10+TE seems to be capable of releasing the dipeptide at double the speed which results in a k_{cat} value about twice as high as for the TycC9-10ΔTE variant. The corresponding mass of the linear dipeptide Orn-Leu was detected by Florian Meyerthaler (Mootz Lab, University of Münster) using LC-MS (**Figure S 12A**). When feeding the substrates individually, I discovered something unexpected: Ornithine as the sole substrate was activated by module 9, while leucine was not activated by its corresponding module 10, neither with nor without the TE domain (**Figure 31D**). The presence of an additional amino acid in the reaction did not change this outcome (**Figure S 9D**). It seems that in TycC9-10, the A domain for leucine (10) is unable to repeatedly activate its own substrate unless the upstream module (9) does so as well. This is the opposite of how the second modules in the dimodular constructs TycC5-6 and TycC8-9 reacted to a lack of substrate for the upstream module.

To exclude that TycC9-10 suffers from an inherent dysfunction in module 10, Florian Meyerthaler performed a thiolation assay (described in section 1.2.5) using radiolabeled leucine and measuring the radioactivity of the precipitated enzyme (in CPM). With that, we could show that TycC9-10+TE can indeed activate and covalently tether leucine to its PPant arm in the presence of ornithine (**Figure 31E**) to the same degree as the positive control (TycC10ΔTE). At the same time, thiolation was reduced significantly ($p < 0.01$) when ornithine was not available to the same construct (**Figure 31E**), but still well about background level (module 10^{PCP*} mutants in TycC9-10+TE and TycC10ΔTE). This suggests that TycC9-10+TE does activate leucine in the absence of ornithine, though not repeatedly.

In both native dimodular constructs, TycC8-9ΔTE and TycC9-10+TE, I inactivated each A and PCP domain individually through the established point mutations and investigated the effect on substrate activation of the affected and its neighboring module by comparing their k_{cat} values. All mutants were purified with good yields (**Figure S 9E**, **F**).

In TycC8-9ΔTE, module 8 does not show activity towards its native substrate valine upon A domain mutation (**Figure 31F**), which is the same effect observed for the di-domain construct $TycC8^{A*}$ (**Figure 30C**). Inactivation of the PCP domain in module 8 lead to a 31% decrease of the activity ($p<0.005$) towards valine(**Figure 31F**), which is a higher drop in activity than in $TycC8^{PCP*}$ (15% decrease, $p=0.016$, **Figure 30C**). Module 9 exhibited a differential effect of A domain mutation depending on whether it was embedded in this dimodular versus the di-domain context: The A domain mutation of module 9 in TycC8-9ΔTE only mildly affected ornithine activation, even in the absence of valine (**Figure 31F**), while in $TycC9^{A*}$ activity was completely abolished (**Figure 30D**). For the PCP domain mutation, the reactions were in line again. In both, TycC8-9^{PCP*} (**Figure 31F**) and $TycC9^{PCP*}$ (**Figure 30D**) the A domain activity for Orn of that module was decreased. To test the hypothesis of ornithine auto release increasing the reaction speed of the upstream A domain, I replaced it by lysine. As expected, the 9^{PCP*} mutant of TycC8-9ΔTE is not affected by the mutation when fed with lysine, but rather maintains the same activity level towards Lys as the "wt" (**Figure S 11A**). This activity level, in turn, is much higher than the activity towards the native substrate Orn, in the presence and absence of Val (compare **Figure S 11A** and **Figure 31F**). Taken together, these data suggest that the presence of upstream elements changes the specificity of the A domain in module 9.

The native TycC9-10+TE dimodular construct was also mutated within each A and PCP domain separately and the effects on k_{cat} values towards various substrates were recorded (**Figure 31G**). In this case, module 9 took the role of an initiation module and both mutations of this module, 9^{A*} and 9^{PCP*}, resulted in loss of continuous A domain activity (**Figure 31G**). This observation is in agreement with what I observed for the single TycC9 module (**Figure 30D**). The effects of mutations of module 10 cannot be assessed individually for this construct, since its activity depends on the presence of Orn (**Figure 31B**). The result of the thiolation assay, however, shows that the module 10^{PCP*} mutant is incapable of leucine thiolation (**Figure 31E**). By comparing the activation rates of "wt", module 9 and module 10 mutants of TycC9-10+TE when both substrates are added to the reaction, it appears that both module 10 mutants do affect A domain activity, though not that of the upstream module (**Figure 31G**).

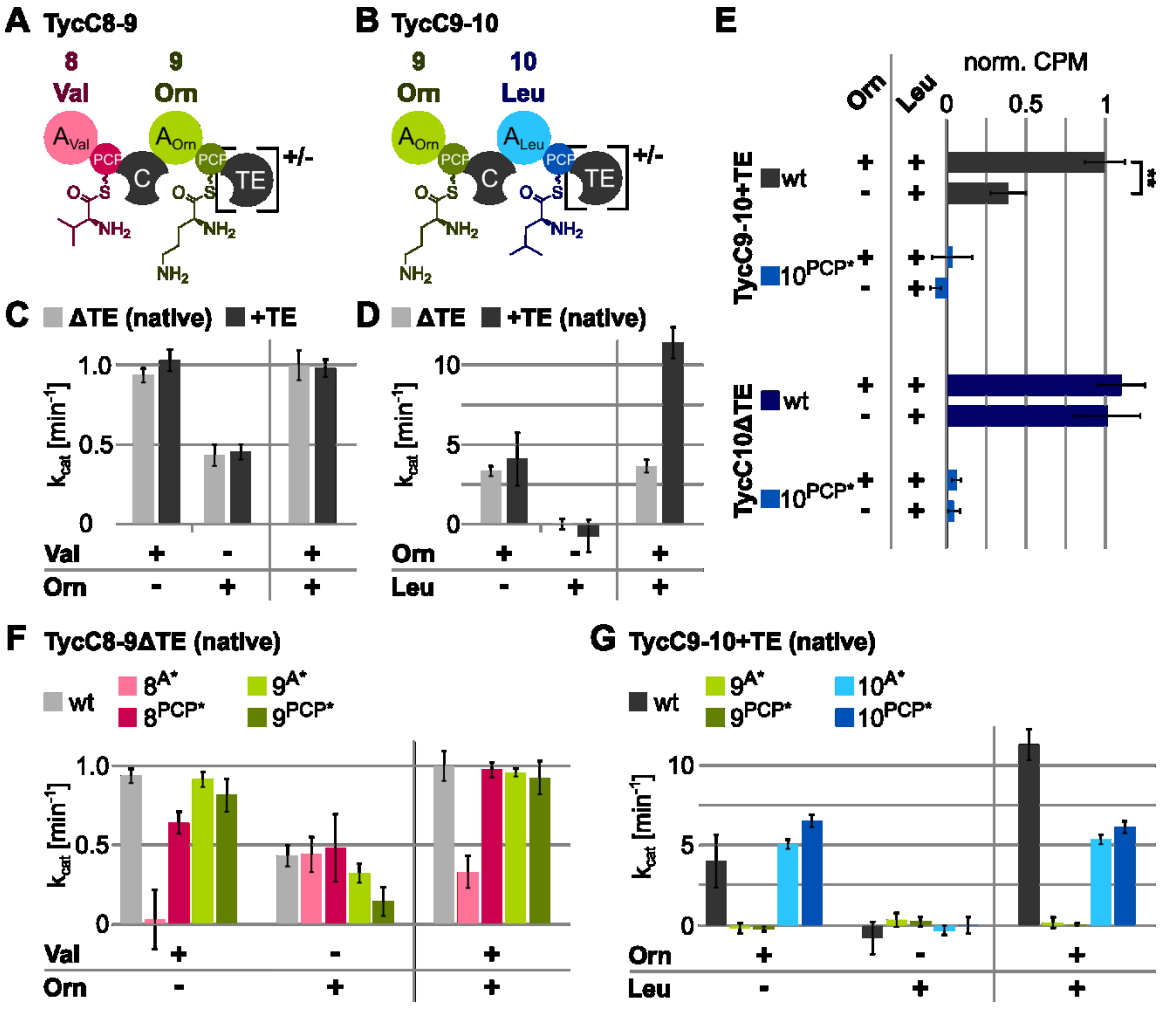

Figure 31 | Activity of A domains within "wt" or mutated dimodular constructs of TycC8-9 and TycC9-10, with or without TE domain. (A, B) Schematic depiction of constructs with amino acids attached to the PPant arms. Brackets highlight the TE domain that is either present (+) or absent (-). Colors as in **Figure 1**, A. **(C, D, F, G)** Graphs showing k_{cat} values for the indicated constructs in the presence of the indicated amino acids. Amino acids were added at 1 mM, enzymes at 0.5 µM, except for TycC9-10ΔTE, which was added at 0.25 µM due to a lower yield in the purification. **(E)** Thiolation assay with [^{3}H]-labelled L-Leu in the presence or absence of Orn. Counts per minute (CPM) were normalized to the value for TycC9-10+TE wt. ** $p<0.01$. Data represent the mean ± standard deviation, SD of three independent experiments.

4.4.4 *Analysis of A domains embedded in the tri-modular construct TycC8-9-10+TE*

Finally, I created a tri-modular construct comprising the last three modules of the tyrocidine synthetase, TycC8-9-10+TE (**Figure 32A**) and the respective A and PCP mutants of all modules. Despite their large size of 326 kDa, I could purify them from BAP1 cells at a good yield (**Figure 32B**). I again analyzed the "wt" and all mutant constructs adding Val, Orn and Leu, either individually or in combination and calculating the respective k_{cat} values. For the "wt", I detected that all modules were capable of repeatedly activating their respective substrate when it was the only one present in the reaction (**Figure 32C**). For module 10, this is contradicting to its performance in the di-module TycC9-10 (**Figure 31D**). The "wt" tri-modular construct also accepted lysine as a substrate instead of ornithine (**Figure S 11C**), also at a higher activation rate than the original ornithine and without any effect of the module 9^{PCP*} mutant. From the overall higher activation rate when all amino acids were added, I assumed that the tripeptide was formed *in vitro*, which was confirmed by LC-MS (**Figure S 12B**).

Analysis of the mutants upon feeding the different amino acids in different combinations illustrates the increasing complexity with increasing NRPS length. Also, drawing clear conclusions becomes more difficult. What I could observe is that embedded in the tri-modular construct, all A domains were still able to repeatedly activate their respective substrate despite carrying the mutation (**Figure 32C-E**) that completely abolished A domain activity in the initiation part of the dimodular constructs (**Figure 31F**, **G**) and in the minimal constructs (**Figure 30C-E**). I speculated whether the presence of an increasing number of modules facilitated a more native fold and thus rescued the A domain mutation by still binding ATP. However, no tripeptide was formed by the TycC8-$9^{A*/PCP*}$-10+TE mutants as the LC-MS measurement confirmed (not shown). In addition, the mutation of module 9^{PCP*} had no effect on Orn (**Figure 32E**) and Lys activation (**Figure S 11C**) as opposed to the di-domain TycC9 construct (TycC9^{PCP*}, **Figure 30D**) and the dimodular TycC8-9 and TycC9-10 constructs (TycC8-9^{PCP*} and TycC9^{PCP*}-10, **Figure 31F**, **G**).

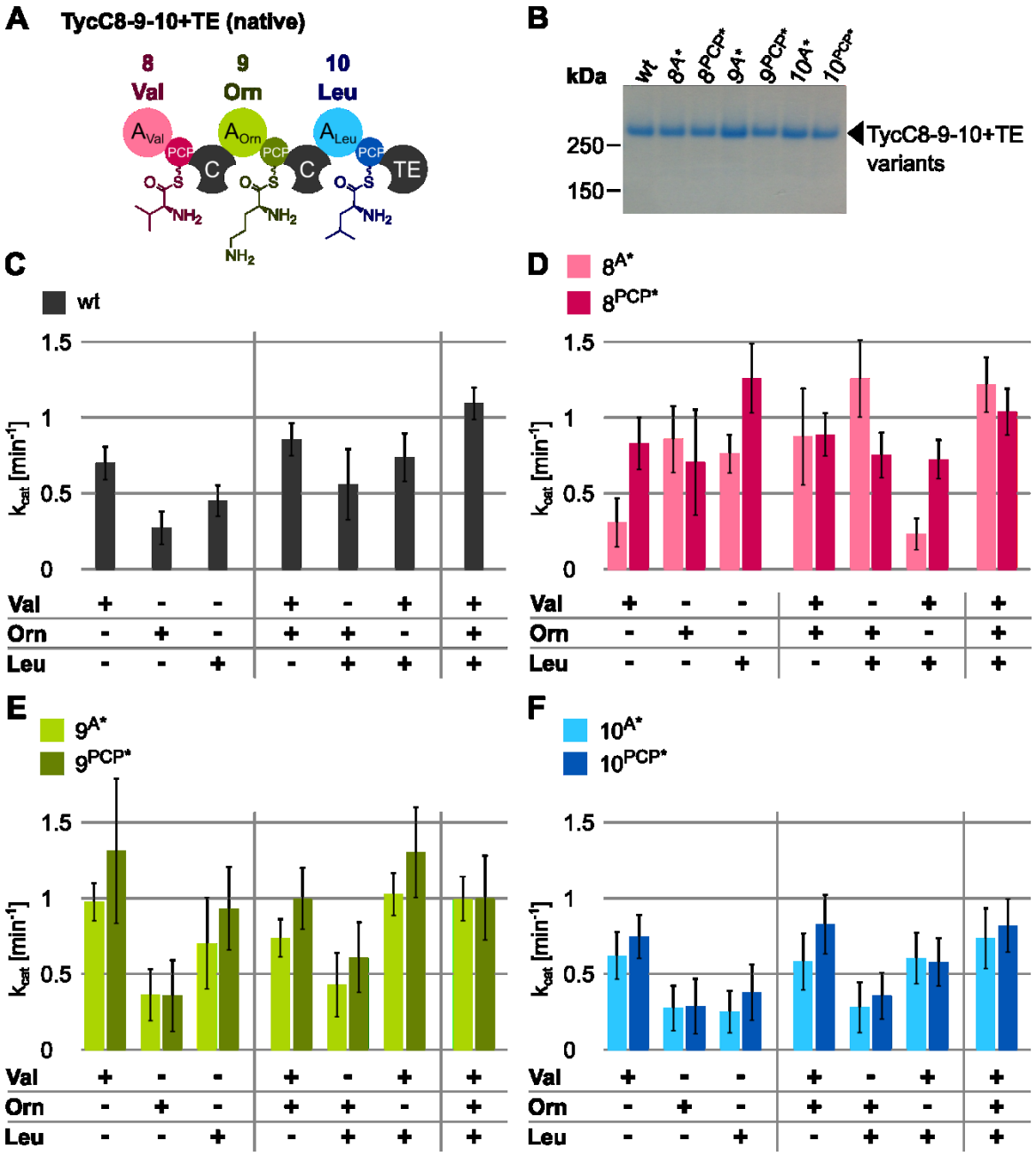

Figure 32 | Activity of A domains within the "wt" or mutated tri-modular construct TycC8-9-10+TE. (A) Schematic depiction of the construct with amino acids attached to the PPant arms. Colors as in Figure 1, A. **(B)** Coomassie-stained SDS-gel showing the indicated purified proteins. **(C-F)** Bar graphs showing k_{cat} values for the indicated constructs in the presence of the indicated amino acids, calculated from the online PP_i release assay after subtraction of the background (=no substrate) value. Amino acids were added at 1 mM, enzymes at 0.5 µM. Data represent the mean (± standard deviation, SD) of three independent experiments.

5. Discussion and Outlook

5.1 Summary of achievements of this work and conclusions for future (engineering) approaches

Through quantification of *in vivo* pigment production and *in vitro* PP_i release, I analyzed a variety of different mutants of the two homologous indigoidine synthetases, IndC and BpsA, and conclude that the proposed biosynthesis mechanism is doubtful. Most notably, I propose that the TE domain plays a crucial role, not acknowledged in the current model, since pigment production *in vivo* and substrate activation *in vitro* are greatly affected by point mutations within this domain, in particular when mutating the conserved serine. Additional experiments, such as thiolation assays, whole enzyme MS detecting attached intermediate substrates and the crystal or cryo-EM structure of an indigoidine synthetase, could provide the missing information to finally elucidate the detailed biosynthesis mechanism.

I continued to test the feasibility of using an indigoidine synthetase as a tag to monitor upstream NRPS module engineering with the premise that a blue pigment is exclusively formed if a module upstream of the indigoidine synthetase is present and functional. For the development of this system, I chose modules from the well-studied tyrocidine synthetase, namely modules 5 and 6, successively incorporating asparagine and glutamine. I screened various fusion sites between TycC6 and IndC, all located within the A_{Gln} domain of both NRPSs. From this screen, I identified a functional chimeric enzyme, which was capable of pigment production. I refer to it as TycC5-6:IndC “wt”. Upon impairing its upstream TycC5 module by a PCP domain mutation, this fusion enzyme almost completely lost its pigment formation ability. I changed the expression backbone of this promising construct and the indigoidine synthetase from IndC to BpsA and could thereby increase the pigment production *in vivo* by about 10 fold for the “wt”, while keeping it close to zero for the upstream PCP_{Asn} mutant. However, in this TycC5-6:BpsA construct, mutation of the upstream A_{Asn} domain did not impair pigment production. LC-MS analysis of the pigment formed by TycC5-6:BpsA *in vitro* revealed that the product is pure indigoidine without any asparagine attached. Taken together, these data suggest that engineered asparagine-indigoidine synthetases form the pigment only due to the activity of the TycC6:BpsA part and not due to the function of the fusion enzyme. Therefore, it is unclear why the N-terminal start site or the PCP domain mutation in the upstream TycC5 module can still have an influence on pigment production. Through these experiments it became evident yet again, how little is known about the cooperation and interaction of domains and modules within the context of larger NRPSs.

For this reason, I investigated the behavior of natively connected NRPS modules of TycC, namely modules 8, 9 and 10 incorporating valine, ornithine and leucine, respectively. I purified all three modules, either individually (TycC8, 9, 10), as dimodules (TycC8-9 or TycC9-10) or trimodules (TycC8-9-10), introduced the same mutations to the A and PCP domains as I did for TycC5 and 6 and analyzed the activation rates via PP_i release feeding different substrate combinations. In this more native context, I found that the individual modules (A-PCP domains) all activated their native substrates and the activation rates differed depending on the N-terminal start site. In the dimodular constructs, the upstream module of the “wt” enzyme always activated its substrate while the downstream module did (in TycC8-9) or did not (in TycC9-10) repeatedly activate its

substrate in the absence of the upstream module substrate. For TycC9-10, the presence of the TE domain seemed to enhance the substrate conversion rate in case both substrates were present. In the trimodular construct, on the other hand, the A domains of all modules activated their substrate irrespective of the presence of other substrates. The mutations in the different modules had different effects on surrounding modules, depending on the context they were in. Sometimes, but not always, these effects matched the effects of not feeding the substrate of the mutated module. From these data I infer that a general conclusion about the cooperation of NRPS modules in their native context cannot be drawn as the module behavior varies extensively depending on its molecular environment.

In summary, my results suggest that NRPS domains and modules of engineered as well as native NRPSs act differently depending on the context they are embedded in. Therefore, generalizations should be done with great caution. This finding also means that for every single engineering approach, the context might need to be evaluated anew, which would make a method for easy and high throughput screening of engineered NRPS function, as explored in this work, all the more valuable.

5.2 Indigoidine synthetases and the proposed roles of the Ox and TE domains in the biosynthesis process

Even though indigoidine and its synthetase have been used as a reporter system in various studies, the detailed biosynthesis mechanism remains elusive. In the currently proposed model, the order of events and the role of the TE domain are unclear.

The investigation of indigoidine synthetases *in vivo* showed that the effects of point mutations can be assessed by measuring the absorption at 600 nm of the expression culture, thereby estimating product formation. Thus, indigoidine is a feasible reporter molecule for *in vivo* screening of NRPS modifications. This approach revealed differences in effects of point mutations between two homologous indigoidine synthetases, IndC and BpsA. In addition, an *in vitro* PP_i release assay provided further information about the indigoidine production mechanism underlining the role of the TE domain and providing insights into why some of the mutants failed to produce indigoidine. Lastly, the attempts of generating an indigoidine-tagged amino acid by feeding BpsA with dipeptides containing a C-terminal glutamine led to the formation of pure indigoidine and the individual AA, suggesting that splitting of the dipeptide and concurrent or subsequent indigoidine production occurred.

Indigoidine-tagged amino acids could not be detected after feeding BpsA with dipeptides, possibly due to the action of the oxidation domain

From *in vitro* substrate activation of the different BpsA mutants, I tried to infer the order of events during indigoidine biosynthesis. As expected, an A domain mutant did not activate glutamine anymore. An oxidation domain mutant in contrast did still activate glutamine, though at a lower speed. This observation indicates that the substrate is activated by the A domain prior to oxidation. It is assumed that thiolation happens immediately upon activation and thus also prior to oxidation. It would be interesting to conduct further experiments, such as thiolation assays or PP_i release assays of the Ox/PCP double mutant to elucidate this.

Feeding dipeptides to wt BpsA revealed that the BpsA A domain is capable of activating a dipeptide even though it is natively intended to accept a single amino acid. This finding is

consistent with data from the Ackerley group[317]. After incubation with the dipeptides glycine-glutamine (GQ) and alanine-glutamine (AQ) for 72 hours (3 days), neither they[314] nor I detected glycine or alanine attached to indigoidine. My hypothesis is that the oxidation domain might play a role in detaching the upstream amino acid from glutamine during the oxidation process, thereby producing pure indigoidine. This hypothesis is based on the observation that for wt IndC, pure indigoidine was only detectable after 3 days of incubation with the dipeptide. In addition, the BpsA oxidation domain mutant was still capable of activating glutamine at a time scale of minutes, not days but did not activate the dipeptide anymore. This observation contradicts the proposed order of events, because oxidation would need to happen before the activation. Further experiments as described above could clarify this ambiguity.

The TE domain of IndC/BpsA differs from the TE domain of other NRPSs and plays a crucial role in indigoidine biosynthesis

Most NRPSs feature a C-terminal thioesterase domain to release the final peptide product via a variety of release mechanisms, e.g. a linear peptide via hydrolysis or a cyclic peptide via intramolecular nucleophilic attack. In most NRPSs, the TE domain features a catalytic triad consisting of the amino acids serine (nucleophile), histidine (base) and aspartate (acid) common among many hydrolases and transferases. Indigoidine synthetases also feature a C-terminal TE domain, however, its contribution to indigoidine production is rather neglected in the current model of indigoidine biosynthesis.

In contrast to other NRPSs, indigoidine synthetases only feature a potentially catalytic dyad consisting of serine and histidine. According to amino acid sequence alignments, the aspartate of other TE domains has been replaced by a small, neutral alanine in indigoidine synthetases. To establish the catalytic role of the conserved serine, I mutated this residue to alanine. This mutation prevented indigoidine formation *in vivo* and greatly reduced substrate activation *in vitro*. Since this conserved serine is not part of any predicted secondary structure within the TE domain, I conclude that it likely has a catalytic role in indigoidine biosynthesis. For a fatty acyl-thioester hydrolase TE domain, it has been shown that the mutation of the conserved serine to a cysteine does not decrease the substrate activation rate k_{cat} (wt = 0.11 s^{-1} vs. S101C = 0.1 s^{-1})[318] presumably because cysteine can also serve as a nucleophile. In contrast, S to C mutation of indigoidine synthetase did have a great impact on substrate activation and decreased compound production by the same level as the S to A loss of function mutation. In a different study, a double S to C and H to R mutant slowed down the reaction kinetics of the polyketide Tautomycetin thioesterase very much to effectively label the TE domain with the product intermediate, whereas the single mutants (S → C or H → R) failed to trap the intermediate product at the TE domain[319]. Single S to C mutation of indigoidine synthetases also failed to trap the intermediate product at the cysteine as assessed by LC-MS by Thomas Ruppert. During the lengthy sample preparation, involving sample digestion into smaller peptides that can be loaded onto the LC-MS, all substrates detached and were not found on the PPant arm or the TE domain. Thus, to elucidate the individual steps and order of events leading to indigoidine synthesis, the different BpsA mutants can serve as a basis to perform a combination of thiolation assays, including radio-labelled dipeptide substrates, and whole enzyme LC-MS to find the status of intermediate products still attached to the enzyme. In further experiments, both components of the catalytic dyad (S and H) could be mutated to test if the intermediate substrate (oxidized glutamine) is loaded onto the TE domain during biosynthesis.

Reconstitution of the catalytic triad of the TE domain in IndC and BpsA through an alanine (A) to aspartate (D) mutation had ambiguous effects: for IndC, the pigment production was completely abrogated *in vivo*, while the BpsA mutant was still able to form indigoidine at a lower level. Replacing this alanine with another small, unpolar amino acid (glycine) in BpsA did not much affect pigment production. Thus, the third position of the triad might also serve a specific purpose in indigoidine synthetases, but in contrast to other NRPSs rather through its absence than its presence.

Indigoidine formation rates inferred from the absorption at 600 nm differ from glutamine activation rates estimated by PP_i release

In vitro indigoidine formation by BpsA measured at 600 nm (~ 6 indigoidine per minute, corresponding to about 12 activated glutamine molecules) is about three times slower than glutamine activation as measured in the PP_i release assay (about 36 molecules per minute) in the very same assay (**Figure 9**). These two values could deviate that much for different reasons: one is that the *in vitro* assay takes place in aqueous buffer, so some indigoidine might be in solution, but some of it might precipitate and thus not be available for proper absorption measurement. The extinction coefficient of indigoidine in buffer might be also different from DMF leading to a miscalculation of indigoidine levels. In addition, the adenylation step might not be the rate-limiting step in indigoidine formation and thus more substrate might be activated than actually thiolated and further modified. This is expensive for the cell since it consumes ATP, but in this way the assembly line is always ready to continue. In favor of the second scenario is the observation that for lower substrate concentrations, k_{cat} of indigoidine formation decreases too (~3 for 1 mM Q, 2.6 for 0.5 mM Q and 1.9 for 0.1 mM Q, as calculated from **Figure 9A**).

BpsA mutants could serve as the basis for a crystal or cryo-EM structure

A deeper insight into the indigoidine biosynthesis mechanism would be provided by a crystal structure of this enzyme. Upon its discovery in 2007, Takahashi *et al.* aimed to provide the crystal structure of BpsA[202], but have not yet succeeded. Owen *et al.*[115] sought to extract a homology model of the BpsA PCP-TE di-domain from the PCP-TE di-domain of EntF (PDB: 2ROQ[165]). However, these two enzymes share less than 24% amino acid identity in the corresponding domains, which is well below the threshold needed to assume structural homology[320,321]. One reason for the difficulty to resolve the crystal structure of BpsA, is the high flexibility of this enzyme, which prevents the formation of crystals unless they are trapped in a certain, rigid state by a mutation or a functional inhibitor. Thus, the BpsA mutants I created, which showed only very low activation rates in the PP_i release assay, could serve as a basis for solving the crystal structures of BpsA trapped in different states. Alternatively, cryo electron microscopy could be used to study the structure of BpsA. An advantage of this method is that it would potenitally allow the analysis of the enzyme in different states[76] in presence of the substrate.

5.3 Exploration of an asparagine-indigoidine fusion synthetase constructed from TycC5-6 and IndC/BpsA

An indigoidine synthetase was successfully reconstituted from the N-terminal part of the TycC6 A_{Gln} domain followed by the Ox, C-terminal A_{Gln}, PCP and TE domain of IndC (**Figure 13**). This shows that the TycC6 A domain is capable of cooperating with the other domains of IndC to produce indigoidine. Further, the TycC6 A domain lies within a native elongation module that can be turned into an initiation module using different start sites. Some of these non-native start sites of the TycC6 A domain do not comprise parts of the upstream TycC6 C domain, which has been proposed to be a pre-requisite to turn an elongation into an initiation A domain[80,81,266]. Thus, to my knowledge, this is the first example of a successful elongation to initiation A domain transition functional *in vivo*, that does not feature a native C-A domain interface.

Through screening different combinations of fusion sites between the TycC6 and indigoidine synthetase A_{Gln} domain, alternative TycC5 start sites, expression vectors and different indigodine synthetases, I identified a TycC5-6:BpsA construct, which was capable of producing a well visible amount of pigment while keeping it reasonably low upon PCP_{Asn} mutation (**Figure 14**). Assuming that the naturally contiguous modules TycC5 and 6 sucessfully produce the dipeptide and TycC6 meanwhile cooperates with BpsA to form the pigment, I suspected that the tagging can work to monitor the success of future NRPS engineering efforts. However, many observations point in the opposite direction: inactivation of the upstream TycC5 module via A_{Asn} domain mutation did not lead to a decrease in pigment formation in comparison to “wt” TycC5-6:BpsA, indicating that the downstream indigoidine producing module is not necessarily dependent on upstream module functionality. Furthermore, the *in vitro* analysis revealed that the downstream module, specific for glutamine, can activate its substrate at high rates in the absence of asparagine, the substrate of the upstream module, while asparagine alone is activated at very low rates. The addition of both, asparagine and glutamine, to the reaction lead to a total k_{cat} similar to when only glutamine was added, indicating that glutamine is converted to indigoidine irrespective of the upstream module input. Pigment production rates, monitored in parallel to the PP_i release, confirmed this hypothesis. Both, the Ox as well as the TE domain, seem to play a crucial role in keeping the indigoidine assembly line running. This is also the case for the TycC5-6:BpsA fusion (**Figure 16**) since upon point mutations in either one of these two domains, the k_{cat} value for glutamine is largely reduced and no pigment is formed anymore. At the same time, these mutations did not lead to a differential activation of asparagine by the upstream module.

Unfortunately, all attempts to detect the indigoidine-tagged asparagine as the predicted product of “wt” TycC5-6:BpsA via ESI-MS (together with Heiko Rudy from the IPMB, Heidelberg) or the intermediate product still attached to the PPant arm of the second module of TycC5-6:BpsA Ox-K600E or TE-S1103A mutants (together with Thomas Ruppert from the ZMBH, Heidelberg) were unsuccesfull. As we only detected the unloaded PPant arm on the PCP domain (not shown), possibly the digest of TycC5-6:BpsA prior to loading the fragments onto the LC-MS took too long causing hydrolyzation of all intermediate substrates. Since in other publications, this approach led to the identification of bound substrates, further development of the experimental setup could help clarifying this issue[118–122]. Finally, the blue pigment formed by the TycC5-6:BpsA fusion construct *in vitro* was identified to be the same as the product of BpsA, i.e. pure indigoidine (**Figure 18**, in collaboration with Aubry Miller, Cancer Drug Development, DKFZ,

Heidelberg). The attachment of an additional amino acid onto indigoidine should lead to a shift in the absorption maximum (as for the attachment of lauric acid, **Figure 5**), which was not observed. One possible explanation why asparagine is not attached to indigoidine, is that, after asparagine and glutamine are successfully fused together at one point of the reaction, the oxidation of glutamine might lead to the elimination of asparagine, resulting in the formation of pure indigoidine. In this scenario, the oxidation domain mutant might produce the colorless dipeptide Asn-Gln, since it does not split off the Asn in the oxidation process. Alternatively, the oxidation of Gln could lead to the rejection by the C domain, which might be only able to process unmodified Gln. However, also the dipeptide hypothetically formed by the TycC5-6:BpsA oxidation domain mutant was not identified in LC-MS from the filtered *in vitro* reaction supernatant. Potentially, an alkaline[34] or cysteamine-HCl[322] treatment to release the intermediate product from the PCP domain and a specialized column optimized for separation of polar substances could still lead to the discovery of Asn-Gln formed by the hybrid enzyme oxidation domain mutant.

Another proof of principle experiment would be to engineer a fusion synthetase where more than one NRPS module is placed upstream of the TycC6:BpsA tagging module. It could then be assessed whether the polypeptide is formed along with pure indigoidine. With the recently discovered, more general fusion sites within C domains[266], this approach becomes feasible.

Finally, I was surprised to see such different effects of the inactivation of module TycC5 on the downstream TycC6:BpsA part, depending on the type of inactivation (A_{Asn} domain mutant versus PCP_{Asn} domain mutant) and the type of analysis (*in vivo* expression and absorption at 600 nm versus *in vitro* PP_i release assay). The discrepancy between *in vivo* and *in vitro* analysis could have a variety of reasons. Enzyme folding might be suboptimal *in vitro*, influenced by the purification process and the storage buffer. Another stabilizing effect might be caused by the presence of other proteins *in vivo*, the so-called crowding effect. I also tried to mimic this effect *in vitro* through the addition of 30% Ficoll 70, which has been shown to decrease the formation of linear side products of the enterobactin synthetase reconstituted *in vitro*[323]. However, this additive did not have an influence on the k_{cat} values of TycC5-6:BpsA (not shown). Another condition that might cause the deviation between absorption at 600 nm of the bacterial expression culture and the *in vitro* reaction is the concentration of reactants. It is very hard to estimate or measure the concentration of e.g. amino acids in bacterial cells. On the other hand, *in vitro* the local concentration of amino acids is set up precisely and is at 1 mM, likely higher than *in vivo*, possibly influencing the reaction of the enzyme. The reaction product also accumulates *in vitro*, which in the case of NRPSs can have a product inhibition effect, e.g. for higher concentrations of PP_i[64]. In the used coupled PP_i release assay, however, the PP_i is also converted and should not cause any problem. The conditions of *in vitro* assays need to be defined carefully with regards to the buffer system, pH and reactant concentrations, to ensure enzyme viability. At the same time, these defined conditions enable the comparison of different enzymes and mutants. Thus, I decided to investigate the same mutations and effects of selective substrate feeding in a more native multi-modular construct to find out whether they do wait for the upstream substrate activation *in vitro,* also in the termination module. These questions were addressed in section 4.4 and will be discussed in section 5.5.

5.4 Other NRP-indigoidine synthetase fusions – an opportunity for the production of indigoidine-tagged NRPs

I undertook several different approaches to still produce an indigoidine-tagged AA. They included changing the post A10 motif within the TycC:BpsA fusions, randomly mutagenizing the hybrid enzyme, externally supplementing the oxidation domain to Ox-K600E mutants or inserting the Ox domain into the TycC6 module, externally supplementing the TE domain to delta TE constructs as well as creating fusions between IndC/BpsA and other known NRPSs modules incorporating glutamine. Unfortunately, none of these approaches lead to substantial progress in the production of an indigoidine-tagged AA.

A putative natural dodecylindigoidine synthetase (dIndS), however, demonstrates that it is indeed possible for an enzyme to fuse an upstream module substrate (here a fatty acid) to the downstream module product indigoidine, though apparently via a different mechanism than I envisioned for the artificial AA-indigoidine fusion synthetase. Thus, in this part of the discussion, I will focus on the initial experiments done to elucidate the functional mechanism of the putative dIndS and how it could be validated for further use in engineering.

Possible reasons for the problems of investigating dIndS and proposed further experiments to elucidate its function

The pigment N,N-dodecylindigoidine produced in *Shewanella violacea* proofs that a fatty acid (and thus maybe also amino acids or peptides) can be tagged with indigoidine. Furthermore, it shows that if another molecule is attached via a peptide bond to the original N-terminus of glutamine, which corresponds to the free primary amino group of indigoidine, this molecule still has an absorption maximum around 600 nm, here 635 nm, and is thus visible by the naked eye. This absorption maximum could differ, if other molecules, e.g. amino acids instead of fatty acids, are fused to indigoidine. For these reasons, it is of interest to understand how N,N-dodecylindigoidine is formed and which NRPS and other enzymes or co-factors are responsible for its production. Nonetheless, Konrad Herbst's and my own attempts to experimentally validate and engineer the putative dIndS to tag amino acids with indigoidine were unsuccessful. In the following, I will speculate why and present some ideas what to consider for future investigations.

In the initial *in silico* identification of the putative dodecylindigoidine synthetase (dIndS) gene, six other proteins exclusively shared between the two pigment producing organisms *S. violacea* and *Rheinmera baltica* were identified[221]. Four of them are part of a gene cluster related to polysaccharide biosynthesis and export[324], while the other two (accession number WP _013051036.1 and WP_041420101.1) are characterized by domains of unknown function (DUF1772 and DUF1302, respectively) according to the conserved domain database[222]. Additionally, it is known that in *S. violacea* the pigment is only produced on solid medium (e.g. supplemented MB agar plates) and not in liquid culture[217]. This could be due to an external trigger needed to activate an internal pathway necessary to initiate pigment production. These associated proteins and pathways might be concertedly responsible for pigment production and make an easier target for gene disruption than just one sole gene.

Likely for that reason, the random transposase-mediated integration of the R6K/kanamycin^r cassette into the *S. violacea* genome yielded a high number of non-pigment producing colonies

(**Figure 25B**), at least four of which did not have an insertion in the putative dIndS gene as confirmed by PCR. Thus, it would be interesting to perform a sequencing reaction from the kanamycin resistance gene outwards to map the integration site back to the genome as originally planned[305] (**Figure 25A**, b). Unfortunately, there were no bacterial colonies after digest and re-ligation of the *S. violacea* genomic DNA of the non-pigment producing clones (scheme in **Figure 25**) and thus no sequencing could be done from the extracted plasmids containing the kan^r interrupted DNA fraction. A technical problem might have compromised the experimental outcome as after BamHI digest, the genomic fragments were too big (10-30 kbp on average, **Figure S 8**) for efficient transformation into *E. coli.* This experimental approach should be reassessed adding a second restriction enzyme to the digest reaction.

A different approach would be to perform a targeted gene disruption of the putative dodecylindigoidine synthetase gene to verify that dIndS is one of the crucial enzymes in dodecylindigoidine production. An advantage of that approach is that all other components of the pathway or other factors would be left untouched and that is why the lack of pigment production could really be attributed to the inactivation of the single gene of interest.

Judging from the bioinformatical predictions, I was nevertheless quite convinced that we had identified the correct NRPS gene responsible for N,N-dodecylindigoidine production. Therefore, I shifted to a bottom-up approach to prove its function. I amplified the gene from *Shewanella violacea*, inserted it into an *E. coli* compatible backbone (pCDFDuet) and expressed it in BAP1 cells. The enzyme was successfully expressed without any further codon optimization (**Figure S 8C**). However, it did not show any pigment production on solid (LB agar plate, not shown) or liquid medium (M9 minimal medium, **Figure 26C**), neither when co-expressed with the PPtase from *Bacillus subtilis* or with the endogenous PPtase (WP_013050488.1) from *S. violacea*.

The replacement of the first A domain of the native dIndS specific for dodecanoic (= lauric) acid by A domains of the tyrocidine synthetase (TycC) specific for valine, ornithine or leucine, likewise did not lead to any pigmentation of the liquid expression culture (**Figure 26C**) despite proper enzyme expression. Therefore, I conclude that additional enzymes of the dodecylindigoidine production pathway need to be present in order to form the final product. An external trigger needed in native *S. violacea* should not be crucial when being expressed under the IPTG inducible promotor of the pCDFDuet vector. However, an external oxidation or condensation domain could be missing to successfully produce dodecylindigoidine in *E. coli*. Putative proteins found in a BLAST search of IndC and BpsA oxidation domain sequence within the *S. violacea DSS12* genome include BAJ02469.1, BAJ01856.1, BAJ03901.1 and BAJ00582.1. They could be tested for co-expression with the native dIndS in *E. coli* and other host organisms more closely related to the marine bacterium *Shewanella*. Once dodecylindigoidine is successfully produced, engineering of the first A domain and addition of further upstream NRPS modules could be tackled.

Overall, experimental identification of the enzyme(s) required for dodecylindigoidine production and its/their successful re-engineering presents a great challenge but worthwhile to be continued.

5.5 A domains activate their substrate dependent on the context

Since during analysis of the TycC5-6:indigoidine synthetase fusions it became evident that A domain activity differs depending on which domains are around and in which status they are (e.g. absence of substrate or mutation), I wanted to investigate whether this is also the case for A domains of modules naturally belonging together. For this experiment I chose the last three modules of TycC from the tyrocidine synthetase and analyzed their A domains in the presence and absence of the native neighboring domains as wt and mutants and when feeding different substrate combinations (section 4.4). Here I rely almost entirely on the PP_i release assay, which is why I will discuss its advantages and drawbacks prior to discussing the data obtained with it.

Advantages and drawbacks of the coupled online PP_i release assay in comparison to the ATP/PP_i exchange assay

The coupled PP_i release assay was recently adapted to be used for analysis of NRPS A domains[70]. This assay is suitable to measure either one distinct round of activation in a photospectrometer where substrates can be added while the absorption measurement continues or, as I used it, for higher throughput measurement of repeated activation cycles in a plate-reader. Since we measure the release of PP_i (subsequently converted by an enzymatic cascade resulting in the oxidation of 2 NADH to 2 NAD^+), the adenylation reaction and the transition of the A domain to the thiolation state need to happen prior to this release. For this reason, a point of criticism of this type of assays is that, if the aminoacyl-AMP cannot be thiolated or further processed, e.g. because of the absence of the downstream module or the mutation of the downstream PCP domain, no repeated activation and thus PP_i release can be detected.Consequently, for these impaired enzymes, what is measured in the PP_i release assay is the frequency of loss of the aminoacyl-AMP from the A domain active site[325] or the hydrolysis from the upstream PCP domain, both prerequisites to clear the way for another cycle of substrate activation.

In contrast, the most commonly used assay to study NRPS A domains, the ATP/PP_i exchange assay[326], measures the reverse reaction of AMP+[32]PP_i after adenylation has taken place and thus does not require a release of the aminoacyl-AMP. Therefore, the ATP/PP_i exchange assay cannot provide information about the concerted action of A domains in intact NRPS systems, which the online PP_i assay is able to provide to some degree (e.g. the effect of the TE domain speeding up substrate activation of the upstream modules). Of course, also with the PP_i release assay it is not possible to determine which A domain of a multimodular NRPS executed the activation; however, one can draw conclusions from selective substrate feeding. Further disadvantages of the ATP/PP_i exchange assay are that it requires radioactive PP_i and that the reaction setup cannot be used to compare the substrate activation rate with product formation. On the other hand, using the PP_i release assay, one could measure the product formation rate using an LC-MS from the very same reaction and compare it to the calculated A domain activation rates, which could demonstrate whether these two values match.

In studies comparing (other) PP_i release assays with the ATP/PP_i exchange assay, the estimated k_{cat} values do differ significantly, often resulting in higher values for the ATP/PP_i exchange assay[325,327]. In one example, the ATP/PP_i exchange assay did not exhibit any substrate activation *in vitro* for the fatty acid adenylating enzyme FadD28, while in a PP_i release assay, a k_{cat} value for the respective module could be determined[328].

In conclusion, no assay is perfect and the best approach is to combine several methods when investigating a biological process. For the activity of A domains this would include a combination of the PP_i release with product formation and thiolation assays, as done for critical samples in section 4.4, the results of which will be discussed next.

Context dependent activity of A domains of the tyrocidine synthetase TycC

A more detailed biochemical characterization of NRPSs, in particular within a larger cluster of NRPS modules naturally belonging together, is needed for a better understanding of the processes during NRP formation. To this end, I measured the A domain activity for different constructs combined from the last three modules of TycC with different substrate combinations in otherwise identical conditions. Moreover, I introduced the A (D to A) and PCP (S to A) domain mutations as described before, preventing ATP binding and conversion to a holo-enzyme, respectively, and assessed their influence on A domain activity.

Mutations of the ATP binding site of A domains led to the loss of adenylation ability in single modules (A-PCP di-domains, **Figure 30**) and the initiation module of both dimodular constructs (**Figure 31**). In contrast, the A domains of the downstream modules of the di- and trimodular constructs were still capable of activating their substrates (**Figure 32**). One could speculate that upstream neighboring modules stabilize the A domain conformationally to still bind ATP since the aspartic acid we mutated is not catalytically active. However, the trimodular construct TycC8-9^{A*}-10+TE, where the A domain of module 9 was mutated, did not yield a tripeptide product as confirmed via LC-MS product formation assay (not shown). Another explanation why I detected adenylation activity for mutated A domains in elongation modules is that a different A domain activates their non-native substrate. However, this explanation is also rather unlikely because for the di-domains I showed that each A domain primarily activates its cognate amino acid. Furthermore, in that case the initial A domains would also be affected.

Mutation of the PCP domain also has different consequences on A domain activity depending on which context they are in. Two out of the three A-PCP di-domains tested could still activate their substrates upon PCP domain mutation, as has been described before for other partial NRPSs without PPant arm[47,316]. The di-domain specific for ornithine fails to activate its substrate upon PCP domain mutation. Possibly this is because ornithine can auto-cyclize and thereby release itself quickly from the PPant arm allowing the enzyme to undergo another round of activation. By feeding an alternative substrate, lysine, I tested this hypothesis, but could not finally resolve the question. The same PCP domain mutation in constructs consisting of two modules impairs the adenylation activity of the cognate A domains (**Figure 31**). Again, there are differences in the quantity of impairment depending on the context, since the A domain of module 8 is less affected by the PCP domain mutation than the A domain of module 9 (32 % decrease versus 67 % decrease, respectively). In addition, I investigated module 9, specific for Orn, when acting as initiation (in TycC9-10) or termination module (in TycC8-9). The effect of the PCP domain mutation of this same module 9 is much more severe when it acts as an initiation module (**Figure 31F** versus **G**). Notably, the same mutation in the TycC8-9^{PCP*}-10+TE trimodular construct does not lead to a decrease in adenylation by its cognate A domain in comparison to the “wt” (**Figure 32**).

Finally, whether a module is repeatedly activating its substrate seems to depend on the state of the upstream module(s) only in rare cases. Most of the time, downstream modules do not seem

to wait for an incoming substrate from their upstream module to activate their own substrate, which is what we observe for the trimodular enzyme. For native NRPSs *in vivo*, this continuous substrate activation would be expensive for the producer organism as it would also continuously consume ATP. Thus, it is likely, that *in vivo*, each A domain activates and thiolates its substrate irrespective of the status of the surrounding modules and is then "locked" in a conformation awaiting the incoming donor amino acid/peptide. With this mechanism, the assembly line is constantly ready for the instant production of the NRP, once all substrates are available. The dimodular construct TycC9-10 represents an exception, as it does not repeatedly activate the substrate of the downstream module 10, leucine, in the absence of the upstream module substrate ornithine, or in case of any of the two mutations (A or PCP domain) of the upstream module (**Figure 31D**, **G**). Possibly, module 10 in this case undergoes only two cycles of adenylation (one to load the PCP domain and another one where the activated substrate remains in the active site pocket A domain) unless the upstream C domain is presented with an acceptor substrate. At least one round of successful thiolation was confirmed by the thiolation assay (**Figure 31E**). A similar observation was recently reported in a different study[70].

In conclusion, I believe that generalizations for domain and modules of NRPSs should be done with great caution, as their behavior appears to be very much dependent on the context they are embedded into.

6. Supplementary information

6.1 Primers

6.1.1 *Sequencing primers*

For inserts into pTrc99 and pCDFDuet

#	name	sequence
42	ACYCDuetUP1	GGATCTCGACGCTCTCCCT
43	DuetDOWN1	GATTATGCGGCCGTGTACAA
44	DuetUP2	TTGTACACGGCCGCATAATC
45	T7 Terminator Primer	GCTAGTTATTGCTCAGCGG
46	pTrc99a Seq fwd	gtgtggaattgtgagcgg
226	FW ptrc99a Seq fwd	GCGCCGACATCATAACGGTTC
227	FW ptrc99a_rv	CGCTACTGCCGCCAGGC
278	226_pTrc99Seq_rv	cagaaccgttatgatgtcggcgc

Within IndC

101	IndC_Ad_rv	cccaatacccaccgaactc
116	IndC(K1014)_fwd	tggattgaacagacagac
117	IndC(C803)_rv	gcagaaataaaatgccgcc
203	IndC_ox-A_linker_fwd	attccacgctgtgtccctg
253	IndC_R524_rv	ggaagtaagaatgtagggcg

Within BpsA

299	BpsA-G362_Seq-fwd	GGGAGAGATTGGGGAACTG
300	BpsA-L709_Seq-fwd	GGACGTGGCTGACGAGG
329	BpsA-L709_Seq-rv	CCTCGTCAGCCACGTCC
334	BpsA_I777_rv	CTGGTACACGCTCTGGTTG

Within the Tyrocidine synthetase

202	TycA-I366_fwd	CGAAGGCGAATTGTGCATC
204	TycA_T841_fwd	ACCGAGCAGTTGTTGAAGCACG
88	TycC_M6-A_fwd	CAG ATG GTT CCG TCC TTG C
88rv	TycC_M6-A_rv	GCAAGGACGGAACCATCTG
90	TycC(A_{Gln})_fwd	GTGGAGGAGCAAGCAG
91	TycC(A_{Gln})_rv	CTGCTTGCTCCTCCAC
104	TycC_$C_{Asn-Gln}$_rv	GCTGATTCCGTTGCCTTC
143	(TycA_S28)TycC(I-5)fwd	ATTTTCGAGCTGATCGCGG
159	TycC(HxxPF+x-A_{Gln})_fwd	GGATCAGTGAGCAGCCAAC
160	TycC(HxxPF+x-A_{Gln})_rv	GTTGGCTGCTCACTGATCC
332	TycC-$C_{Asn-Gln}$-L140_rv	CTCGGCAAATTCACGCAGC
333	TycC-$C_{Asn-Gln}$-HxxxPF_fwd	CCAGCTCACATTCGCAACC
349	TycC($C_{Orn-Leu}$_F278)_fwd	CGTGAACACCTTGGCGATGC
350	TycC($C_{Orn-Leu}$ _F278)_rv	GCATCGCCAAGGTGTTCACG

369	TycC-A_{Val}(L239)_fwd	GTCGCCGAAGCACATCAACC
370	TycC-A_{Orn}(P296)_fwd	CGCTTCCGAACATGACGATG
371	TycC-A_{Leu}(Q234)_fwd	GTTGCGTTGCGTCACTTTGG
352	TycC(-41L_A_{Orn})_fwd	CTGCACAGCTTCCTCGCAACC
353	TycC(-41L_A_{Orn})_rv	GGTTGCGAGGAAGCTGTGCAG
397	TycC-A_{Orn}(E135)_fwd	GAGCACCGCAGCTATGCC
398	TycC-A_{Leu}(E61)_fwd	GAGCGCATTCAGTACCTGCTC

Within LchAA

138	LchAA_A_{Gln}_657_fwd	GCTGCCGATGATCTTTCAG
139	LchAA(A_{Gln}_F2)_fwd	TTCGAACAAAAAGCAGC

Within the dIndS

20	a Sviola 4553224 fwd	CGATGTAAACCACCAAGTGC
58	a Sviola 4552808 fwd	ATCTGGCTCCGCATACTGG
21	a Sviola 4555179 rv	CTGGGTGTGCGATTGTTAGC
22	b Sviola 4555179 fwd	GCTAACAATCGCACACCCAG
23	b Sviola 4557222 rv	GCACTAGCGAGTATGAGTTTCC
24	c Sviola 4557222 fwd	GGAAACTCATACTCGCTAGTGC
25	c Sviola 4559501 rv	CGCTGTTAGTCGTCATCAATGG
62	Tn903_Kan_fwd	CTCTTCCGACCATCAAGC
63	Tn903_Kan_rv	GCTTGATGGTCGGAAGAG

6.1.2 ***Primers for Gibson assembly***

To prepare empty vectors pTrc99, pCDFDuet(MCS1) and pCDFDuet(MCS2)

39	pTrc99_Orf_fwd	tgagagaagattttcagcctg
40	pTrc99_Orf_rv	catggtctgtttcctgtgtg
32	pCDFDuet_MCS1_rv	catggtatatctccttattaaagttaaac
33	pCDFDuet_MCS1_fwd	taatgcttaagtcgaacagaaag
26	pCDFDuet_MCS2_fwd	taattaacctaggctgctg
27	pCDFDuet_MCS2_rv	catatgtatatctccttcttatacttaac

IndC in pCDFDuet or pTrc99

93	IndC_MCS2_fwd	gttaagtataagaaggagatatacatatgttagaaaataatattacac
94	IndC_MCS2_rv	cagcagcctaggttaattagattattttctcaatctc
106	IndC(TE_GSG-His6):(MCS2)_rv50C	cagcagcctaggttaatta*gtggtgatggtgatgatg*ACCTGATCCgattatttctcaatctca
107	IndC(TE_GSG-His6):(MCS2)_rv63C	cagcagcctaggttaatta*gtggtgatggtgatgatg*ACCTGATCCgattatttctcaatctcagcaacac
95	IndC(MCS2-M27)_fwd	gttaagtataagaaggagatatacatatgcttgagagtcaact
109	IndC(MCS2-A21)	gtataagaaggagatatacatatggcaataacattgatgg
206	IndC(T1023)_MCS2_fwd	gaaggagatatacatatgacaatatcaagattaattttattg
207	IndC(His6-GSG)_MCS1_fwd	ctttaataaggagatataccatgcatcatcaccatcaccacGGATCAGGTttagaaaataatattacacaatgtgac
208	IndC(K1022)_MCS1_rv	ctttctgttcgacttaagcattatttagagtctgtctgttcaatcc

225	IndC(pTrc99)_fwd	cacacaggaaacagaccatgttagaaaataatattacacaatgtgactc

BpsA in pTrc99, BpsA TE domain in pCDFDuet(MCS1) and BpsA Ox domain in pCDFDuet(MCS1)

279	BpsA(pMM64)_pTrc99_fwd	cacacaggaaacagaccatgACTAGTACACTGCAGGAAACAAGC
280	BpsA(pMM64)_(GSG-His6)pTrc99_rv	ggctgaaaatcttctctca*gtggtgatggtgatgatg*ACCTGATCCAGCTTCCCCCAGCAGGTATC
339	BpsA(E1015)-pCDF1_fwd	ctttaataaggagatataccatgGAGTCTAGTCGCTTCGTCCG
340	BpsA(pMM64)_(GSG-His6)pCDF1_rv	ctttctgttcgacttaagcatta*gtggtgatggtgatgatg*ACCTGATCCAGCTTCCCCCAGCAGGTATC
341	BpsA(Q1014)-pTrc_rv	caggctgaaaatcttctctcaCTGAGCCACTTCTCTCTCCAG
346	BpsA(Y569)-pCDf1_fwd	ctttaataaggagatataccatgTACTCTCGGAAGGCAGCCG
347	BpsA(L695)-pCDF1_rv	ctttctgttcgacttaagcattaCCAGGCCATAGGCTGGC

TycA with TycB-Com$_D$ in pCDFDuet(MCS1)

114	TycA(MCS1)_fwd	*gtttaactttaataaggagatatacc*atgTTAGCAAATCAGGC
113	TycA(Ed1+L135)-TycB_Com$_D$):(MCS1)	CAATTGAGCCAAAGTTATCAAAAGCATCTGCTTGGCATCATCGAGCATTGCATGGCAAAAGAAGAAGGCGAGTACACCCCGAGCGACCTGGGGGATGAAGAGCTGTCCATGGAGGAGCTGGAAAACATCCTGGAATGGATTtaatgcttaagtcgaacagaaagtaatcgta
115	TycA(L1046)_rv2	GCAGATGCTTTTGATAAC

TycC modules Asn (5) and Gln (6) different start sites in pCDFDuet(MCS2)

112	TycC(M1_C$_{D\text{-}Phe\text{-}Asn}$):(MCS2)_fwd	gttaagtataagaaggagatatacatatgAAAAAGCAGGAAAACATCGC
85	TycC(Q-167_A$_{Asn}$)_(MCS2)_fwd	gttaagtataagaaggagatatacatatgCAGACGAACAAACAACAGACGTTCAGC
192	TycC(I-30_A$_{Asn}$):(MCS2)_fwd	gtataagaaggagatatacatatgATCACCGAAGAGGAAAAGCAG
193	TycC(ML-E-27_A$_{Asn}$)-(MCS2)_fwd	gtataagaaggagatatacatatgCTGGAGGAAAAGCAGCAACTGCTC
147	(MCS2)TycC(MLA-Y-18_A$_{Asn}$)fwd	gtataagaaggagatatacatatgCTGGCCTACAACGACACGGCTGCTG
141	TycC(Y-18_A$_{Asn}$):(MCS2)_fwd	gttaagtataagaaggagatatacatatgTACAACGACACGGCTGCTG
144	(MCS2)TycA(M1-S28):(I-5_A$_{Asn}$)TycC	aagtataagaaggagatatacatatgTTAGCAAATCAGGCCAATCTCATCGACAACAAGCGGGAACTGGAGCAGCATGCGCTAGTTCCATATGCACAGGGCAAGTCGATTTTCGAGCTGATCGCGG
142	TycC(I-5_ A$_{Asn}$)-(MCS2)_fwd	aagtataagaaggagatatacatatgATTTTCGAGCTGATCGCGG
131	TycC(Q-167_A$_{Gln}$):(MCS2)_fwd	gtataagaaggagatatacatatgCAGCCTGCCAGCAGCAAAAC
105	TycC(S-42_A$_{Gln}$):(MCS2)_fwd	G TAT AAG AAG GAG ATA TAC ATA TGA GTG AGC AGC CAA CGG CAAG
96	IndC(M1-27):(V1_A$_{Gln}$ _TycC)_rv	CTGCTTGCTCCTCCACcatatccatcaatgttatt
148	(MCS2)TycC(L-19_A$_{Gln}$)fwd	gtataagaaggagatatacatatgCTGAACGTGAACGATACGTTTGTCG

108	TycC(A-7_A_{Gln}):(MCS2)_fwd	gtataagaaggagatatacatatgGCGACCGCTTTGCATCAATTAG
81	TycC(A_{Gln}_V1)_(MCS2)_fwd	gttaagtataagaaggagatatacatatgGTGGAGGAGCAAGCAGCACGC

TycC:IndC different fusion sites

145	TycC(V-37_ A_{Gln}):(N3_IndC)rv	gtaatattattCACGTTCAGAAGGAGTGTGCG
146	IndC(N3)_(V-17_A_{Gln})_fwd	TCTGAACGTGaataatattacacaatgtgactc
149	TycC(A20_ A_{Gln}):(L48_IndC)_rv	gcgtaactgagGGCATATTCTTCGTACACGACGG
150	IndC(L48)_(A20_A_{Gln})TycC_fwd	AGAATATGCCctcagttacgctgatttc
151	TycC(L173_ A_{Gln}):(G210)IndC_rv	gaatgacatccCAGCGGGAACGTCTCTCTC
152	IndC(G210)_(L173_A_{Gln})TycC_fwd	GTTCCCGCTGggatgtcattcccggattttac
153	TycC(F270_ A_{Gln}):(L306)IndC_rv	gtaaaactattcaaAAAACGCGAGACGAGGGC
154	IndC(L306):(F270_A_{Gln})_fwd	CTCGCGTTTTttgaatagttttactcactg
155	TycC(K309_ A_{Gln}):(P349)IndC_rv	tatctacaggTTTGCCGATCGGAATTGC
156	IndC(P349):(K309_A_{Gln})TycC_fwd	GATCGGCAAAcctgtagataataccgaatac
97	TycC(A_{Gln}_D372):(L415)IndC_rv	*ctggttaccagGTCTCCGGTAC*GGTACATGGTGC
98	IndC(L415):(D372_ A_{Gln})TycC_fwd	*GTA CCG GAG ACC TGG TAA CCA G*AG GGG CTG
82	TycC(A_{Gln}_G425)_(R468)IndC_rv	gagattttggaaacccgttctGCCATCCTGGCGGGCCATG
83	IndC(R468):(G425_A_{Gln})TycC)_fwd	CATGGCCCGCCAGGATGGCagaacgggtttccaaaatctc
194	TycC(K426_ A_{Gln}):(G470)IndC_rv	gattttggaaaccTTTGCCATCCTGGCGGG
195	IndC(G470):(K426_A_{Gln})TycC_fwd	GGATGGCAAAggtttccaaaatctcatcgc
243	TycC(A_{Gln} _K426):(T469)IndC_rv	ggaaacccgtTTTGCCATCCTGGCGG
244	IndC(T469):(K426_A_{Gln})TycC_fwd	GGATGGCAAAacgggtttccaaaatctcatc
157	TycC(R395_ A_{Gln}):(G438)IndC_rv	tacggtaaccTCTGATTTTGACTTGATGGTCG
158	IndC(G438):(R395_A_{Gln})TycC_fwd	CAAAATCAGAggttaccgtattgagc
196	TycC(A437_ A_{Gln}):(D481)IndC_rv	ctttctcatcCGCTACGACATAGGCGTAC
197	IndC(D481):(A437_A_{Gln})TycC_fwd	TGTCGTAGCGgatgagaaagaagctgcattg
245	IndC(E876):(G427_A_{Gln})TycC_rv	CTGTTTTGCCCttcacacatatattgcgcttgg
246	TycC(G427_	atatgtgtgaaGGGCAAAACAGCCTGTACG

	A_{Gln}):(E876)IndC_fwd	
247	TycC(PCP$_{Gln}$-A58):(K1014)IndC_rv	caatccacttCGCGATCGTTGGCGAT
248	IndC(K1014):(PCP$_{Gln}$ -A58)TycC_fwd	AACGATCGCGaagtggattgaacagacagactc
118	IndC(K922)-TycC(A_{Gln}_V477-PCP_A565):(W1015)IndC	gatagacctatggcggcattttatttctgcataggtgggggtattagccaagcgcaatatatgtgtgaaggcatgaaagaagatgttgttcatatgaaagggccagttgaaatcattaaagatgatcttcaacaacaactccctcaatatatgattccaaataaggtattagttttcgataaattacctttgacggccaatggaaaaGTGGATCGCAAGGCATTGCCTCAACCGGAGGATGCCGCCGCCTCTGCTGCCGTGTATGTGGCGCCGCGCAACGAATGGGAAGCCAAGCTCGCAGCGATATGGGAAAGTGTGCTTGGAGTCGAGCCGATCGGGGTTCACGATCATTTCTTTGAACTGGGCGGACATTCTTTGAAAGCGATGCACGTCATTTCTTTGCTCCAGCGCAGCTTCCAGGTGGACGTACCGTTGAAAGTCCTGTTTGAATCGCCAACGATCGCGGGCCTGGCCtggattgaacagacagactctaaaac
119	IndC(I864)-TycC(A_{Gln}_K426-PCP_A565):(W1015)IndC	cttgatagacctatggcggcattttatttctgcataAAAGGGCAAAACAGCCTGTACGCCTATGTCGTAGCGGAGCAGGACATCCAGACAGCGGAGCTGAGAACGTACCTGTCTGCCACCTTGCCAGCCTACATGGTTCCGTCCGCTTTTGTTTTCTTGGAGCAGCTGCCGCTTTCAGCGAACGGCAAAGTGGATCGCAAGGCATTGCCTCAACCGGAGGATGCCGCCGCCTCTGCTGCCGTGTATGTGGCGCCGCGCAACGAATGGGAAGCCAAGCTCGCAGCGATATGGGAAAGTGTGCTTGGAGTCGAGCCGATCGGGGTTCACGATCATTTCTTTGAACTGGGCGGACATTCTTTGAAAGCGATGCACGTCATTTCTTTGCTCCAGCGCAGCTTCCAGGTGGACGTACCGTTGAAAGTCCTGTTTGAATCGCCAACGATCGCGGGCCTGGCCtggattgaacagacagactctaaaacaatatcaag

TycC:IndC/BpsA in pTrc99a

209	TycC(Q-167_A_{Gln})_pTrc99_fwd	cacacaggaaacagaccatgCAGACGAACAAACAACAG
395	TycC(-29T_A_{Asn})_pTrc_fwd	cacacaggaaacagaccatgACCGAAGAGGAAAAGCAGC
372	TycC(-21-A_{Asn})-pTrc_fwd	cacacaggaaacagaccatgCTCGTCGCCTACAACGACAC
381	TycC(PCP$_{Gln}$_A5):(GSG-H6)_rv	gatgatgACCTGATCCGCGGGCAGCCGCAAC
383	TycC(PCP$_{Asn}$_+E4)_rv	gatgatgACCTGATCCTTCCGACCGCGAAAGGAAG
396	TycC(-29T_A_{Gln})_pTcr_fwd	cacacaggaaacagaccatgACGGAAGCGGAAAAACGC
242	TycC(L-19_A_{Gln})_pTrc99_fwd	cacacaggaaacagaccatgCTGAACGTGAACGATACGTTTG
218	TycC(P157-$C_{Dphe-Asn}$)_pTrc99_fwd	cacacaggaaacagaccatgCCGCGGGAAGCGAC
219	TycC(-I6-$C_{Dphe-Asn}$)_pTrc99_fwd	cacacaggaaacagaccatgATTCAACCCGTGGCAGC
281	BpsA(G434)- (Ad6-R395)TycC_fwd	GTCAAAATCAGAGGATATAGAGTCGAGCTGGATG
282	TycC(A_{Gln}-R395):(G434)BpsA_rv	CGACTCTATATCCTCTGATTTTGACTTGATGGTCG
342	TycC(PCP$_{Gln}$-A58):(GSG-H6)pTrc_rv	caggctgaaaatcttctctca*gtggtgatggtgatgatg*ACCTGATCCCGCGATCGTTGGCGATTC

343	TycC(PCP$_{Gln}$-A58):(K1005-BpsA)_rv	CGGGCCAGCTTCGCGATCGTTGGCGATTC
344	BpsA(K1005):(PCP$_{Gln}$-A58)TycC_fwd	CAACGATCGCGAAGCTGGCCCGACGGC
345	BpsA(E1004):(GSG-H6)pTrc_rv	caggctgaaaatcttctctca*gtggtgatggtgatgatg*ACCTGATCCTTCAATTGTTGGTGACTCCAGG
210	Ghis6_pTrc99_rv	ggctgaaaatcttctctcagtggtgatggtgatgatgACC

Dodecylindigoidine synthetase, S. violacea Pptase and TycC-dIndS fusions

166	SVDodecInd-(MCS2)_fwd	gtataagaaggagatatacatatgGAACCTAAGTCGTTCAAC
167	SVDodecInd-(MCS2)_rv	cagcagcctaggttaattaGCTCTCGTTAATTACCACCTG
168	SV-Pptase-(pTRc99)_fwd	cacacaggaaacagaccatgGTTCGTGGTTACCTTAG
169	SV-Pptase-(pTRc99)_rv	aggctgaaaatcttctctcaAAAAGTACTTGATACCAGTTTAGC
373	pSV1(A$_{Lauric\ acid}$_R541)_fwd	CGTTTAACCCGTCAAATGTTGG
374	pSV1(E11)_rv	CGTTGTTTGTTCCGCTAAGTTGAAC
375	TycC(A$_{Val}$_-26T):(11T_dIndS)_fwd	CTTAGCGGAACAAACAACGATTCAGGAGCTGTTCGAGC
376	TycC(A$_{Val}$_60R):(R541_A$_{lauric\ acid}$-dIndS)_rv	CATTTGACGGGTTAAACGGCGGTCGACCTTTCCGTTC
377	TycC(A$_{Orn}$_T-26):(11T_dIndS)_fwd	CTTAGCGGAACAAACAACGTTCCAGAAGCTGTTCGAG
378	TycC(A$_{Orn}$_60R):(R541_A$_{lauric\ acid}$-dIndS)_rv	CATTTGACGGGTTAAACGTCTGTCTACTTTGCCATTGGCAG
379	TycC(A$_{Leu}$_-26T):(11T_dIndS)_fwd	CTTAGCGGAACAAACAACGCTGCACCAGCTATTCG
380	TycC(A$_{Leu}$_61R):(R541_A$_{lauric\ acid}$-dIndS)_rv	CATTTGACGGGTTAAACGGCGATCCACTTTTCCATTGG

Tyrocidine synthetase with IndC and dIndS

172	IndC(A1013):(A1651)DindS_1rv	CGACCACGGCagccaattctgctatattag
173	IndC(A1013):(A1651)DindS_2rv	CGACCACGGCagccaattctgctatattaggagattg
174	DindS(A1651):(A1013)IndC_fwd	gcagaattggctGCCGTGGTCGAAAATAATG
175	DindS-His6-(MCS2)_rv	cagcagcctaggttaattagtggtgatggtgatgatgACCTGATCCGCTCTCGTTAATTACCACCTGTG
176	DindS(S1536):(L896)IndC_fwd	ttaaagatgatcttTCCCAAGCCTTGCCTAG
177	IndC(L896):(S1536)DindS_1rv	GGCTTGGGAaagatcatctttaatgatttc

ppsD:IndC fusions

249	ppsD(-174V_A$_{Gln}$)-Ptrc99_fwd	tcacacaggaaacagaccatgGTGCAGGGGGCAAAAACG
250	ppsD(A$_{Gln}$_R393):(G437)IndC_rv	tacggtaaccGCGGATTTTCACCTGATCATC
251	IndC(G437):(A$_{Gln}$_R393)ppsD_fwd	GAAAATCCGCggttaccgtattgagcttgatg

252	ppsD(-60L_A_{Gln})-Ptrc99_fwd	cacacaggaaacagaccatgCTTTTCACCACGCTTGTGG
262	Ybdz_pCDF(MCS2)_fwd	gtataagaaggagatatacatatgGCAAATCCTTTTGAAAATGCG
263	Ybdz_pCDF(MCS2)_rv	cagcagcctaggttaattaCACATTTTCAACAGTCTTTAGGC

LchAA:IndC fusions

120	LchAA(MCS2_M1)_fwd	gtataagaaggagatatacatATGGGAAACACATGTTACCC
140	IndC(MCS2_LchAA_F2-A_{Gln})_rv	GCTGCTTTTTGTTCGAAcatatccatcaatgttatt
123	LchAA(A_{Gln}_T368)_(IndC_T424)_rv	caacaaaataagtCGTACCGTCAGGAAGCC
124	IndC(T424):(T368_A_{Gln})LchAA_fwd	CTGACGGTACGacttattttgttggtcgggttg
125	LchAA(MCS2_ A_{Gln}_M-27)_fwd	gtataagaaggagatatacatATGACCCTGCACGGTTTATTC
126	LchAA(MCS2_H323)_fwd	GTA TAA GAA GGA GAT ATA CAT ATG CAT CAA AAG TAT CCG TAC AAT C
127	yngIA(MCS1)_fwd	ctttaataaggagatataccatgGTTACATTAAAAAAACAAACG
128	yngIA(MCS1)_rv	ctgttcgacttaagcattaCTGCACCGGGCTC

ItuB-BpsA fusions in pTrc99

303	ItuB(M1)_pTrc99_fwd	cacacaggaaacagaccatgTCGGTATTTAGAAATCAAGAAACGTACTGGG
304	ItuB(A_{Asn}_T-18)_rv	GCGCAGACGTGTCGTTAAGCG
305	ItuB(A_{Asn} _T-18)_fwd	CGCTTAACGACACGTCTGCGC
306	ItuB(A_{Gln}_H406):(G434)BpsA_rv	GACTCTATATCCGTGAATTTTCACCTGATGGTCAATCCG
307	BpsA(G434)-ItuB(A_{Gln}_H406)_fwd	GGTGAAAATTCACGGATATAGAGTCGAGCTGGATG
310	ItuB(Epi_{Asn}_A-7)_fwd	GCCGATCAGGGTGAAGTCAAAGG
311	ItuB(Epi_{Asn}_A-7)_rv	CCTTTGACTTCACCCTGATCGGC
312	ItuB(A_{Gln} _E212)_fwd	GAGGTCGTGTCAGACGGTAC
313	ItuB(A_{Gln} _E212)_rv	GTACCGTCTGACACGACCTC
314	ItuB(A_{Tyr}_M-79)_pTrc99_fwd	cacacaggaaacagaccatgGAGAACAACGGGGTTCACTTG
315	ItuB(A_{Tyr}_N-156)_pTrc99_fwd	cacacaggaaacagaccatgAATACGCTTACAAGCGAAAGCACGTTC
330	ItuB(A_{Asn}_I-175)_pTrc99_fwd	cacacaggaaacagaccatgATTAACACCATTCCCGTCCGAATCC
331	ItuB(A_{Asn} _P-77)_pTrc99_fwd	cacacaggaaacagaccatgCCGGGTGACGCGATGAAC

TycC8, 9, 10 ±TE and combinations thereof in pTrc99

355	TycC(-185E_A_{Val})-pTrc_fwd	cacacaggaaacagaccatgGAGACGAAACACTTCCAGGCATTTTTG
392	TycC(-49S_A_{Val})_pTrc_fwd	cacacaggaaacagaccatgAGCGAGGAAGAGCGCCGAATTG
356	TycC(-40V_ A_{Val})-pTrc_fwd	cacacaggaaacagaccatgGTTGATTTCAACAACACGTTTGCCG
384	TycC(PCP_{Val}_A4)_rv	gatgatgacctgatcctgcgctctcggcaatatgcg
354	TycC(-185G_A_{Orn})-pTrc_fwd	cacacaggaaacagaccatgGGCAGCCTGAGCTTCCGC

393	TycC(-49S_A_{Orn})_pTrc_fwd	cacacaggaaacagaccatgTCCAAAGCAGAGACGGAGCACATG
351	TycC(-41L_ A_{Orn})-pTrc99_fwd	cacacaggaaacagaccatgCTGCACAGCTTCCTCGC
385	TycC(PCP_{Orn}_A4)_rv	gatgatgACCTGATCCGGCTGTCTCTTCGATGAACGCC
388	TycC_M_{Orn}+TE1_rv	GCCGTATCTGCTCTCAAAGGCTGTCTCTTCGATGAACGC
389	TycC M_{Orn}+TE1_fwd	TTTGAGAGCAGATACGGCAC
390	TycC M_{Orn}+TE2_rv	CTCAAAGCGCTTATGCGTGATGAACGCCGCCAGTTC
391	TycC M_{Orn}+TE2_fwd	ACGCATAAGCGCTTTGAG
394	TycC(-49L_A_{Leu})_pTrc_fwd	cacacaggaaacagaccatgTTGCCAGAAGAAAAACAGCAGATTTTGG
386	TycC(-40L_ A_{Leu})-pTrc_fwd	cacacaggaaacagaccatgGCCGGGTTCAACGATACGG
387	TycC(PCP_{Leu} _R4)_rv	gatgatgACCTGATCCGCGCTTATGCGTGATGAAATCG
382	pTrc99-GSG-H6_fwd	GGATCAGGTcatcatcacc
348	TycC(TE):(GSG-H6)Ptrc99_rv	caggctgaaaatcttctctcagtggtgatggtgatgatgACCTGATCCTTTCAGGATGAACAGTTCTTGCAGG

6.1.3 ***Primers for site-directed mutagenesis***

All primers used to introduce point mutations are complementary to each other. In the following, only the forward primer of the pair is listed.

IndC site-directed mutations

228	IndC-Ad-TS179VA_fwd	GTCTGGCTTATATTATTTATGTGGCGGGTAGCACGGGTAAGCCAAAG
230	IndC-Ad-D414A_fwd	GTTATATCGAACGGGAGCACTGGTAACCAGAGGGG
221	IndC-Oxd-N571S_fwd	GACAGCCACTCAATCGAGCGAAATTAGCTCTTTGC
223	IndC-Oxd-N571A_fwd	GACAGCCACTCAATCGGCGGAAATTAGCTCTTTGCC
186	IndC-OxD-K605A_fwd	CATCAACGTTTATTGCCCGCGTATGCCTATGCTTCACC
188	IndC-OxD-K605E_fwd	CCATCAACGTTTATTGCCCGAATATGCCTATGCTTCACCG
200	IndC-Ox-Y615A_fwd	GCTTCACCGGGTGCTCTCGCGGCGACACAAATGTATTTTG
232	IndC-Ad-K922A_fwd	GACGGCCAATGGAGCGGTGGATTATCAATC
274	IndC-pA10_1_fwd	cgtaaagcgttaccggaaccgaaagccgtggagaatgtttcaacacagcg
275	IndC-pA10_1_rv	cggttccggtaacgctttacgatccacttttccattggccgtcaaagg
276	IndC-pA10_2_fwd	cgtaaagcgttaaccggctttaaagccgtggagaatgtttcaacacagcg
277	IndC-pA10_2_rv	aaagccggttaacgctttacgatccacttttccattggccgtcaaagg
240	IndC-PCP-K13G_fwd	GAAAAATTTGGATGGAAGTACTGGGCTGGGATTCAGTATCTGCCCTCG
238	IndC-PCP-S18G_fwd	CTGAAATGGGATTCAGTAGGCGCCCTCGATGATTTTTTC
234	IndC-PCP-N979D_fwd	CGAAAGTGGGGGTGATTCTTTGATGGCC
236	IndC-PCP-N979H_fwd	CGAAAGTGGGGGTCATTCTTTGATGGCC
136	IndC-PCP_S980A_fwd	GAAAGTGGGGGTAATGCGTTGATGGCCGTTGC
190	IndC-TE-S1108A_fwd	GCCATATATATTGTGGGGATATGCGTTTGGTGCCCGAGTAGCATTTG
254	IndC-TE-S1108C_fwd	CCATATATATTGTGGGGATATTGCTTTGGTGCCCGAGTAGCATTTG
178	IndC-TE_A1135D_fwd	GCATTGAATTTATTGGATCCGGGATCTCCTCATC
180	IndC-TE_APG98DAE_fwd	GTTAACGCATTGAATTTATTGGATGCGGAATCTCCTCATCTTGATATGAAGC

BpsA site-directed mutations

285	BpsA-D410L_fwd	CGGCTGTACAAGACCGGGCTGCTGGGCCAGTGGAACAATG
301	BpsA-K600E_fwd	GAACGCCTGCTGCCTGAATACGGCTATGCCTC
287	BpsA-Y603F_fwd	CTAAGTACGGCTTTGCCTCCCCAGG
289	BpsA-Y603A_fwd	CTGCTGCCTAAGTACGGCGCGGCCTCCCCAGGAGCACTG
291	BpsA-S974A_fwd	CTTTGAGTCCGGCGGAAACGCGCTGATCGCCGTCGGCCTG
293	BpsA-S1103A_fwd	CCCTGTGGGGATATGCGTTCGGAGCTCGCGTG
337	BpsA-S1103C_fwd	CCTGTGGGGATATTGCTTCGGAGCTCG
295	BpsA-A1131G_fwd	CAACCTGTTCCTGATCGGCCCTGGGTCACCAAAAG
297	BpsA-A1131D_fwd	CAACCTGTTCCTGATCGATCCTGGGTCACCAAAAG

DindS site-directed mutations

182	DindS-TE_D106A_fwd	GTGTTTATCTCCATAGCGGCTGAGGCACCCTATG
184	DindS-TE_DAE106APG_fwd	GAGGTGTTTATCTCCATAGCGCCGGGCGCACCCTATGTGCCAAAAG

Tyrocidine synthetase site-directed mutations

198	TycA($PCP_{D\text{-}Phe}$-S27A)_fwd	GCTCGGCGGAGATGCGATCCAAGCGATC
256	TycC-A_{Asn}-TS142VA_fwd	GATTTGGCGTACATGATCTACGTGGCGGGTTCTACCGGCAATCCAAAAG
258	TycC- A_{Asn} -D370A_fwd	GTATCGAACAGGCGCGCTGGCAAAATGGCTG
260	TycC- A_{Asn} -K474A_fwd	CCGCTGACCGCGAACGGAGCCGTAGAGCGAAAAAAATTGC
129	TycC(PCP_{Asn} _S28A)_fwd	CTTGGGCGGACACGCGTTGAAGGCGATG
99	TycC($C_{Asn\text{-}Gln}$)_HaxxxDG_fwd	GTTTTTGCTCGACATGCACGCGATCATCTCGGACGGCGTTTC
110	TycC($C_{Asn\text{-}Gln}$)_Haxxx*AA*_fwd	CATGCACGCGATCATCTCGGCGGCGGTTTCTTCGCAAATTTTGC
335	TycC-A_{Gln}_D372A_fwd	CATGTACCGTACCGGAGCGATGGTCCGCTATTTGCC
357	TycC-A_{Val}-D349A_fwd	GATGTATCGCACCGGAGCGTTGGCGAGATGGCTGC
359	TycC- PCP_{Val} -S28A_fwd	GAAATCGGCGGTCATGCGTTGAAGGCGATGAACG
361	TycC-A_{Orn}-D352A_fwd	GTACCGCACAGGTGCGCTGGCGAAGTGGC
363	TycC- PCP_{Orn} -S28A_fwd	GCTCGGCGGTCACGCTTTGCGTGCCATGC
365	TycC-A_{Leu}-D354A_fwd	CATGTACAAAACAGGCGCGTTGGTAAAATGGCGGAC
367	TycC- PCP_{Leu} -S29A_fwd	GAACTCGGAGGACATGCCTTAAAAGCTACGC

6.2 Plasmids

6.2.1 *(Dodecyl-) and Indigoidine Synthetases*

dIndS and TycC A domain insertion variants

name	backbone	insert	origin
pSV1	pCDFDuet – MCS1	dIndS($A_{lauric\ acid}$ -PCP-C-A_{Gln}-PCP-TE (WP 013053246.1)	This work
pSV2	pTrc99a	Shewanella violacea Pptase (WP_013050488.1)	This work
pST4	pCDFDuet-MCS1	(E1-T11_dIndS):(T-26_A_{Val})TycC(+60R):(R541_$A_{lauric\ acid}$)dIndS	This work

pST5	pCDFDuet-MCS1	(E1-T11_dIndS):(T-26_A_{Orn})TycC(+60R):(R541_$A_{lauric\ acid}$)dIndS	This work
pST6	pCDFDuet-MCS1	(E1-T11_dIndS):(T-26_A_{Leu})TycC(+61R):(R541_$A_{lauric\ acid}$)dIndS	This work

IndC and BpsA variants

pID8H	pCDFDuet-MCS2	IndC(A_{Gln}OxA-PCP-TE)- H_6	This work
pID9aH	pCDFDuet-MCS2	IndC(A21_A_{Gln}OxA-PCP-TE)- H_6	This work
pID9H	pCDFDuet-MCS2	IndC(M27_A_{Gln}OxA-PCP-TE)- H_6	This work
pTrc8H	pTrc99a	IndC(A_{Gln}OxA-PCP-TE)- H_6	This work
pMM64	pETDuet – MCS1	BpsA(A_{Gln}OxA-PCP-TE)	Fussenegger Lab[199]
pTrc64H	pTrc99a	BpsA(A_{Gln}OxA-PCP-TE)-H_6	This Work
pTrc64dTE	pTrc99a	H_6-GSG-BpsA(A_{Gln}OxA-PCP_Q1014)	This Work
pCDF64Ox	pCDFDuet-MCS1	BpsA-Ox(Y569-L695)	This Work
pCDF64TE	pCDFDuet-MCS1	BpsA-TE(E1015-end)-H_6	This work

6.2.2 *Tyrocidine-Indigoidine Synthetase fusion constructs*

TycC(A_{Gln}):IndC(Ox):TycC(A_{Gln}PCP):IndC(TE) fusion constructs

pID12eH	pCDFDuet-MCS2	TycC(Q-167_A_{Gln}-D372):(L415)IndC(K862):(A_{Gln}_V477)TycC(PCP_{Gln}_A61):(W1015)IndC-H_6	This work
pID12aH	pDCFDuet-MCS2	IndC(M1-M27)-TycC(V1_A_{Gln}-D372):(L415)IndC(K862):(A_{Gln}_V477)TycC(PCP_A61):(W1015)IndC-H_6	This work
pID12dH	pDCFDuet-MCS2	TycC(A-7_A_{Gln}_D372):(L415)IndC(K862):(A_{Gln}_V477)TycC(PCP_A61):(W1015)IndC	This work
pID12H	pCDFDuet-MCS2	TycC(V1_A_{Gln}_D372):(L415)IndC(K862):(A_{Gln}_V477)TycC(PCP_A61):(W1015)IndC-H_6	This work
pID13eH	pCDFDuet-MCS2	TycC(Q-167_A_{Gln}-D372):(L415)IndC(I804):(A_{Gln}_K426)TycC(PCP_A61):(W1015)IndC-H_6	This work
pID13aH	pCDFDuet-MCS2	IndC(M1-M27)-TycC(V1_A_{Gln}_D372):(L415)IndC(I804):(A_{Gln}_K426)TycC(PCP_+A61):(W1015)IndC-H_6	This work
pID13dH	pCDFDuet-MCS2	TycC(A-7_A_{Gln}_D372):(L415)IndC(I804):(A_{Gln}_K426)TycC(PCP_A61):(W1015)IndC-H_6	This work
pTrc22c2H+Td	pTrc99	TycC(L-19_A_{Gln}_K426):(T469)IndC(E876):(G427_A_{Gln})TycC(PCP-A58):(K1014)IndC-H_6	This work

IndC:dIndS fusion constructs

pIS1H	pCDFDuet-MCS2	IndC(L896):(S1536)DindS-H_6	This work
pIS2H	pCDFDuet-MCS2	IndC(A1013):(A1651)DindS-H_6	This work
pIS3H	pCDFDuet-MCS2	TycA(M1-S28)-TycC(I-5_A_{Asn}-PCP-C-A_{Gln}_D372):(L415)IndC(L896):(S1536)DindS-H_6	This work

TycC(A_{Gln}):IndC fusion constructs

Different **start sites** for fusion site "10"=TycC(A_{Gln}_D372):(L415)IndC

pID10eH	pCDFDuet-MCS2	TycC(**Q-167**_A_{Gln}_D372):(L415)IndC-H_6	This work
pID10cH	pCDFDuet-MCS2	TycC(**S-42**_A_{Gln}_D372):(L415)IndC-H_6	This work
pID10aH	pCDFDuet-MCS2	**IndC(M1-M27):(V1)**TycC(A_{Gln}_D372):(L415)IndC-H_6	This work
pID10c2H	pCDFDuet-MCS2	TycC(**L-19**_A_{Gln}_D372):(L415)IndC-H_6	This work
pID10dH	pCDFDuet-MCS2	TycC(**A-7**_A_{Gln}_D372):(L415)IndC-H_6	This work
pID10H	pCDFDuet-MCS2	TycC(**V1**_A_{Gln}_D372):(L415)IndC-H_6	This work

TycC(A_{Asn}-PCP-C-A_{Gln}):IndC fusion constructs

Different **start sites** for fusion site "14"=TycC(A_{Asn}-PCPC-C-A_{Gln}_V-17):(N3)IndC

pID14bH	pCDFDuet-MCS2	TycC(**Q-167**_A_{Asn}-PCP-C-A_{Gln}_V-17):(N3_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID14hH	pCDFDuet-MCS2	**TycA(M1-S28)-TycC(I-5**_A_{Asn}-PCP-C-A_{Gln}_ V-17):(N3_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID14iH	pCDFDuet-MCS2	TycC(**LA-Y-18**_A_{Asn}-PCP-C-A_{Gln}_ V-17):(N3_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID14fH	pCDFDuet-MCS2	TycC(**Y-18**_A_{Asn}-PCP-C-A_{Gln}_ V-17):(N3_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID14gH	pCDFDuet-MCS2	TycC(**I-5**_A_{Asn}-PCP-C-A_{Gln}_ V-17):(N3_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work

Different **start sites** for fusion site "10"=TycC(A_{Asn}-PCPC-C-A_{Gln}_D372):(L415)IndC

pID2-10H	pCDFDuet-MCS2	TycC(**M1_C**-A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10bH	pCDFDuet-MCS2	TycC(**Q-167**_A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10hH	pCDFDuet-MCS2	**TycA(M1-S28)-TycC(I-5**_A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10iH	pCDFDuet-MCS2	TycC(**LA-Y-18**_A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10fH	pCDFDuet-MCS2	TycC(**Y-18**_A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10gH	pCDFDuet-MCS2	TycC(**I-5**_A_{Asn}-PCP-C-A_{Gln}_D372):(L415_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work

Different **start sites** for fusion site "19"=TycC(A_{Asn}-PCPC-C-A_{Gln}_R395):(G437)IndC

pID2-19H	pCDFDuet-MCS2	TycC(**M1_C**-A_{Asn}-PCP-C-A_{Gln}_R395):(G437_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID19bH	pCDFDuet-MCS2	TycC(**Q-167**_A_{Asn}-PCP-C-A_{Gln}_R395):(G437_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID19hH	pCDFDuet-MCS2	**TycA(M1-S28)-TycC(I-5**_A_{Asn}-PCP-C-	This work

		A_{Gln}_R395):(G437_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	
pID19jH	pCDFDuet-MCS2	TycC(**I-30**_A_{Asn}-PCP-C-A_{Gln}_R395):(G437_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
Different fusion sites for start site "b"=(Q-167)A_{Asn}			
pPW04	pSB1C3	TycC(Q-167-A_{Asn}-PCP-C-A_{Gln}_**T-161):(M1**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	P. Walch, iGem Team Heidelberg, 2013
pID14bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**V-17):(N3**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID15bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**A20):(L48**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID16bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**L173):(G210**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID17bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**F270):(L306**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID18bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**K309):(P349**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID10bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**D372):(L415**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID19bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**R395):(G437**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID20bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**G425):(R468**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID21bH	pCDFDuet-MCS2	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**K426):(G470**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
TycC:IndC fusions in pTrc99a			
pTrc10c2H	pTrc99a	**IndC(M1-M27):(V1)**TycC(A_{Gln}_D372):(L415)IndC-H_6	This work
pTrc10aH	pTrc99a	TycC(**L-19**_A_{Gln}_D372):(L415)IndC-H_6	This work
pTrc19c2H	pTrc99a	**IndC(M1-M27):(V1)**TycC(A_{Gln}_R395):(G437)IndC-H_6	This work
pTrc19aH	pTrc99a	TycC(**L-19**_A_{Gln}_ R395):(G437)IndC-H_6	This work
pTrc2-19bH	pTrc99a	TycC(**M1_C**-A_{Asn}-PCP-C-A_{Gln}_ R395):(G437_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pTrc19bH	pTrc99a	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**R395):(G437**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pTrc19bH-pA10-1	pTrc99a	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G437_A_{Gln}-Ox-A_**pA10-1: RKALPEP**_PCP-TE)IndC-H_6	This work
pTrc19bH-pA10-2	pTrc99a	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_R395):(G437_A_{Gln}-Ox-A_**pA10-2: RKALTGF**_PCP-TE)IndC-H_6	This work
pID22bH	pTrc99a	TycC(Q-167_A_{Asn}-PCP-C-A_{Gln}_**A437):(D481**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work

TycC-BpsA fusions in pTrc99a

TycC(A_{Gln})-BpsA fusion constructs

pSA19**a**H	pTrc99a	**IndC(M1-M27):(V1)**TycC(A_{Gln}_R395):(G434)BpsA-H_6	This work
pSA19**o2**H	pTrc99a	TycC(**T-29**_A_{Gln}_R395):(G434)BpsA-H_6	This work
pSA19**c2**H	pTrc99a	TycC(**L-19**_A_{Gln}_R395):(G434)BpsA-H_6	This work
Different start sites of TycC(A_{Asn}-PCPC-C-A_{Gln}_R395):(G434)BpsA fusion constructs			
pSA2-19H	pTrc99a	TycC(**M1_C**-A_{Asn}-PCP-C-A_{Gln}_ R395):(G434_A_{Gln}-Ox-A-PCP-TE)BpsA-H_6	This work
pSA19bH	pTrc99a	TycC(**Q-167**_A_{Asn}-PCP-C-A_{Gln}_ R395):(G434_A_{Gln}-Ox-A-PCP-TE)BpsA-H_6	This work
pSA19oH	pTrc99a	TycC(**T-29**_A_{Asn}-PCP-C-A_{Gln}_ R395):(G434_A_{Gln}-Ox-A-PCP-TE)BpsA-H_6	This work
pSA19**c3**H	pTrc99a	TycC(**L-21**_A_{Asn}-PCP-C-A_{Gln}_ R395):(G434_A_{Gln}-Ox-A-PCP-TE)BpsA-H_6	This work
pSA19bH-dTE	pTrc99a	TycC(**Q-167**_A_{Asn}-PCP-C-A_{Gln}_ R395):(G434_A_{Gln}-Ox-A-PCP_**Q1014**)BpsA-H_6	This work

TycA(B_{ComD})-TycC(B_{ComD}_C-A_{Asn}-PCP-C-A_{Gln}):IndC fusion

pID3-10H	pCDFDuet	MCS1: TycA(M1_A_{Phe}-PCP-E_+L135):(+G128_E)TycB MCS2: TycC(M1_C-A_{Asn}-PCP-C-A_{Gln}_**D372):(L415**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work
pID3-19H	pCDFDuet	MCS1: TycA(M1_A_{Phe}-PCP-E_+L135):(+G128_E)TycB MCS2: TycC(M1_C-A_{Asn}-PCP-C-A_{Gln}_**R395):(G437**_A_{Gln}-Ox-A-PCP-TE)IndC-H_6	This work

6.2.3 *Other NRPS-Indigoidine Synthetase Fusions*

LchAA:IndC fusions

pLI3H	pCDFDuet-MCS2	LchAA(M1_C-A_{Gln}_T368):(T424)IndC-H_6	This work
pLI3bH	pCDFDuet-MCS2	LchAA(H-163_A_{Gln}_T368):(T424)IndC-H_6	This work
pLI3aH	pCDFDuet-MCS2	LchAA(M-27_A_{Gln}_T368):(T424)IndC-H_6	This work

ppsD:IndC fusions

pPS1	pCDFDuet-MCS2	Ybdz (MbtH protein)	This work
pPS19c2H	pTrc99a	ppsD(-60L_A_{Gln}_R393):(G437)IndC-H_6	This work
pPS19bH	pTrc99a	ppsD(V-174_Apro-PCP-C-A_{Gln}_R393):(G437)IndC-H_6	This work

ItuB:BpsA fusions

pItu19bH	pTrc99a	ItuB(N-175_ A_{Asn}-PCP -Epi_{Asn} -C-A_{Gln}_H406):(G434)BpsA-H_6	This work
pItu25bH	pTrc99a	ItuB(N-156_ A_{Tyr}- PCP-Epi_{Tyr}-C-A_{Asn}-PCP -Epi_{Asn} -C-A_{Gln}_H406):(G434)BpsA- H_6	This work
pItu25zH	pTrc99a	ItuB(M1_A_{Tyr}- PCP-Epi_{Tyr}-C-A_{Asn}-PCP -Epi_{Asn} -C-A_{Gln}_H406):(G434)BpsA- H_6	This work

Others

pID5	pTrc99a	TycF (TEII)	This work
pID4	pDCFDuet-MCS1	TycA(M1_ A_{Phe}-PCP-Epi_+L135):(+G128_E)TycB	This work

6.2.4 *TycC single-, di-, and trimodular constructs*

TycC single module constructs = di-domain A-PCP			
pTycC1c2H	pTrc99a	TycC(L-21_A_{Asn}-PCP_+E4)-H_6	This work
pTycC2c2H	pTrc99a	TycC(L-19_A_{Gln}-PCP_+R6)-H_6	This work
pTycC4c2H	pTrc99a	TycC(V-40_A_{Val}-PCP_+A4)-H_6	This work
pTycC4oH	pTrc99a	TycC(S-49_A_{Val}-PCP_+A4)-H_6	This work
pTycC5c2H	pTrc99a	TycC(L-41_A_{Orn}-PCP_+A4)-H_6	This work
pTycC5oH	pTrc99a	TycC(S-49_A_{Orn}-PCP_+A4)-H_6	This work
pTycC6c2H	pTrc99a	TycC(L-40_A_{Leu}-PCP_+R4)-H_6	This work
pTycC6oH	pTrc99a	TycC(L-49_A_{Leu}-PCP_+R4)-H_6	This work
pTycC6c2+TE-H	pTrc99a	TycC(L-40_A_{Leu}-PCP-TE)-H_6	This work
TycC dimodular constructs			
pTycC12c2H	pTrc99a	TycC(L-21_A_{Asn}-PCP-C-A_{Gln}-PCP_+R6)-H_6	This work
pTycC45c2H	pTrc99a	TycC(V-40_A_{Val}-PCP-C-A_{Orn}-PCP_+A4)-H_6	This work
pTycC45c2H +TE1	pTrc99a	TycC(V-40_A_{Val}-PCP-C-A_{Orn}-PCP_+A4):(F-16)TE-H_6	This work
pTycC45c2H +TE2	pTrc99a	TycC(V-40_A_{Val}-PCP-C-A_{Orn}-PCP):(T-20)TE-H_6	This work
pTycC56bH	pTrc99a	TycC(G-185_A_{Orn}-PCP-C- A_{Leu}-PCP-TE)-H_6	This work
pTycC56c2H-ΔTE	pTrc99a	TycC(L-41_A_{Orn}-PCP-C- A_{Leu}-PCP_+R4)-H_6	This work
pTycC56c2+TE	pTrc99a	TycC(L-41_A_{Orn}-PCP-C- A_{Leu}-PCP-TE)-H_6	This work
TycC trimodular constructs			
pTycC456bH	pTrc99a	TycC(E-185_A_{Val}-PCP-C-A_{Orn}-PCP-C-A_{Leu}-PCP-TE)-H_6	This work
pTycC456c2H	pTrc99a	TycC(V-40_A_{Val}-PCP-C-A_{Orn}-PCP-C-A_{Leu}-PCP-TE)-H_6	This work

6.3 Supplementary figures

6.3.1 *Indigoidine production by IndC, BpsA and their mutants*

LR-MS (ESI-):

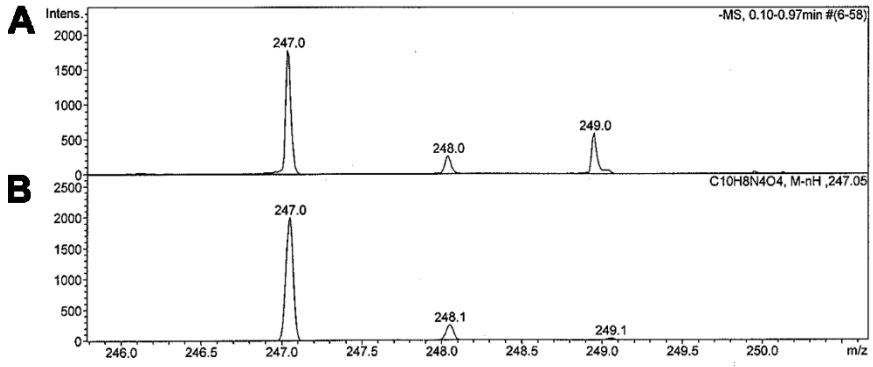

Figure S 1 | ESI-MS of indigoidine purified from *E. coli*. (A) m/z detected in negative mode ESI-MS of indigoidine in DMSO:MeOH (1:1) plus 1% formic acid matches the **(B)** expected m/z for indigoidine ($C_{10}H_8N_4O_4$).

Exact mass calculated for $C_{10}H_8N_4O_4$ $[M - H]^-$ 247.05. Found: 247.0.

NMR

^{1}H NMR (500 MHz, $(CD_3)_2SO$): 11.29 ppm (s; NH), 8.17 ppm (s; CH), and 6.46 ppm (s; NH_2).

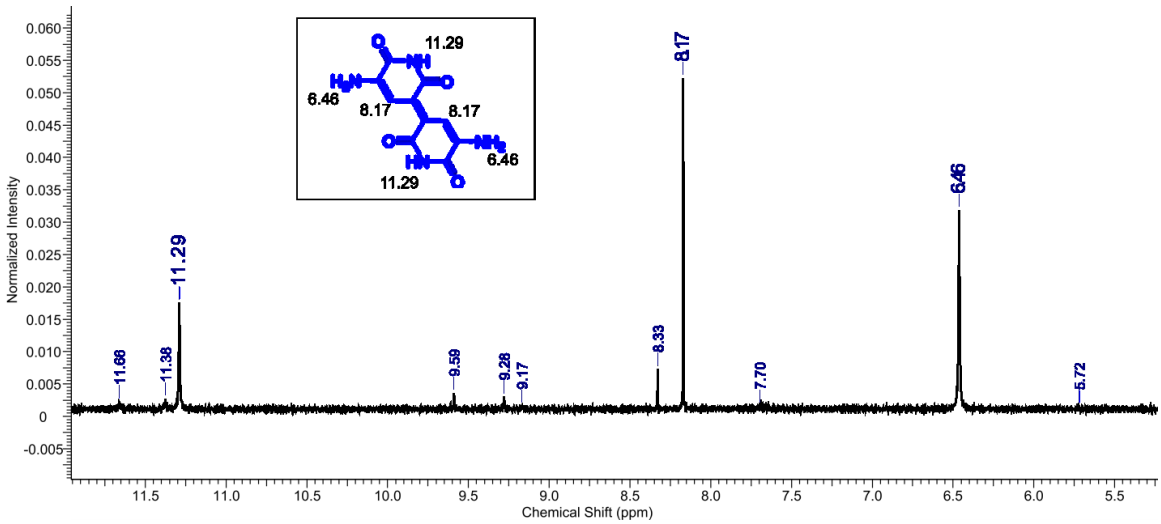

Figure S 2 | ^{1}H NMR spectrum at 500 MHz of indigoidine extracted from IndC expression in BAP1. Indigoidine was dissolved in DMSO-d6. Relevant peaks are at 11.29 ppm (s; NH), 8.17 ppm (s; CH), and 6.46 ppm (s; NH_2).

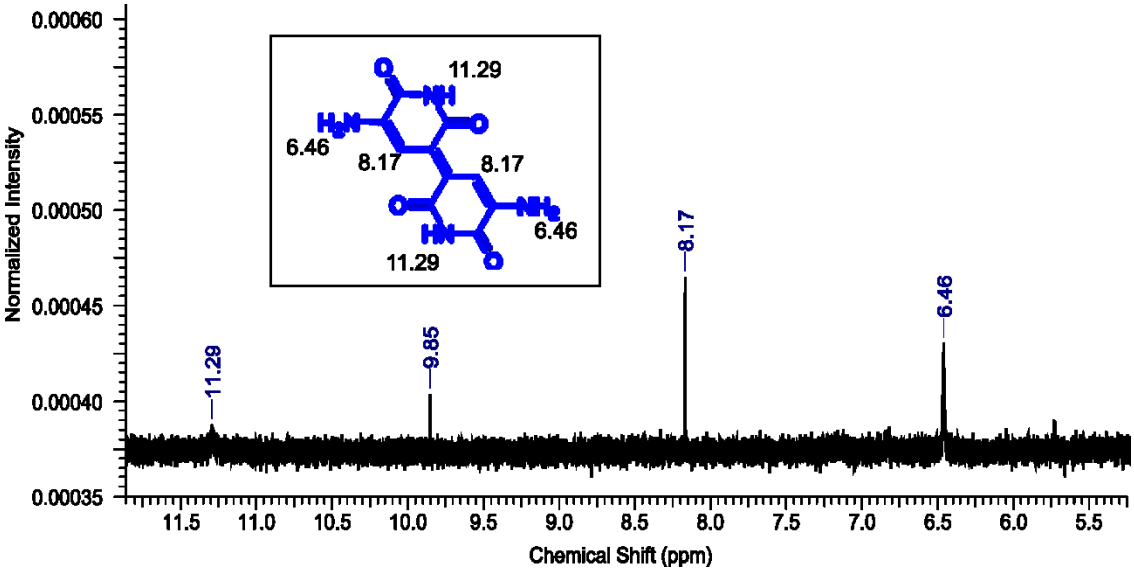

Figure S 3 | ^{1}H NMR spectrum at 500 MHz of indigoidine produced by purified IndC in vitro. Indigoidine pellet was dissolved in DMSO-d6. Relevant peaks are at 11.29, 8.17 and 6.46 ppm.

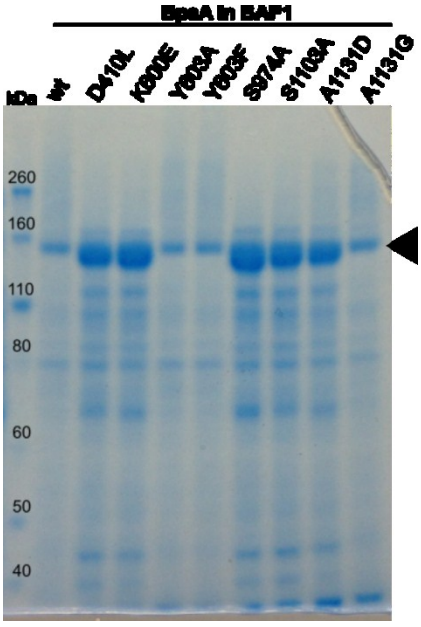

Figure S 4 | SDS-PAGE of BpsA wt and mutants. A 7% Tris-Acetate gel was loaded with equal amounts of the overnight expression of the indicated constructs and stained with Coomassie. The arrowhead indicates the position of the BpsA variants at an approximate expected size of ~150 kDa.

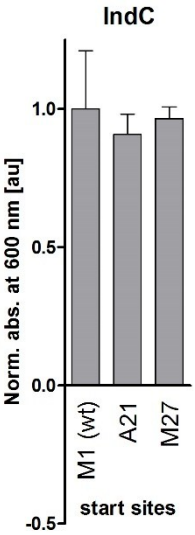

Figure S 5 | Influence of truncating the IndC N-terminus on pigment production. Absorption of three overnight BAP1 cultures expressing the indicated proteins was measured, the background (PCP mutant) was subtracted, and values were normalized to IndC wt absorption at 600 nm.

6.3.2 *Development of a tyrocidine-indigoidine synthetase*

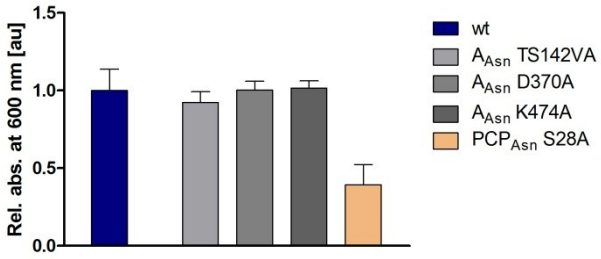

Figure S 6 | Effects of point mutations in the Asn incorporating module on pigment production of TycC(Q-167_A$_{Asn}$-PCP-C-A$_{Gln}$_R395):(G438)IndC. The "wt" and the indicated mutants were expressed in BAP1 cells from the pTrc99 vector and the absorption at 600 nm was calculated relative to the negative control (BpsA PCP mutant, S974A) and normalized to the "wt" levels. n = 3 +SD.

6.3.3 *Further engineering approaches on NRP-indigoidine fusion synthetases*

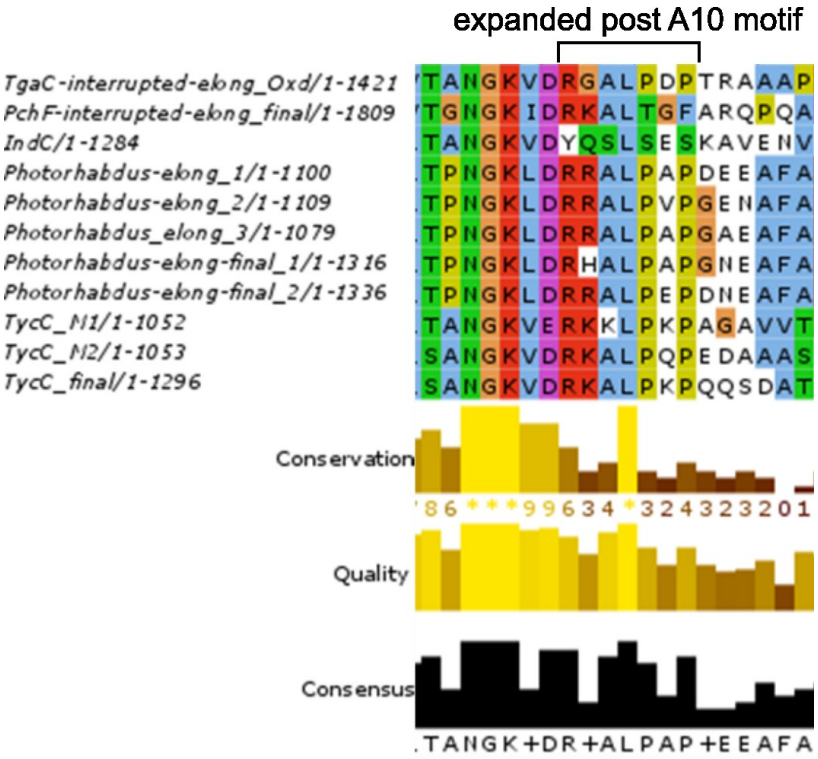

Figure S 7 | The post A10 motif differs between elongation, interrupted elongation and interrupted termination modules analyzed here. Protein sequence alignment of the post A10 motif of different NRPS modules from *Photorhabdus luminescence*, *Brevibacillus Brevis*, IndC, one interrupted elongation (TgaC) and one interrupted termination module A domain (PchF).

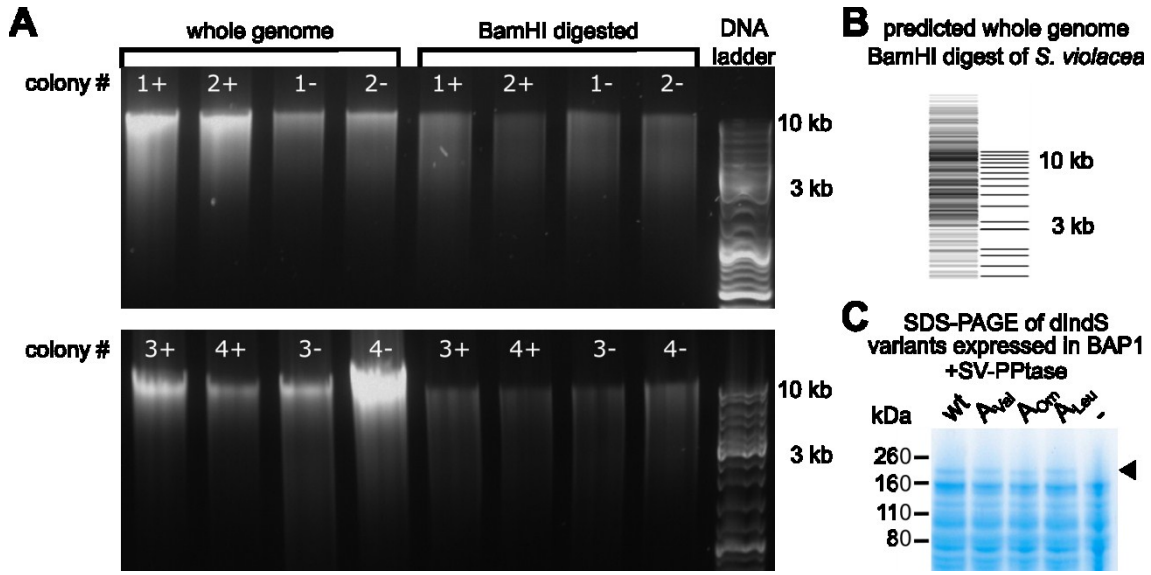

Figure S 8 | Quality control of BamH1 digest of whole S. violacea genomes and control of dIndS variant expression in BAP1. (A) Agarose gel of the intact versus BamHI digested whole genomes of pigment producing (1-4 +) and non-pigment producing (1-4 -) *S. violacea*. For the intact whole genomes, a distinct band at a very high molecular weight is visible, which disappears for the BamHI digested version. This indicates a successful digest. **(B)** Predicted band pattern for the whole genome digest with BamHI. The prediction was done in ApE and the size distribution of resulting genomic fragments ranges between 30 kb and 3 kb, which is what we also observed in A. **(C)** 4-12 % Bis-TRIS SDS-PAGE of dIndS wt and engineered variants expressed in BAP1 along with the putative S. violacea PPtase. The dIndS wt and engineered versions where the dodecanoic acid specific A domain was replaced by TycC A domains specific for Val, Orn or Leu, respectively, are indicated by the black arrow. For the negative control, empty pTrc99 and pCDFDuet vectors were transformed.

6.3.4 *Context-dependent activity of A domains in the tyrocidine synthetase*

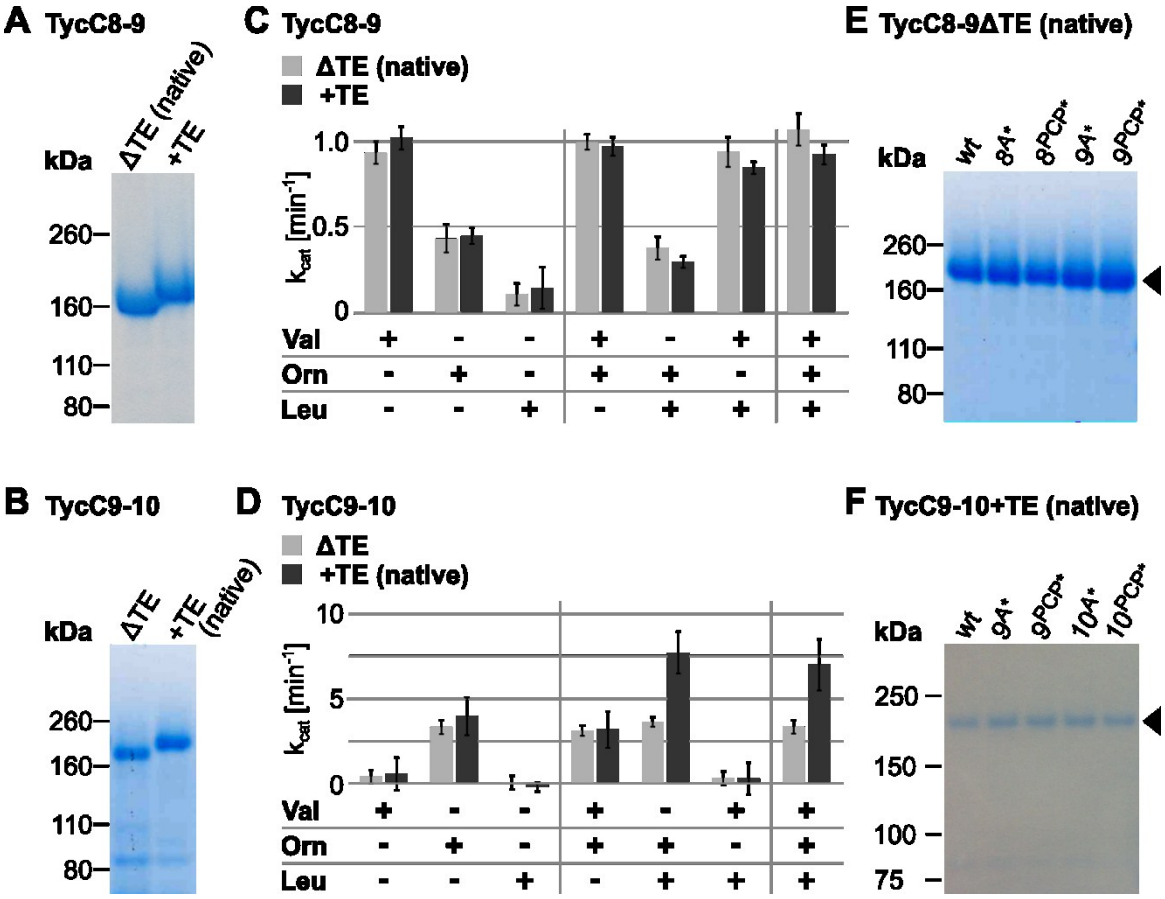

Figure S 9 | Activity of A domains within "wt" dimodular constructs of TycC8-9 and TycC9-10 plus or minus TE domain. (A, B and E, F) Coomassie-stained SDS-gels showing the indicated purified proteins. **(C, D)** Bar graphs showing k_{cat} values for the indicated constructs in the presence of the indicated amino acids, calculated from the online PP_i release assay after background (=no substrate) value subtraction. Amino acids were added at 1 mM, enzymes at 0.5 µM, except for TycC9-10ΔTE, which was added at 0.25 µM. Data represent the mean (± standard deviation, SD) of three independent experiments.

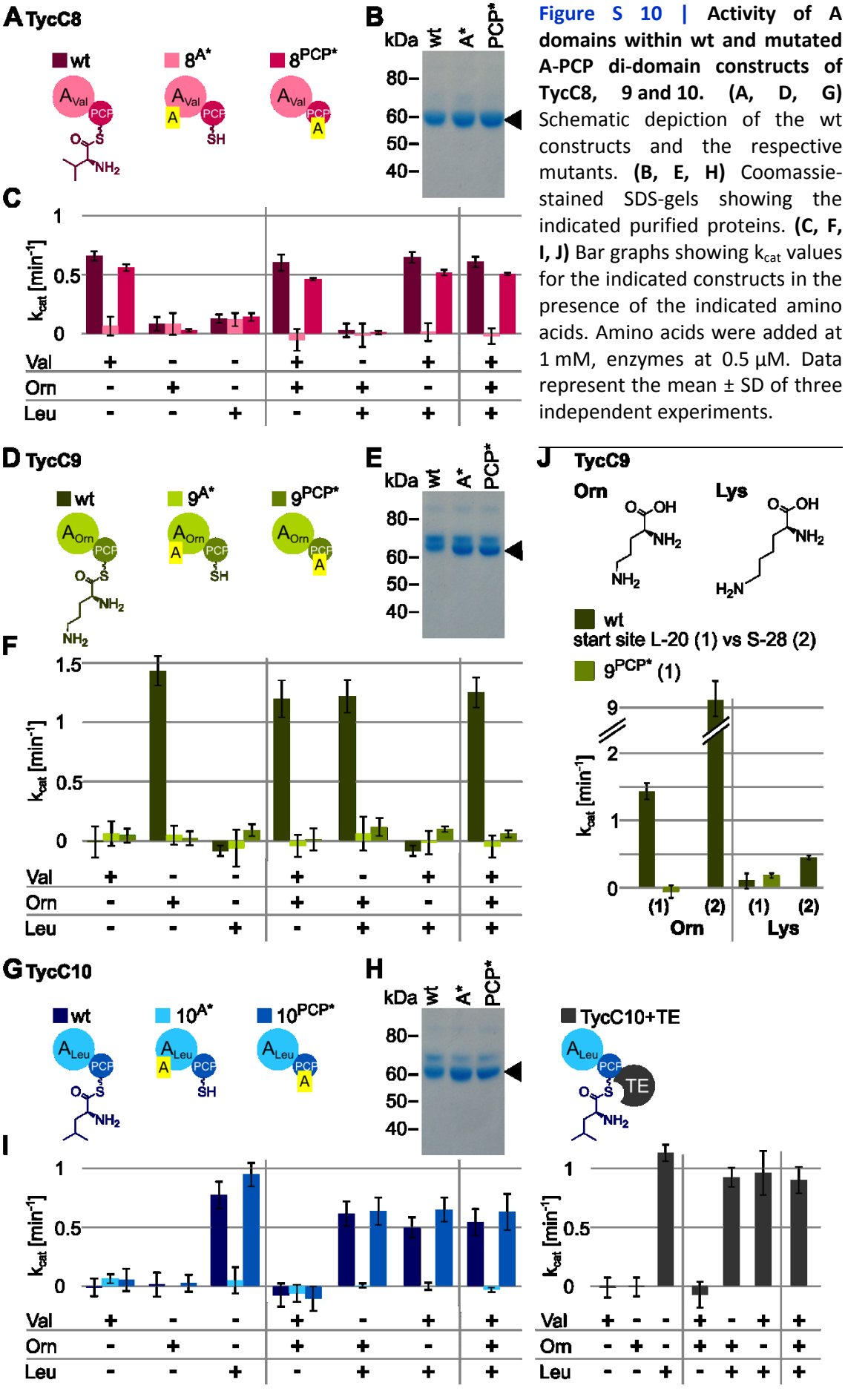

Figure S 10 | Activity of A domains within wt and mutated A-PCP di-domain constructs of TycC8, 9 and 10. (A, D, G) Schematic depiction of the wt constructs and the respective mutants. **(B, E, H)** Coomassie-stained SDS-gels showing the indicated purified proteins. **(C, F, I, J)** Bar graphs showing k_{cat} values for the indicated constructs in the presence of the indicated amino acids. Amino acids were added at 1 mM, enzymes at 0.5 µM. Data represent the mean ± SD of three independent experiments.

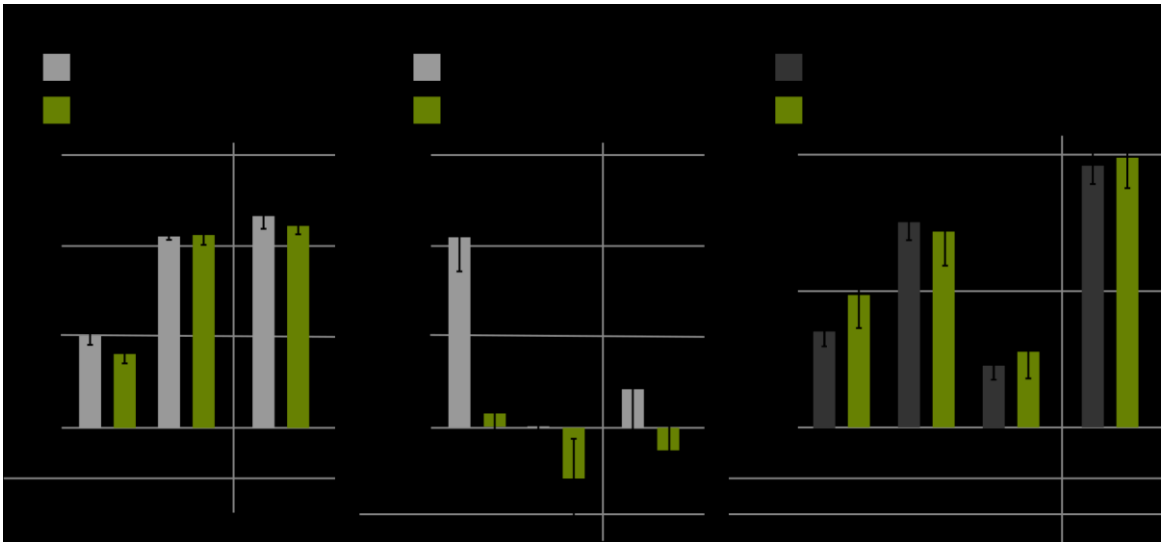

Figure S 11 | Activity of A domains of different di-, and trimodular TycC constructs when replacing the native substrate Orn by the similar Lys. (A-C) Bar graphs showing kcat values for the indicated constructs in the presence of the indicated amino acids, calculated using the online PPi release assay after subtraction of the background (=no substrate) value. Amino acids were added at 1 mM, enzymes at 0.5 µM, except TycC9-10ΔTE, which was added at 0.25 µM. Data represent the mean (± standard deviation, SD) of three independent experiments.

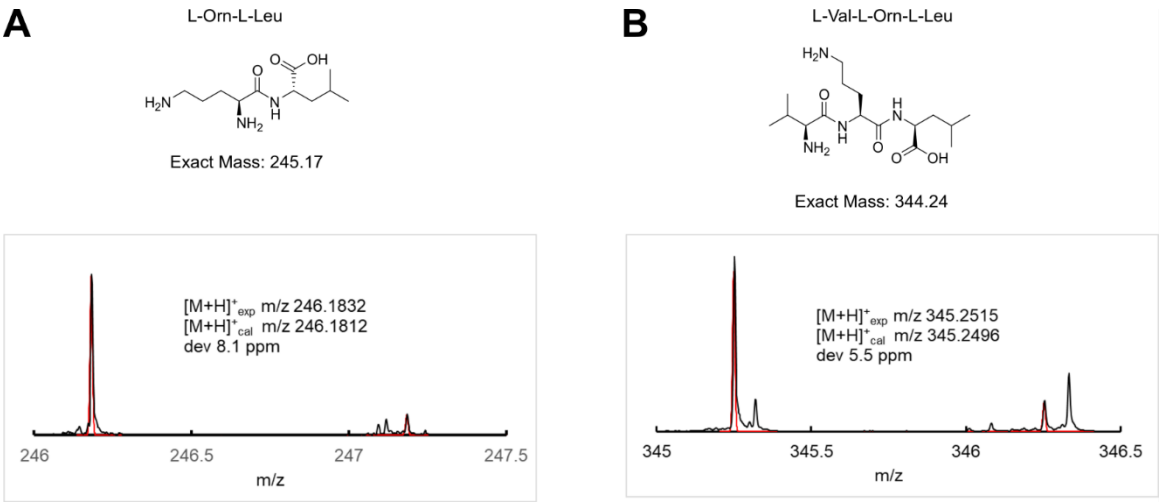

Figure S 12 | Product formation measured by mass spectrometry. (A) MS spectrum of the peptide product Orn-Leu of TycC9-10+TE. **(B)** MS spectrum of the peptide product Val-Orn-Leu of TycC8-9-10+TE. Only in the presence of enzyme, ATP, and all substrates, the peptides could be detected. Black represents experimental data and red a simulation using mMass for the expected signal. The measured signal shows strong resemblance of the simulation indicating product formation.

7. List of Abbreviations

7.1 Amino acids

A - Ala	alanine	H - His	histidine	S- Ser	serine
R - Arg	arginine	I - Ile	isoleucine	T - Thr	threonine
N - Asn	asparagine	L - Leu	leucine	W- Trp	tryptophane
D - Asp	aspartic acid	K - Lys	lysine	Y - Tyr	tyrosine
C - Cys	cysteine	M - Met	methionine	V - Val	valine
G - Gly	glycine	P -Pro	proline		
E - Glu	glutamic acid	F - Phe	phenylalanine	Non-	proteinogenic
Q - Gln	glutamine	U - Sec	selenocysteine	O -Orn	ornithine

7.2 Chemicals

ATP	adenosine triphosphate
BisTris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
$CaCl_2$	calcium chloride
CDA	calcium-dependent antibiotic
DHAP	dihydroxyacetone phosphate
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
F1,6P	fructose-1,6-diphosphate
F6P	d-fructose-6- phosphate
FMN	flavin mononucleotide
GAP	glyceraldehyde-3-phosphate
GDH	glycererophosphate dehydrogenase
$H_{(2)}$	hydrogen
HCl	hydrochloric acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MES	2-(N-morpholino)ethanesulfonic acid
MesG	2-amino-6-mercapto-7-methylpurine ribonucleoside
MOPS	3-(N-morpholino)propanesulfonic acid
NaCl	sodium chloride
NAD(H)	nicotinamide adenine dinucleotide
Ni-NTA	nickel - nitrilotriacetic acid
PNP	purine nucleoside phosphorylase

PP	polypropylene
PPant	4’-phosphopantethein
PP_i	pyrophosphate
PP_iase	pyrophosphatase
PP_i-PFK	phosphofructokinase (pyrophosphate dependent)
PVDF	polyvinylidene fluoride
TA	TRIS-acetate
TEA	TRIS-acetate-EDTA
TEAB	triethylammonium bicarbonate buffer
TPI	triosephosphate isomerase
TRIS	tris(hydroxymethyl)aminomethane

7.3 Units and more

(k)Da	(kilo) Dalton
(m)M	(mili-)Molar
°C	degrees celcius
µ	micro
ATCC	American Type Culture Collection
BBH	bidirectional best hits
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BpsA	blue pigment synthetase A
CCD	charge-coupled device
dIndS	N,N-dodecylindigoidine synthetase
DUF	domain of unknown function
EM	electron microscopy
ESI-MS	electronspray ionization - mass spectrometry
FACS	fluorescence activated cell sorting
g	g-force
IM-2	[(2R,3R,1’R)-2-1’-hydroxybutyl-3-hydroxymethyl-γ-butanolide]
IndC	Indigoidine Synthetase C
IPMB	Institute of Pharmacy and Molecular Biotechnology
IR	inverted repeats
k_{cat}	catalytic constant (kinetic constant of the theoretical maximum rate of catalysis, specific for an enzyme towards a substrate under the conditions tested)
l	liter
m	mili
m/z	atomic mass number / charge number of the ion
min	minutes

MS	mass spectrometry
MWCO	molecular weight cut off
NMR	nuclear magnetic resonance
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
o/n	over night
OD	optical density
ORF	open reading frame
p	pico
PCR	polymerase chain reaction
PKS	polyketide synthetase
PPTase	4’-phosphopantetheinyl transferase
rpm	rounds per minute
s	seconds
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
wt	wild type

8. List of figures

9. List of Tables

10. References

(1) WHO | United Nations high-level meeting on antimicrobial resistance http://www.who.int/antimicrobial-resistance/events/UNGA-meeting-amr-sept2016/en/ (accessed Jun 6, 2017).

(2) WHO | Antimicrobial resistance http://www.who.int/antimicrobial-resistance/en/ (accessed Nov 14, 2018).

(3) World Health Organisation, W. Global Action Plan on Antimicrobial Resistance. May 2016.

(4) Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; McMahon, M. D.; Thomas, M. G. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol. Pharm.* **2008**, *5* (2), 191–211.

(5) Might, M. The Illustrated Guide to a Ph.D.

(6) Weissman, K. J. The Structural Biology of Biosynthetic Megaenzymes. *Nat. Chem. Biol.* **2015**, *11* (9), 660–670.

(7) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.

(8) Bayly, C. L.; Yadav, V. G. Towards Precision Engineering of Canonical Polyketide Synthase Domains: Recent Advances and Future Prospects. *Mol. Basel Switz.* **2017**, *22* (2).

(9) Kalkreuter, E.; Williams, G. J. Engineering Enzymatic Assembly Lines for the Production of New Antimicrobials. *Curr. Opin. Microbiol.* **2018**, *45*, 140–148.

(10) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108–160.

(11) Lee, S. G.; Roskoski, R.; Bauer, K.; Lipmann, F. Purification of the Polyenzymes Responsible for Tyrocidine Synthesis and Their Dissociation into Subunits. *Biochemistry* **1973**, *12* (3), 398–405.

(12) Maiya, S.; Grundmann, A.; Li, S.-M.; Turner, G. The Fumitremorgin Gene Cluster of Aspergillus Fumigatus: Identification of a Gene Encoding Brevianamide F Synthetase. *ChemBioChem* **2006**, *7* (7), 1062–1069.

(13) Scholz-Schroeder, B. K.; Soule, J. D.; Gross, D. C. The SypA, SypS, and SypC Synthetase Genes Encode Twenty-Two Modules Involved in the Nonribosomal Peptide Synthesis of Syringopeptin by Pseudomonas Syringae Pv. Syringae B301D. *Mol. Plant-Microbe Interact. MPMI* **2003**, *16* (4), 271–280.

(14) Arima, K.; Kakinuma, A.; Tamura, G. Surfactin, a Crystalline Peptidelipid Surfactant Produced by Bacillus Subtilis: Isolation, Characterization and Its Inhibition of Fibrin Clot Formation. *Biochem. Biophys. Res. Commun.* **1968**, *31* (3), 488–494.

(15) Kraas, F. I.; Helmetag, V.; Wittmann, M.; Strieker, M.; Marahiel, M. A. Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation. *Chem. Biol.* **2010**, *17* (8), 872–880.

(16) Roongsawang, N.; Washio, K.; Morikawa, M. Diversity of Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Lipopeptide Biosurfactants. *Int. J. Mol. Sci.* **2010**, *12* (1), 141–172.

(17) Schwarzer, D.; Finking, R.; Marahiel, M. A. Nonribosomal Peptides: From Genes to Products. *Nat. Prod. Rep.* **2003**, *20* (3), 275–287.

(18) Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chem. Int. Ed.* **2013**, *52* (28), 7098–7124.

(19) Caboche, S.; Leclère, V.; Pupin, M.; Kucherov, G.; Jacques, P. Diversity of Monomers in Nonribosomal Peptides: Towards the Prediction of Origin and Biological Activity. *J. Bacteriol.* **2010**, *192* (19), 5143–5150.

(20) Agrawal, S.; Acharya, D.; Adholeya, A.; Barrow, C. J.; Deshmukh, S. K. Nonribosomal Peptides from Marine Microbes and Their Antimicrobial and Anticancer Potential. *Front. Pharmacol.* **2017**, *8*, 828.

(21) Galm, U.; Hager, M. H.; Van Lanen, S. G.; Ju, J.; Thorson, J. S.; Shen, B. Antitumor Antibiotics: Bleomycin, Enediynes, and Mitomycin. *Chem. Rev.* **2005**, *105* (2), 739–758.

(22) Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev. MMBR* **2015**, *80* (1), 1–43.

(23) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed Engl.* **2017**, *56* (14), 3770–3821.

(24) Karpiński, T. M.; Adamczak, A. Anticancer Activity of Bacterial Proteins and Peptides. *Pharmaceutics* **2018**, *10* (2).

(25) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed Engl.* **2017**, *56* (14), 3770–3821.

(26) Dubos, R. J.; Hotchkiss, R. D. THE PRODUCTION OF BACTERICIDAL SUBSTANCES BY AEROBIC SPORULATING BACILLI. *J. Exp. Med.* **1941**, *73* (5), 629–640.

(27) Gordon, A. H.; Martin, A. J. P.; Synge, R. L. M. The Amino-Acid Composition of Tyrocidine. *Biochem. J.* **1943**, *37* (3), 313–318.

(28) Mach, B.; Reich, E.; Tatum, E. L. Separation of the Biosynthesis of the Antibiotic Polypeptide Tyrocidine from Protein Biosynthesis. *Proc. Natl. Acad. Sci.* **1963**, *50* (1), 175–181.

(29) Roskoski, R.; Gevers, W.; Kleinkauf, H.; Lipmann, F. Tyrocidine Biosynthesis by Three Complementary Fractions from Bacillus Brevis (ATCC 8185). *Biochemistry* **1970**, *9* (25), 4839–4845.

(30) Roskoski, R.; Kleinkauf, H.; Gevers, W.; Lipmann, F. Isolation of Enzyme-Bound Peptide Intermediates in Tyrocidine Biosynthesis. *Biochemistry* **1970**, *9* (25), 4846–4851.

(31) Mootz, H. D.; Marahiel, M. A. The Tyrocidine Biosynthesis Operon of Bacillus Brevis: Complete Nucleotide Sequence and Biochemical Characterization of Functional Internal Adenylation Domains. *J. Bacteriol.* **1997**, *179* (21), 6843–6850.

(32) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. Catalytic Role of the Condensation Domain. *J. Biol. Chem.* **1998**, *273* (35), 22773–22781.

(33) Belshaw, P. J. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science* **1999**, *284* (5413), 486–489.

(34) Mootz, H. D.; Schwarzer, D.; Marahiel, M. A. Construction of Hybrid Peptide Synthetases by Module and Domain Fusions. *Proc. Natl. Acad. Sci.* **2000**, *97* (11), 5848–5853.

(35) Bergendahl, V.; Linne, U.; Marahiel, M. A. Mutational Analysis of the C-Domain in Nonribosomal Peptide Synthesis. *Eur. J. Biochem. FEBS* **2002**, *269* (2), 620–629.

(36) Linne, U.; Stein, D. B.; Mootz, H. D.; Marahiel, M. A. Systematic and Quantitative Analysis of Protein-Protein Recognition between Nonribosomal Peptide Synthetases Investigated in the Tyrocidine Biosynthetic Template. *Biochemistry* **2003**, *42* (17), 5114–5124.

(37) Schmelz, S.; Naismith, J. H. Adenylate-Forming Enzymes. *Curr. Opin. Struct. Biol.* **2009**, *19* (6), 666–671.

(38) Gevers, W.; Kleinkauf, H.; Lipmann, F. Peptidyl Transfers in Gramicidin S Bisoynthesis from Enzyme-Bound Thioester Intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **1969**, *63* (4), 1335–1342.

(39) Kleinkauf, H.; Gevers, W.; Lipmann, F. Interrelation Between Activation and Polymerization in Gramicidin S Biosynthesis. *Proc. Natl. Acad. Sci.* **1969**, *62* (1), 226–233.
(40) Dieckmann, R.; Lee, Y. O.; van Liempt, H.; von Döhren, H.; Kleinkauf, H. Expression of an Active Adenylate-Forming Domain of Peptide Synthetases Corresponding to Acyl-CoA-Synthetases. *FEBS Lett.* **1995**, *357* (2), 212–216.
(41) Haese, A.; Pieper, R.; von Ostrowski, T.; Zocher, R. Bacterial Expression of Catalytically Active Fragments of the Multifunctional Enzyme Enniatin Synthetase. *J. Mol. Biol.* **1994**, *243* (1), 116–122.
(42) Stachelhaus, T.; Marahiel, M. A. Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme GrsA. *J. Biol. Chem.* **1995**, *270* (11), 6163–6169.
(43) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. Structural Basis for the Activation of Phenylalanine in the Non-Ribosomal Biosynthesis of Gramicidin S. *EMBO J.* **1997**, *16* (14), 4174–4183.
(44) Gulick, A. M.; Starai, V. J.; Horswill, A. R.; Homick, K. M.; Escalante-Semerena, J. C. The 1.75 Å Crystal Structure of Acetyl-CoA Synthetase Bound to Adenosine-5‘-Propylphosphate and Coenzyme A. *Biochemistry* **2003**, *42* (10), 2866–2873.
(45) Du, L.; He, Y.; Luo, Y. Crystal Structure and Enantiomer Selection by D-Alanyl Carrier Protein Ligase DltA from Bacillus Cereus. *Biochemistry* **2008**, *47* (44), 11473–11480.
(46) Lee, T. V.; Johnson, L. J.; Johnson, R. D.; Koulman, A.; Lane, G. A.; Lott, J. S.; Arcus, V. L. Structure of a Eukaryotic Nonribosomal Peptide Synthetase Adenylation Domain That Activates a Large Hydroxamate Amino Acid in Siderophore Biosynthesis. *J. Biol. Chem.* **2010**, *285* (4), 2415–2427.
(47) Mitchell, C. A.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* **2012**, *51* (15), 3252–3263.
(48) Sundlov, J. A.; Shi, C.; Wilson, D. J.; Aldrich, C. C.; Gulick, A. M. Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. *Chem. Biol.* **2012**, *19* (2), 188–198.
(49) Yonus, H.; Neumann, P.; Zimmermann, S.; May, J. J.; Marahiel, M. A.; Stubbs, M. T. Crystal Structure of DltA. Implications for the Reaction Mechanism of Non-Ribosomal Peptide Synthetase Adenylation Domains. *J. Biol. Chem.* **2008**, *283* (47), 32484–32491.
(50) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Leigh Allen, C.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. *Nature* **2016**, *529* (7585), 235–238.
(51) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97* (7), 2651–2674.
(52) Wu, R.; Cao, J.; Lu, X.; Reger, A. S.; Gulick, A. M.; Dunaway-Mariano, D. Mechanism of 4-Chlorobenzoate: Coenzyme A Ligase Catalysis. *Biochemistry* **2008**, *47* (31), 8026–8039.
(53) Gulick, A. M. Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem. Biol.* **2009**, *4* (10), 811–827.
(54) Miller, B. R.; Sundlov, J. A.; Drake, E. J.; Makin, T. A.; Gulick, A. M. Analysis of the Linker Region Joining the Adenylation and Carrier Protein Domains of the Modular Non-Ribosomal Peptide Synthetases. *Proteins* **2014**, *82* (10), 2691–2702.
(55) Labby, K. J.; Watsula, S. G.; Garneau-Tsodikova, S. Interrupted Adenylation Domains: Unique Bifunctional Enzymes Involved in Nonribosomal Peptide Biosynthesis. *Nat Prod Rep* **2015**, *32* (5), 641–653.
(56) Bučević-Popović, V.; Sprung, M.; Soldo, B.; Pavela-Vrančič, M. The A9 Core Sequence from NRPS Adenylation Domain Is Relevant for Thioester Formation. *Chembiochem Eur. J. Chem. Biol.* **2012**, *13* (13), 1913–1920.

(57) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chem. Biol.* **1999**, *6* (8), 493–505.

(58) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, Structure-Based Model of Amino Acid Recognition by Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Biol.* **2000**, *7* (3), 211–224.

(59) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. Specificity Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Transductive Support Vector Machines (TSVMs). *Nucleic Acids Res.* **2005**, *33* (18), 5799–5808.

(60) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39* (Web Server issue), W362–W367.

(61) Prieto, C.; García-Estrada, C.; Lorenzana, D.; Martín, J. F. NRPSsp: Non-Ribosomal Peptide Synthase Substrate Predictor. *Bioinformatics* **2012**, *28* (3), 426–427.

(62) Baranašić, D.; Zucko, J.; Diminic, J.; Gacesa, R.; Long, P. F.; Cullum, J.; Hranueli, D.; Starcevic, A. Predicting Substrate Specificity of Adenylation Domains of Nonribosomal Peptide Synthetases and Other Protein Properties by Latent Semantic Indexing. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 461–467.

(63) Reger, A. S.; Carney, J. M.; Gulick, A. M. Biochemical and Crystallographic Analysis of Substrate Binding and Conformational Changes in Acetyl-CoA Synthetase. *Biochemistry* **2007**, *46* (22), 6536–6546.

(64) Alfermann, J.; Sun, X.; Mayerthaler, F.; Morrell, T. E.; Dehling, E.; Volkmann, G.; Komatsuzaki, T.; Yang, H.; Mootz, H. D. FRET Monitoring of a Nonribosomal Peptide Synthetase. *Nat. Chem. Biol.* **2017**.

(65) Quadri, L. E.; Sello, J.; Keating, T. A.; Weinreb, P. H.; Walsh, C. T. Identification of a Mycobacterium Tuberculosis Gene Cluster Encoding the Biosynthetic Enzymes for Assembly of the Virulence-Conferring Siderophore Mycobactin. *Chem. Biol.* **1998**, *5* (11), 631–645.

(66) Lautru, S.; Oves-Costales, D.; Pernodet, J.-L.; Challis, G. L. MbtH-like Protein-Mediated Cross-Talk between Non-Ribosomal Peptide Antibiotic and Siderophore Biosynthetic Pathways in Streptomyces Coelicolor M145. *Microbiology* **2007**, *153* (5), 1405–1412.

(67) Drake, E. J.; Cao, J.; Qu, J.; Shah, M. B.; Straubinger, R. M.; Gulick, A. M. The 1.8 Å Crystal Structure of PA2412, an MbtH-like Protein from the Pyoverdine Cluster of Pseudomonas Aeruginosa. *J. Biol. Chem.* **2007**, *282* (28), 20425–20434.

(68) Felnagle, E. A.; Barkei, J. J.; Park, H.; Podevels, A. M.; McMahon, M. D.; Drott, D. W.; Thomas, M. G. MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases†. *Biochemistry* **2010**, *49* (41), 8815–8817.

(69) Davidsen, J. M.; Bartley, D. M.; Townsend, C. A. Non-Ribosomal Propeptide Precursor in Nocardicin A Biosynthesis Predicted from Adenylation Domain Specificity Dependent on the MbtH Family Protein NocI. *J. Am. Chem. Soc.* **2013**, *135* (5), 1749–1759.

(70) Kittilä, T.; Schoppet, M.; Cryle, M. J. Online Pyrophosphate Assay for Analyzing Adenylation Domains of Nonribosomal Peptide Synthetases. *Chembiochem Eur. J. Chem. Biol.* **2016**, *17* (7), 576–584.

(71) Lee, K. S.; Lee, B. M.; Ryu, J. H.; Kim, D. H.; Kim, Y. H.; Lim, S.-K. Increased Vancomycin Production by Overexpression of MbtH-like Protein in Amycolatopsis Orientalis KFCC10990P. *Lett. Appl. Microbiol.* **2016**.

(72) Baltz, R. H. Function of MbtH Homologs in Nonribosomal Peptide Biosynthesis and Applications in Secondary Metabolite Discovery. *J. Ind. Microbiol. Biotechnol.* **2011**, *38* (11), 1747–1760.

(73) Herbst, D. A.; Boll, B.; Zocher, G.; Stehle, T.; Heide, L. Structural Basis of the Interaction of MbtH-like Proteins, Putative Regulators of Nonribosomal Peptide Biosynthesis, with Adenylating Enzymes. *J. Biol. Chem.* **2013**, *288* (3), 1991–2003.

(74) Schomer, R. A.; Park, H.; Barkei, J. J.; Thomas, M. G. Alanine Scanning of YbdZ, an MbtH-like Protein, Reveals Essential Residues for Functional Interactions with Its Nonribosomal Peptide Synthetase Partner EntF. *Biochemistry* **2018**.

(75) Miller, B. R.; Drake, E. J.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-Like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **2016**.

(76) Tarry, M. J.; Haque, A. S.; Bui, K. H.; Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Struct. Lond. Engl. 1993* **2017**, *25* (5), 783-793.e4.

(77) Gehring, A. M.; Mori, I.; Walsh, C. T. Reconstitution and Characterization of the Escherichia Coli Enterobactin Synthetase from EntB, EntE, and EntF. *Biochemistry* **1998**, *37* (8), 2648–2659.

(78) Hoffmann, D.; Hevel, J. M.; Moore, R. E.; Moore, B. S. Sequence Analysis and Biochemical Characterization of the Nostopeptolide A Biosynthetic Gene Cluster from Nostoc Sp. GSV224. *Gene* **2003**, *311*, 171–180.

(79) Kaljunen, H.; Schiefelbein, S. H. H.; Stummer, D.; Kozak, S.; Meijers, R.; Christiansen, G.; Rentmeister, A. Structural Elucidation of the Bispecificity of A Domains as a Basis for Activating Non-Natural Amino Acids. *Angew. Chem. Int. Ed Engl.* **2015**, *54* (30), 8833–8836.

(80) Meyer, S.; Kehr, J.-C.; Mainz, A.; Dehm, D.; Petras, D.; Süssmuth, R. D.; Dittmann, E. Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chem. Biol.* **2016**, *23* (4), 462–471.

(81) Li, R.; Oliver, R. A.; Townsend, C. A. Identification and Characterization of the Sulfazecin Monobactam Biosynthetic Gene Cluster. *Cell Chem. Biol.* **2017**, *24* (1), 24–34.

(82) Lee, S. G.; Lipmann, F. Tyrocidine Synthetase System. *Methods Enzymol.* **1975**, *43*, 585–602.

(83) Linne, U.; Marahiel, M. A. Reactions Catalyzed by Mature and Recombinant Nonribosomal Peptide Synthetases. *Methods Enzymol.* **2004**, *388*, 293–315.

(84) Otten, L. G.; Schaffer, M. L.; Villiers, B. R. M.; Stachelhaus, T.; Hollfelder, F. An Optimized ATP/PP(i)-Exchange Assay in 96-Well Format for Screening of Adenylation Domains for Applications in Combinatorial Biosynthesis. *Biotechnol. J.* **2007**, *2* (2), 232–240.

(85) Phelan, V. V.; Du, Y.; McLean, J. A.; Bachmann, B. O. Adenylation Enzyme Characterization Using γ-18O4-ATP Pyrophosphate Exchange. *Chem. Biol.* **2009**, *16* (5), 473–478.

(86) Kadi, N.; Challis, G. L. Chapter 17. Siderophore Biosynthesis a Substrate Specificity Assay for Nonribosomal Peptide Synthetase-Independent Siderophore Synthetases Involving Trapping of Acyl-Adenylate Intermediates with Hydroxylamine. *Methods Enzymol.* **2009**, *458*, 431–457.

(87) Webb, M. R. A Continuous Spectrophotometric Assay for Inorganic Phosphate and for Measuring Phosphate Release Kinetics in Biological Systems. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (11), 4884–4887.

(88) Upson, R. H.; Haugland, R. P.; Malekzadeh, M. N.; Haugland, R. P. A Spectrophotometric Method to Measure Enzymatic Activity in Reactions That Generate Inorganic Pyrophosphate. *Anal. Biochem.* **1996**, *243* (1), 41–45.

(89) McQuade, T. J.; Shallop, A. D.; Sheoran, A.; Delproposto, J. E.; Tsodikov, O. V.; Garneau-Tsodikova, S. A Nonradioactive High-Throughput Assay for Screening and Characterization of Adenylation Domains for Nonribosomal Peptide Combinatorial Biosynthesis. *Anal. Biochem.* **2009**, *386* (2), 244–250.

(90) Duckworth, B. P.; Wilson, D. J.; Aldrich, C. C. Measurement of Nonribosomal Peptide Synthetase Adenylation Domain Activity Using a Continuous Hydroxylamine Release Assay. *Methods Mol. Biol. Clifton NJ* **2016**, *1401*, 53–61.

(91) Lloyd, A. J.; Thomann, H. U.; Ibba, M.; Söll, D. A Broadly Applicable Continuous Spectrophotometric Assay for Measuring Aminoacyl-TRNA Synthetase Activity. *Nucleic Acids Res.* **1995**, *23* (15), 2886–2892.

(92) Bergmeyer, H. U. [New values for the molar extinction coefficients of NADH and NADPH for the use in routine laboratories (author's transl)]. *Z. Klin. Chem. Klin. Biochem.* **1975**, *13* (11), 507–508.

(93) Volkman, B. F.; Zhang, Q.; Debabov, D. V.; Rivera, E.; Kresheck, G. C.; Neuhaus, F. C. Biosynthesis of D-Alanyl-Lipoteichoic Acid: The Tertiary Structure of Apo-D-Alanyl Carrier Protein. *Biochemistry* **2001**, *40* (27), 7964–7972.

(94) Lohman, J. R.; Ma, M.; Cuff, M. E.; Bigelow, L.; Bearden, J.; Babnigg, G.; Joachimiak, A.; Phillips, G. N.; Shen, B. The Crystal Structure of BlmI as a Model for Nonribosomal Peptide Synthetase Peptidyl Carrier Proteins. *Proteins* **2014**, *82* (7), 1210–1218.

(95) Zimmermann, S.; Pfennig, S.; Neumann, P.; Yonus, H.; Weininger, U.; Kovermann, M.; Balbach, J.; Stubbs, M. T. High-Resolution Structures of the D-Alanyl Carrier Protein (Dcp) DltC from Bacillus Subtilis Reveal Equivalent Conformations of Apo- and Holo-Forms. *FEBS Lett.* **2015**, *589* (18), 2283–2289.

(96) Koglin, A.; Mofid, M. R.; Löhr, F.; Schäfer, B.; Rogov, V. V.; Blum, M.-M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dötsch, V. Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science* **2006**, *312* (5771), 273–276.

(97) Haslinger, K.; Redfield, C.; Cryle, M. J. Structure of the Terminal PCP Domain of the Non-Ribosomal Peptide Synthetase in Teicoplanin Biosynthesis: Structure of PCP7 $_{Tei}$. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (4), 711–721.

(98) Jaremko, M. J.; Lee, D. J.; Opella, S. J.; Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *J. Am. Chem. Soc.* **2015**, *137* (36), 11546–11549.

(99) Harden, B. J.; Frueh, D. P. Molecular Cross-Talk between Nonribosomal Peptide Synthetase Carrier Proteins and Unstructured Linker Regions. *Chembiochem Eur. J. Chem. Biol.* **2017**, *18* (7), 629–632.

(100) Schlumbohm, W.; Stein, T.; Ullrich, C.; Vater, J.; Krause, M.; Marahiel, M. A.; Kruft, V.; Wittmann-Liebold, B. An Active Serine Is Involved in Covalent Substrate Amino Acid Binding at Each Reaction Center of Gramicidin S Synthetase. *J. Biol. Chem.* **1991**, *266* (34), 23135–23141.

(101) Gehring, A. M.; Lambalot, R. H.; Vogel, K. W.; Drueckhammer, D. G.; Walsh, C. T. Ability of Streptomyces Spp. Aryl Carrier Proteins and Coenzyme A Analogs to Serve as Substrates in Vitro for E. Coli Holo-ACP Synthase. *Chem. Biol.* **1997**, *4* (1), 17–24.

(102) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A New Enzyme Superfamily — the Phosphopantetheinyl Transferases. *Chem. Biol.* **1996**, *3* (11), 923–936.

(103) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Posttranslational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31* (1), 61–108.

(104) Reimer, J. M.; Aloise, M. N.; Harrison, P. M.; Martin Schmeing, T. Synthetic Cycle of the Initiation Module of a Formylating Nonribosomal Peptide Synthetase. *Nature* **2016**, *529* (7585), 239–242.

(105) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. Characterization of Sfp, a Bacillus Subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry* **1998**, *37* (6), 1585–1595.

(106) Gaudelli, N. M.; Long, D. H.; Townsend, C. A. β-Lactam Formation by a Non-Ribosomal Peptide Synthetase during Antibiotic Biosynthesis. *Nature* **2015**, *520* (7547), 383–387.
(107) Sunbul, M.; Marshall, N. J.; Zou, Y.; Zhang, K.; Yin, J. Catalytic Turnover-Based Phage Selection for Engineering the Substrate Specificity of Sfp Phosphopantetheinyl Transferase. *J. Mol. Biol.* **2009**, *387* (4), 883–898.
(108) Owen, J. G.; Robins, K. J.; Parachin, N. S.; Ackerley, D. F. A Functional Screen for Recovery of 4'-Phosphopantetheinyl Transferase and Associated Natural Product Biosynthesis Genes from Metagenome Libraries. *Environ. Microbiol.* **2012**, *14* (5), 1198–1209.
(109) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. Coli. *Science* **2001**, *291* (5509), 1790–1792.
(110) Gruenewald, S.; Mootz, H. D.; Stehmeier, P.; Stachelhaus, T. In Vivo Production of Artificial Nonribosomal Peptide Products in the Heterologous Host Escherichia Coli. *Appl. Environ. Microbiol.* **2004**, *70* (6), 3282–3291.
(111) Liu, T.; Mazmouz, R.; Neilan, B. A. An In Vitro and In Vivo Study of Broad-Range Phosphopantetheinyl Transferases for Heterologous Expression of Cyanobacterial Natural Products. *ACS Synth. Biol.* **2018**, *7* (4), 1143–1151.
(112) Izoré, T.; Cryle, M. J. The Many Faces and Important Roles of Protein-Protein Interactions during Non-Ribosomal Peptide Synthesis. *Nat. Prod. Rep.* **2018**.
(113) Kittilä, T.; Mollo, A.; Charkoudian, L. K.; Cryle, M. J. New Structural Data Reveal the Motion of Carrier Proteins in Nonribosomal Peptide Synthesis. *Angew. Chem. Int. Ed.* **2016**, n/a-n/a.
(114) Goodrich, A. C.; Meyers, D. J.; Frueh, D. P. Molecular Impact of Covalent Modifications on Nonribosomal Peptide Synthetase Carrier Protein Communication. *J. Biol. Chem.* **2017**.
(115) Owen, J. G.; Calcott, M. J.; Robins, K. J.; Ackerley, D. F. Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem. Biol.* **2016**, *23* (11), 1395–1406.
(116) Beer, R.; Herbst, K.; Ignatiadis, N.; Kats, I.; Adlung, L.; Meyer, H.; Niopek, D.; Christiansen, T.; Georgi, F.; Kurzawa, N.; et al. Creating Functional Engineered Variants of the Single-Module Non-Ribosomal Peptide Synthetase IndC by T Domain Exchange. *Mol. Biosyst.* **2014**, *10* (7), 1709–1718.
(117) Ullrich, C.; Kluge, B.; Palacz, Z.; Vater, J. Cell-Free Biosynthesis of Surfactin, a Cyclic Lipopeptide Produced by Bacillus Subtilis. *Biochemistry* **1991**, *30* (26), 6503–6508.
(118) Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry †. *Biochemistry* **2006**, *45* (42), 12756–12766.
(119) Meluzzi, D.; Zheng, W. H.; Hensler, M.; Nizet, V.; Dorrestein, P. C. Top-down Mass Spectrometry on Low-Resolution Instruments: Characterization of Phosphopantetheinylated Carrier Domains in Polyketide and Non-Ribosomal Biosynthetic Pathways. *Bioorg. Med. Chem. Lett.* **2008**, *18* (10), 3107–3111.
(120) Bumpus, S. B.; Kelleher, N. L. Accessing Natural Product Biosynthetic Processes by Mass Spectrometry (Truncated Title: MS Analysis of Thiotemplate Biosynthesis). *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 475–482.
(121) Lohr, F.; Jenniches, I.; Frizler, M.; Meehan, M. J.; Sylvester, M.; Schmitz, A.; Gütschow, M.; Dorrestein, P. C.; König, G. M.; Schäberle, T. F. α,β → β,γ Double Bond Migration in Corallopyronin A Biosynthesis. *Chem. Sci.* **2013**, *4* (11), 4175.

(122) Henderson, J. C.; Fage, C. D.; Cannon, J. R.; Brodbelt, J. S.; Keatinge-Clay, A. T.; Trent, M. S. Antimicrobial Peptide Resistance of Vibrio Cholerae Results from an LPS Modification Pathway Related to Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **2014**, *9* (10), 2382–2392.

(123) De Crécy-Lagard, V.; Marlière, P.; Saurin, W. Multienzymatic Non Ribosomal Peptide Biosynthesis: Identification of the Functional Domains Catalysing Peptide Elongation and Epimerisation. *Comptes Rendus Académie Sci. Sér. III Sci. Vie* **1995**, *318* (9), 927–936.

(124) Samel, S. A.; Schoenafinger, G.; Knappe, T. A.; Marahiel, M. A.; Essen, L.-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **2007**, *15* (7), 781–792.

(125) Leslie, A. G. Refined Crystal Structure of Type III Chloramphenicol Acetyltransferase at 1.75 A Resolution. *J. Mol. Biol.* **1990**, *213* (1), 167–186.

(126) Lewendon, A.; Murray, I. A.; Shaw, W. V.; Gibbs, M. R.; Leslie, A. G. Replacement of Catalytic Histidine-195 of Chloramphenicol Acetyltransferase: Evidence for a General Base Role for Glutamate. *Biochemistry* **1994**, *33* (7), 1944–1950.

(127) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. The Structure of VibH Represents Nonribosomal Peptide Synthetase Condensation, Cyclization and Epimerization Domains. *Nat. Struct. Biol.* **2002**, *9* (7), 522–526.

(128) Roche, E. D.; Walsh, C. T. Dissection of the EntF Condensation Domain Boundary and Active Site Residues in Nonribosomal Peptide Synthesis. *Biochemistry* **2003**, *42* (5), 1334–1344.

(129) Linne, U.; Marahiel, M. A. Control of Directionality in Nonribosomal Peptide Synthesis: Role of the Condensation Domain in Preventing Misinitiation and Timing of Epimerization. *Biochemistry* **2000**, *39* (34), 10439–10447.

(130) Bloudoff, K.; Rodionov, D.; Schmeing, T. M. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J. Mol. Biol.* **2013**, *425* (17), 3137–3150.

(131) Ehmann, D. E.; Trauger, J. W.; Stachelhaus, T.; Walsh, C. T. Aminoacyl-SNACs as Small-Molecule Substrates for the Condensation Domains of Nonribosomal Peptide Synthetases. *Chem. Biol.* **2000**, *7* (10), 765–772.

(132) Stein, D. B.; Linne, U.; Marahiel, M. A. Utility of Epimerization Domains for the Redesign of Nonribosomal Peptide Synthetases. *FEBS J.* **2005**, *272* (17), 4506–4520.

(133) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A.; Walsh, C. T. Chirality of Peptide Bond-Forming Condensation Domains in Nonribosomal Peptide Synthetases: The C5 Domain of Tyrocidine Synthetase Is a DCL Catalyst. *Biochemistry* **2003**, *42* (41), 12095–12104.

(134) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. Phylogenetic Analysis of Condensation Domains in NRPS Sheds Light on Their Functional Evolution. *BMC Evol. Biol.* **2007**, *7* (1), 78.

(135) Miao, V.; Coëffet-LeGal, M.-F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M.; et al. Daptomycin Biosynthesis in Streptomyces Roseosporus: Cloning and Analysis of the Gene Cluster and Revision of Peptide Stereochemistry. *Microbiology* **2005**, *151* (5), 1507–1523.

(136) Miao, V.; Brost, R.; Chapple, J.; She, K.; Gal, M.-F. C.-L.; Baltz, R. H. The Lipopeptide Antibiotic A54145 Biosynthetic Gene Cluster from Streptomyces Fradiae. *J. Ind. Microbiol. Biotechnol.* **2006**, *33* (2), 129–140.

(137) Bloudoff, K.; Schmeing, T. M. Structural and Functional Aspects of the Nonribosomal Peptide Synthetase Condensation Domain Superfamily: Discovery, Dissection and Diversity. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2017**, *1865* (11, Part B), 1587–1604.

(138) Samel, S. A.; Czodrowski, P.; Essen, L.-O. Structure of the Epimerization Domain of Tyrocidine Synthetase A. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70* (Pt 5), 1442–1452.
(139) Dowling, D. P.; Kung, Y.; Croft, A. K.; Taghizadeh, K.; Kelly, W. L.; Walsh, C. T.; Drennan, C. L. Structural Elements of an NRPS Cyclization Domain and Its Intermodule Docking Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (44), 12432–12437.
(140) Long, D.; Townsend, C. A. Mechanism of Integrated β-Lactam Formation by a Non-Ribosomal Peptide Synthetase during Antibiotic Synthesis. *Biochemistry* **2018**.
(141) Haslinger, K.; Peschke, M.; Brieke, C.; Maximowitsch, E.; Cryle, M. J. X-Domain of Peptide Synthetases Recruits Oxygenases Crucial for Glycopeptide Biosynthesis. *Nature* **2015**, *advance online publication*.
(142) Gao, X.; Haynes, S. W.; Ames, B. D.; Wang, P.; Vien, L. P.; Walsh, C. T.; Tang, Y. Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-like Domain. *Nat. Chem. Biol.* **2012**, *8* (10), 823–830.
(143) Zhang, J.; Liu, N.; Cacho, R. A.; Gong, Z.; Liu, Z.; Qin, W.; Tang, C.; Tang, Y.; Zhou, J. Structural Basis of Nonribosomal Peptide Macrocyclization in Fungi. *Nat. Chem. Biol.* **2016**, *12* (12), 1001–1003.
(144) Hoffmann, K.; Schneider-Scherzer, E.; Kleinkauf, H.; Zocher, R. Purification and Characterization of Eucaryotic Alanine Racemase Acting as Key Enzyme in Cyclosporin Biosynthesis. *J. Biol. Chem.* **1994**, *269* (17), 12710–12714.
(145) Tang, G.-L.; Cheng, Y.-Q.; Shen, B. Chain Initiation in the Leinamycin-Producing Hybrid Nonribosomal Peptide/Polyketide Synthetase from Streptomyces Atroolivaceus S-140. Discrete, Monofunctional Adenylation Enzyme and Peptidyl Carrier Protein That Directly Load D-Alanine. *J. Biol. Chem.* **2007**, *282* (28), 20273–20282.
(146) von Döhren, H.; Keller, U.; Vater, J.; Zocher, R. Multifunctional Peptide Synthetases. *Chem. Rev.* **1997**, *97* (7), 2675–2706.
(147) Chen, W.-H.; Li, K.; Guntaka, N. S.; Bruner, S. D. Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **2016**, *11* (8), 2293–2303.
(148) Chatterjee, J.; Rechenmacher, F.; Kessler, H. N-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew. Chem. Int. Ed Engl.* **2013**, *52* (1), 254–269.
(149) Zolova, O. E.; Garneau-Tsodikova, S. KtzJ-Dependent Serine Activation and O-Methylation by KtzH for Kutznerides Biosynthesis. *J. Antibiot. (Tokyo)* **2014**, *67* (1), 59–64.
(150) Mori, S.; Garzan, A.; Tsodikov, O. V.; Garneau-Tsodikova, S. Deciphering Nature's Intricate Way of N,S-Dimethylating l-Cysteine: Sequential Action of Two Bifunctional Adenylation Domains. *Biochemistry* **2017**, *56* (46), 6087–6097.
(151) Miller, D. A.; Walsh, C. T.; Luo, L. C-Methyltransferase and Cyclization Domain Activity at the Intraprotein PK/NRP Switch Point of Yersiniabactin Synthetase. *J. Am. Chem. Soc.* **2001**, *123* (34), 8434–8435.
(152) Shi, R.; Lamb, S. S.; Zakeri, B.; Proteau, A.; Cui, Q.; Sulea, T.; Matte, A.; Wright, G. D.; Cygler, M. Structure and Function of the Glycopeptide N-Methyltransferase MtfA, a Tool for the Biosynthesis of Modified Glycopeptide Antibiotics. *Chem. Biol.* **2009**, *16* (4), 401–410.
(153) Ansari, M. Z.; Sharma, J.; Gokhale, R. S.; Mohanty, D. In Silico Analysis of Methyltransferase Domains Involved in Biosynthesis of Secondary Metabolites. *BMC Bioinformatics* **2008**, *9* (1), 454.
(154) Mori, S.; Pang, A. H.; Lundy, T. A.; Garzan, A.; Tsodikov, O. V.; Garneau-Tsodikova, S. Structural Basis for Backbone N-Methylation by an Interrupted Adenylation Domain. *Nat. Chem. Biol.* **2018**, *14* (5), 428–430.

(155) Perlova, O.; Fu, J.; Kuhlmann, S.; Krug, D.; Stewart, A. F.; Zhang, Y.; Müller, R. Reconstitution of the Myxothiazol Biosynthetic Gene Cluster by Red/ET Recombination and Heterologous Expression in Myxococcus Xanthus. *Appl. Environ. Microbiol.* **2006**, *72* (12), 7485–7494.

(156) Schneider, T. L.; Shen, B.; Walsh, C. T. Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation. *Biochemistry* **2003**, *42* (32), 9722–9730.

(157) Chen, H.; O'Connor, S.; Cane, D. E.; Walsh, C. T. Epothilone Biosynthesis: Assembly of the Methylthiazolylcarboxy Starter Unit on the EpoB Subunit. *Chem. Biol.* **2001**, *8* (9), 899–912.

(158) Schneider, A.; Marahiel, M. A. Genetic Evidence for a Role of Thioesterase Domains, Integrated in or Associated with Peptide Synthetases, in Non-Ribosomal Peptide Biosynthesis in Bacillus Subtilis. *Arch. Microbiol.* **1998**, *169* (5), 404–410.

(159) Keating, T. A.; Ehmann, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T. Chain Termination Steps in Nonribosomal Peptide Synthetase Assembly Lines: Directed Acyl-S-Enzyme Breakdown in Antibiotic and Siderophore Biosynthesis. *Chembiochem Eur. J. Chem. Biol.* **2001**, *2* (2), 99–107.

(160) Conway, K. R.; Boddy, C. N. ClusterMine360: A Database of Microbial PKS/NRPS Biosynthesis. *Nucleic Acids Res.* **2013**, *41* (D1), D402–D407.

(161) Hari, T. P. A.; Labana, P.; Boileau, M.; Boddy, C. N. An Evolutionary Model Encompassing Substrate Specificity and Reactivity of Type I Polyketide Synthase Thioesterases. *ChemBioChem* **2014**, *15* (18), 2656–2661.

(162) Kopp, F.; A. Marahiel, M. Macrocyclization Strategies in Polyketide and Nonribosomal Peptide Biosynthesis. *Nat. Prod. Rep.* **2007**, *24* (4), 735–749.

(163) Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. Polyketide Synthase and Non-Ribosomal Peptide Synthetase Thioesterase Selectivity: Logic Gate or a Victim of Fate? *Nat. Prod. Rep.* **2015**.

(164) Samel, S. A.; Wagner, B.; Marahiel, M. A.; Essen, L.-O. The Thioesterase Domain of the Fengycin Biosynthesis Cluster: A Structural Base for the Macrocyclization of a Non-Ribosomal Lipopeptide. *J. Mol. Biol.* **2006**, *359* (4), 876–889.

(165) Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G. Dynamic Thiolation–Thioesterase Structure of a Non-Ribosomal Peptide Synthetase. *Nature* **2008**, *454* (7206), 903–906.

(166) Liu, Y.; Zheng, T.; Bruner, S. D. Structural Basis for Phosphopantetheinyl Carrier Domain Interactions in the Terminal Module of Nonribosomal Peptide Synthetases. *Chem. Biol.* **2011**, *18* (11), 1482–1488.

(167) Koglin, A.; Löhr, F.; Bernhard, F.; Rogov, V. V.; Frueh, D. P.; Strieter, E. R.; Mofid, M. R.; Güntert, P.; Wagner, G.; Walsh, C. T.; et al. Structural Basis for the Selectivity of the External Thioesterase of the Surfactin Synthetase. *Nature* **2008**, *454* (7206), 907–911.

(168) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science* **2008**, *321* (5889), 659–663.

(169) Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T. Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE. *Structure* **2002**, *10* (3), 301–310.

(170) Kohli, R. M.; Walsh, C. T. Enzymology of Acyl Chain Macrocyclization in Natural Product Biosynthesis. *Chem. Commun.* **2003**, *0* (3), 297–307.

(171) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed.; W.H. Freeman: New York, 1985.

(172) Holmquist, M. Alpha/Beta-Hydrolase Fold Enzymes: Structures, Functions and Mechanisms. *Curr. Protein Pept. Sci.* **2000**, *1* (2), 209–235.

(173) Korman, T. P.; Crawford, J. M.; Labonte, J. W.; Newman, A. G.; Wong, J.; Townsend, C. A.; Tsai, S.-C. Structure and Function of an Iterative Polyketide Synthase Thioesterase Domain Catalyzing Claisen Cyclization in Aflatoxin Biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (14), 6246–6251.
(174) Gaudelli, N. M.; Townsend, C. A. Epimerization and Substrate Gating by a TE Domain in β-Lactam Antibiotic Biosynthesis. *Nat. Chem. Biol.* **2014**, *10* (4), 251–258.
(175) Rizo, J.; Gierasch, L. M. Constrained Peptides: Models of Bioactive Peptides and Protein Substructures. *Annu. Rev. Biochem.* **1992**, *61*, 387–418.
(176) Baldwin, A. J.; Kay, L. E. NMR Spectroscopy Brings Invisible Protein States into Focus. *Nat. Chem. Biol.* **2009**, *5* (11), 808–814.
(177) Heikinheimo, P.; Goldman, A.; Jeffries, C.; Ollis, D. L. Of Barn Owls and Bankers: A Lush Variety of Alpha/Beta Hydrolases. *Struct. Lond. Engl. 1993* **1999**, *7* (6), R141-146.
(178) Meier, J. L.; Barrows-Yano, T.; Foley, T. L.; Wike, C. L.; Burkart, M. D. The Unusual Macrocycle Forming Thioesterase of Mycolactone. *Mol. Biosyst.* **2008**, *4* (6), 663–671.
(179) Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. Regeneration of Misprimed Nonribosomal Peptide Synthetases by Type II Thioesterases. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (22), 14083–14088.
(180) Kotowska, M.; Pawlik, K. Roles of Type II Thioesterases and Their Application for Secondary Metabolite Yield Improvement. *Appl. Microbiol. Biotechnol.* **2014**, *98* (18), 7735–7746.
(181) Li, J.; Jaitzig, J.; Theuer, L.; Legala, O. E.; Süssmuth, R. D.; Neubauer, P. Type II Thioesterase Improves Heterologous Biosynthesis of Valinomycin in Escherichia Coli. *J. Biotechnol.* **2015**, *193*, 16–22.
(182) Hahn, M.; Stachelhaus, T. Selective Interaction between Nonribosomal Peptide Synthetases Is Facilitated by Short Communication-Mediating Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (44), 15585–15590.
(183) Hahn, M.; Stachelhaus, T. Harnessing the Potential of Communication-Mediating Domains for the Biocombinatorial Synthesis of Nonribosomal Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (2), 275–280.
(184) Dehling, E.; Volkmann, G.; Matern, J. C. J.; Dörner, W.; Alfermann, J.; Diecker, J.; Mootz, H. D. Mapping of the Communication-Mediating Interface in Nonribosomal Peptide Synthetases Using a Genetically Encoded Photocrosslinker Supports an Upside-Down Helix-Hand Motif. *J. Mol. Biol.* **2016**, *428* (21), 4345–4360.
(185) Koglin, A.; Walsh, C. T. Structural Insights into Nonribosomal Peptide Enzymatic Assembly Lines. *Nat. Prod. Rep.* **2009**, *26* (8), 987–1000.
(186) Hoyer, K. M.; Mahlert, C.; Marahiel, M. A. The Iterative Gramicidin s Thioesterase Catalyzes Peptide Ligation and Cyclization. *Chem. Biol.* **2007**, *14* (1), 13–22.
(187) Yu, D.; Xu, F.; Zhang, S.; Zhan, J. Decoding and Reprogramming Fungal Iterative Nonribosomal Peptide Synthetases. *Nat. Commun.* **2017**, *8*, 15349.
(188) Haynes, S. W.; Challis, G. L. Non-Linear Enzymatic Logic in Natural Product Modular Mega-Synthases and -Synthetases. *Curr. Opin. Drug Discov. Devel.* **2007**, *10* (2), 203–218.
(189) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (25), 9259–9264.
(190) Ali, H.; Ries, M. I.; Lankhorst, P. P.; van der Hoeven, R. A. M.; Schouten, O. L.; Noga, M.; Hankemeier, T.; van Peij, N. N. M. E.; Bovenberg, R. A. L.; Vreeken, R. J.; et al. A Non-Canonical NRPS Is Involved in the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in Penicillium Chrysogenum. *PLoS ONE* **2014**, *9* (6).

(191) Muro, U. C.; Gracida, R. J.; Abreu, C. A.; Arana, C. A.; Téllez, J. A. Pigments from Fungi, an Opportunity of Production for Diverse Applications. *Biologia (Bratisl.)* **2016**, *71* (10), 1067–1079.

(192) Caro, Y.; Venkatachalam, M.; Lebeau, J.; Fouillaud, M.; Dufossé, L. Pigments and Colorants from Filamentous Fungi. In *Fungal Metabolites*; Mérillon, J.-M., Ramawat, K. G., Eds.; Reference Series in Phytochemistry; Springer International Publishing: Cham, 2017; pp 499–568.

(193) Narsing Rao, M. P.; Xiao, M.; Li, W.-J. Fungal and Bacterial Pigments: Secondary Metabolites with Wide Applications. *Front. Microbiol.* **2017**, *8*.

(194) Newsome, A. G.; Culver, C. A.; van Breemen, R. B. Nature's Palette: The Search for Natural Blue Colorants. *J. Agric. Food Chem.* **2014**, *62* (28), 6498–6511.

(195) Newsome, A. G.; Culver, C. A.; van Breemen, R. B. Nature's Palette: The Search for Natural Blue Colorants. *J. Agric. Food Chem.* **2014**, *62* (28), 6498–6511.

(196) Claessen, H. Über Einen Indigoblauen Farbstoff Erzeugenden Bacillus Aus Wasser. **1890**, *Zbl. Bakt.* (I. Abt. Orig. VII), 13–17.

(197) Kuhn, R.; Starr, M. P.; Kuhn, D. A.; Bauer, H.; Knackmuss, H. J. INDIGOIDINE AND OTHER BACTERIAL PIGMENTS RELATED TO 3,3'-BIPYRIDYL. *Arch. Mikrobiol.* **1965**, *51*, 71–84.

(198) Kuhn, R.; Bauer, H.; Knackmuss, H.-J. Struktur Und Synthesen Des Bakterienfarbstoffs Indigoidin. *Chem. Ber.* **1965**, *98* (7), 2139–2153.

(199) Müller, M.; Ausländer, S.; Ausländer, D.; Kemmer, C.; Fussenegger, M. A Novel Reporter System for Bacterial and Mammalian Cells Based on the Non-Ribosomal Peptide Indigoidine. *Metab. Eng.* **2012**, *14* (4), 325–335.

(200) Starr, M. P.; Cosens, G.; Knackmuss, H. J. Formation of the Blue Pigment Indigoidine by Phytopathogenic Erwinia. *Appl. Microbiol.* **1966**, *14* (6), 870–872.

(201) Reverchon, S.; Rouanet, C.; Expert, D.; Nasser, W. Characterization of Indigoidine Biosynthetic Genes in Erwinia Chrysanthemi and Role of This Blue Pigment in Pathogenicity. *J. Bacteriol.* **2002**, *184* (3), 654–665.

(202) Takahashi, H.; Kumagai, T.; Kitani, K.; Mori, M.; Matoba, Y.; Sugiyama, M. Cloning and Characterization of a Streptomyces Single Module Type Non-Ribosomal Peptide Synthetase Catalyzing a Blue Pigment Synthesis. *J. Biol. Chem.* **2007**, *282* (12), 9073–9081.

(203) Andreani, N. A.; Carraro, L.; Martino, M. E.; Fondi, M.; Fasolato, L.; Miotto, G.; Magro, M.; Vianello, F.; Cardazzo, B. A Genomic and Transcriptomic Approach to Investigate the Blue Pigment Phenotype in Pseudomonas Fluorescens. *Int. J. Food Microbiol.* **2015**, *213*, 88–98.

(204) McFadden, B. A.; Howes, W. V. PSEUDOMONAS INDIGOFERA. *J. Bacteriol.* **1961**, *81* (6), 858–862.

(205) Brachmann, A. O.; Kirchner, F.; Kegler, C.; Kinski, S. C.; Schmitt, I.; Bode, H. B. Triggering the Production of the Cryptic Blue Pigment Indigoidine from Photorhabdus Luminescens. *J. Biotechnol.* **2012**, *157* (1), 96–99.

(206) Gibson, Q. H.; Hastings, J. W. The Oxidation of Reduced Flavin Mononucleotide by Molecular Oxygen. *Biochem. J.* **1962**, *83*, 368–377.

(207) Kurniawan, Y. N.; Kitani, S.; Maeda, A.; Nihira, T. Differential Contributions of Two SARP Family Regulatory Genes to Indigoidine Biosynthesis in Streptomyces Lavendulae FRI-5. *Appl. Microbiol. Biotechnol.* **2014**, *98* (23), 9713–9721.

(208) Pait, I. G. U.; Kitani, S.; Kurniawan, Y. N.; Asa, M.; Iwai, T.; Ikeda, H.; Nihira, T. Identification and Characterization of LbpA, an Indigoidine Biosynthetic Gene in the γ-Butyrolactone Signaling System of Streptomyces Lavendulae FRI-5. *J. Biosci. Bioeng.* **2017**.

(209) Cude, W. N.; Mooney, J.; Tavanaei, A. A.; Hadden, M. K.; Frank, A. M.; Gulvik, C. A.; May, A. L.; Buchan, A. Production of the Antimicrobial Secondary Metabolite Indigoidine

Contributes to Competitive Surface Colonization by the Marine Roseobacter Phaeobacter Sp. Strain Y4I. *Appl. Environ. Microbiol.* **2012**, *78* (14), 4771–4780.

(210) Gromek, S. M.; Suria, A. M.; Fullmer, M. S.; Garcia, J. L.; Gogarten, J. P.; Nyholm, S. V.; Balunas, M. J. Leisingera Sp. JC1, a Bacterial Isolate from Hawaiian Bobtail Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios. *Front. Microbiol.* **2016**, *7*.

(211) Day, P. A.; Villalba, M. S.; Herrero, O. M.; Arancibia, L. A.; Alvarez, H. M. Formation of Indigoidine Derived-Pigments Contributes to the Adaptation of Vogesella Sp. Strain EB to Cold Aquatic Iron-Oxidizing Environments. *Antonie Van Leeuwenhoek* **2016**, 1–14.

(212) Xie, Z.; Zhang, Z.; Cao, Z.; Chen, M.; Li, P.; Liu, W.; Qin, H.; Zhao, X.; Tao, Y.; Chen, Y. An External Substrate-Free Blue/White Screening System in Escherichia Coli. *Appl. Microbiol. Biotechnol.* **2017**.

(213) Ji, C.-H.; Kim, J.-P.; Kang, H.-S. Library of Synthetic Streptomyces Regulatory Sequences for Use in Promoter Engineering of Natural Product Biosynthetic Gene Clusters. *ACS Synth. Biol.* **2018**.

(214) Sun, Y.-Q.; Busche, T.; Rückert, C.; Paulus, C.; Rebets, Y.; Novakova, R.; Kalinowski, J.; Luzhetskyy, A.; Kormanec, J.; Sekurova, O. N.; et al. Development of a Biosensor Concept to Detect the Production of Cluster-Specific Secondary Metabolites. *ACS Synth. Biol.* **2017**.

(215) Brown, A. S.; Robins, K. J.; Ackerley, D. F. A Sensitive Single-Enzyme Assay System Using the Non-Ribosomal Peptide Synthetase BpsA for Measurement of L-Glutamine in Biological Samples. *Sci. Rep.* **2017**, *7*, 41745.

(216) Calcott, M. J.; Owen, J. G.; Lamont, I. L.; Ackerley, D. F. Biosynthesis of Novel Pyoverdines by Domain Substitution in a Nonribosomal Peptide Synthetase of Pseudomonas Aeruginosa. *Appl. Environ. Microbiol.* **2014**, *80* (18), 5723–5731.

(217) Kobayashi, H.; Nogi, Y.; Horikoshi, K. New Violet 3,3'-Bipyridyl Pigment Purified from Deep-Sea Microorganism Shewanella Violacea DSS12. *Extremophiles* **2007**, *11* (2), 245–250.

(218) Nogi, Y.; Kato, C.; Horikoshi, K. Taxonomic Studies of Deep-Sea Barophilic Shewanella Strains and Description of Shewanella Violacea Sp. Nov. *Arch. Microbiol.* **1998**, *170* (5), 331–338.

(219) Aono, E.; Baba, T.; Ara, T.; Nishi, T.; Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Hasegawa, M.; Takai, Y.; Okumura, Y.; et al. Complete Genome Sequence and Comparative Analysis of Shewanella Violacea, a Psychrophilic and Piezophilic Bacterium from Deep Sea Floor Sediments. *Mol. Biosyst.* **2010**, *6* (7), 1216–1226.

(220) Grossart, H.-P.; Thorwest, M.; Plitzko, I.; Brinkhoff, T.; Simon, M.; Zeeck, A. Production of a Blue Pigment (Glaukothalin) by Marine *Rheinheimera* Spp. *Int. J. Microbiol.* **2009**, *2009*, 1–7.

(221) Herbst, K. *Characterization of an Hypothetical Gene from the Extromophile Shewanella Violacea DSS12*; Protocol; 2015.

(222) Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C. J.; Lu, S.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; et al. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45* (Database issue), D200–D203.

(223) Marahiel, M. A. A Structural Model for Multimodular NRPS Assembly Lines. *Nat. Prod. Rep.* **2015**.

(224) Reimer, J. M.; Haque, A. S.; Tarry, M. J.; Schmeing, T. M. Piecing Together Nonribosomal Peptide Synthesis. *Curr. Opin. Struct. Biol.* **2018**, *49*, 104–113.

(225) Sidebottom, A. M.; Carlson, E. E. A Reinvigorated Era of Bacterial Secondary Metabolite Discovery. *Curr. Opin. Chem. Biol.* **2015**, *24*, 104–111.

(226) Woodhouse, J. N.; Fan, L.; Brown, M. V.; Thomas, T.; Neilan, B. A. Deep Sequencing of Non-Ribosomal Peptide Synthetases and Polyketide Synthases from the Microbiomes of Australian Marine Sponges. *ISME J.* **2013**, *7* (9), 1842–1851.

(227) Bourbouli, M.; Katsifas, E. A.; Papathanassiou, E.; Karagouni, A. D. The Kolumbo Submarine Volcano of Santorini Island Is a Large Pool of Bacterial Strains with Antimicrobial Activity. *Arch. Microbiol.* **2015**.

(228) Graça, A. P.; Calisto, R.; Lage, O. M. Planctomycetes as Novel Source of Bioactive Molecules. *Front. Microbiol.* **2016**, *7*, 1241.

(229) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; et al. A New Antibiotic Kills Pathogens without Detectable Resistance. *Nature* **2015**.

(230) Maffioli, S. I.; Zhang, Y.; Degen, D.; Carzaniga, T.; Del Gatto, G.; Serina, S.; Monciardini, P.; Mazzetti, C.; Guglierame, P.; Candiani, G.; et al. Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **2017**, *169* (7), 1240-1248.e23.

(231) Wang, X.; Zhang, X.; Liu, L.; Xiang, M.; Wang, W.; Sun, X.; Che, Y.; Guo, L.; Liu, G.; Guo, L.; et al. Genomic and Transcriptomic Analysis of the Endophytic Fungus Pestalotiopsis Fici Reveals Its Lifestyle and High Potential for Synthesis of Natural Products. *BMC Genomics* **2015**, *16* (1), 28.

(232) Helfrich, E. J. N.; Vogel, C. M.; Ueoka, R.; Schäfer, M.; Ryffel, F.; Müller, D. B.; Probst, S.; Kreuzer, M.; Piel, J.; Vorholt, J. A. Bipartite Interactions, Antibiotic Production and Biosynthetic Potential of the Arabidopsis Leaf Microbiome. *Nat. Microbiol.* **2018**, *3* (8), 909–919.

(233) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. Discovery of a New Peptide Natural Product by Streptomyces Coelicolor Genome Mining. *Nat. Chem. Biol.* **2005**, *1* (5), 265–269.

(234) Iftime, D.; Kulik, A.; Härtner, T.; Rohrer, S.; Niedermeyer, T. H. J.; Stegmann, E.; Weber, T.; Wohlleben, W. Identification and Activation of Novel Biosynthetic Gene Clusters by Genome Mining in the Kirromycin Producer Streptomyces Collinus Tü 365. *J. Ind. Microbiol. Biotechnol.* **2015**.

(235) Carro, L.; Nouioui, I.; Sangal, V.; Meier-Kolthoff, J. P.; Trujillo, M. E.; Montero-Calasanz, M. D. C.; Sahin, N.; Smith, D. L.; Kim, K. E.; Peluso, P.; et al. Genome-Based Classification of Micromonosporae with a Focus on Their Biotechnological and Ecological Potential. *Sci. Rep.* **2018**, *8* (1), 525.

(236) Schrimpe-Rutledge, A. C.; Sherrod, S. D.; McLean, J. A. Improving the Discovery of Secondary Metabolite Natural Products Using Ion Mobility–Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 160–166.

(237) Flissi, A.; Dufresne, Y.; Michalik, J.; Tonon, L.; Janot, S.; Noé, L.; Jacques, P.; Leclère, V.; Pupin, M. Norine, the Knowledgebase Dedicated to Non-Ribosomal Peptides, Is Now Open to Crowdsourcing. *Nucleic Acids Res.* **2016**, *44* (D1), D1113-1118.

(238) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Res.* **2011**, *39* (suppl_2), W339–W346.

(239) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; et al. AntiSMASH 3.0—a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2015**, gkv437.

(240) Jaitzig, J.; Li, J.; Süssmuth, R. D.; Neubauer, P. Reconstituted Biosynthesis of the Nonribosomal Macrolactone Antibiotic Valinomycin in Escherichia Coli. *ACS Synth. Biol.* **2014**, *3* (7), 432–438.

(241) Watanabe, K.; Hotta, K.; Praseuth, A. P.; Koketsu, K.; Migita, A.; Boddy, C. N.; Wang, C. C. C.; Oguri, H.; Oikawa, H. Total Biosynthesis of Antitumor Nonribosomal Peptides in Escherichia Coli. *Nat. Chem. Biol.* **2006**, *2* (8), 423–428.

(242) Praseuth, A. P.; Wang, C. C. C.; Watanabe, K.; Hotta, K.; Oguri, H.; Oikawa, H. Complete Sequence of Biosynthetic Gene Cluster Responsible for Producing Triostin A and Evaluation of Quinomycin-Type Antibiotics from Streptomyces Triostinicus. *Biotechnol. Prog.* **2008**, *24* (6), 1226–1231.

(243) Tang, Y.; Frewert, S.; Harmrolfs, K.; Herrmann, J.; Karmann, L.; Kazmaier, U.; Xia, L.; Zhang, Y.; Müller, R. Heterologous Expression of an Orphan NRPS Gene Cluster from Paenibacillus Larvae in Escherichia Coli Revealed Production of Sevadicin. *J. Biotechnol.* **2015**, *194*, 112–114.

(244) Richter, L.; Wanka, F.; Boecker, S.; Storm, D.; Kurt, T.; Vural, Ö.; Süßmuth, R.; Meyer, V. Engineering of Aspergillus Niger for the Production of Secondary Metabolites. *Fungal Biol. Biotechnol.* **2014**, *1*, 4.

(245) Heneghan, M. N.; Yakasai, A. A.; Halo, L. M.; Song, Z.; Bailey, A. M.; Simpson, T. J.; Cox, R. J.; Lazarus, C. M. First Heterologous Reconstruction of a Complete Functional Fungal Biosynthetic Multigene Cluster. *Chembiochem Eur. J. Chem. Biol.* **2010**, *11* (11), 1508–1512.

(246) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345.

(247) Kouprina, N.; Larionov, V. Transformation-Associated Recombination (TAR) Cloning for Genomics Studies and Synthetic Biology. *Chromosoma* **2016**, *125* (4), 621–632.

(248) Joska, T. M.; Mashruwala, A.; Boyd, J. M.; Belden, W. J. A Universal Cloning Method Based on Yeast Homologous Recombination That Is Simple, Efficient, and Versatile. *J. Microbiol. Methods* **2014**, *100*, 46–51.

(249) Birch, A. J. The Biosynthesis of Antibiotics. *Pure Appl. Chem.* **1963**, *7* (4).

(250) Rinehart, K. L. Mutasynthesis of New Antibiotics. *Pure Appl. Chem.* **1977**, *49* (9), 1361–1384.

(251) Stachelhaus, T.; Schneider, A.; Marahiel, M. A. Rational Design of Peptide Antibiotics by Targeted Replacement of Bacterial and Fungal Domains. *Science* **1995**, *269* (5220), 69–72.

(252) Eppelmann, K.; Stachelhaus, T.; Marahiel, M. A. Exploitation of the Selectivity-Conferring Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics. *Biochemistry* **2002**, *41* (30), 9718–9726.

(253) Uguru, G. C.; Milne, C.; Borg, M.; Flett, F.; Smith, C. P.; Micklefield, J. Active-Site Modifications of Adenylation Domains Lead to Hydrolysis of Upstream Nonribosomal Peptidyl Thioester Intermediates. *J. Am. Chem. Soc.* **2004**, *126* (16), 5032–5033.

(254) Evans, B. S.; Chen, Y.; Metcalf, W. W.; Zhao, H.; Kelleher, N. L. Directed Evolution of the Nonribosomal Peptide Synthetase AdmK Generates New Andrimid Derivatives in Vivo. *Chem. Biol.* **2011**, *18* (5), 601–607.

(255) Zhang, K.; Nelson, K. M.; Bhuripanyo, K.; Grimes, K. D.; Zhao, B.; Aldrich, C. C.; Yin, J. Engineering the Substrate Specificity of the DhbE Adenylation Domain by Yeast Cell Surface Display. *Chem. Biol.* **2013**, *20* (1), 92–101.

(256) Villiers, B.; Hollfelder, F. Directed Evolution of a Gatekeeper Domain in Nonribosomal Peptide Synthesis. *Chem. Biol.* **2011**, *18* (10), 1290–1299.

(257) Kries, H.; Wachtel, R.; Pabst, A.; Wanner, B.; Niquille, D.; Hilvert, D. Reprogramming Nonribosomal Peptide Synthetases for “Clickable” Amino Acids. *Angew. Chem. Int. Ed.* **2014**, *53* (38), 10105–10108.

(258) Niquille, D. L.; Hansen, D. A.; Mori, T.; Fercher, D.; Kries, H.; Hilvert, D. Nonribosomal Biosynthesis of Backbone-Modified Peptides. *Nat. Chem.* **2018**, *10* (3), 282–287.

(259) Han, J. W.; Kim, E. Y.; Lee, J. M.; Kim, Y. S.; Bang, E.; Kim, B. S. Site-Directed Modification of the Adenylation Domain of the Fusaricidin Nonribosomal Peptide Synthetase for Enhanced Production of Fusaricidin Analogs. *Biotechnol. Lett.* **2012**, *34* (7), 1327–1334.

(260) Crüsemann, M.; Kohlhaas, C.; Piel, J. Evolution-Guided Engineering of Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Sci.* **2013**, *4* (3), 1041–1045.

(261) Kries, H.; Niquille, D. L.; Hilvert, D. A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides. *Chem. Biol.* **2015**, *22* (5), 640–648.

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

(263) Schauwecker, F.; Pfennig, F.; Grammel, N.; Keller, U. Construction and in Vitro Analysis of a New Bi-Modular Polypeptide Synthetase for Synthesis of N-Methylated Acyl Peptides. *Chem. Biol.* **2000**, *7* (4), 287–297.

(264) Duerfahrt, T.; Doekel, S.; Sonke, T.; Quaedflieg, P. J. L. M.; Marahiel, M. A. Construction of Hybrid Peptide Synthetases for the Production of Alpha-l-Aspartyl-l-Phenylalanine, a Precursor for the High-Intensity Sweetener Aspartame. *Eur. J. Biochem. FEBS* **2003**, *270* (22), 4555–4563.

(265) Bozhüyük, K. A. J.; Fleischhacker, F.; Linck, A.; Wesche, F.; Tietze, A.; Niesert, C.-P.; Bode, H. B. De Novo Design and Engineering of Non-Ribosomal Peptide Synthetases. *Nat. Chem.* **2018**, *10* (3), 275–281.

(266) Bozhüyük, K. A. J.; Linck, A.; Tietze, A.; Wesche, F.; Nowak, S.; Fleischhacker, F.; Bode, H. B. Modification and de Novo Design of Non-Ribosomal Peptide Synthetases (NRPS) Using Specific Assembly Points within Condensation Domains. *bioRxiv* **2018**, 354670.

(267) Doekel, S.; Marahiel, M. A. Dipeptide Formation on Engineered Hybrid Peptide Synthetases. *Chem. Biol.* **2000**, *7* (6), 373–384.

(268) Ackerley, D. F.; Lamont, I. L. Characterization and Genetic Manipulation of Peptide Synthetases in Pseudomonas Aeruginosa PAO1 in Order to Generate Novel Pyoverdines. *Chem. Biol.* **2004**, *11* (7), 971–980.

(269) Linne, U.; Doekel, S.; Marahiel, M. A. Portability of Epimerization Domain and Role of Peptidyl Carrier Protein on Epimerization Activity in Nonribosomal Peptide Synthetases. *Biochemistry* **2001**, *40* (51), 15824–15834.

(270) Chiocchini, C.; Linne, U.; Stachelhaus, T. In Vivo Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex. *Chem. Biol.* **2006**, *13* (8), 899–908.

(271) Liu, H.; Gao, L.; Han, J.; Ma, Z.; Lu, Z.; Dai, C.; Zhang, C.; Bie, X. Biocombinatorial Synthesis of Novel Lipopeptides by COM Domain-Mediated Reprogramming of the Plipastatin NRPS Complex. *Front. Microbiol.* **2016**, *7*.

(272) de Ferra, F.; Rodriguez, F.; Tortora, O.; Tosi, C.; Grandi, G. Engineering of Peptide Synthetases. Key Role of the Thioesterase-like Domain for Efficient Production of Recombinant Peptides. *J. Biol. Chem.* **1997**, *272* (40), 25304–25309.

(273) Mootz, H. D.; Kessler, N.; Linne, U.; Eppelmann, K.; Schwarzer, D.; Marahiel, M. A. Decreasing the Ring Size of a Cyclic Nonribosomal Peptide Antibiotic by In-Frame Module Deletion in the Biosynthetic Genes. *J. Am. Chem. Soc.* **2002**, *124* (37), 10980–10981.

(274) Gao, L.; Liu, H.; Ma, Z.; Han, J.; Lu, Z.; Dai, C.; Lv, F.; Bie, X. Translocation of the Thioesterase Domain for the Redesign of Plipastatin Synthetase. *Sci. Rep.* **2016**, *6*, 38467.

(275) Gao, L.; Guo, J.; Fan, Y.; Ma, Z.; Lu, Z.; Zhang, C.; Zhao, H.; Bie, X. Module and Individual Domain Deletions of NRPS to Produce Plipastatin Derivatives in Bacillus Subtilis. *Microb. Cell Factories* **2018**, *17* (1).

(276) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. Nonribosomal Peptide Biosynthesis: Point Mutations and Module Skipping Lead to Chemical Diversity. *Angew. Chem. Int. Ed.* **2006**, *45* (14), 2296–2301.

(277) Butz, D.; Schmiederer, T.; Hadatsch, B.; Wohlleben, W.; Weber, T.; Süssmuth, R. D. Module Extension of a Non-Ribosomal Peptide Synthetase of the Glycopeptide Antibiotic Balhimycin Produced by Amycolatopsis Balhimycina. *ChemBioChem* **2008**, *9* (8), 1195–1200.

(278) Baltz, R. H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* **2014**, *3* (10), 748–758.

(279) Nguyen, K. T.; Ritz, D.; Gu, J.-Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. Combinatorial Biosynthesis of Novel Antibiotics Related to Daptomycin. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (46), 17462–17467.

(280) Nguyen, K. T.; He, X.; Alexander, D. C.; Li, C.; Gu, J.-Q.; Mascio, C.; Van Praagh, A.; Mortin, L.; Chu, M.; Silverman, J. A.; et al. Genetically Engineered Lipopeptide Antibiotics Related to A54145 and Daptomycin with Improved Properties. *Antimicrob. Agents Chemother.* **2010**, *54* (4), 1404–1413.

(281) Doekel, S.; Coëffet-Le Gal, M.-F.; Gu, J.-Q.; Chu, M.; Baltz, R. H.; Brian, P. Non-Ribosomal Peptide Synthetase Module Fusions to Produce Derivatives of Daptomycin in Streptomyces Roseosporus. *Microbiol. Read. Engl.* **2008**, *154* (Pt 9), 2872–2880.

(282) Yakimov, M. M.; Giuliano, L.; Timmis, K. N.; Golyshin, P. N. Recombinant Acylheptapeptide Lichenysin: High Level of Production by Bacillus Subtilis Cells. *J. Mol. Microbiol. Biotechnol.* **2000**, *2* (2), 217–224.

(283) Konz, D.; Doekel, S.; Marahiel, M. A. Molecular and Biochemical Characterization of the Protein Template Controlling Biosynthesis of the Lipopeptide Lichenysin. *J. Bacteriol.* **1999**, *181* (1), 133–140.

(284) Zobel, S.; Boecker, S.; Kulke, D.; Heimbach, D.; Meyer, V.; Süssmuth, R. D. Reprogramming the Biosynthesis of Cyclodepsipeptide Synthetases to Obtain New Enniatins and Beauvericins. *Chembiochem Eur. J. Chem. Biol.* **2016**, *17* (4), 283–287.

(285) Bode, H. B.; Reimer, D.; Fuchs, S. W.; Kirchner, F.; Dauth, C.; Kegler, C.; Lorenzen, W.; Brachmann, A. O.; Grün, P. Determination of the Absolute Configuration of Peptide Natural Products by Using Stable Isotope Labeling and Mass Spectrometry. *Chem. – Eur. J.* **2012**, *18* (8), 2342–2348.

(286) Nannemann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O. Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis. *Future Med. Chem.* **2011**, *3* (7), 809–819.

(287) Gibney, E.; Noorden, R. V.; Ledford, H.; Castelvecchi, D.; Warren, M. ' Test-Tube' Evolution Wins Chemistry Nobel Prize. *Nature* **2018**, *562*, 176.

(288) Goodrich, A. C.; Harden, B. J.; Frueh, D. P. Solution Structure of a Nonribosomal Peptide Synthetase Carrier Protein Loaded with Its Substrate Reveals Transient, Well-Defined Contacts. *J. Am. Chem. Soc.* **2015**.

(289) Amann, E.; Ochs, B.; Abel, K. J. Tightly Regulated Tac Promoter Vectors Useful for the Expression of Unfused and Fused Proteins in Escherichia Coli. *Gene* **1988**, *69* (2), 301–315.

(290) Radeck, J.; Kraft, K.; Bartels, J.; Cikovic, T.; Dürr, F.; Emenegger, J.; Kelterborn, S.; Sauer, C.; Fritz, G.; Gebhard, S.; et al. The Bacillus BioBrick Box: Generation and Evaluation of Essential Genetic Building Blocks for Standardized Work with Bacillus Subtilis. *J. Biol. Eng.* **2013**, *7* (1), 29.

(291) Miller, J. H. *Experiments in Molecular Genetics*, 11. print.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1972.

(292) Davis, M. W. ApE- A plasmid Editor http://biologylabs.utah.edu/jorgensen/wayned/ape/ (accessed Mar 12, 2018).
(293) Lapid, C.; Gao, Y. PrimerX http://www.bioinformatics.org/primerx/ (accessed Mar 14, 2018).
(294) Edelheit, O.; Hanukoglu, A.; Hanukoglu, I. Simple and Efficient Site-Directed Mutagenesis Using Two Single-Primer Reactions in Parallel to Generate Mutants for Protein Structure-Function Studies. *BMC Biotechnol.* **2009**, *9* (1), 61.
(295) McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y. M.; Buso, N.; Cowley, A. P.; Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41* (W1), W597–W600.
(296) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Humana Press, 2005; pp 571–607.
(297) Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* **2012**, *40* (Web Server issue), W597-603.
(298) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
(299) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nat. Protoc.* **2012**, *7* (8), 1511–1522.
(300) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
(301) Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44* (Database issue), D7–D19.
(302) Harrington, B. Draw Freely | Inkscape https://inkscape.org/de/ (accessed Mar 14, 2018).
(303) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25* (9), 1189–1191.
(304) Zotero | Your personal research assistant https://www.zotero.org/ (accessed Mar 14, 2018).
(305) Bouhenni, R.; Gehrke, A.; Saffarini, D. Identification of Genes Involved in Cytochrome c Biogenesis in Shewanella Oneidensis, Using a Modified Mariner Transposon. *Appl. Environ. Microbiol.* **2005**, *71* (8), 4935–4937.
(306) Myers, J. A.; Curtis, B. S.; Curtis, W. R. Improving Accuracy of Cell and Chromophore Concentration Measurements Using Optical Density. *BMC Biophys.* **2013**, *6* (1), 4.
(307) Yu, D.; Xu, F.; Valiente, J.; Wang, S.; Zhan, J. An Indigoidine Biosynthetic Gene Cluster from Streptomyces Chromofuscus ATCC 49982 Contains an Unusual IndB Homologue. *J. Ind. Microbiol. Biotechnol.* **2013**, *40* (1), 159–168.
(308) Stewart, E. J.; Aslund, F.; Beckwith, J. Disulfide Bond Formation in the Escherichia Coli Cytoplasm: An in Vivo Role Reversal for the Thioredoxins. *EMBO J.* **1998**, *17* (19), 5543–5550.
(309) Prinz, W. A.; Aslund, F.; Holmgren, A.; Beckwith, J. The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the Escherichia Coli Cytoplasm. *J. Biol. Chem.* **1997**, *272* (25), 15661–15667.
(310) Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res.* **2016**, *44* (D1), D279–D285.
(311) Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol.* **2011**, *498*, 349–361.

(312) Team:Heidelberg - 2013.igem.org http://2013.igem.org/Team:Heidelberg (accessed Dec 3, 2018).
(313) Heumann, W.; Young, D.; Gottlich, C. Leucoindigoidine Formation by an Arthrobacter Species and Its Oxidation to Indigoidine by Other Micro-Organisms. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1968**, *156* (2), 429–431.
(314) Brown, A. Engineering the Indigoidine-Synthesising Enzyme BpsA for Diverse Applications in Biotechnology, 2018.
(315) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
(316) Wilson, D. J.; Shi, C.; Teitelbaum, A. M.; Gulick, A. M.; Aldrich, C. C. Characterization of AusA: A Dimodular Nonribosomal Peptide Synthetase Responsible for the Production of Aureusimine Pyrazinones. *Biochemistry* **2013**, *52* (5), 926–937.
(317) Ackerley, D. F.; Brown, A. S.; Robins, K. J. Methods of Detecting and Measuring Glutamine and Analogues Thereof, and Methods Related Thereto, December 5, 2014.
(318) Witkowski, A.; Witkowska, H. E.; Smith, S. Reengineering the Specificity of a Serine Active-Site Enzyme. Two Active-Site Mutations Convert a Hydrolase to a Transferase. *J. Biol. Chem.* **1994**, *269* (1), 379–383.
(319) Scaglione, J. B.; Akey, D. L.; Sullivan, R.; Kittendorf, J. D.; Rath, C. M.; Kim, E.-S.; Smith, J. L.; Sherman, D. H. Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angew. Chem. Int. Ed Engl.* **2010**, *49* (33), 5726–5730.
(320) Sander, C.; Schneider, R. Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment. *Proteins* **1991**, *9* (1), 56–68.
(321) Rost, B. Twilight Zone of Protein Sequence Alignments. *Protein Eng.* **1999**, *12* (2), 85–94.
(322) Belecki, K.; Townsend, C. A. Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Enediyne Polyketide Synthase CalE8. *J. Am. Chem. Soc.* **2013**, *135* (38).
(323) Guo, Z.-F.; Jiang, M.; Zheng, S.; Guo, Z. Suppression of Linear Side Products by Macromolecular Crowding in Nonribosomal Enterobactin Biosynthesis. *Org. Lett.* **2008**, *10* (4), 649–652.
(324) Raedts, J.; Lundgren, M.; Kengen, S. W. M.; Li, J.-P.; Oost, J. van der. A Novel Bacterial Enzyme with D-Glucuronyl C5-Epimerase Activity. *J. Biol. Chem.* **2013**, *288* (34), 24332–24339.
(325) Couch, R.; O'Connor, S. E.; Seidle, H.; Walsh, C. T.; Parry, R. Characterization of CmaA, an Adenylation-Thiolation Didomain Enzyme Involved in the Biosynthesis of Coronatine. *J. Bacteriol.* **2004**, *186* (1), 35–42.
(326) Gevers, W.; Kleinkauf, H.; Lipmann, F. The Activation of Amino Acids for Biosynthesis of Gramicidin S. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *60* (1), 269–276.
(327) Luo, L.; Burkart, M. D.; Stachelhaus, T.; Walsh, C. T. Substrate Recognition and Selection by the Initiation Module PheATE of Gramicidin S Synthetase. *J. Am. Chem. Soc.* **2001**, *123* (45), 11208–11218.
(328) Wilson, D. J.; Aldrich, C. C. A Continuous Kinetic Assay for Adenylation Enzyme Activity and Inhibition. *Anal. Biochem.* **2010**, *404* (1), 56–63.