Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis

Sofie Dekimpe[1,2,3] and Joleen Masschelein[1,2,*]

[1]*Laboratory for Biomolecular Discovery & Engineering, Department of Biology, KU Leuven, Leuven, Belgium*
[2]*VIB-KU Leuven Center for Microbiology, Flanders Institute for Biotechnology, Leuven, Belgium*
[3]*Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium*
**Corresponding author*

Abstract

Nonribosomal peptide synthetases are remarkable molecular machines that produce a wide range of structurally complex peptide natural products with important applications in medicine and agriculture. Condensation domains play a central role in these biosynthetic pathways by catalysing amide bond formation between various aminoacyl substrates. In recent years, however, it has become increasingly clear that the catalytic repertoire of C domains extends far beyond conventional peptide bond formation. C domains have been shown to perform highly diverse functions during nonribosomal peptide assembly, such as β-lactam formation, dehydration, hydrolysis, chain length control, cycloaddition, Pictet-Spengler cyclization, Dieckmann condensation and recruitment of auxiliary enzymes. In this review, a comprehensive overview of the multifaceted role of C domains in the biosynthesis of specialized metabolites in bacteria and fungi is presented. Different perspectives are also offered on how the exceptional functional versatility of C domains may be exploited for bioengineering approaches to expand the chemical diversity of nonribosomal peptides and other natural products.

1. Introduction: the biosynthesis of nonribosomal peptides

Microorganisms produce a wealth of structurally diverse specialized metabolites with a remarkable range of biological activities to improve their growth and survival in competitive environments.[1,2] Many of these natural products also have important applications in medicine and agriculture due to their antibiotic, antiparasitic, insecticidal, cholesterol-lowering, immunomodulatory or anticancer properties.[3,4] Nonribosomal peptides represent a major class of structurally diverse and therapeutically relevant natural products (**Fig. 1A**). They are produced by large enzymatic assembly lines, known as nonribosomal peptide synthetases (NRPSs). Type I NRPSs are modular multienzyme complexes operate in a step-wise fashion to sequentially select, activate and link diverse amino acid building blocks. Each module is responsible for the incorporation of one building block and minimally consists of three core domains: an adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation (C) domain. First, the A domain selects a specific amino acid substrate and activates it as an aminoacyl adenylate before loading it onto the adjacent PCP domain (**Fig. 1B**). In addition to the standard set of 20 proteinogenic α-amino acids, A domains are capable of activating a variety of non-proteinogenic amino acids, carboxylic acids, α-hydroxy acids and even keto acids, thereby increasing the structural diversity of the final peptide product. PCP domains are post-translationally modified with a coenzyme A (CoA)-derived 4'-phosphopantetheinyl (4'-Ppant) prosthetic group in a reaction catalysed by phosphopantetheinyl transferases. This Ppant moiety is equipped with a reactive thiol group that binds the biosynthetic intermediates via a thioester bond. It also functions as a swinging arm to transfer the intermediates to the catalytic sites of adjacent domains.

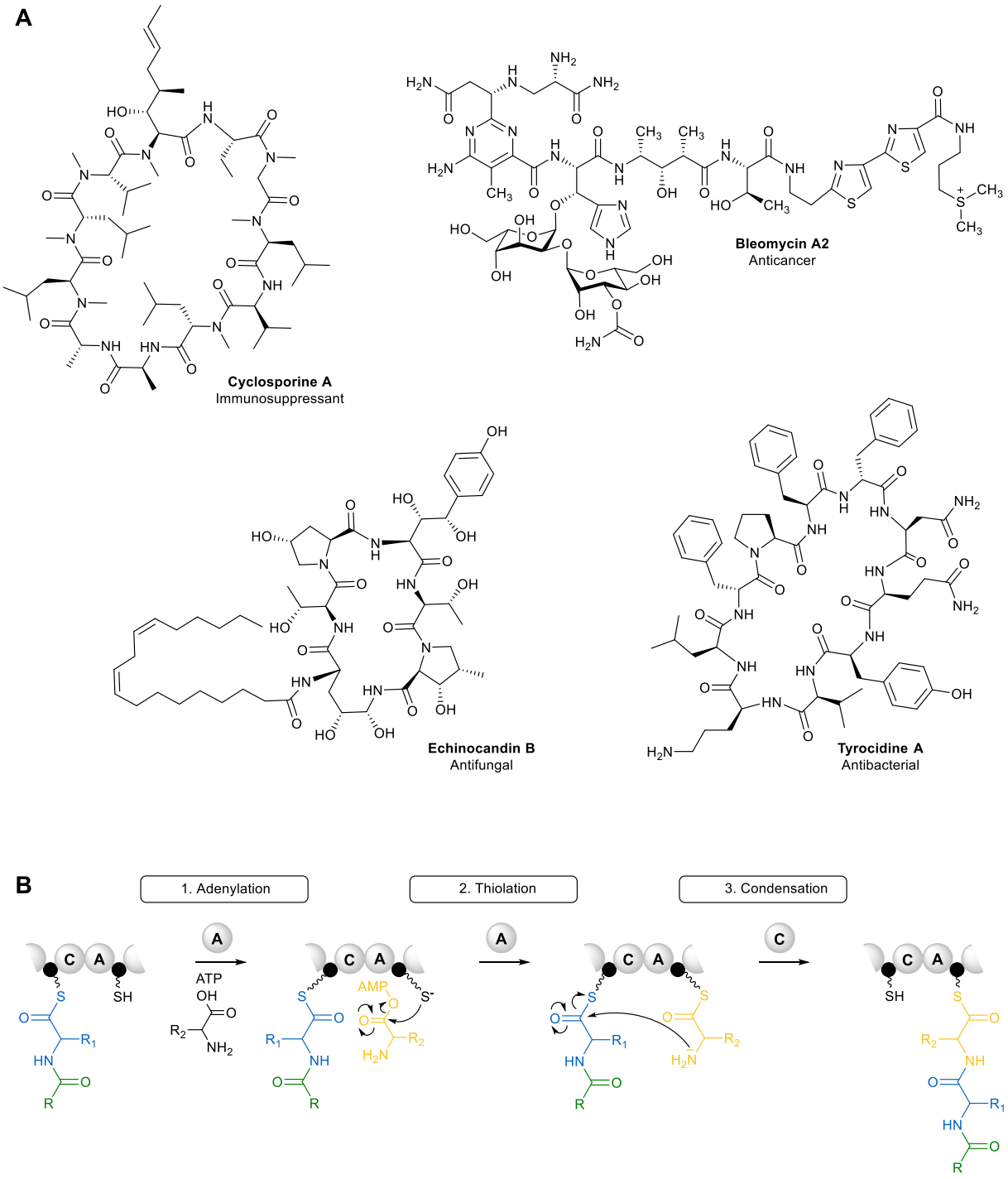

Fig. 1. (A) Examples of nonribosomal peptides with therapeutically important activities. (B) General principle of an NRPS elongation cycle. The A domain first activates a cognate amino acid with ATP to form an aminoacyl adenylate. The PCP domain binds the activated amino acid substrate as a thioester through the thiol moiety of its 4'-Ppant prosthetic group. The nucleophilic amine group of the PCP-bound amino acid then attacks the electrophilic thioester carbonyl carbon of the aminoacyl or peptidyl substrate from the upstream module, resulting in peptide bond formation. This chain elongation reaction is catalysed by the C domain. Carrier proteins and their associated phosphopantetheinyl groups are shown in as black dots. Each incorporated building block is presented in a distinct colour.

After a new amino acid building block is loaded onto the PCP domain, the C domain catalyses peptide bond formation between the PCP-bound aminoacyl substrate and the upstream peptidyl thioester intermediate bound to the PCP domain of the preceding module (**Fig. 1B**). The elongated peptidyl intermediate is then used as a substrate in the next round of chain extension catalysed by the C domain within the downstream module. The final module often contains a thioesterase (TE) domain at its C-terminus which releases the fully assembled peptide via hydrolysis or macrocyclization. In addition to the C, A and PCP core domains, NRPS modules can contain a variety of auxiliary domains, such as epimerisation (E), C-/N-methyltransferase (C-/N-MT) or oxidation (OX) domains, that further modify the PCP-bound aminoacyl or peptidyl thioester intermediates. Moreover, *trans*-acting tailoring enzymes, such as glycosyltransferases and cytochrome P450 monooxygenases, typically introduce additional modifications after the peptide is released from the assembly line. These post-assembly modifications further increase the structural complexity and diversity, and are often essential for biological activity. Unlike type I NRPSs that involve large multifunctional enzymes, type II NRPS systems are composed of discrete monofunctional or didomain proteins that form multienzyme complexes together with specialised tailoring enzymes. They typically provide modified and unconventional amino acid building blocks to other biosynthetic pathways, such as fatty acid synthases, polyketide synthases and type I NRPSs.[5] For detailed information on the genetics, enzymology, structure and engineering of NRPSs, the reader is referred to several excellent reviews that have been published over the last few years.[6–13] In this review, we will highlight the remarkably versatile role of C domains in natural product biosynthesis. Indeed, as the number of characterized NRP biosynthetic pathways increases constantly, more and more C superfamily domains are identified whose function extends beyond conventional peptide bond formation. These enzymes do not only represent promising biocatalytic tools, but also offer exciting prospects as starting points for engineering biosynthetic assembly lines and producing novel natural product derivatives.

2. The catalytic role of canonical condensation domains

Most C domains involved in NRP biosynthesis are responsible for chain elongation by catalysing the condensation reaction between PCP-bound substrates of adjacent modules.[14] Condensation is initiated by a nucleophilic attack of the α-amino group of the PCP-tethered aminoacyl monomer on the thioester carbonyl carbon of the aminoacyl or peptidyl thioester intermediate bound to the PCP domain in the preceding module (**Fig. 1B**). C domains are pseudo-dimeric enzymes of approximately 450 amino acids. They are composed of two subunits of similar size that form a V-like structure (**Fig. 2**).[15–20] The catalytic site lies at the interface of both subdomains and is defined by a highly conserved HHxxxDG motif.[21] Both subdomains are linked via a small floor loop at the bottom of the V and an additional bridging latch covering the active site.[15–20] Condensation takes place when both the upstream and downstream PCP domain simultaneously interact with the C domain, as recently visualised in X-ray crystal structures of the dimodular linear gramicidin synthetase subunit A.[22] The nearly symmetrical architecture of C domains allows insertion of the Ppant arm into the substrate tunnel, enabling access to the active site from two opposite sides.[22] The first binding pocket, also referred to as the acceptor site, accommodates the downstream PCP-linked aminoacyl monomer. This site has been shown to exhibit a high degree of

selectivity towards the side chain and chirality of amino acid substrates.[23–25] C domains have therefore been proposed to act as gatekeepers in NRP biosynthesis to avoid erroneous substrate incorporation, controlling the selection by the A domain[26]. However, recent work using natural evolution analyses and combinatorial exchange strategies indicates that this C domain selectivity is not a general precept.[27] The second binding site, called the donor site, binds the peptidyl intermediate linked to the PCP domain of the upstream module. Although C domains generally exhibit a broader substrate specificity at their donor site, enantiomeric selectivity has been observed.[23–25,28] C domains can therefore be categorized into two distinct groups, depending on their stereoselectivity. $^{L}C_{L}$ domains link two L-configured amino acids, while C domains that couple a D-configured donor to an L-configured acceptor are referred to as $^{D}C_{L}$ domains.[29] A series of elegant studies over the past few decades have provided many insights into the reaction mechanism of C domains. However, the catalytic roles of the active site residues remain a subject of debate. Initially, the second histidine residue in the HHxxxDG active site motif was proposed to act as a general base to deprotonate the α-amino group of the acceptor amino acid.[14,30,31] However, mutational analyses have shown that this residue is not essential for catalytic activity in some C domains.[15,32] Moreover, theoretical estimations of pK values have suggested that the second histidine residue is protonated under physiological conditions, which would refute its role as general base.[16] Instead, it has been proposed to stabilize the tetrahedral transition state.[16] More recently, however, the crystal structure of an NRPS C domain in complex with an acceptor substrate mimic has revealed a hydrogen bond network involving the second histidine residue and the α-amino group of the substrate. These results strongly suggest that the histidine residue is not positively charged and is mainly responsible for ensuring correct substrate positioning.[19]

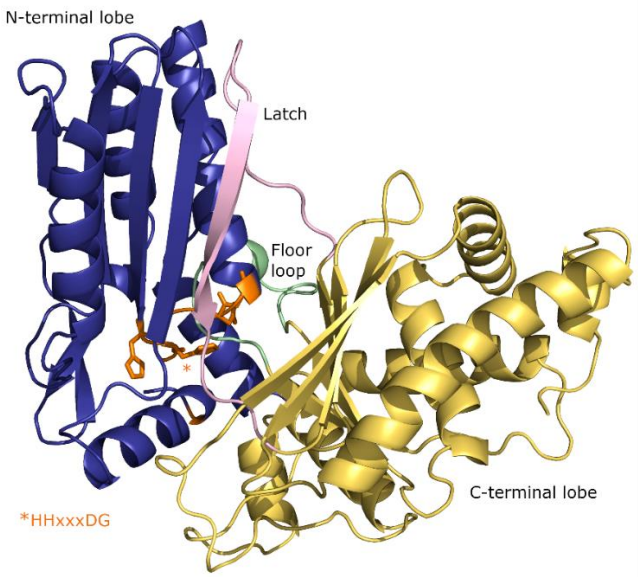

Figure 2. Cartoon representation of the stand-alone C domain VibH. The N-terminal lobe is coloured in blue. The C-terminal part is shown in yellow. The conserved HHxxxDG motif is presented as orange sticks, with the catalytic histidine residue marked by an asterisk. The latch and floor loop are highlighted in pink and green, respectively. Structure rendered from PDB entry 1L5A.

3. Functional diversity within the C domain superfamily

3.1 Starter C domains

NRPS loading modules typically lack C domains because no condensation reaction is required for loading of the starter unit. A notable exception are lipopeptide synthetases, such as those that assemble daptomycin and surfactin. These assembly lines harbour an N-terminal starter C (Cs) domain. Cs domains cluster in a separate phylogenetic clade, but have the same HHxxxDG active site motif as canonical C domains.[29] They are responsible for the conjugation of an activated fatty acid to the aminoacyl thioester bound to the PCP domain of the first module, a process called lipoinitiation (**Fig. 3**).[29,33–36] These fatty acid moieties typically play an important role in the bioactivity of lipopeptides by increasing their hydrophobicity and their ability to interact with bacterial membranes.[37] They are either derived from primary metabolism, like in daptomycin[38] and surfactin biosynthesis[34], or made by dedicated enzymes encoded in the biosynthetic gene cluster, as in the calcium-dependent lipopeptide antibiotics (CDA).[39,40] The acyl chains are often loaded onto a *holo*-acyl carrier protein (ACP) for delivery to the Cs domain. This activation and transfer reaction is catalysed by fatty acyl ligases.[38,41] In other cases, the fatty acid chains are directly assembled on a trans-acting ACP domain in a separate pathway, like in CDA biosynthesis.[39,40] In the biosynthetic pathways of the surfactin, WAP-8294A and glidobactin lipopeptides, on the other hand, the Cs domains directly use CoA-activated free fatty acids.[34,35,42] Some Cs domains are proposed to catalyse multiple simultaneous N-acylation reactions. In the biosynthesis of the isonitrile lipopeptides from *Mycobacterium* and *Streptomyces*, for example, the ScoA Cs domain is predicted to condense two isonitrile-modified fatty acids to both amino groups of a PCP-tethered lysine residue.[43]

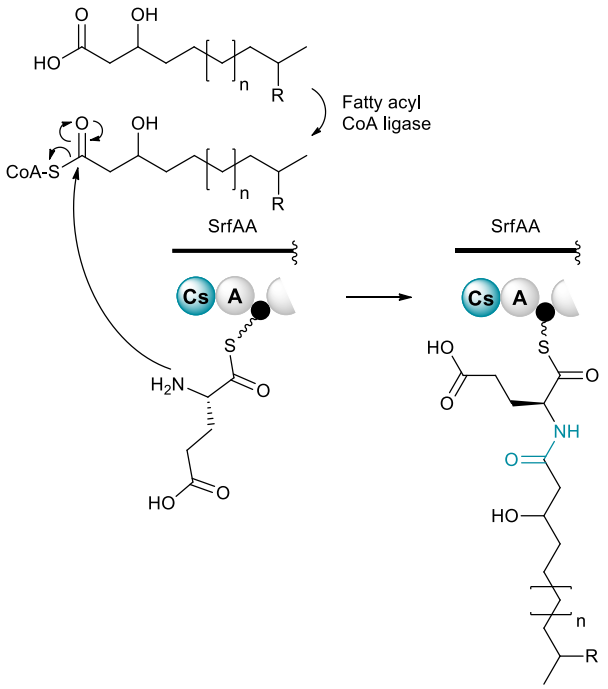

Figure 3. Fatty acid incorporation in surfactin biosynthesis. 3-hydroxy myristic acid is activated by a fatty acid CoA ligase and subsequently condensed with the PCP-bound glutamate starter unit by the N-terminal Cs domain of SrfAA. The Cs domain is coloured in turquoise.

Besides fatty acids, other substrates are also be incorporated by Cs domains. For instance, in the biosynthesis of the heptapeptide antibiotics JBIR-78 and JBIR-95 in *Kibdelosporangium sp*. AK-AA56, the Cs domain condenses phenylacetyl-CoA with the N-terminal valine thioester starter unit.[44] In the

biosynthesis of the siderophore enterobactin, the EntF C domain iteratively couples 2,3-dihydroxybenzoic acid units to PCP-bound seryl residues until a trimer is assembled and macrocyclized by the TE domain. The 2,3-dihydroxybenzoic acid residues are first activated by the stand-alone A domain EntE and loaded onto the type II aryl carrier protein within EntB prior to the condensation reaction.[45] Although the involvement of a type II NRPS differs from the typical mechanism of lipoinitiation, the EntF C domain falls into the same phylogenetic clade as starter C domains.[29,46] Similar examples of aryl acid attachment are found in the biosynthesis of actinomycin and bacillibactin.[47,48]

Among these various types of starter C domains, there are two that stand out due to their unique functionality: the bifunctional AebF Cs domain and the internal IcoA starter C domain in the biosynthetic pathways of amphi-enterobactin and icosalide A, respectively.

3.1.1 Bifunctional activity

Amphi-enterobactin is an amphiphilic enterobactin-like siderophore produced by the marine bacterium *Vibrio campbelli* BAA-1116.[49] Its peptide backbone is composed of four serine residues. One is decorated with a fatty acid chain, while the other three are connected to 2,3-dihydroxybenzoic acid moieties. Based on heterologous expression and *in vitro* biochemical experiments, a biosynthetic pathway was proposed in which the AebF Cs domain first catalyses lipoinitiation using a fatty acyl-CoA thioester substrate and the L-serine starter unit bound to the downstream PCP domain (**Fig. 4**).[49] The acylated serinyl thioester intermediate is then transferred to the TE domain, which mediates oligomerization. The L-serine residues that are loaded onto AebF during the next three iterative rounds are condensed with a PCP-linked 2,3-dihydrobenzoic acid unit by the same Cs domain in a process similar to the formation of enterobactin (see Section 3.1). The N-terminal AebF Cs domain therefore displays remarkable bifunctional activity, catalysing condensation reactions with two different thioester donor substrates.

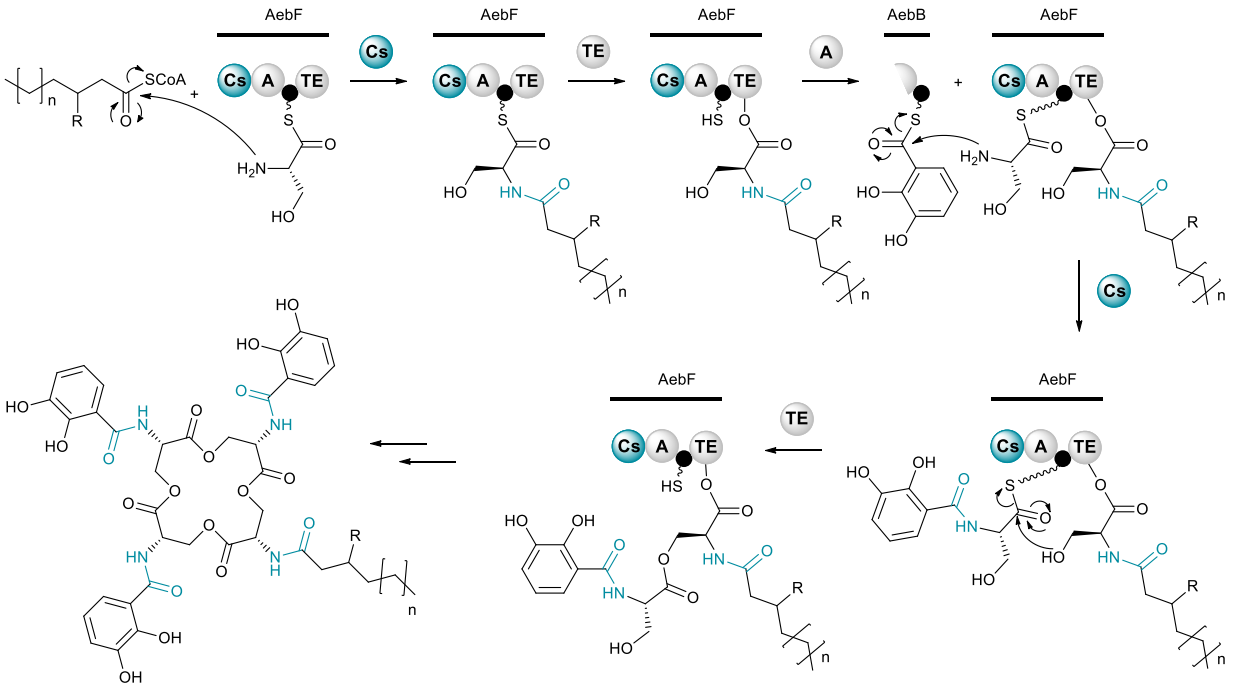

Figure 4. Proposed biosynthetic pathway for amphi-enterobactin. TE = Thioesterase domain. The starter C domain Is highlighted in turquoise.

3.1.2 Internal starter C domains

The icosalides are unusual lipopeptidiolide antibiotics and bacterial swarming inhibitors that were originally isolated from fungal cultures.[50] Recently, *Burkholderia* symbionts were identified as the true producers of these compounds.[51,52] The icosalides are biosynthesized from two serine, two leucine and two β-hydroxy fatty acid units by an unprecedented NRPS assembly line (IcoA) that contains two Cs domains. One is located at the N-terminus of the first module and has an HHxxxDG active site motif, while the other is embedded in the middle of the assembly line in module 3 and is characterized by a PHxxxDG motif. (**Fig. 5A**). The first Cs domain is proposed to initiate chain assembly by condensing β-hydroxyoctanoic acid with to a PCP-bound L-leucinyl thioester. Subsequent epimerisation and chain elongation with L-serine by the dual E/C domain within module 2 gives rise to a β-hydroxyacyl-D-leucinyl-L-serinyl thioester intermediate. *In vitro* reconstitution and intact protein mass spectrometry experiments have revealed that the Cs domain in module 3 then initiates the assembly of a second lipopeptide chain by loading either a β-hydroxydecanoyl or a β-hydroxyoctanoyl moiety onto the amine of the L-serinyl extender unit bound to the module 3 PCP domain (**Fig. 5B**).[113] Next, the N-terminal C domain of module 3 is proposed to catalyse ester bond formation between both N-acylated thioester intermediates. Following the incorporation of an additional L-leucine building block, the C-terminal TE domain performs a macrolactonization reaction to yield the final lipopeptidiolide antibiotics.[51,52] Icosalide assembly therefore uniquely involves two separate chain initiation events within a single NRPS subunit.

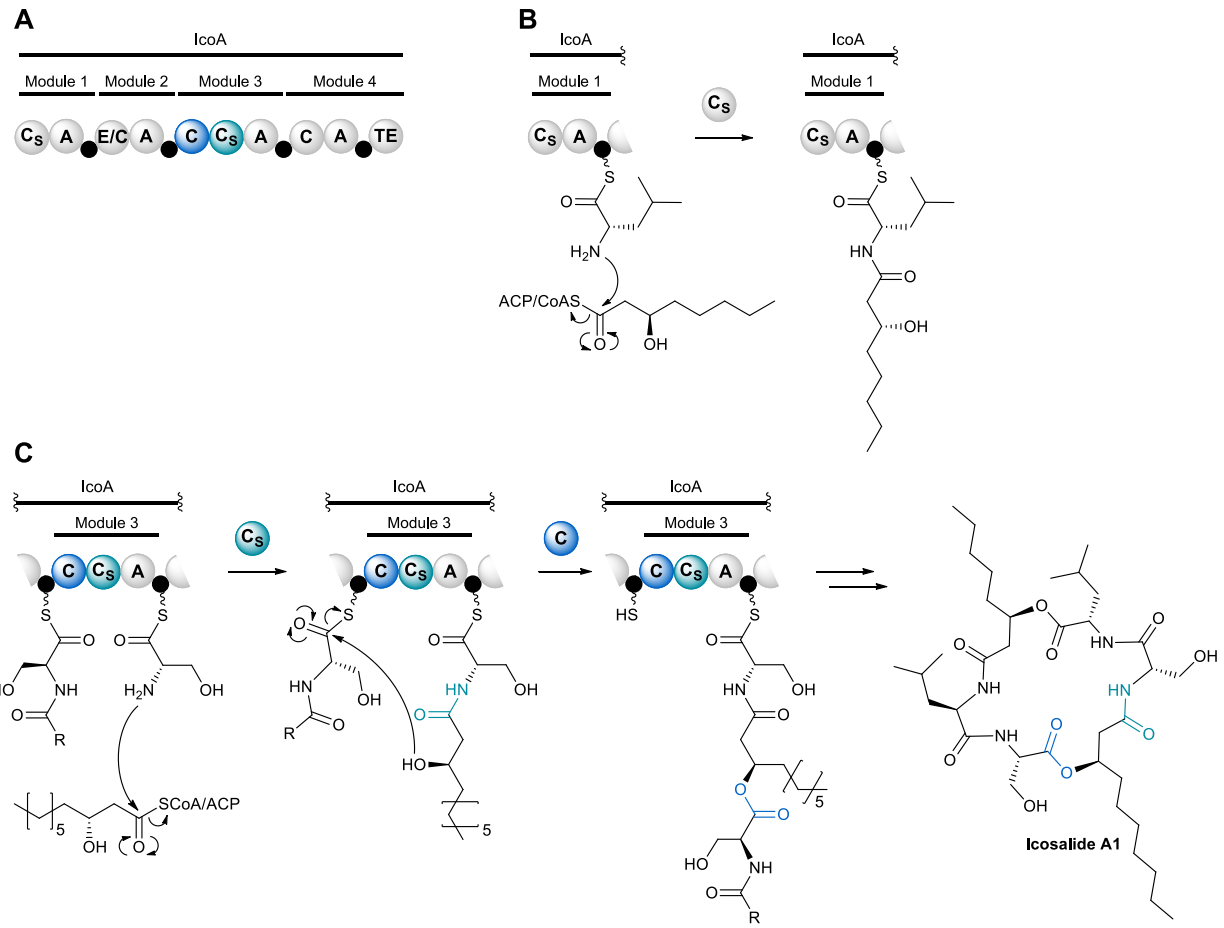

Figure 5. (A) The icosalide NRPS has an unusual architecture, featuring two Cs domains: one in module 1 and one embedded in module 3. (B) Mechanism of chain initiation by the Cs domain in module 1. (C) The internal Cs domain catalyses a second chain initiation event by loading either a β-hydroxydecanoyl or a β-hydroxyoctanoyl unit onto the L-serine residue bound to the module 3 PCP domain. The acylating C domain is highlighted in turquoise, while the ester bond forming C domain is coloured blue.

3.2 Epimerisation

To increase the structural diversity of their bioactive peptide arsenal, microorganisms often incorporate D-configured amino acids.[53,54] The presence of D-amino acids not only improves the resistance of peptides to proteolytic degradation, but it also allows them to adopt specific conformations that are required for further processing or for biological activity.[55,56] NRPSs use diverse strategies to introduce D-amino acids into the peptide scaffold (**Fig. 6A-C**). In some cases, such as cyclosporin A[57] and HC-toxin[58] assembly, D-amino acids are provided directly by external racemases. More often, however, D-amino acids are formed by modules that contain an additional epimerisation (E) domain inserted between the PCP domain and the downstream C domain.[56] Phylogenetically, E domains belong to the C domain superfamily and adopt a highly similar structural conformation, with two subdomains organized in a V-shape.[59,60] However, the function and sequence of E and C domains differ significantly, even though E domains share the conserved HHxxxDG motif.[21,29] E domains act on the elongated PCP-bound peptidyl intermediate and isomerize the α-position of the C-terminal amino acid residue through de- and reprotonation, producing a racemic mixture.[25,55,56,59,61] The downstream C domain, typically a $^{D}C_{L}$ catalyst (see Section 2), ensures that the correct stereoisomer is used in the subsequent elongation reaction.[28] The second histidine residue in the HHxxxDG active site motif is positioned across a highly conserved glutamate residue in E domains. This

glutamate residue is proposed to act as a general base in the epimerisation reaction to remove the Cα proton. According to structure-based computational predictions, the catalytic histidine residue is doubly protonated in the active site and likely responsible for stabilizing the enolate intermediate (**Fig. 6D**). [59,62] However, further biochemical evidence is needed to confirm the presumed roles of the active site residues.

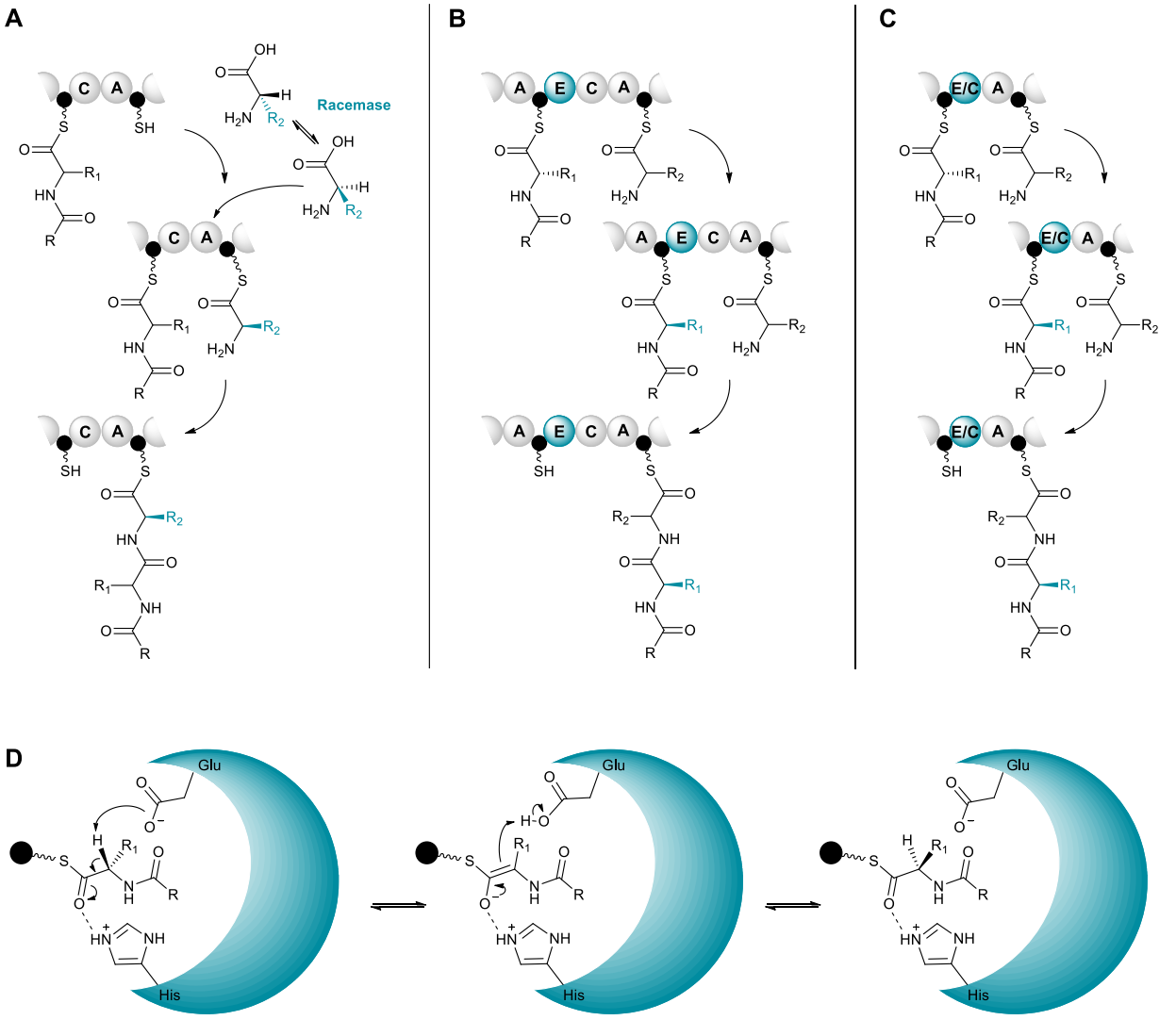

Figure 6. Schematic representation of the three alternative strategies to incorporate D-amino acids into nonribosomal peptides. (A) D-amino acids are provided by external racemases. The A domain is selective for D-configured monomers. (B) The epimerisation reaction is performed by an E domain, prior to condensation by the downstream C domain. (C) A dual E/C domain performs both the epimerisation and condensation reaction. The racemising enzymes/domains along with the associated epimerized groups are highlighted in turquoise. (D) Proposed reaction mechanism of E domains. The conserved glutamate and histidine residues are indicated. The conserved glutamate serves as general base/acid to catalyse the de- and reprotonation of the Cα, respectively. The racemising enzymes/domains along with the associated epimerized groups are highlighted in turquoise.

A third strategy to incorporate D-configured amino acids into the growing peptide chain is the use of dual E/C domains, which catalyse both epimerization and condensation. These domains frequently replace canonical C domains in the biosynthetic pathways of cyclic lipopeptides in *Pseudomonas, Burkholderia*

and *Xanthomonas* spp.[63–66] Dual E/C domains were first discovered in the arthrofactin biosynthetic pathway in *Pseudomonas sp.* MIS38.[63] The biosurfactant arthrofactin is assembled by a modular NRPS (ArfA-C) comprised of 11 modules. Although seven amino acids in the final lipopeptide product are D-configured, no E domains could be identified in the pathway.[67] Instead, a novel type of C domain with dual epimerisation and condensation activity was discovered, based on *in vitro* biochemical assays, sequence analyses and phylogenetic sorting.[29,63] Following chain elongation with an L-configured amino acid, the dual E/C domains are proposed to isomerise the newly-incorporated building blocks. After epimerisation, they are believed to act as a $^{D}C_{L}$ catalyst by coupling the isomerized peptide to the downstream aminoacyl extender unit.[29,63] *In vitro* biochemical studies on the arthrofactin synthetase have shown that the epimerization reaction only takes place after the downstream PCP domain is loaded with a new extender unit. It has therefore been postulated that dual E/C domains adopt distinct conformational states to ensure the correct timing of the epimerisation and condensation reactions.[63] Only after interaction with both the upstream and downstream peptidyl- or aminoacyl-S-PCP, the dual E/C domain is able to isomerize the C-terminal amino acid residue of the incoming peptide chain. E/C domains harbour an HH[I/L]xxxxGD motif at their N-terminus but lack to the conserved glutamate residue found in E domains.[29] Future biochemical studies may shed light on the exact epimerisation and condensation mechanism of dual E/C domains.

3.3 Heterocyclization

In some NRPS modules, C domains are replaced by homologous heterocyclization (Cy) domains that incorporate heterocyclic rings into the peptide backbone.[68] Cy domains catalyse chain elongation with cysteine, serine or threonine residues, followed by a distinct two-step cyclodehydration reaction to form five-membered thiazoline or (methyl)oxazoline rings.[69–73] The resulting heterocycles can undergo further reduction[74] or oxidation[75] by trans-acting reductases or integrated OX domains, respectively. The presence of heterocyclic rings is essential for the biological activity of several peptide natural products, such as pyochelin, vibriobactin, epothilone and bacillamide[69,72,74,76,77] (**Fig. 7A**). They introduce a high degree of structural rigidity that protects the peptides against proteolytic degradation and enables their interaction with biological targets[78], or the chelation of metal ions.[32,79,80] Heterocyclization is proposed to occur in three consecutive steps (**Fig. 7B**). First, the Cy domain catalyses condensation between the peptidyl thioester intermediate and the new cysteine, serine or threonine extender unit. Subsequently, the side chain thiol or hydroxyl group carries out a nucleophilic attack on the carbonyl carbon of the newly formed amide bond. The resulting hemiaminal then undergoes dehydration to form the final five-membered rings.[69–72] Cy domains are characterized by a conserved DxxxxD motif that replaces the HHxxxDG motif in canonical C domains. Mutational analyses have shown that both aspartate residues are important for condensation as well as heterocyclization.[32,76,80,81] However, crystal structures of Cy domains from the bacillamide and epothilone synthetases have revealed that they play a structural, rather than a catalytic role.[72,73] Instead, a catalytically important aspartate, asparagine and threonine residue were identified through structure-guided mutational analyses and *in vitro* activity assays. Of these three, the aspartate is believed to act as the catalytic base/acid in the heterocyclization reaction.[72,73]

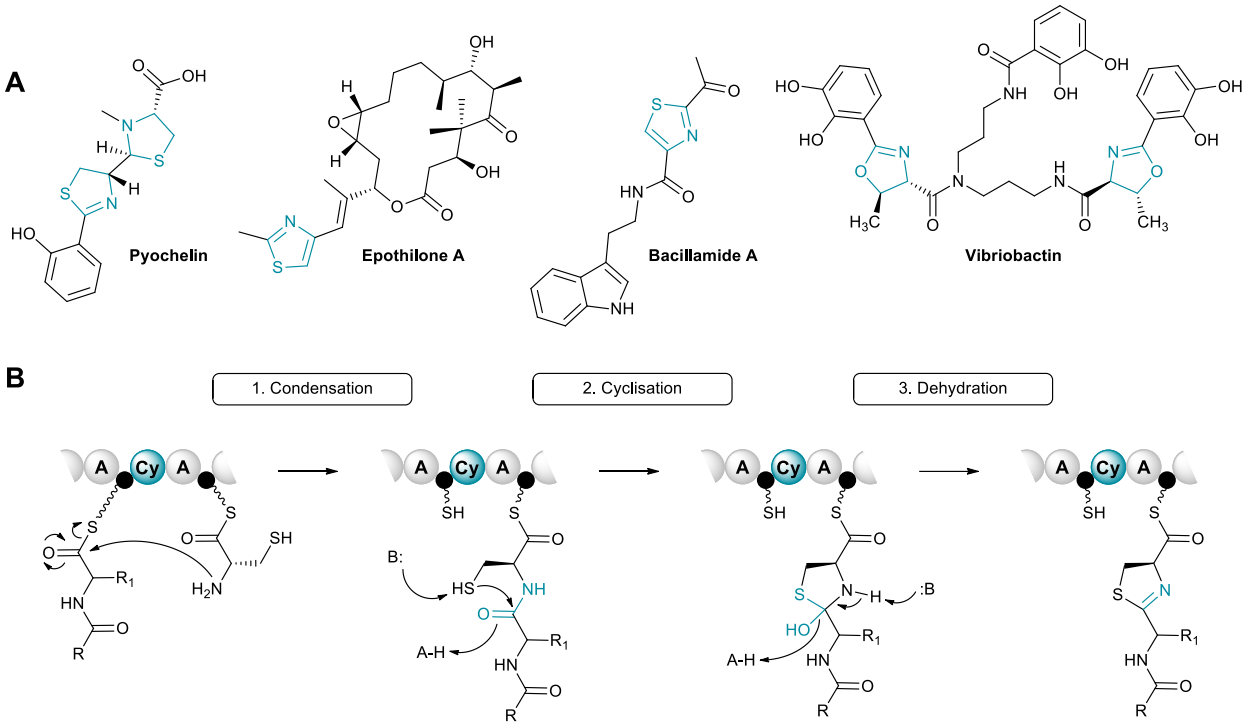

Figure 7. (A) Examples of heterocycle-containing peptides. (B) Schematic representation of the cyclisation reaction by a heterocyclisation (Cy) domain forming a thiazoline ring. The Cy domain is highlighted in turquoise.

3.4 Ester bond formation

While most C domains are known to catalyse amide bond formation, some are capable of generating ester bonds between PCP-tethered substrates. In bacterial depsipeptide synthetases, ester bonds are formed by specialized modules that contain an additional ketoreductase (KR) domain.[82] Instead of amino acids, the A domains in these modules select and load α-keto acids onto the adjacent PCP domain. The PCP-linked α-keto acyl monomers are stereoselectively reduced by the KR domain to generate α-hydroxyacyl thioester intermediates, which are then used by the C domain for ester bond formation. Examples of such (C)–A–KR–PCP depsipeptide modules are found in the biosynthetic pathways of the antifungal hectochlorin, the piscicide antimycin and the cryptophycin anticancer agents.[83–86] *In vitro* biochemical assays with the final module of the cryptophycin synthetase and synthetic N-acetyl cysteamine (SNAC) thioester substrate mimics confirmed activation and transfer of α-ketoisocaproic acid to the CrpD PCP domain, followed by ketoreduction and C domain-mediated ester bond formation (**Fig. 8A**).[84] In the biosynthesis of the cereulide, valinomycin and kutzneride depsipeptides, the C and the A domain within the ester bond-forming module are located on two separate NRPS subunits, causing the C domain to work *in trans* to mediate condensation.[87–92]

In fungal depsipeptides synthetases, such as those involved in the biosynthesis of the bassianolides, beauvericins and enniatins, α-hydroxy acids are directly selected and activated by the A domains. These NRPS systems typically consist of two modules that iteratively condense D-configured α-hydroxy acid and N-methyl-L-amino acid monomers to form cyclic oligomers. The biosynthetic pathway features three distinct C domains: an N-terminal C domain (C1) that is catalytically inactive, a canonical C domain (C2) that catalyses the formation of amide bonds, and a C-terminal C domain (C3) that controls the chain length

and catalyses ester bond formation during the assembly of the oligomer, as well as the final macrocyclization reaction. During the assembly process, the growing peptidyl thioester intermediate is believed to shuttle between the different PCP domains, enabling C2 and C3 to alternately form amide and ester bonds until the peptide reaches its full length (**Fig. 8B**).[93,94]

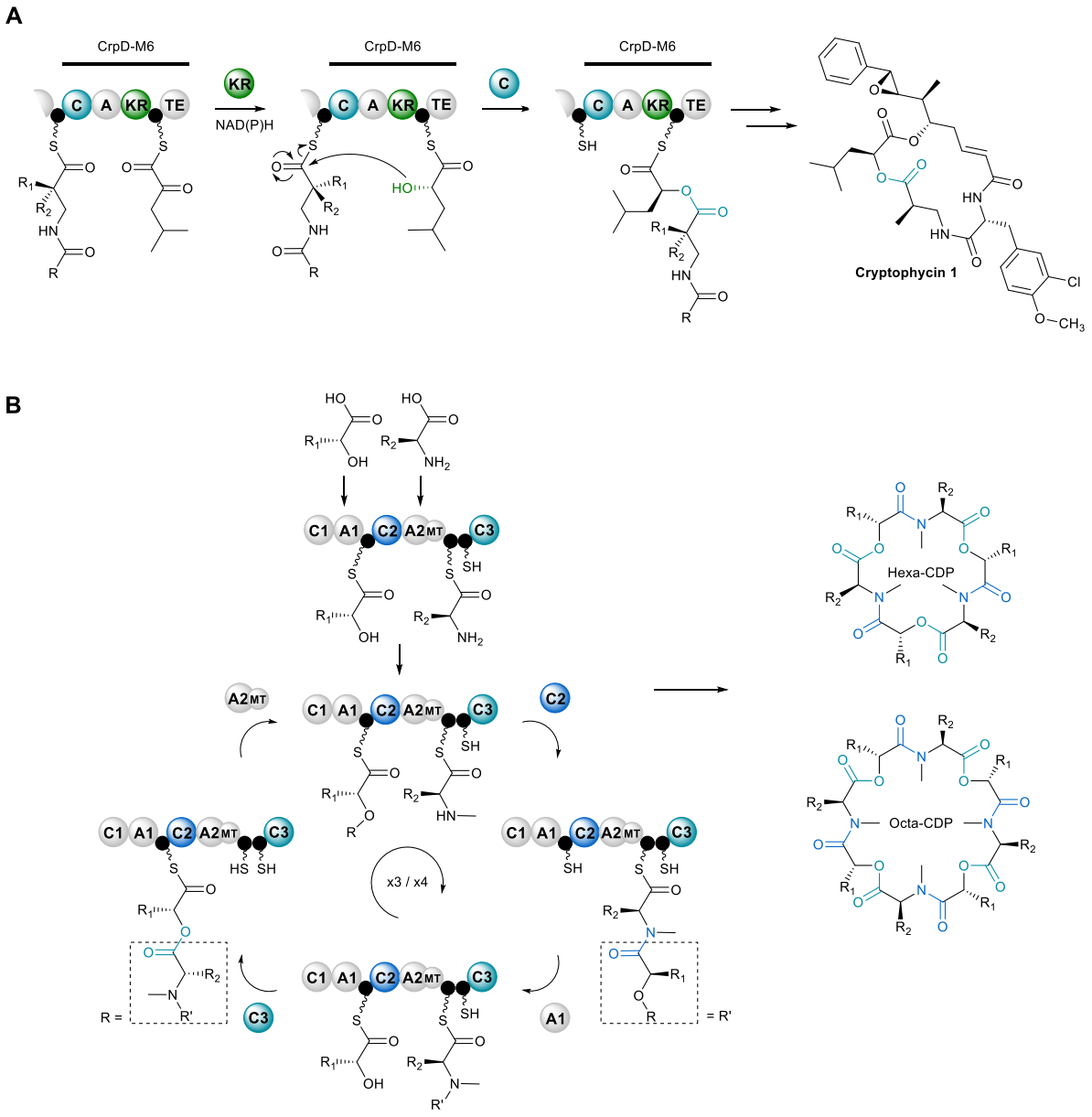

Figure 8. (A) Schematic representation of ester bond formation in the biosynthesis of cryptophycin 1. The A domain selects and activates α-ketoisocaproic acid, which is stereoselectively reduced by the ketoreductase (KR, green) domain. The resulting hydroxyl group serves as a nucleophile for ester bond formation, catalysed by the C domain (turquoise). TE = Thioesterase. (B) Proposed biosynthetic pathway for cyclic depsipeptides (CDPs) in filamentous fungi. D-α-hydroxycarboxylic acids and L-amino acids are respectively activated by A1 and A2, followed by transfer to the PCP domain. Module 2 contains an N-MT domain which methylates the incorporated L-amino acid. C2 (blue) and C3 (turquoise) alternately catalyse amide and ester bond formation to elongate the growing peptide. Meanwhile, the intermediate is shuttled between PCP1 and PCP2a/b. This process is repeated three or four times to form the final hexa- or octa-CDPs.

Finally, the bimodular NRPS AdxA involved in the assembly of the fungal metabolite acu-dioxomorpholine is predicted to feature an atypical ester bond-forming C domain (C_R) that contains an arginine residue in place of the highly conserved histidine residue. The exact role of the arginine residue in the condensation reaction remains to be elucidated. Interestingly, active site arginine residues have also been found in X domains (see Section 3.11), two amide bond-forming C domains in the biosynthetic pathway of D-lysergic acid peptides and in a C domain with Diels-Alderase activity from the lovastatin pathway (see Section 3.9).[95,96]

3.5 Isopeptide bond formation

Occasionally, A domains have been shown to activate the carboxylic acid side chain instead of the α-carboxyl group of aminoacyl extender units. In the biosynthetic pathway of the microcystin toxins, for example, the A domains within McyE and the second module of McyB activate the carboxylic side chains of D-glutamate and β-methyl-D-aspartate, respectively. Consequently, the downstream C domains do not catalyse the formation of a conventional peptide bond, but generate an isopeptide bond instead.[97] The A domains within these isopeptide bond-forming modules contain signature amino acid residues in their binding pocket that can be used as a bioinformatics fingerprint to distinguish them from α-amino acid activating domains.[73]

On the other hand, isopeptide bonds between α-carboxyl groups and side chain amines are also known to occur, e.g. in the biosynthesis of erythrochelin. Erythrochelin is assembled by the tetramodular NRPS EtcD. In the proposed biosynthetic pathway, module 3 is predicted to activate L-δ-N-hydroxyornithine, which is then believed to attack the upstream PCP-bound peptidyl thioester intermediate with its δ-amine group (**Fig. 9**).[98,99] However, direct biochemical evidence for the involvement of the C domain in isopeptide bond formation is currently lacking.

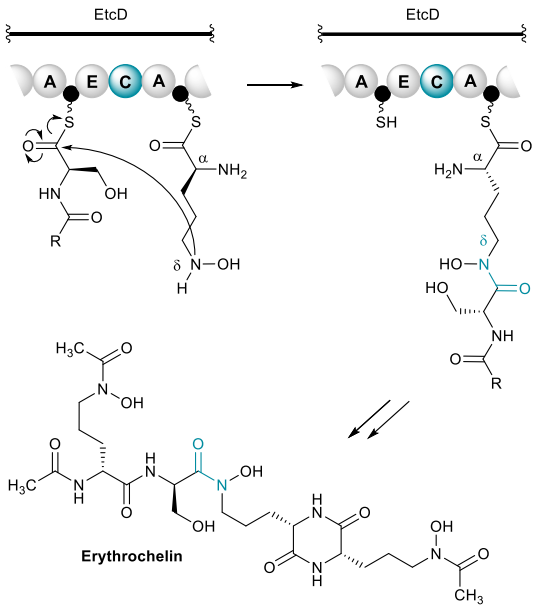

Figure 9. Proposed formation of an isopeptide bond by the third module of EtcD in the biosynthesis of erythrochelin. The isopeptide bond-forming C domain is highlighted in turquoise.

Isopeptide bonds are also found in ε-poly-L-lysine (ε-PL), a broad spectrum antimicrobial agent from *Streptomyces albulus*. ε-PL is composed of 25-35 lysine residues, all connected via isopeptide bonds between the α-carboxyl and ε-amine groups.[100,101] The polymers are assembled by an unusual monomodular NRPS composed of an N-terminal A and PCP domain, followed by three tandem soluble subdomains which are interspersed by six transmembrane regions (**Fig. 10**). Although there is no significant sequence similarity to canonical C domains, *in vitro* biochemical assays with truncated variants of the ε-PL synthetase have indicated that all three subdomains are essential for isopeptide bond formation.[100,102] They have therefore been tentatively designated as C1, C2 and C3.[100] In the proposed biosynthetic pathway, the A domain iteratively adenylates L-lysine monomers and loads them onto the PCP domain. The first PCP-bound lysine residue is believed to be condensed with a soluble L-lysine monomer, resulting in a lysine homodimer. ε-PL polymers of diverse chain length are then generated by iterative elongation of the L-lysine dimer with a varying number of additional PCP-bound lysine residues, until chain elongation is terminated (**Fig. 10**). Whether and how the growing ε-PL chain remains associated with the synthetase during the assembly process remains to be investigated. Analogous NRPS systems have been implicated in the biosynthesis of poly-L-diaminopropionic acid and its mirror-image polymer poly-D-diaminopropionic acid in *Streptomyces* and *Streptoalloteichus* spp., respectively.[103,104]

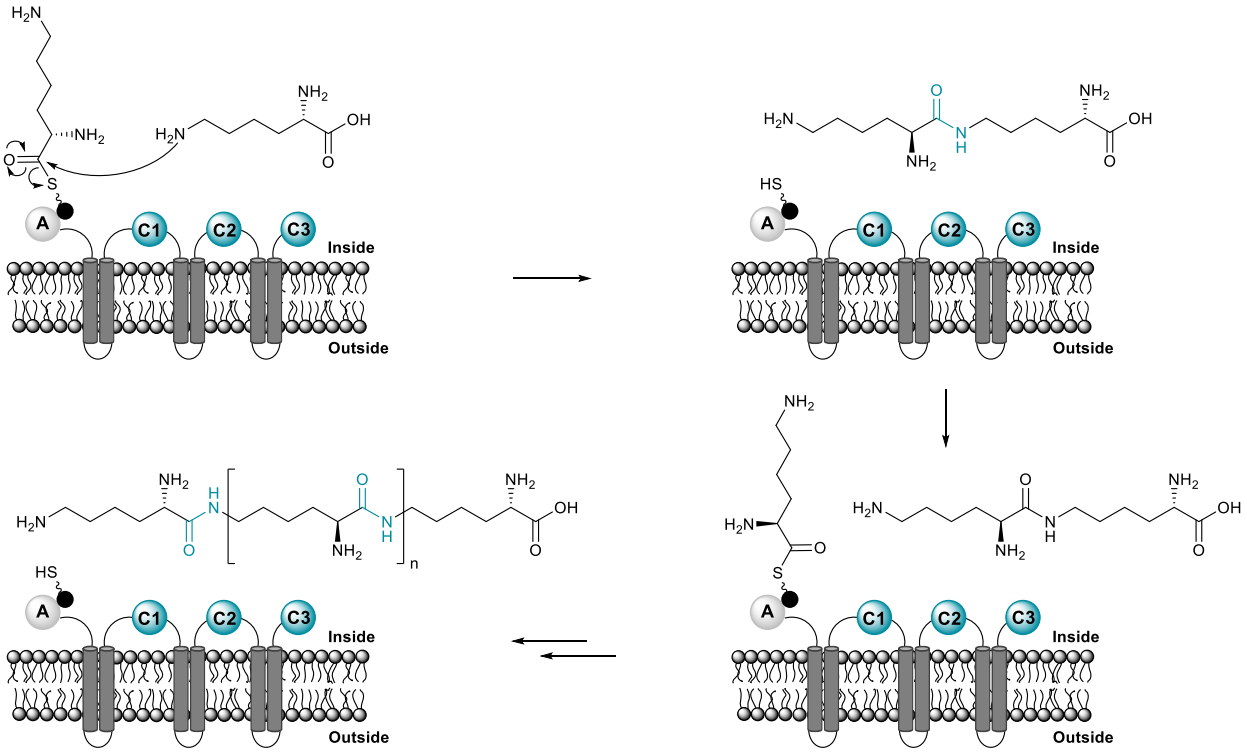

Figure 10. Schematic representation of the proposed biosynthetic pathway of ε-poly-L-lysine in the cytoplasmic membrane in *Streptomyces albulus*. The tandem soluble subdomains that catalyse isopeptide bond formation are highlighted in turquoise.

3.6 β-lactam formation

NRPS C domains have also been shown to fulfil remarkably different functions, such as β-lactam ring formation. β-lactam rings in antibiotics are typically installed by specialized enzymes. The formation of the β-lactam group in penicillins and cephalosporins, for example, is catalysed by a non-heme iron (II)-dependent oxidase, while carbapenems and clavams are formed by ATP-driven ring closure in a process

mediated by dedicated synthetases.[105,106] In the biosynthesis of the monocyclic β-lactam antibiotic nocardicin A, however, the functional lactam ring is introduced by an unconventional NRPS C domain. The nocardicins are assembled by two NRPS subunits, NocA and NocB, which are comprised of five modules. Together, they form the pentapeptide precursor pro-nocardicin G **(Fig. 11A)**.[107–109] β-lactam ring formation is mediated by NocB-C5, the C domain within the final module.[110] This C domain was found to harbour an unusual active site motif with an additional histidine residue directly upstream from the conserved HHxxxDG sequence.[110] Mutational analyses and *in vitro* enzymatic assays with synthetic substrate analogues have indicated the importance of this additional histidine residue in the formation of the β-lactam ring. A mechanism was proposed in which the histidine acts as a base to catalyse β-elimination of a hydroxide from the L-seryl residue of the upstream PCP-bound tetrapeptidyl-thioester, resulting in the formation of an electrophilic dehydroalanyl intermediate **(Fig. 11B)**. Subsequent β-addition of the downstream PCP-tethered L-*p*-hydroxyphenylglycine monomer generates a secondary amine bond, with the additional histidine serving a as catalytic acid in this case. The NocB-C5 domain shows features of $^{D}C_{L}$ domains, which might be explained by the temporary loss of the stereocenter and inversion of the configuration at the seryl β-carbon during this step. Finally, nucleophilic attack of the secondary amine onto the thioester carbonyl completes the formation of the β-lactam ring.[110,111] Interestingly, the extended HHHxxxDG signature motif of the C domain, in combination with an adjacent serine or threonine-activating A domain, may be used as a bioinformatic handle to search for NRPSs that assemble novel β-lactam-containing compounds.[110]

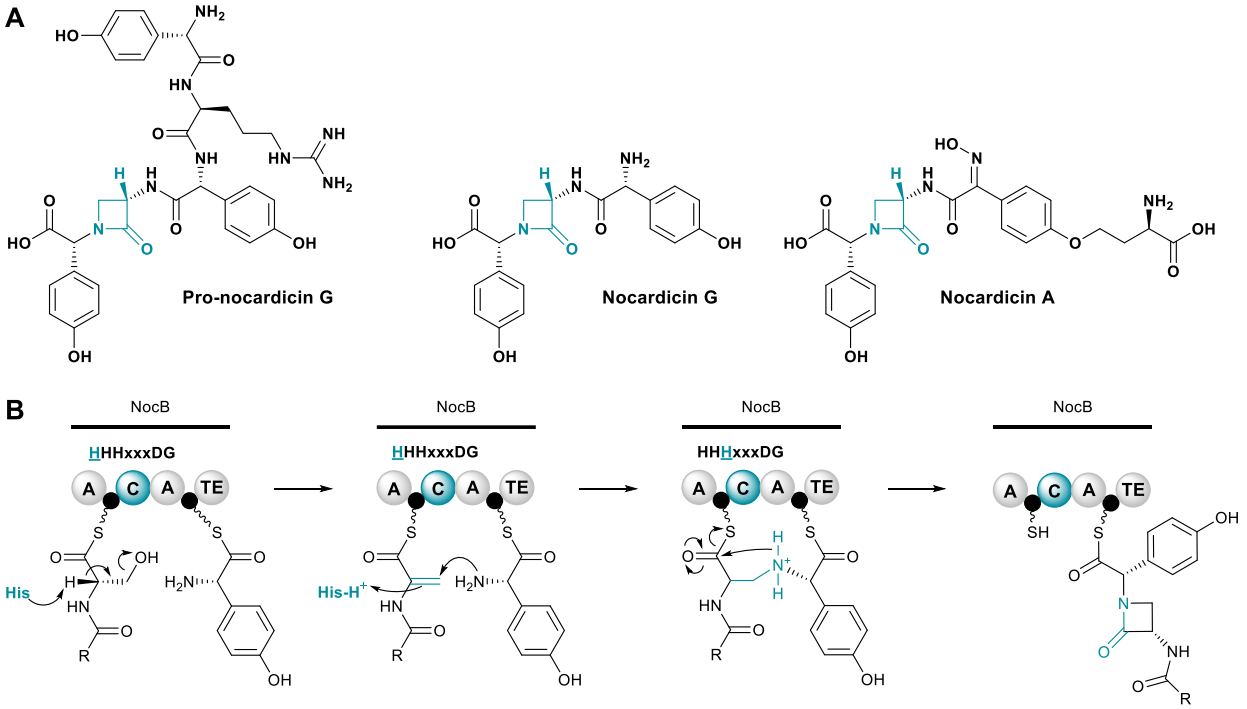

Figure 11. (A) Structures of pro-nocardicin G, nocardicin G and nocardicin A. (B) Proposed reaction mechanism of β-lactam formation in the assembly of the nocardicins by the NocB C domain, shown in turquoise. For each reaction step, the presumed catalytic residue is highlighted, as discussed in the main text.

3.7 Pictet-Spengler cyclization

Another type of C domain that is capable of performing complex chemical transformations is found in the biosynthetic pathways of tetrahydroisoquinoline antitumor antibiotics, such as saframycin A and ecteinascidin 743. The pentacyclic tetrahydroisoquinoline framework of saframycin A is assembled from a fatty acyl chain, an L-alanine, a glycine and two 3-hydroxy-5-methyl-O-methyltyrosine residues by an unusual NRPS (SfmA-C) (**Fig. 12**). Extensive *in vitro* biochemical assays with purified recombinant enzymes and synthetic substrates have indicated that SfmA and SfmB form a PCP-bound N-acyldipeptidyl thioester intermediate that is reductively released by the C-terminal thioester reductase (R) domain within SfmC.[112–114] R domains typically operate at the end of assembly lines to permanently off-load the fully assembled peptide chains. However, in saframycin A biosynthesis, the free aldehyde intermediate serves as substrate for the SmfC C domain. *In vitro* assays with truncated versions of SmfC showed that this unusual C domain functions as a Pictet Spenglerase. It catalyses the condensation of the aldehyde intermediate with a 3-hydroxy-5-methyl-O-methyltyrosine residue attached to the SfmC PCP domain to form an imine, followed by a 6-endo-trig cyclization. Strikingly, the SmfC C domain contains a signature HxxxxD motif in place of the canonical HHxxxDG. The resulting intermediate is then again off-loaded by the R domain and reused as a substrate for a second Pictet-Spengler reaction catalysed by the SmfC C domain. Finally, a third R domain-mediated reduction releases the saframycin A precursor molecule.[112]

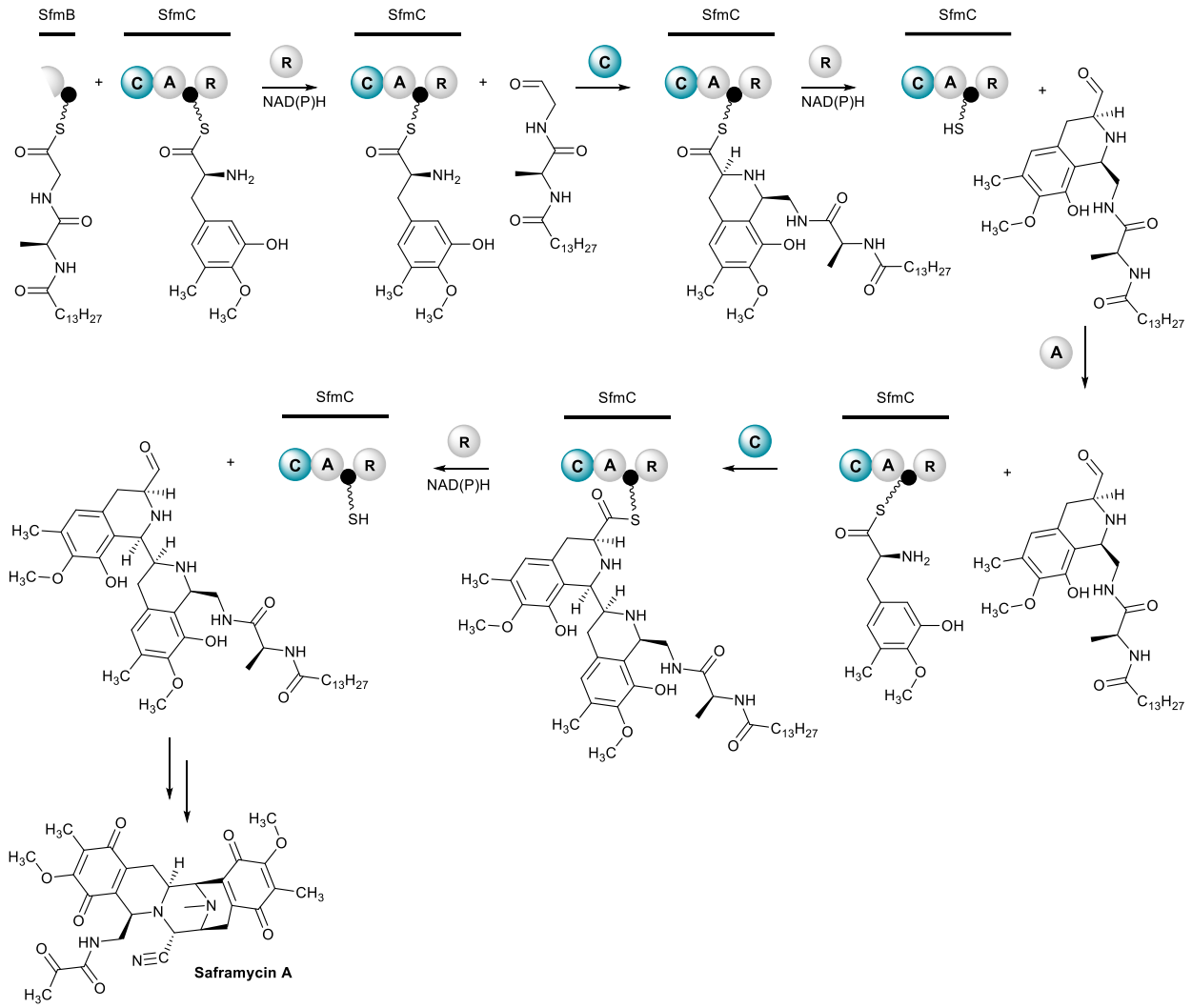

Figure 12. Proposed biosynthetic pathway of saframycin A. R = Reductase domain. The Pictet-Spenglerase C domain is highlighted in turquoise.

3.8 Dehydration

Besides catalysing chain elongation, some C domains are believed to dehydrate the newly incorporated amino acid, particularly in the case of serine and threonine residues. Assembly lines that incorporate dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues typically lack a candidate tailoring enzyme to which this dehydration activity can be attributed. However, phylogenetic analyses have shown that the C domains directly downstream of the serine/threonine-activating A domains all cluster together in a distinct clade. This has led to suggestions that the dehydration activity is embedded within these C domains.[46] Biochemical evidence for this was recently provided by *in vitro* studies of the albopeptide biosynthetic pathway (AlbA-AlbB) in *Streptomyces albofaciens*. *In vitro* reconstitution experiments with purified recombinant enzymes and chemical capture of biosynthetic intermediates revealed that the C2 and C3 domain within AlbB are responsible for the formation of Dha and (*E*)-Dhb, respectively.[115] A similar dehydration is predicted to take place in the microcystin biosynthetic pathway. The McyA A1 domain selects and incorporates a serine residue, which undergoes N-methylation by the adjacent N-MT domain.

Next, the downstream C domain is proposed to catalyse the formation of N-methyldehydroalanine by dehydration of the seryl side chain (**Fig. 13A**).[116,117] A Dha residue is also believed to be installed by the second C domain within BlmVI in the bleomycin biosynthetic pathway.[118] Likewise, N-methylthreonine is presumably dehydrated to N-methyldehydrothreonine in the nodularin biosynthetic pathway by the action of the NdaA C domain.[119] Future structural investigations may shed light on the exact reaction mechanism of these C domains.

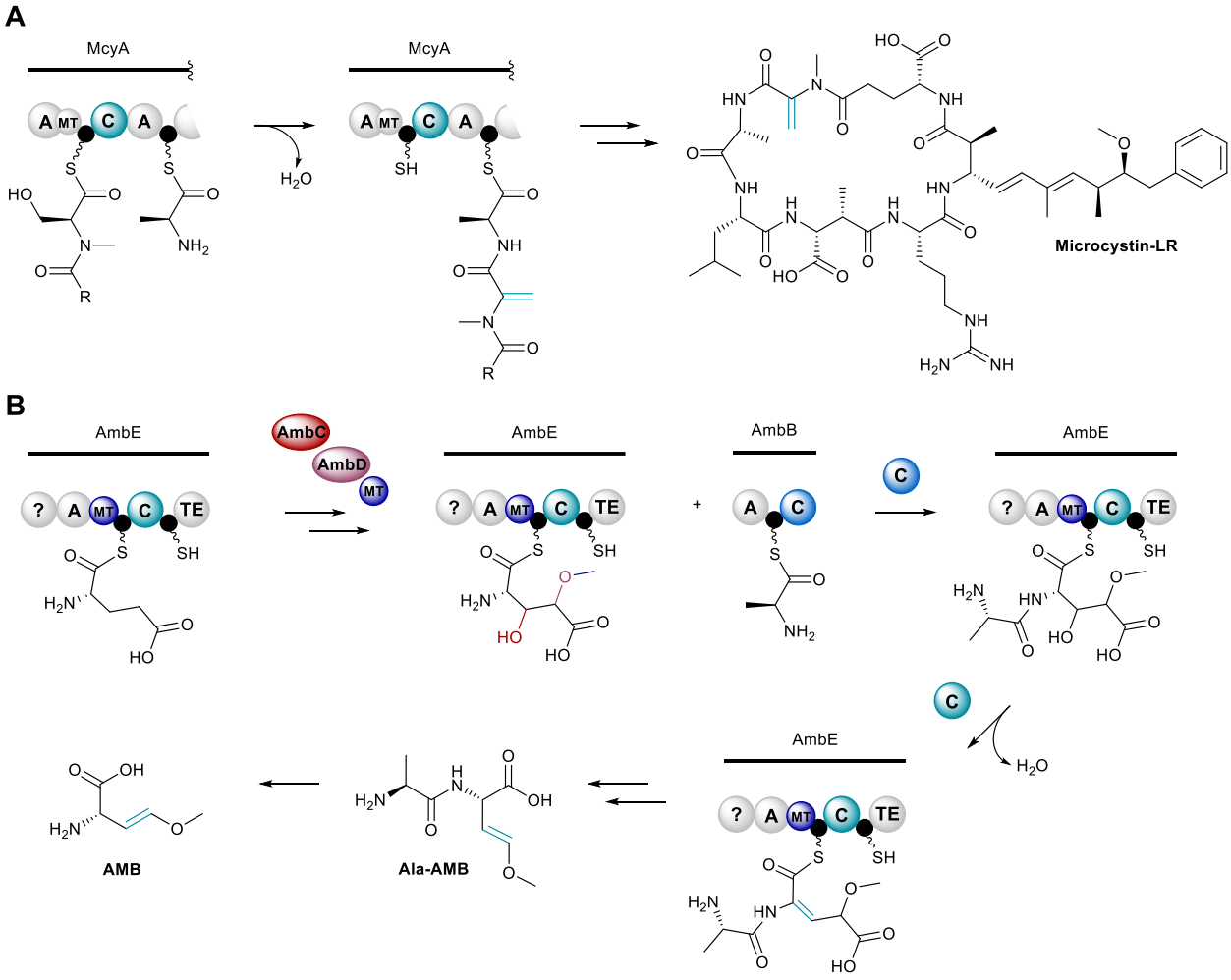

Figure 13. (A) Proposed dehydration reaction catalysed by the first C domain of McyA in the biosynthesis of microcystin-LR. (B) Proposed biosynthetic pathway of L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB). The question mark indicates a domain of unknown function. The C domains that are presumed to have dehydration activity are highlighted in turquoise.

Another remarkable dehydration event takes place in the biosynthesis of oxyvinylglycines, a family of nonproteinogenic amino acids that shows antagonistic activity against a wide range of bacteria and plants by inhibition of PLP-dependent enzymes.[120] A well-known example is methoxyvinylglycine, also known as L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB). AMB was first isolated from *Pseudomonas aeruginosa* PAO1 and exhibits potent antibiotic activity against a variety of bacterial strains (**Fig. 13B**).[121,122] The *amb* biosynthetic gene cluster encodes an NRPS assembly line which consists of two NRPS subunits, AmbE and AmbB. AMB is produced from glutamate as a dipeptide coupled to alanine. Based on *in vitro* reconstitution

and isotopic labelling experiments, a biosynthetic pathway was proposed that starts with the loading of L-glutamate and L-alanine onto the first AmbE and the AmbB PCP domain, respectively. The glutamyl thioester undergoes two hydroxylations and a methylation reaction, and is subsequently condensed to the PCP-linked alanyl residue by the action of the AmbB C domain. The resulting alanyl-3-hydroxy-4-methoxy-glutamate thioester intermediate is then proposed to undergo dehydration, forming a double bond between C2 and C3. This reaction is presumably catalysed by the AmbE C domain since phylogenetic analysis revealed that this non-canonical C domain clusters with the McyA and BlmVI C domains.[123] The 2,3-double bond is believed to promote subsequent decarboxylation and isomerization, and hydrolytic chain release then affords the biologically inactive Ala-AMB. The N-terminal alanine residue masks the functionally important α-amino group of AMB and is presumably cleaved off during export as part of a self-protection strategy.

3.9 Diels-Alder cyclization

The catalytic capabilities of NRPS C domains also extend to Diels-Alder cyclizations. A well-known example is found in the biosynthesis of the fungal metabolite lovastatin. The highly reducing iterative lovastatin polyketide synthase (PKS) LNKS is fused to a non-canonical C domain.[124,125] The catalytically important histidine residue in the active site HHxxxDG motif of this C domain is substituted by arginine, and its crystal structure indicates that the acceptor site is blocked by a flexible loop.[96,124] The same phenomenon has also been observed in other C domains that do not utilise downstream acceptor substrates, such as X domains (see Section 3.10) and C-terminal C_T domains of fungal NRPSs (see Section 3.12.1.1).[126,127] Both *in vitro* and *in vivo* assays have shown that the LNKS C domain is indispensable for the production of the lovastatin precursor dihydromonacolin L, but does not participate in chain release.[125] Instead, it is proposed to catalyse a unique cycloaddition reaction on the LNKS-tethered hexaketide triene intermediate to form the decalin ring system.[128] This was confirmed by heterologous expression of C domains from other fungal PKS-NRPSs with a truncated variant of LNKS lacking the C domain. These experiments only yielded non-cyclized products, while *in-trans* expression of the LNKS C domain was able to restore dihydromonacolin L production.[124,129]

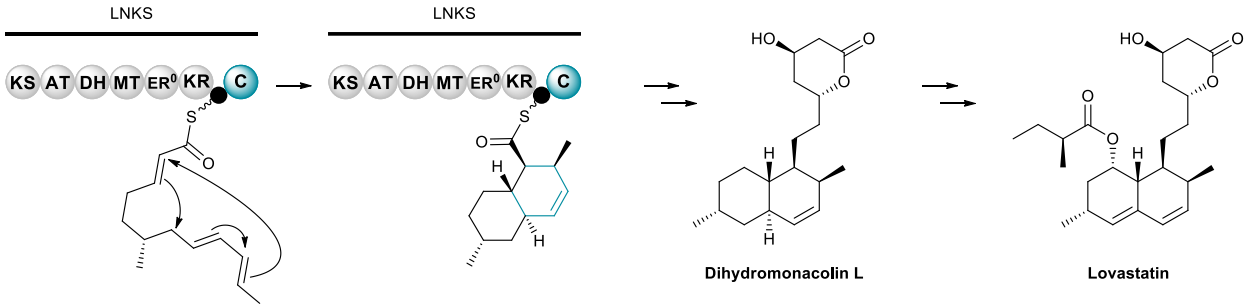

Figure 14. Proposed Diels-Alder reaction catalysed by the C-terminal C domain of LNKS in the biosynthesis of lovastatin. KS = Ketosynthase. AT = Acyltransferase. DH = Dehydratase. ER^0 = Inactive enoyl reductase.

3.10 Recruitment of auxiliary enzymes

Glycopeptide antibiotics (GPAs), such as vancomycin and teicoplanin, are an important class of structurally complex nonribosomal peptides that effectively kill Gram-positive bacteria by inhibiting peptidoglycan biosynthesis. They consist of a highly crosslinked heptapeptide core that can be extensively modified by glycosylation, halogenation, methylation, acylation and sulfation.[130] The crosslinks between the aromatic amino acid side chains introduce a high degree of structural rigidity that is essential for the biological activity of the GPAs. *In vivo* gene disruption experiments have shown that the oxidative crosslinking reactions are catalysed by a cascade of trans-acting cytochrome P450 monooxygenases (known as Oxy enzymes) while the heptapeptide substrate is still attached to the NRPS assembly line.[131–138] A comprehensive study by Haslinger and Peschke *et al.* revealed that an atypical C domain within the final module of the NRPS, called the X domain, is reponsible for recruiting and aligning the oxygenase enzymes towards the adjacent PCP domain (**Fig. 15**).[127] Structural characterization of the teicoplanin X domain by X-ray crystallography showed that it adopts the same overall fold as canonical C domains, but harbours a modified HRxxxDD active site motif. The arginine residue in this motif was shown to obstruct the substrate binding tunnel, rendering the X-domain catalytically inactive.[127] The structure of the X domain in complex with the cytochrome P450 enzyme OxyB further allowed the interaction interface between these two proteins to be delineated.[127] Subsequent *in vitro* kinetic experiments revealed that the different oxygenase enzymes continuously compete for the same interaction interface on the X domain and are able to 'scan' the cyclization state of the peptidyl thioester intermediate.[137–139] Recognition of the correct substrate leads to a conformational change in the enzymes that enhances their affinity for the substrate and initiates catalysis.[139] X domains as enzyme recruitment platforms are therefore yet another testament to the remarkable functional versatility of C domains.[140]

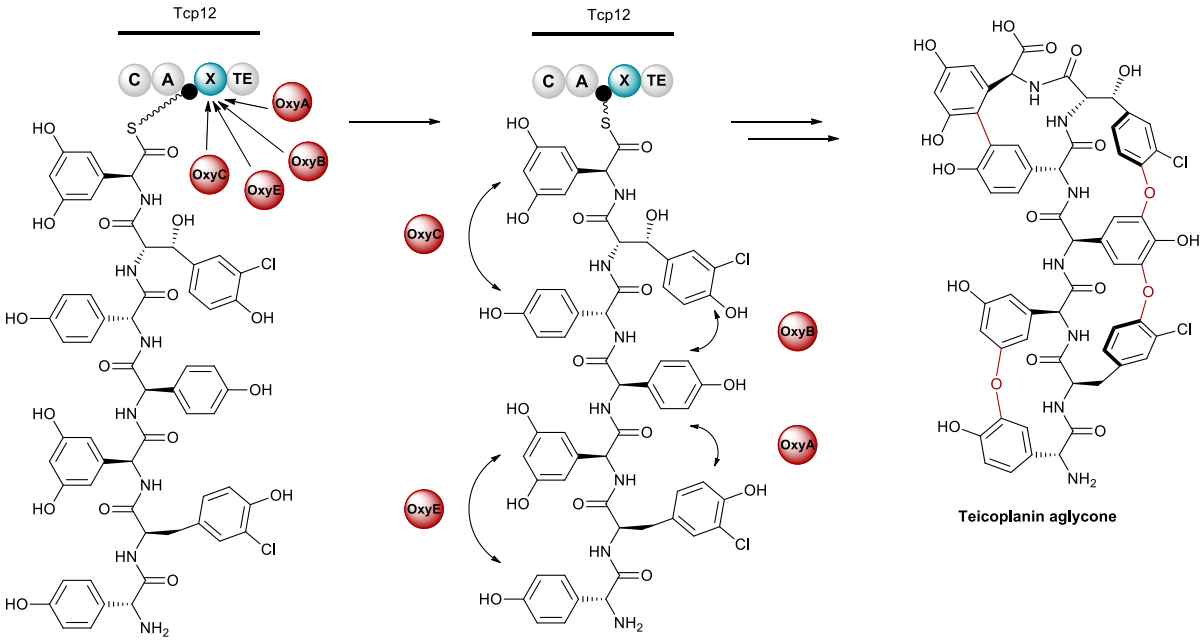

Figure 15. Recruitment of cytochrome P450 monooxygenases by the X domain in the final module of the teicoplanin NRPS. The X domain is highlighted in turquoise, while the P450 oxygenases are coloured in red.

3.11 Interfacing

Recently, a novel C domain function was predicted based on bioinformatic analysis of bacterial siderophore biosynthetic pathways that employ β-hydroxylases (βHs).[141] βH enzymes are known to hydroxylate PCP-bound L-Asp residues in the biosynthesis of peptide siderophores, such as serobactin, cupriachelin and alterobactin.[141,142] The resulting β-hydroxyaspartic acid residues play a key role in iron-chelation. β-hydroxylases belong to several different enzyme families and can either act as stand-alone enzymes, or as embedded domains fused to the N-terminus of NRPS subunits. The latter typically appear alongside an interface (I) domain in the assembly line. I domains are homologous to canonical C domains, but cluster in a separate phylogenetic clade and lack the conserved HHxxxDG motif. Only the active site aspartate residue is preserved, for which a structural role is predicted. I domains are found directly upstream from the A domain that selects and loads the aspartate residue. Similar to the function of X domains, the I domains are proposed to correctly position the βH domain and the aminoacyl thioester substrate for β-hydroxylation (**Fig. 16**). However, direct biochemical evidence for this is still missing. In a few rare cases, I domains have also been found in pathways with a discrete histidine β-hydroxylase, e.g. in the biosynthesis of histicorrugatin.[141]

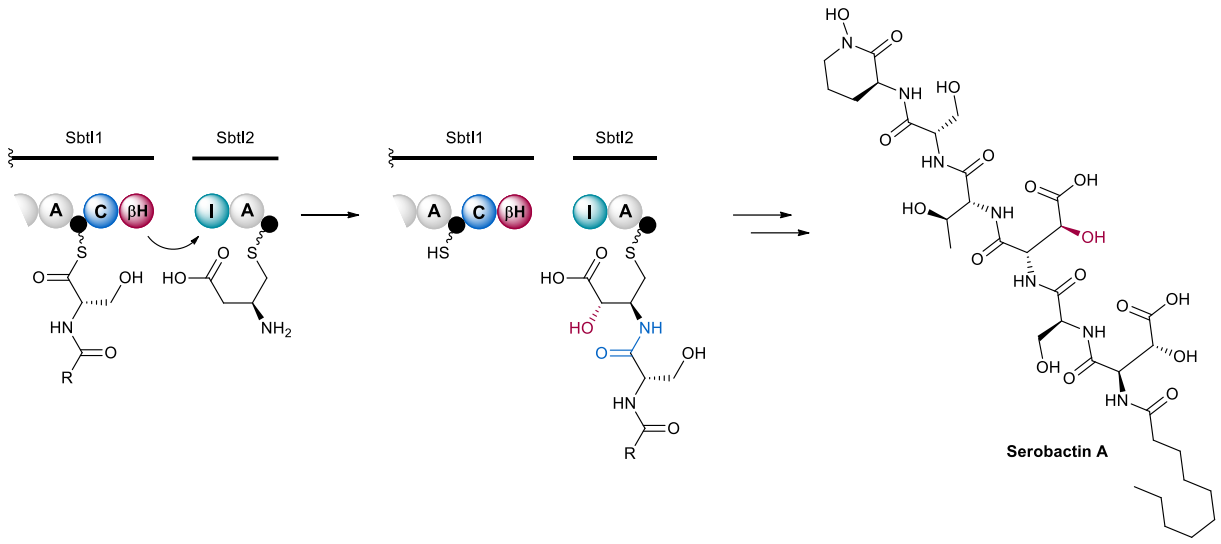

Figure 16. Proposed hydroxylation of the L-aspartyl thioester intermediate during the biosynthesis of serobactin A. The interface (I) domain is believed to be important for correct positioning of the β-hydroxylase (βH) towards the aminoacyl substrate. The βH, I domain and C domain are highlighted in red, turquoise and blue respectively.

3.12 Chain release catalysts

At the C-terminal end of NRPS assembly lines, C domains have also been implicated in various types of chain release reactions, ranging from intra- and intermolecular cyclizations to hydrolysis (**Fig. 17**).

3.12.1 Intramolecular cyclization

3.12.1.1 Macrolactamization

Macrocyclic nonribosomal peptides in filamentous fungi are typically released from their assembly lines by a C-terminal C_T domain, rather than the canonical TE domain observed in many bacterial NRPSs.[126,143] Well-known examples include cyclosporin A,[144] echinocandin,[145] apicidin[146] and the fungal siderophores ferrichrome A[147] and ferricrocin[148]. C_T domains catalyse macrolactamization by using an internal amine group as a nucleophile in the chain release reaction (**Fig. 17A**).[143] The role of the C_T domain in the

biosynthesis of the multicyclic alkaloid fumiquinazoline F has been extensively studied *in vitro* by Gao *et. al.* (2012).[143] The TqaA NRPS responsible for the assembly of fumiquinazoline F was shown to generate an anthranilate-L-tryptophan-L-alanine tripeptidyl thioester which is cyclized and released by the C-terminal C_T domain to form a 10-membered macrolactam product. Subsequent transannular cyclizations and aromatizing dehydrations yield fumiquinazoline F.[143] A similar pathway has been postulated for asperlicin biosynthesis.[149] Gao *et al* also observed that the TqaA C_T domain is unable to accept peptidyl-SNAC or peptidyl-CoA substrates, indicating the importance of the protein-protein interactions with the adjacent PCP domain.[143] The crystal structure of the TqaA C_T domain revealed a similar fold as in canonical C domains. Like in the case of the LNKS C domain and X domains[96,127], the entrance to the acceptor substrate channel was found to be blocked, consistent with the absence of a PCP-bound acceptor substrate.[126]

3.12.1.2 Macrolactone formation

Aside from amine groups, C_T domains are also known to utilise internal hydroxyl groups as nucleophiles in macrolactonization reactions (**Fig. 17B**). Prominent examples are found in the biosynthesis of the fungal metabolite aureobasidin A[150] and the polyketide-nonribosomal peptide hybrids apratoxin A [151] and thermolides A-F.[152] The latter are formed by the concerted action of the highly reducing iterative monomodular PKS ThmA, which forms the polyketide core, and the monomodular NRPS ThmB (C-A-PCP-C_T). The N-terminal C domain of ThmB couples the PCP-linked aminoacyl thioester to the fully assembled polyketide chain, while the C-terminal C_T domain, which harbours an HHxxxD active site motif, mediates macrolactonization via ester bond formation.[152] Remarkably, the C-terminal C domains of the iterative NRPSs involved in the biosynthesis of the enniatin, beauvericin and bassianolide cyclodepsipeptides do not only catalyse internal ester bonds, but are also responsible for chain length determination and peptide macrolactonization.[93,94] In contrast to C_T domains from modular fungal NRPSs, these C domains have a functional acceptor site.[94]

Macrolactone-forming C domains of bacterial origin have been reported in the hybrid PKS-NRPS pathways that assemble the immunosuppressants rapamycin,[153,154] FK506,[155] FK520[156] and the structurally related meridamycin in *Streptomyces* spp.[157] These metabolites harbour an unusual NRPS-derived pipecolate moiety, which is conjugated to the fully assembled polyketide chain by the action of RapP/FkbP/MerP, a monomodular NRPS with C-A-PCP-C domain organization. The first C domain condenses the linear polyketide scaffold to the PCP-tethered pipecolate unit, while the C-terminal C domain is believed to catalyse macrolactonization and concomitant chain release.[158]

3.12.1.3 Regioselective cyclizations

Among the more complex cyclization reactions that C domains may perform during chain termination are benzazepine-dione formation, diketopiperazine formation and Dieckmann cyclisation.

A unique C_T domain capable of catalysing benzazepine-dione formation was recently identified in the biosynthetic pathway for the peptidyl alkaloid nanangelenin A from *Aspergillus nanangensis*.[159] The core anthranilic acid-L-kynurenine cyclic dipeptide (nanangelenin B) is formed by the bimodular NRPS NanA, which consists of seven domains in the order A-PCP-C-A-PCP-C_T-PCP. The C_T domain harbours an SHxxxD

active site motif and catalyses a regioselective lactamization between the side chain amine group of the PCP-tethered L-kynurenine residue and the thioester carbonyl to generate the unusual benzazepine scaffold in nanangelenin A (**Fig. 17C**). Heterologous expression of NanA mutants *in vivo* showed that the presence of the C-terminal PCP domain greatly enhances the formation of nanangelenin B, indicating that the peptide is presumably transferred to the final PCP domain prior to cyclization.[159]

Natural products with 2,5-diketopiperazine (DKP) moieties are known for their structural diversity and their impressive range of biological activities.[160] The formation of NRPS-derived diketopiperazines involves the assembly of a peptidyl thioester precursor, which undergoes regioselective cyclization and subsequent chain release (**Fig. 17D**). A well-known example is the fungal virulence factor gliotoxin, produced by the notorious human pathogen *A. fumigatus*. The bimodular NRPS converts L-phenylalanine and L-serine into the DKP cyclo-(L-phenylalanyl-L-seryl), and is organised into seven domains in the order A-PCP-C-A-PCP-C_T-PCP. Although DKP formation appears spontaneous *in vitro*, *in vivo* mutational analysis has shown that the final C_T and PCP domain are critical for cyclorelease.[161] Like the C_T domain in nanangelenin biosynthesis, the GliP C_T domain is characterized by an SHXXXD active site motif. A mechanism has been proposed in which the linear dipeptide is transferred to the terminal PCP domain, followed by C_T-catalysed cyclization. The C-terminal PCP domain is hypothesized to serve as an anchor point for additional tailoring enzymes, prior to cyclization and release of the diketopiperazine moiety. Similar transformations occur in the NRPS-derived diketopiperazines brevianamide F[162], roquefortine C[163] and erythrochelin[98,99].

Finally, C domains have also been proposed to mediate chain release by catalysing a Dieckmann cyclization **(Fig. 17E**). In the biosynthesis of the antibiotic malonomycin in *Streptomyces rimosus*, the hybrid PKS-NRPS MloJ is proposed to elongate the L-serine-L-γ-aspartate dipeptidyl thioester tethered to MloI with a malonyl and L-diaminopropionate unit, followed by Dieckmann cyclization and chain release to form premalonomycin.[164] However, further studies are needed to add weight to this hypothesis.

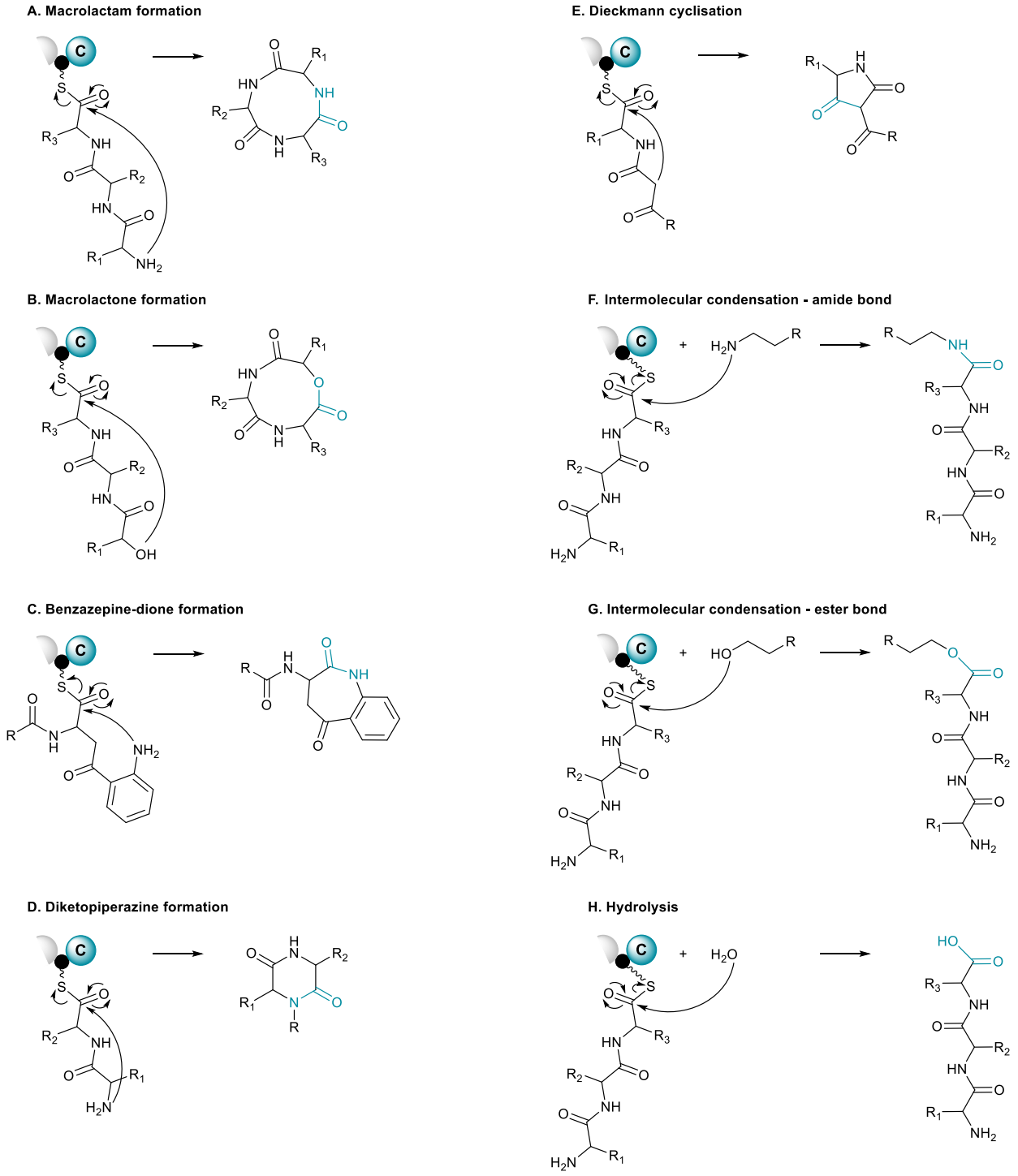

Figure 17. Schematic overview of C domain-mediated chain release reactions.

3.12.2 Intermolecular condensation

C domains do not only mediate chain release by intramolecular cyclizations, but they also catalyse intermolecular condensation reactions between fully-assembled PCP-bound peptidyl or polyketide

thioesters and a diverse range of soluble acceptor molecules (**Fig. 17F and Fig. 18**). The first biochemically characterized example of such a C domain is VibH, a stand-alone C domain that plays an important role in the biosynthesis of the siderophore vibriobactin in *Vibrio cholera*.[165] Vibriobactin is a triacylated triamine that is constructed from two molecules of L-threonine, three molecules of 2,3-dihydroxybenzoic acid and one molecule of norspermidine by NRPS enzymes (**Fig. 18A**).[166] VibH selectively monoacylates a primary amine of norspermidine by catalysing the condensation with a PCP-linked dihydroxybenzoyl thioester.[165] The resulting N1-acylated norspermidine product then undergoes two further acylations on its two remaining free amines by the NRPS VibF. In two iterative rounds, VibF activates two L-threonine molecules and condenses them to a dihydroxybenzoyl residue. The resulting thioester intermediates undergo heterocyclization and dehydration, and are then transferred onto the remaining N9 primary amine and N5 secondary amine group of N1-(2,3-dihydroxybenzoyl)-norspermidine by the C-terminal VibF C domain to yield vibriobactin.[32,166] Interestingly, *in vitro* biochemical assays have demonstrated that the condensation reactions catalysed by VibH and the C-terminal C domain of VibF are both reversible. Hence, these enzymes are capable of reloading their cognate PCP domains using acylated norspermidine as a substrate, albeit with reduced efficiency compared to the forward reaction.[167] Other examples of C domains that mediate chain release by condensing PCP-bound peptidyl thioesters to soluble amine substrates are found in the biosynthetic pathways that assemble the *Burkholderia* siderophores malleobactin and ornibactin[168,169], the rhabdopeptide/xenortide peptides in nematode symbiotic *Photorhabdus* and *Xenorhabdus* strains[170–172], congocidine, distamycin and disgocidine[173] and the bleomycins, tallysomycins and zorbamycins.[118,174,175]

C domains have also been found fused to a PKS assembly line where they release the fully-assembled polyketide chain by condensation with a small amine substrate. This has been observed in the production of the fungal wortmanamides by the hybrid PKS-C TwmB in *Talaromyces wortmanii*. In this pathway, the PKS-derived octa- and nonaketide chains are condensed to the ω-amine group of a soluble 5-aminopentanoic acid acceptor substrate (**Fig. 18A**). The catalytically important histidine residue in the active site motif of this C domain is replaced with a proline residue, which is proposed to play a role in substrate positioning.[176]

Another remarkable collaboration between a PKS assembly line and a stand-alone C domain is found in the biosynthesis of the broad spectrum zeamine antibiotics in *Serratia plymuthica*.[177–179] The zeamines result from the condensation of a 40-carbon polyaminoalcohol to hexapeptide-mono-/diketide thioesters generated by a hybrid PKS-NRPS. The polyaminoalcohol zeamine II is formed by a polyunsaturated fatty acid synthase-like assembly line. The free-standing C domain Zmn19 releases the ACP-bound hexapeptide-mono-/diketide thioesters via intermolecular condensation with the primary amine group of zeamine II (**Fig. 18A**).[177,178,180] A similar chain release mechanism takes place in the biosynthesis of the fabclavines, which are structurally related to the zeamines.[181,182]

In addition to polyamines, C domains are capable of using a highly diverse range of soluble primary amine-containing molecules as acceptor nucleophiles for chain release, ranging from nucleosides[183,184] and aminoglycosides[185] to pteridines[186,187]. Interestingly, a rare example of an NRPS C domain that catalyses intermolecular condensation has been found in higher eukaryotes. Ebony is a monomodular NRPS (A-PCP-C) from *Drosophila melanogaster* that plays a key role in the regulation of the amine neurotransmitters histamine and dopamine in various cell tissues. By condensation of a PCP-bound β-alanyl thioester to histamine or dopamine, the neurotransmitters become inactivated. The condensation reaction is catalysed by the C-terminal C domain of Ebony. Crystal structures have revealed that, in contrast to

canonical C domains, this domain does not adopt the typical V shape, but rather a rigid aryl-alkylamine-*N*-acetyl transferase fold. This fold might facilitate the extremely rapid condensation reaction, which is critical for its regulatory role *in vivo*. β-Alanine might be preloaded onto the Ebony PCP domain, preparing the enzyme for a neurotransmitter burst and allowing fast inactivation.[188,189]

Aside from generating amide bonds, C domains have also been shown to mediate chain release by intermolecular ester bond formation (**Fig. 17G and 18B**). The first esterifying C domain was found in the biosynthesis of the polyketide-derived mycotoxin fumonisin in filamentous fungi. Studies on *Fusarium verticillioides* demonstrated that the PCP-C didomain NRPS Fum14p is responsible for the formation of two tricarballylic esters on the fumonisin polyketide backbone.[190,191] The tricarballylic acids, or precursors thereof, are presumably activated by the action of Fum10p, a standalone acyl-CoA synthetase or A domain. The tricarballylic monothioesters are then transferred to the Fum14p PCP domain and subsequently condensed to the C14 and C15 hydroxyl groups of the freely diffusible fumonisin backbone by the Fum14p C domain **(Fig. 18B)**. The resulting tricarballylic esters are critical for the toxicity of fumonisin.

Similarly, SgcC5, a discrete C domain in the biosynthetic pathway of the enediyne antitumor antibiotic C-1027, catalyses the regio- and stereoselective esterification of SgcC2 PCP-bound (S)-3-chloro-5-hydroxy-β-tyrosine to the enediyne core. The interaction between SgcC5 and SgcC2 was proven to essential in this respect, since SgcC5 is incapable of catalysing condensation with a free donor substrate. Remarkably, *in vitro* experiments showed that SgcC5 is also capable of conjugating substrates via an amide bond, albeit to a lesser extent. SgcC5 is therefore the first C domain discovered that is able to catalyse both ester and amide bond formation.[192] Strikingly, SgcC5 consists as a homodimer in solution, contrasting the monomeric character of canonical NRPSs.[193] Phylogenetic analysis has revealed that ester bond forming C domains involved in enediyne biosynthesis cluster in a separate clade[193], suggesting similar reaction mechanisms for other members of this class, such as the kedarcidin (KedY5)[194] and sporolide (SpoT10)[195] enediynes.

Other acceptor substrates that have been used for C domain-mediated ester bond formation in chain release reactions include dihydroxycyclohexane carboxylic acid (DHCCA) and trihydroxy confertifolin in the biosynthesis of the polyketide antibiotic enacyloxin and the astellolide sesquiterpenes, respectively (**Fig. 18B**).[196–198] In enacyloxin biosynthesis, a non-elongating ketosynthase (KS^0) domain first transfers the fully assembled polyketide chain from the final ACP domain in the assembly line to a separate PCP domain. The Bamb_5915 C domain then offloads the chain from the PCP domain by condensing it with (1*S*,3*R*,4*S*)-3,4-DHCCA. Extensive genetic and biochemical experiments have revealed that the KS^0 domain-mediated transacylation reaction circumvents the inability of the C domain to communicate with the ACP domain.[197] The crystal structure of the Bamb_5915 C domain was elucidated and the molecular basis for its interaction with the PCP domain was fully characterized using a combination of solution NMR spectroscopy, carbene footprinting and molecular dynamics simulations.[196] Interestingly, the Bamb_5915 C domain responsible for the intermolecular ester bond formation has been shown to have a relaxed substrate specificity, both at the acceptor and the donor site. This opens the door to producing a wide range of novel enacyloxin derivatives via biosynthetic engineering.[197]

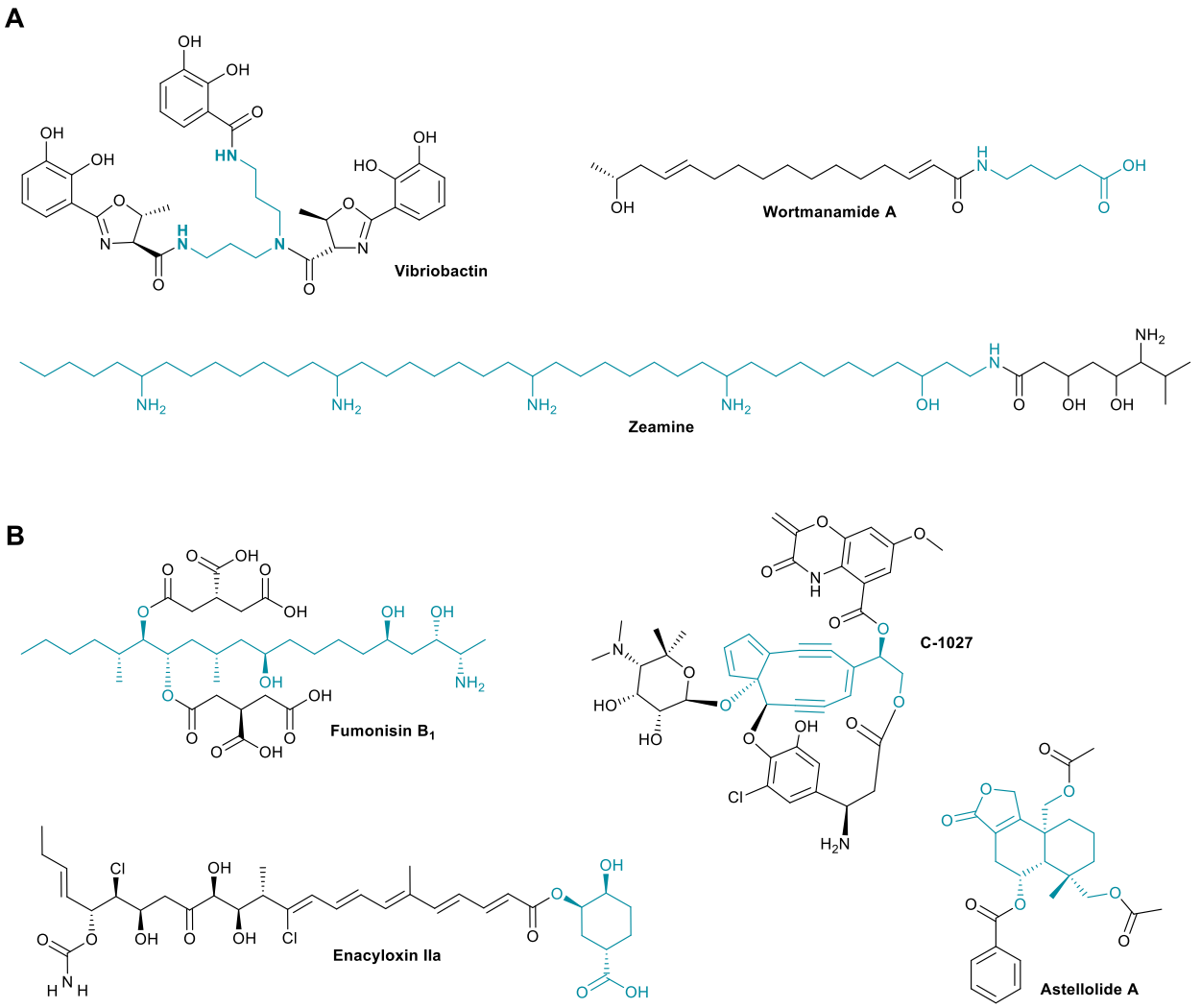

Figure 18. Examples of natural products that are assembled *via* intermolecular condensation with the amino (A) or hydroxyl (B) group of a nucleophilic acceptor substrate. The acceptor substrates used in these C domain catalysed reactions are indicated in turquoise.

3.12.3 Hydrolysis

On very rare occasions, terminal C domains have been found to release fully assembled peptidyl thioesters by hydrolysis (**Fig. 17H**). The first example was reported in the biosynthesis of the antifungal and cytotoxic crocacins in *Chondromyces crocatus* Cm c5. The final NRPS module, encoded by *croK*, has a C-A-PCP-C architecture. Deletion of the C-terminal C domain completely abolishes crocacin production *in vivo*. Further *in vitro* biochemical assays with purified CroK-C2 and crocacin-SNAC substrates validated its role in the release of a linear compound by hydrolysis.[199] Recently, a second example has been discovered in IvoA, a monomodular NRPS associated with pigment formation in *Aspergillus nidulans*. As shown by *in vitro* experiments, this NRPS presumably functions as a catalyst for unidirectional stereoconversion of L-tryptophan to D-tryptophan. After L-tryptophan activation by the A domain and transfer to the PCP, the E domain inverts the chirality of the substrate. The C domain subsequently acts as stereochemical gatekeeper to only release D-configured tryptophan by hydrolysis.[200]

4. Condensation domains as a starting point for engineering nonribosomal peptide synthetases

The impressive chemical diversity of nonribosomal peptides is translated into a wide variety of biological activities and therapeutic applications. Despite their promising biological activities, however, natural products often require structural modifications to circumvent issues with toxicity and resistance development, or to improve their activity, stability and/or bioavailability. The structural complexity of most nonribosomal peptides makes the preparation of derivatives via semi- or total synthesis extremely challenging. Rational biosynthetic pathway engineering has the potential to circumvent these issues and to generate novel natural product derivatives with improved pharmaceutical properties. The modular architecture of NRPSs greatly facilitates such approaches, offering possibilities to modify, insert or remove domains and/or complete modules within assembly lines. Multiple research groups have therefore attempted to rationally modify nonribosomal peptide synthetases, and this has been the topic of various reviews.[9,201–205] So far, the main focus of these engineering efforts has been on the substitution or engineering of A domains, due to their primary gatekeeping role in the selection of biosynthetic precursors. Unfortunately, many of these of engineering attempts have been unsuccessful, leading to drastic drops in the desired titres or complete loss of peptide production. Over the past decade, it has become increasingly clear that rational biosynthetic pathway engineering is a complex puzzle where many factors need to be taken into account, including the impact of the C domain on the rate[206] and specificity[26,207] of amino acid activation by the A domain. Due to their central role as chain elongation catalysts, C domains represent a promising yet underexplored starting point for biosynthetic engineering strategies. Here, we will discuss several examples of C domain-based engineering approaches that have been undertaken so far. While many efforts have been focused on canonical C domains, it is clear that harnessing the exceptional functional versatility of C domains will offer tremendous opportunities in the future for expanding the chemical diversity of nonribosomal peptides and other natural products (**Fig. 19**).

4.1 Whole C domain swapping

The biosynthesis of fungal cyclic depsipeptides (CDPs) offers exciting possibilities for engineering by C domain swapping (**Fig. 19A**). As mentioned in section 3.12.1.2, the C-terminal ester bond forming C3 domains in the enniatin, bassianolide and beauvericin biosynthetic pathways control the chain length of the final products. Exchange of these C3 domains or PCP-C3 didomains has been shown to trigger the production of compounds with modified chain lengths, in accordance with the cognate substrate specificity of the different C3 domains.[93,94] This has enabled the production of octa-enniatin B, octa-beauvericin and hexa-bassianolide in substantial yields by heterologous expression in *Aspergillus niger*.[94]

In the biosynthesis of the fumiquinazolines and tryptoquialanine, an alanyl or aminoisobutyryl unit is condensed with oxidized fumiquinazoline F by the monomodular NRPSs Af12050 and TqaB, respectively. These final imidazoindolone scaffolds differ in the stereochemistry of one C-N bond. The stereochemical outcome is determined by the C-terminal C domain of the NRPS. Swapping of these highly homologous C domains has allowed the stereochemistry of these peptidyl alkaloids to be inverted *in vitro*.[208]

4.2 Combinatorial biosynthesis

To overcome limitations associated with the presumed gatekeeping role of C domains, the impact of C domains on A domain selectivity or the importance of protein-protein interactions at the interface

between C and A domains, several different multidomain substitutions have been performed, including C-A pairs, PCP-C-A units or whole modules. However, even when the cognate C-A interface was preserved, success was not guaranteed.[201,204] Recently, the Bode group published a novel swapping approach, using A-PCP-C or A-PCP-C/E exchange units (XUs) instead of traditional C-A-PCP modules for combinatorial biosynthesis. The borders of each XU were set at a fixed point within the flexible C-A linker region. Although this method enabled the production of novel peptides, the obtained titres were found to drop with the number of coupled XUs. Moreover, acceptor site specificity had to be respected in the selection of sequential XUs, creating drawbacks for combinatorial biosynthesis.[209] To resolve these issues, the group developed a second strategy to combine units from different assembly lines, placing the borders of the exchange units within the linker region that connects both C subdomains. This divides the C domain into an N- and C-terminal region, harbouring the donor (C_{Don}) and acceptor site (C_{Acc}), respectively. Combining various C_{Acc}-A-PCP-C_{Don} units (XUC units) allowed the construction of completely novel NRP assembly lines without limitations associated with C domain specificity, and without disrupting C-A protein-protein interfaces.[210] This strategy can therefore be employed for the incorporation of non-proteinogenic amino acids or for the targeted production of a specific metabolite of interest. Bozhüyük *et al*. also showed that elongation units can replace starter units, as long as the associated C_{Acc} or upstream C domain remains fused to the XUC. One potential drawback may be the intolerance of TE domains towards the cyclization or release of non-native peptides.[210] To circumvent this, elongating C domains may be used as chain release catalysts. Such C domains have been successfully utilized for the formation of both linear and cyclised peptides, albeit with reduced efficiency compared to the original TE domains.[209] Although the XUC approach has been shown to work best for XUCs of closely related bacterial strains, it offers great potential for future NRPS redesign and the formation of randomized peptide libraries.[210]

Fungal NRPS XUs composed of (C-)A-PCP(-C) units and auxiliary domains have been designed by exploiting specific recombination points in adjacent linker regions.[211] In order to combine XUs with an N-or C-terminal C domain, the acceptor site specificity of the upstream C domain or the donor site substrate tolerance of the downstream C domain, respectively, needs to be taken into account. By following these rules, Steiniger *et al.*, were able to produce CDPs from hybrid CDP synthetases containing XUs from the linear cyclosporine synthetase.[94,211] The use of combined (C-)A-PCP(-C) units may further expand our abilities to form large NRP libraries via combinatorial biosynthesis, but this concept still needs to be validated in practice.[211] It is interesting to note that recent work of Calcott *et. al.* has countered the assumption that C domains impact substrate selectivity. According to their study, combinatorial biosynthesis may be effected by swapping A domains together with the associated C-A linker region, thereby placing the recombination border at the C-terminal end of C domains (**Fig. 19B**).[27]

In other efforts to form hybrid fungal cyclic depsipeptides, different iterative synthetase parts from the enniatin, beauvericin and bassianolide pathways have been fused.[212] In this process, the terminal C3 domains were converted into canonical chain-elongating C domains that mediate ester bond formation. Although mainly wild-type metabolites were produced by these hybrid pathways in *E. coli*, novel non-natural compounds were also detected. In addition to the expected linearly processed peptides, products resulting from a combined linear and iterative operation were observed. These results indicate that each CDP synthetase subunit can serve as an initiation point for biosynthesis. At each internal C3 domain, biosynthesis can either continue linearly, or the intermediate can be transferred upstream for iterative processing. Presumably, the system offers sufficient flexibility to enable transfer of the peptide to any of the upstream modules. Whenever the peptide reaches a C3 domain at the length of the native CDP, either

internal or C-terminal, the product can be cyclized and released. Following this biosynthetic scheme, minor amounts of a wide variety of hybrid compounds could be produced.[212]

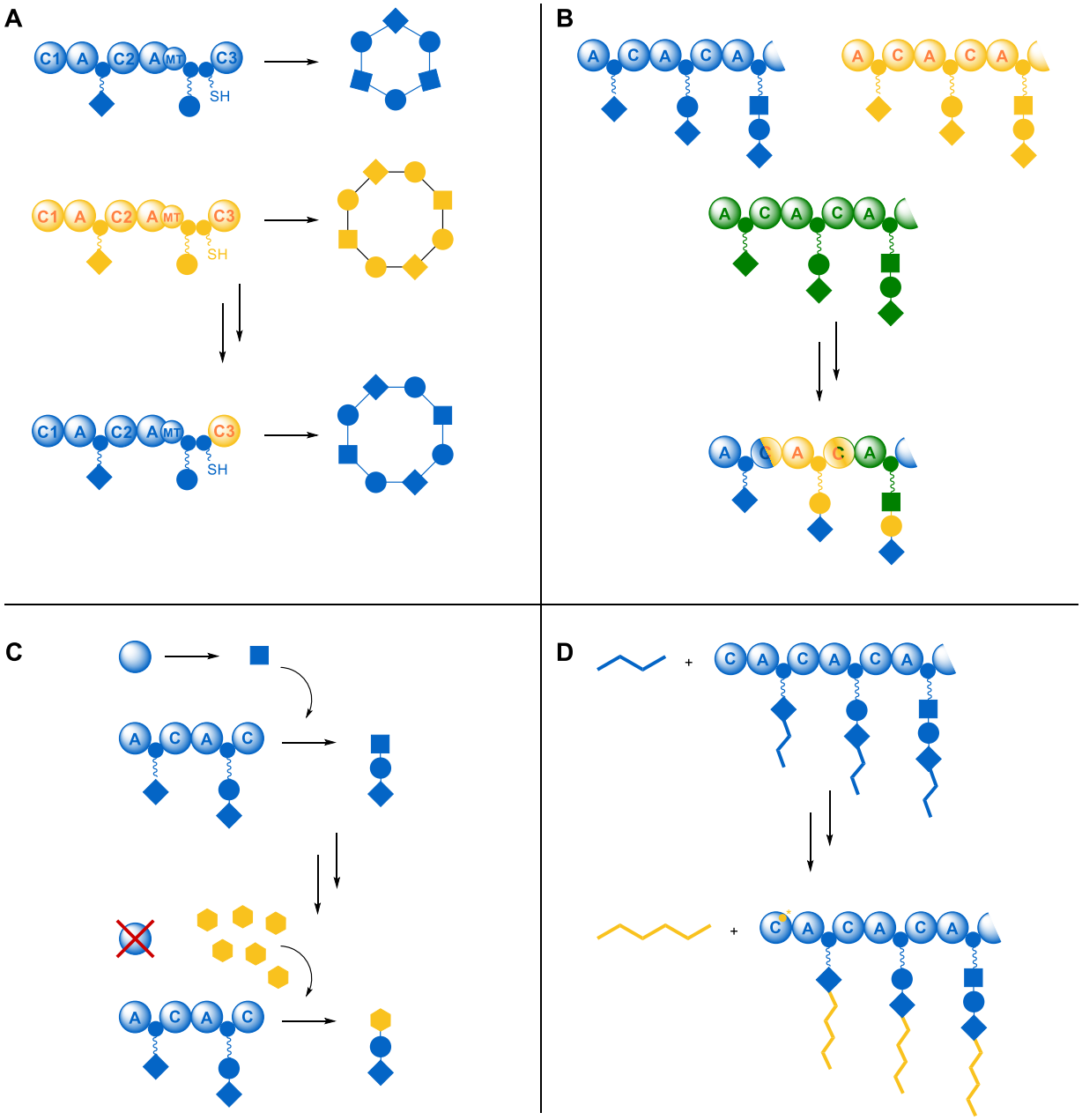

Figure 19: Possible approaches for C domain based pathway engineering. (A) C domain swapping. Schematic representation of chain length alteration in the biosynthesis of CDPs by swapping the C3 domain. (B) Combinatorial biosynthesis. Schematic representation of the formation of novel nonribosomal peptides by the combination of modules from various biosynthetic assembly lines according to the XUC principle. (C) Mutasynthesis. Schematic representation of the formation of novel biosynthetic compounds exploiting the substrate flexibility of chain releasing C domains. (D) Site specific mutagenesis. Schematic representation of the alteration of fatty acid chain attachment by the mutagenesis of specific residues in the starter C domain.

4.3 Mutasynthesis

A powerful approach to exploit the relaxed substrate specificity of natural product biosynthetic enzymes is mutasynthesis. By inactivating the assembly of a natural precursor *in vivo* and feeding a substrate analogue to such an engineered bacterial or fungal strain, natural product biosynthesis can be redirected to the production of novel derivatives. Mutasynthesis has already been successfully applied for the incorporation of several alternative amino acids, *e.g.* in CDA biosynthesis.[39] However, this A domain-targeted approach often only allows for the incorporation of a few closely-related building blocks due to the substrate specificity of associated domains.

C domains that catalyse chain release via intermolecular condensation with a soluble acceptor substrate, on the other hand, represent a more promising starting point for mutasynthesis, given that no A domain selectivity has to be overcome (**Fig. 19C**). Although this approach has not often been applied yet in practice, many different C domains have shown potential for this due to their relaxed substrate specificity. One example is the stand-alone vibriobactin C domain VibH, which is able to catalyse condensation of VibB-bound 2,3-dihydroxybenzoate to various amine substrates *in vitro*.[165] Similarly, several distinct amines are incorporated into the C-terminal moiety of bleomycin.[118] Cross-feeding studies on the fabclavine pathway have revealed the ability of the FclL C domain to utilise longer polyamine substrates, leading to novel fabclavine variants.[181] Meanwhile, heterologous expression of the rhabdopeptide/xenortide biosynthetic gene cluster from *Xenorhabdus* KJ12.1 in *E. coli* and feeding of various amine substrates has resulted in the production of 86 distinct peptides, incorporating 16 different amines.[170] The ester-bond forming SgcC5 has also been shown to have a broad substrate tolerance *in vitro*, both at the acceptor and donor site. Although SgcC5 shows strict stereospecificity by favouring the S-enantiomer of the tyrosine substrate, various (S)-tyrosine analogues are accepted as a substrate. These are regio- and stereospecifically condensed to the C-2 of the (R)-enediyne core. SgcC5 is even capable of catalysing both ester and amide bond formation.[192,213] The same holds true for the Bamb_5915 C domain of the enacyloxin biosynthetic pathway. This enzyme has also been shown to have a broad substrate tolerance, both at the acceptor and donor site. *In vitro* biochemical assays have revealed that various cyclic and linear analogues of DHCCA can be used as substrates in the intermolecular condensation reaction, as long as a nucleophilic amine/hydroxyl group, and a carboxylic acid group are appropriately juxtaposed. At the donor site, a relaxed specificity for different polyketide chain lengths and branches was observed. These observations open up promising avenues for producing novel enacyloxin analogues with modifications in both the polyketide backbone and the DHCCA moiety by using a mutasynthesis approach.[197]

Starter C domains in lipopeptide biosynthetic pathways also represent an interesting starting point for mutasynthesis due to their relaxed substrate tolerance. Lipopeptides are often produced as a mixture of closely-related compounds with slight variations in the length and structure of the conjugated fatty acids.[214] Blocking the production or activation of the native acyl chain and feeding non-cognate fatty acids may therefore enable the biosynthesis of novel lipopeptide derivatives. The starter C domain of the glidobactin biosynthetic pathway, for example, has been shown to accept a wide range of fatty acyl chains.[35] Furthermore, active site modification of the ketosynthase FabF3 from the CDA biosynthetic pathway has resulted in the production of shorter fatty acyl chains, which were subsequently accepted as a substrate by the starter C domain of the NRPS and incorporated into the final lipopeptide products.[215]

In practice, it may be difficult to inactivate the biosynthesis of a natural acceptor substrate or fatty acyl chain. An alternative approach is precursor directed biosynthesis. In this case, analogous substrates are fed to an organism without abolishing the biosynthesis of the native substrate. This, however, does not prevent the efficient production of the original natural products. To avoid this issue, a second alternative approach could be the expression of the key biosynthetic genes in a heterologous host, combined with feeding of various alternative precursors.

4.4 Site-specific mutagenesis

Engineering biosynthetic assembly lines can also be accomplished by site-specific mutagenesis of C domains. Starter C domains in lipopeptide biosynthetic pathways have shown to be a good starting point for this approach (**Fig. 19D**). Since the fatty acid chains have a huge impact on the biological activity of lipopeptides, modifying these moieties may generate compounds with altered functionalities. While the enzymes responsible for activation of fatty acids can display a certain degree of substrate selectivity[38,216], additional selectivity is conferred by the starter C domain. Recently, *in silico* analysis of the Cs domain from the lipopeptide A54145 NRPS has identified candidate amino acid residues which likely interact with the acyl chain in the substrate tunnel. To assess the impact of these residues on substrate selection, mutated versions of the starter C domain were generated and evaluated using *in vitro* biochemical assays. While some mutants completely lost all catalytic activity, others showed a shift in substrate specificity for either shorter or longer chains.[216] Furthermore, the impact of residues involved in protein-protein interactions with the associated carrier proteins on substrate specificity was analysed. Molecular dynamics simulations indicated that the mutations have an effect on the substrate binding pocket, the preferred chain length and the structural organisation of the enzyme.[217]

In the biosynthesis of the N-glycoside antibiotic streptothricin, the ORF18 C domain catalyses intermolecular amide bond formation between PCP-bound L-β-lysine oligopeptide thioesters and the aminoglycoside streptothrisamine. [185] The catalytic efficiency of ORF18 has been shown to impact the final chain length of the L-β-lysine oligopeptide. The ORF18 C domain has a modified HQXXXDM signature motif instead of the canonical HHxxxDG motif and mutation of the active site Gln residue to Ala was shown to decrease the catalytic rate of the enzyme. This shifted the production from short to longer β-lysine oligopeptides *in vitro*, some of which were not produced using wild-type enzymes.[185]

5. Conclusion and perspectives

Nonribosomal peptides are a remarkable family of structurally diverse and complex natural products with a wide range of biological activities and important applications in medicine, agriculture and biotechnology. C domains play a key role in the biosynthesis of nonribosomal peptides by catalysing amide bond formation between the PCP-bound substrates of adjacent modules. However, C domains play a much larger role in generating structural diversity and complexity than initially anticipated. Over the years, it has become increasingly clear that the functions and catalytic abilities of C domains greatly exceed conventional peptide bond formation. C domains have been shown to exhibit remarkably diverse catalytic activities, ranging from β-lactam and diketopiperazine formation to hydrolysis, chain length control, cycloaddition, and Pictet-Spengler cyclization. The continuous progress in sequencing technology, genome mining and advanced bioinformatics will undoubtedly further expand our knowledge of these

atypical C domain functions. As indicated in this review, noncanonical behaviour is often reflected by changes in the conserved signature motif of C domains. Further identification and characterization of atypical C domains will therefore continue to improve our ability to predict the structures of novel natural products.

In order to exploit the full catalytic potential of C domains, it will be of great importance to characterise novel variants by biochemically validating their function and elucidating their reaction mechanism. Indeed, biosynthetic pathway engineering holds great promise for generating novel natural product derivatives with improved pharmaceutical properties. However, current engineering efforts are still hampered by limited substrate flexibility of enzymes and tight interactions with accompanying domains. C domain-based engineering approaches may overcome some of these issues. In light of their central catalytic role and impressive functional versatility, C domains represent a promising yet underexplored starting point for engineering. Our increasing knowledge on the structure and function of atypical variants should allow their catalytic capabilities to be more fully exploited in the near future. The ongoing development and improvement of molecular biology tools for the genetic manipulation and heterologous expression of biosynthetic pathways will also greatly facilitate these efforts.

References

1 J. O'Brien and G. D. Wright, *Curr. Opin. Biotechnol.*, 2011, **22**, 552–558.
2 O. Tyc, C. Song, J. S. Dickschat, M. Vos and P. Garbeva, *Trends Microbiol.*, 2017, **25**, 280–292.
3 E. A. Felnagle, E. E. Jackson, Y. A. Chan, A. M. Podevels, A. D. Berti, M. D. McMahon and M. G. Thomas, *Mol. Pharm.*, 2008, **5**, 191–211.
4 J. V. Pham, M. A. Yilma, A. Feliz, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park and Y. J. Yoon, *Front. Microbiol.*, 2019, **10**, 1404.
5 M. J. Jaremko, T. D. Davis, J. C. Corpuz and M. D. Burkart, *Nat. Prod. Rep.*, 2020, **37**, 355–379.
6 G. H. Hur, C. R. Vickery and M. D. Burkart, *Nat. Prod. Rep.*, 2012, **29**, 1074–1098.
7 K. Bloudoff and T. M. Schmeing, *Biochim. Biophys. acta. Proteins proteomics*, 2017, **1865**, 1587–1604.
8 K. J. Weissman, *Nat. Prod. Rep.*, 2015, **32**, 436–53.
9 R. D. Süssmuth and A. Mainz, *Angew. Chemie - Int. Ed.*, 2017, **56**, 3770–3821.
10 B. R. Miller and A. M. Gulick, *Methods Mol. Biol.*, 2016, **1401**, 3–29.
11 M. McErlean, J. Overbay and S. Van Lanen, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 493–513.
12 J. A. E. Payne, M. Schoppet, M. H. Hansen and M. J. Cryle, *Mol. BioSyst.*, 2017, **13**, 9–22.
13 J. M. Reimer, A. S. Haque, M. J. Tarry and T. M. Schmeing, *Curr. Opin. Struct. Biol.*, 2018, **49**, 104–113.
14 T. Stachelhaus, H. D. Mootz, V. Bergendahl and M. A. Marahiel, *J. Biol. Chem.*, 1998, **273**, 22773–22781.
15 T. A. Keating, C. G. Marshall, C. T. Walsh and A. E. Keating, *Nat. Struct. Biol.*, 2002, **9**, 522–526.
16 S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel and L.-O. Essen, *Structure*, 2007, **15**, 781–792.
17 K. Bloudoff, D. Rodionov and T. M. Schmeing, *J. Mol. Biol.*, 2013, **425**, 3137–3150.
18 A. Tanovic, S. A. Samel, L.-O. Essen and M. A. Marahiel, *Science (80-.).*, 2008, **321**, 659–663.
19 E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skiniotis, C. C. Aldrich and A. M. Gulick, *Nature*, 2016, **529**, 235–238.
20 K. Bloudoff, D. A. Alonzo and T. M. Schmeing, *Cell Chem. Biol.*, 2016, **23**, 331–339.

21 V. De Crécy-Lagard, P. Marlière and W. Saurin, *C. R. Acad. Sci. III.*, 1995, **318**, 927–36.
22 J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt and T. Martin Schmeing, *Science (80-.).*, , DOI:10.1126/science.aaw4388.
23 P. J. Belshaw, C. T. Walsh and T. Stachelhaus, *Science (80-.).*, 1999, **284**, 486–489.
24 D. E. Ehmann, J. W. Trauger, T. Stachelhaus and C. T. Walsh, *Chem. Biol.*, 2000, **7**, 765–772.
25 U. Linne and M. A. Marahiel, *Biochemistry*, 2000, **39**, 10439–10447.
26 S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 426–471.
27 M. J. Calcott, J. G. Owen and D. F. Ackerley, *Nat. Commun.*, 2020, **11**, 4554.
28 S. L. Clugston, S. A. Sieber, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2003, **42**, 12095–12104.
29 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.
30 V. Bergendahl, U. Linne and M. A. Marahiel, *Eur. J. Biochem.*, 2002, **269**, 620–629.
31 E. D. Roche and C. T. Walsh, *Biochemistry*, 2003, **42**, 1334–1344.
32 C. G. Marshall, N. J. Hillson and C. T. Walsh, *Biochemistry*, 2002, **41**, 244–250.
33 L. Robbel and M. A. Marahiel, *J. Biol. Chem.*, 2010, **285**, 27501–27508.
34 F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker and M. A. Marahiel, *Chem. Biol.*, 2010, **17**, 872–880.
35 H. J. Imker, D. Krahn, J. Clerc, M. Kaiser and C. T. Walsh, *Chem. Biol.*, 2010, **17**, 1077–1083.
36 F. I. Kraas, T. W. Giessen and M. A. Marahiel, *FEBS Lett.*, 2012, **586**, 283–288.
37 R. H. Baltz, V. Miao and S. K. Wrigley, *Nat. Prod. Rep.*, 2005, **22**, 717–741.
38 M. Wittmann, U. Linne, V. Pohlmann and M. A. Marahiel, *FEBS J.*, 2008, **275**, 5343–5354.
39 Z. Hojati, C. Milne, B. Harvey, L. Gordon, M. Borg, F. Flett, B. Wilkinson, P. J. Sidebottom, B. A. M. Rudd, M. A. Hayes, C. P. Smith and J. Micklefield, *Chem. Biol.*, 2002, **9**, 1175–1187.
40 A. Powell, M. Borg, B. Amir-Heidari, J. M. Neary, J. Thirlway, B. Wilkinson, C. P. Smith and J. Micklefield, *J. Am. Chem. Soc.*, 2007, **129**, 15182–15191.
41 S. Paul, H. Ishida, L. T. Nguyen, Z. Liu and H. J. Vogel, *Protein Sci.*, 2017, **26**, 946–959.
42 H. Chen, A. S. Olson, W. Su, P. H. Dussault and L. Du, *RSC Adv.*, 2015, **5**, 105753–105759.
43 N. C. Harris, M. Sato, N. A. Herman, F. Twigg, W. Cai, J. Liu, X. Zhu, J. Downey, R. Khalaf, J. Martin, H. Koshino and W. Zhang, *Proc. Natl. Acad. Sci.*, 2017, **114**, 7025–7030.
44 K. Takeda, K. Kemmoku, Y. Satoh, Y. Ogasawara, K. Shin-Ya and T. Dairi, *ACS Chem. Biol.*, 2017, **12**, 1813–1819.
45 Z. L. Reitz, M. Sandy and A. Butler, *Metallomics*, 2017, **9**, 824–839.
46 N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen and P. R. Jensen, *PLoS One*, 2012, **7**, e34064.
47 M. Liu, Y. Jia, Y. Xie, C. Zhang, J. Ma, C. Sun and J. Ju, *Mar. Drugs*, 2019, **17**, 240.
48 J. J. May, T. M. Wendrich and M. A. Marahiel, *J. Biol. Chem.*, 2001, **276**, 7209–7217.
49 H. K. Zane, H. Naka, F. Rosconi, M. Sandy, M. G. Haygood and A. Butler, *J. Am. Chem. Soc.*, 2014, **136**, 5615–5618.
50 C. Boros, C. J. Smith, Y. Vasina, Y. Che, A. B. Dix, B. Darveaux and C. Pearce, *J. Antibiot. (Tokyo).*, 2006, **59**, 486–494.
51 B. Dose, S. P. Niehs, K. Scherlach, L. V. Flórez, M. Kaltenpoth and C. Hertweck, *ACS Chem. Biol.*, 2018, **13**, 2414–2420.
52 M. Jenner, X. Jian, Y. Dashti, J. Masschelein, C. Hobson, D. M. Roberts, C. Jones, S. Harris, J. Parkhill, H. A. Raja, N. H. Oberlies, C. J. Pearce, E. Mahenthiralingam and G. L. Challis, *Chem. Sci.*, 2019, **10**, 5489–5494.
53 S. Caboche, V. Leclère, M. Pupin, G. Kucherov and P. Jacques, *J. Bacteriol.*, 2010, **192**, 5143–5150.
54 T. Caradec, M. Pupin, A. Vanvlassenbroeck, M.-D. Devignes, M. Smaïl-Tabbone, P. Jacques and V. Leclère, *PLoS One*, 2014, **9**, e85667.
55 L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2002, **41**,

9184–9196.
56 T. Stachelhaus and C. T. Walsh, *Biochemistry*, 2000, **39**, 5775–5787.
57 K. Hoffmann, E. Schneider-Scherzer, H. Kleinkauf and R. Zocher, *J. Biol. Chem.*, 1994, **269**, 12710–12714.
58 Y. Cheng and J. D. Walton, *J. Biol. Chem.*, 2000, **275**, 4906–4911.
59 S. A. Samel, P. Czodrowski and L.-O. Essen, *Acta Crystallogr.*, 2014, **D70**, 1442–1452.
60 W.-H. Chen, K. Li, N. S. Guntaka and S. D. Bruner, *ACS Chem. Biol.*, 2016, **11**, 2293–2303.
61 U. Linne, S. Doekel and M. A. Marahiel, *Biochemistry*, 2001, **40**, 15824–15834.
62 W. E. Kim, A. Patel, G. H. Hur, P. Tufar, M. G. Wuo, J. A. McCammon and M. D. Burkart, *ChemBioChem*, 2019, **20**, 147–152.
63 C. J. Balibar, F. H. Vaillancourt and C. T. Walsh, *Chem. Biol.*, 2005, **12**, 1189–1200.
64 M. Royer, R. Koebnik, M. Marguerettaz, V. Barbe, G. P. Robin, C. Brin, S. Carrere, C. Gomez, M. Hügelland, G. H. Völler, J. Noëll, I. Pieretti, S. Rausch, V. Verdier, S. Poussier, P. Rott, R. D. Süssmuth and S. Cociancich, *BMC Genomics*, 2013, **14**, 658.
65 J. B. Biggins, H.-S. Kang, M. A. Ternei, D. Deshazer and S. F. Brady, *J. Am. Chem. Soc.*, 2014, **136**, 9484–9490.
66 Q. Esmaeel, M. Pupin, N. P. Kieu, G. Chataigné, M. Béchet, J. Deravel, F. Krier, M. Höfte, P. Jacques and V. Leclère, *Microbiologyopen*, 2016, **5**, 512–526.
67 N. Roongsawang, K. Hase, M. Haruki, T. Imanaka, M. Morikawa and S. Kanaya, *Chem. Biol.*, 2003, **10**, 869–80.
68 A. M. Gehring, I. Mori, R. D. Perry and C. T. Walsh, *Biochemistry*, 1998, **37**, 11637–11650.
69 C. G. Marshall, M. D. Burkart, T. A. Keating and C. T. Walsh, *Biochemistry*, 2001, **40**, 10655–10663.
70 T. Duerfahrt, K. Eppelmann, R. Müller and M. A. Marahiel, *Chem. Biol.*, 2004, **11**, 261–271.
71 D. A. Miller and C. T. Walsh, *Biochemistry*, 2001, **40**, 5313–5321.
72 K. Bloudoff, C. D. Fage, M. A. Marahiel and T. M. Schmeing, *Proc. Natl. Acad. Sci.*, 2017, **114**, 95–100.
73 D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh and C. L. Drennan, *Proc. Natl. Acad. Sci.*, 2016, **113**, 12432–12437.
74 H. M. Patel and C. T. Walsh, *Biochemistry*, 2001, **40**, 9023–9031.
75 T. L. Schneider, B. Shen and C. T. Walsh, *Biochemistry*, 2003, **42**, 9722–9730.
76 W. L. Kelly, N. J. Hillson and C. T. Walsh, *Biochemistry*, 2005, **44**, 13385–13393.
77 R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw and C. T. Walsh, *Nat. Prod. Rep.*, 1999, **16**, 249–263.
78 L. Du, M. Chen, Y. Zhang and B. Shen, *Biochemistry*, 2003, **42**, 9731–9740.
79 D. A. Miller, L. Luo, N. Hillson, T. A. Keating and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 333–344.
80 M. Di Lorenzo, M. Stork, H. Naka, M. E. Tolmasky and J. H. Crosa, *BioMetals*, 2008, **21**, 635–648.
81 T. A. Keating, D. A. Miller and C. T. Walsh, *Biochemistry*, 2000, **39**, 4729–4739.
82 D. A. Alonzo, C. Chiche-Lapierre, M. J. Tarry, J. Wang and T. M. Schmeing, *Nat. Chem. Biol.*, 2020, **16**, 493–496.
83 N. A. Magarvey, Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore and D. H. Sherman, *ACS Chem. Biol.*, 2006, **1**, 766–779.
84 Y. Ding, C. M. Rath, K. L. Bolduc, K. Hakansson and D. H. Sherman, *J. Am. Chem. Soc.*, 2011, **133**, 14492–14495.
85 A. V. Ramaswamy, C. M. Sorrels and W. H. Gerwick, *J. Nat. Prod.*, 2007, **70**, 1977–1986.
86 M. Sandy, Z. Rui, J. Gallagher and W. Zhang, *ACS Chem. Biol.*, 2012, **7**, 1956–1961.
87 M. Ehling-Schulz, M. Fricker, H. Grallert, P. Rieck, M. Wagner and S. Scherer, *BMC Microbiol.*, 2006, **6**, 20.
88 G. W. Heberlig and C. N. Boddy, *J. Nat. Prod.*, 2020, **83**, 1990–1997.

89 N. A. Magarvey, M. Ehling-schulz and C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 10698–10699.
90 Y. Q. Cheng, *ChemBioChem*, 2006, **7**, 471–477.
91 N. Huguenin-Dezot, D. A. Alonzo, G. W. Heberlig, M. Mahesh, D. P. Nguyen, M. H. Dornan, C. N. Boddy, T. M. Schmeing and J. W. Chin, *Nature*, 2019, **565**, 112–117.
92 D. G. Fujimori, S. Hrvatin, C. S. Neumann, M. Strieker, M. A. Marahiel and C. T. Walsh, *Proc. Natl. Acad. Sci.*, 2007, **104**, 16498–16503.
93 D. Yu, F. Xu, S. Zhang and J. Zhan, *Nat. Commun.*, 2017, **8**, 15349.
94 C. Steiniger, S. Hoffmann, A. Mainz, M. Kaiser, K. Voigt, V. Meyer and R. D. Süssmuth, *Chem. Sci.*, 2017, **8**, 7834–7843.
95 M. T. Robey, R. Ye, J. W. Bok, K. D. Clevenger, M. N. Islam, C. Chen, R. Gupta, M. Swyers, E. Wu, P. Gao, P. M. Thomas, C. C. Wu, N. P. Keller and N. L. Kelleher, *ACS Chem. Biol.*, 2018, **13**, 1142–1147.
96 L. Wang, M. Yuan and J. Zheng, *Synth. Syst. Biotechnol.*, 2019, **4**, 10–15.
97 S. Meyer, A. Mainz, J.-C. Kehr, R. D. Süssmuth and E. Dittmann, *ChemBioChem*, 2017, **18**, 2376–2379.
98 O. Lazos, M. Tosin, A. L. Slusarczyk, S. Boakes, J. Cortés, P. J. Sidebottom and P. F. Leadlay, *Chem. Biol.*, 2010, **17**, 160–173.
99 L. Robbel, T. A. Knappe, U. Linne, X. Xie and M. A. Marahiel, *FEBS J.*, 2010, **277**, 663–676.
100 K. Yamanaka, C. Maruyama, H. Takagi and Y. Hamano, *Nat. Chem. Biol.*, 2008, **4**, 766–772.
101 E. Purev, T. Kondo, D. Takemoto, J. T. Niones and M. Ojika, *Molecules*, 2020, **25**, 1032.
102 N. Kito, C. Maruyama, K. Yamanaka, Y. Imokawa, T. Utagawa and Y. Hamano, *J. Biosci. Bioeng.*, 2013, **115**, 523–526.
103 Z. Xu, Z. Sun, S. Li, Z. Xu, C. Cao, Z. Xu, X. Feng and H. Xu, *Sci. Rep.*, 2015, **5**, 17400.
104 K. Yamanaka, H. Fukumoto, M. Takehara, Y. Hamano and T. Oikawa, *ACS Chem. Biol.*, 2020, **15**, 1964–1973.
105 K. Tahlan and S. E. Jensen, *J. Antibiot. (Tokyo).*, 2013, **66**, 401–410.
106 R. B. Hamed, J. R. Gomez-Castellanos, L. Henry, C. Ducho, M. A. McDonough and C. J. Schofield, *Nat. Prod. Rep.*, 2013, **30**, 21–107.
107 J. M. Davidsen, D. M. Bartley and C. A. Townsend, *J. Am. Chem. Soc.*, 2013, **135**, 1749–1759.
108 J. M. Davidsen and C. A. Townsend, *Chem. Biol.*, 2012, **19**, 297–306.
109 M. Gunsior, S. D. Breazeale, A. J. Lind, J. Ravel, J. W. Janc and C. A. Townsend, *Chem. Biol.*, 2004, **11**, 927–938.
110 N. M. Gaudelli, D. H. Long and C. A. Townsend, *Nature*, 2015, **520**, 383–387.
111 D. H. Long and C. A. Townsend, *Biochemistry*, 2018, **57**, 3353–3358.
112 K. Koketsu, K. Watanabe, H. Suda, H. Oguri and H. Oikawa, *Nat. Chem. Biol.*, 2010, **6**, 408–410.
113 Y. Mikami, K. Takahashi, K. Yazawa, T. Arai, M. Namikoshi, S. Iwasaki and S. Okuda, *J. Biol. Chem.*, 1985, **260**, 344–348.
114 L. Li, W. Deng, J. Song, W. Ding, Q.-F. Zhao, C. Peng, W.-W. Song, G.-L. Tang and W. Liu, *J. Bacteriol.*, 2008, **190**, 251–263.
115 S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law and H. Deng, *Angew. Chemie Int. Ed.*, 2021, **60**, 3229–3237.
116 D. Tillett, E. Dittmann, M. Erhard, H. Von Döhren, T. Börner and B. A. Neilan, *Chem. Biol.*, 2000, **7**, 753–764.
117 G. Christiansen, J. Fastner, M. Erhard, T. Börner and E. Dittmann, *Microbiology*, 2003, **185**, 564–572.
118 L. Du, C. Sánchez, M. Chen, D. J. Edwards and B. Shen, *Chem. Biol.*, 2000, **7**, 623–642.
119 M. C. Moffitt and B. A. Neilan, *Appl. Environ. Microbiol.*, 2004, **70**, 6353–6362.
120 D. B. Berkowitz, B. D. Charette, K. R. Karukurichi and J. M. McFadden, *Tetrahedron: Asymmetry*,

2006, **17**, 869–882.
121 J. P. Scannell, D. L. Pruess, T. C. Demny, L. H. Sello, T. Williams and A. Stempel, *J. Antibiot. (Tokyo).*, 1972, **25**, 122–127.
122 U. Sahm, G. Knobloch and F. Wagner, *J. Antibiot. (Tokyo).*, 1973, **26**, 389–390.
123 J. B. Patteson, Z. D. Dunn and B. Li, *Angew. Chemie - Int. Ed.*, 2018, **57**, 6780–6785.
124 D. Boettger, H. Bergmann, B. Kuehn, E. Shelest and C. Hertweck, *ChemBioChem*, 2012, **13**, 2363–2373.
125 S. M. Ma, J. W. Li, J. W. Choi, H. Zhou, K. K. M. Lee, V. A. Moorthie, X. Xie, J. T. Kealey, N. A. Da Silva, J. C. Vederas and Y. Tang, *Science (80-.).*, 2009, **326**, 589–592.
126 J. Zhang, N. Liu, R. A. Cacho, Z. Gong, Z. Liu, W. Qin, C. Tang, Y. Tang and J. Zhou, *Nat. Chem. Biol.*, 2016, **12**, 1001–1003.
127 K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch and M. J. Cryle, *Nature*, 2015, **521**, 105–109.
128 K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van den Heever, C. R. Hutchinson and J. C. Vederas, *J. Am. Chem. Soc.*, 2000, 122, 11519–11520.
129 T. B. Kakule, Z. Lin and E. W. Schmidt, *J. Am. Chem. Soc.*, 2014, **136**, 17882–17890.
130 G. Yim, M. N. Thaker, K. Koteva and G. Wright, *J. Antibiot. (Tokyo).*, 2014, **67**, 31–41.
131 D. Bischoff, S. Pelzer, B. Bister, G. J. Nicholson, S. Stockert, M. Schirle, W. Wohlleben, G. Jung and R. D. Süssmuth, *Angew. Chemie - Int. Ed.*, 2001, **40**, 4688–4691.
132 D. Bischoff, S. Pelzer, A. Höltzel, G. J. Nicholson, S. Stockert, M. Schirle, W. Wohlleben, G. Jung and R. D. Süssmuth, *Angew. Chemie - Int. Ed.*, 2001, **40**, 1693–1696.
133 R. D. Süssmuth, S. Pelzer, G. Nicholson, T. Walk, W. Wohlleben and G. Jung, *Angew. Chemie - Int. Ed.*, 1999, **38**, 1976–1979.
134 B. Hadatsch, D. Butz, T. Schmiederer, J. Steudle, W. Wohlleben, R. Süssmuth and E. Stegmann, *Chem. Biol.*, 2007, **14**, 1078–1089.
135 E. Stegmann, S. Pelzer, D. Bischoff, O. Puk, S. Stockert, D. Butz, K. Zerbe, J. Robinson, R. D. Süssmuth and W. Wohlleben, *J. Biotechnol.*, 2006, **124**, 640–653.
136 M. Schoppet, M. Peschke, A. Kirchberg, V. Wiebach, R. D. Süssmuth, E. Stegmann and M. J. Cryle, *Chem. Sci.*, 2019, **10**, 118–133.
137 A. Greule, T. Izoré, D. Iftime, J. Tailhades, M. Schoppet, Y. Zhao, M. Peschke, I. Ahmed, A. Kulik, M. Adamek, R. J. A. Goode, R. B. Schittenhelm, J. A. Kaczmarski, C. J. Jackson, N. Ziemert, E. H. Krenske, J. J. De Voss, E. Stegmann and M. J. Cryle, *Nat. Commun.*, 2019, **10**, 2613.
138 M. Peschke, C. Brieke and M. J. Cryle, *Sci. Rep.*, 2016, **6**, 35584.
139 M. Peschke, K. Haslinger, C. Brieke, J. Reinstein and M. J. Cryle, *J. Am. Chem. Soc.*, 2016, **138**, 6746–6753.
140 Y. Zhao, Y. T. C. Ho, J. Tailhades and M. Cryle, *ChemBioChem*, 2021, **22**, 43–51.
141 Z. L. Reitz, C. D. Hardy, J. Suk, J. Bouvet and A. Butler, *Proc. Natl. Acad. Sci.*, 2019, **116**, 19805–19814.
142 M. F. Kreutzer, H. Kage and M. Nett, *J. Am. Chem. Soc.*, 2012, **134**, 5415–5422.
143 X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh and Y. Tang, *Nat. Chem. Biol.*, 2012, **8**, 823–30.
144 X. Yang, P. Feng, Y. Yin, K. Bushley, J. W. Spatafora and C. Wang, *MBio*, 2018, **9**, e01211-18.
145 R. A. Cacho, W. Jiang, Y.-H. Chooi, C. T. Walsh and Y. Tang, *J. Am. Chem. Soc.*, 2012, **134**, 16781–16790.
146 J. M. Jin, S. Lee, J. Lee, S.-R. Baek, J.-C. Kim, S.-H. Yun, S.-Y. Park, S. Kang and Y.-W. Lee, *Mol. Microbiol.*, 2010, **76**, 456–466.
147 B. Winterberg, S. Uhlmann, U. Linne, F. Lessing, M. A. Marahiel, H. Eichhorn, R. Kahmann and J. Schirawski, *Mol. Microbiol.*, 2010, **75**, 1260–1271.
148 M. Eisendle, H. Oberegger, I. Zadra and H. Haas, *Mol. Microbiol.*, 2003, **49**, 359–375.

149 X. Gao, W. Jiang, G. Jiménez-Osés, M. S. Choi, K. N. Houk, Y. Tang and C. T. Walsh, *Chem. Biol.*, 2013, **20**, 870–878.
150 J. L. Slightom, B. P. Metzger, H. T. Luu and A. P. Elhammer, *Gene*, 2009, **431**, 67–79.
151 R. V. Grindberg, T. Ishoey, D. Brinza, E. Esquenazi, R. C. Coates, W. Liu, L. Gerwick, P. C. Dorrestein, P. Pevzner, R. Lasken and W. H. Gerwick, *PLoS One*, 2011, **6**, e18565.
152 J.-M. Zhang, H.-H. Wang, X. Liu, C.-H. Hu and Y. Zou, *J. Am. Chem. Soc.*, 2020, **142**, 1957–1965.
153 A. König, T. Schwecke, I. Molnár, G. A. Böhm, P. A. S. Lowden, J. Staunton and P. F. Leadlay, *Eur. J. Biochem.*, 1997, **247**, 526–534.
154 T. Schwecke, J. F. Aparicio, I. Molnár, A. König, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortés, J. B. Lester, G. A. Böhm, J. Staunton and P. F. Leadlay, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 7839–7843.
155 H. Motamedi and A. Shafiee, *Eur. J. Biochem.*, 1998, **256**, 528–534.
156 K. Wu, L. Chung, W. P. Revill, L. Katz and C. D. Reeves, *Gene*, 2000, **251**, 81–90.
157 M. He, B. Haltli, M. Summers, X. Feng and J. Hucul, *Gene*, 2006, **377**, 109–118.
158 G. J. Gatto, S. M. Mcloughlin, N. L. Kelleher and C. T. Walsh, *Biochemistry*, 2005, **44**, 5993–6002.
159 H. Li, C. L. M. Gilchrist, C.-S. Phan, H. J. Lacey, D. Vuong, S. A. Moggach, E. Lacey, A. M. Piggott and Y.-H. Chooi, *J. Am. Chem. Soc.*, 2020, **142**, 7145–7152.
160 P. Borgman, R. D. Lopez and A. L. Lane, *Org. Biomol. Chem.*, 2019, **17**, 2305–2314.
161 J. A. Baccile, H. H. Le, B. T. Pfannenstiel, J. W. Bok, C. Gomez, E. Brandenburger, D. Hoffmeister, N. P. Keller and F. C. Schroeder, *Angew. Chemie - Int. Ed.*, 2019, **58**, 14589–14593.
162 S. Maiya, A. Grundmann, S.-M. Li and G. Turner, *ChemBioChem*, 2006, **7**, 1062–1069.
163 H. Ali, M. I. Ries, J. G. Nijland, P. P. Lankhorst, T. Hankemeier, R. A. L. Bovenberg, R. J. Vreeken and A. J. M. Driessen, *PLoS One*, 2013, **8**, e65328.
164 B. J. C. Law, Y. Zhuo, M. Winn, D. Francis, Y. Zhang, M. Samborskyy, A. Murphy, L. Ren, P. F. Leadlay and J. Micklefield, *Nat. Catal.*, 2018, **1**, 977–984.
165 T. A. Keating, C. G. Marshall and C. T. Walsh, *Biochemistry*, 2000, **39**, 15513–15521.
166 T. A. Keating, C. G. Marshall and C. T. Walsh, *Biochemistry*, 2000, **39**, 15522–15530.
167 C. J. Balibar and C. T. Walsh, *ChemBioChem*, 2008, **9**, 42–45.
168 K. Agnoli, C. A. Lowe, K. L. Farmer, S. I. Husnain and M. S. Thomas, *J. Bacteriol.*, 2006, **188**, 3631–3644.
169 M. J. Vargas-Straube, B. Cámara, M. Tello, F. Montero-Silva, F. Cárdenas and M. Seeger, *PLoS One*, 2016, **11**, e0151273.
170 X. Cai, S. Nowak, F. Wesche, I. Bischoff, M. Kaiser, R. Fürst and H. B. Bode, *Nat. Chem.*, 2017, **9**, 379–386.
171 D. Reimer, K. N. Cowles, A. Proschak, F. I. Nollmann, A. J. Dowling, M. Kaiser, R. ffrench Constant, H. Goodrich-Blair and H. B. Bode, *ChemBioChem*, 2013, **14**, 1991–1997.
172 D. Reimer, F. I. Nollmann, K. Schultz, M. Kaiser and H. B. Bode, *J. Nat. Prod.*, 2014, **77**, 1976–1980.
173 A. Vingadassalon, F. Lorieux, M. Juguet, G. Le Goff, C. Gerbaud, J.-L. Pernodet and S. Lautru, *ACS Chem. Biol.*, 2015, **10**, 601–610.
174 M. Tao, L. Wang, E. Wendt-Pienkowski, N. P. George, U. Galm, G. Zhang, J. M. Coughlin and B. Shen, *Mol. Biosyst.*, 2007, **3**, 60–74.
175 U. Galm, E. Wendt-Pienkowski, L. Wang, N. P. George, T.-J. Oh, F. Yi, M. Tao, J. M. Coughlin and B. Shen, *Mol. Biosyst.*, 2009, **5**, 77–90.
176 Y. Hai and Y. Tang, *J. Am. Chem. Soc.*, 2018, **140**, 1271–1274.
177 J. Masschelein, C. Clauwers, U. R. Awodi, K. Stalmans, W. Vermaelen, E. Lescrinier, A. Aertsen, C. Michiels, G. L. Challis and R. Lavigne, *Chem. Sci.*, 2015, **6**, 923–929.
178 J. Masschelein, W. Mattheus, L.-J. Gao, P. Moons, R. van Houdt, B. Uytterhoeven, C. Lamberigts, E. Lescrinier, J. Rozenski, P. Herdewijn, A. Aertsen, C. Michiels and R. Lavigne, *PLoS One*, 2013, **8**,

e54143.
179 R. van Houdt, D. Van der Lelie, J. A. Izquierdo, A. Aertsen, J. Masschelein, R. Lavigne, C. W. Michiels and S. Taghavi, *Genome Announc.*, 2014, **2**, e00021-14.
180 Y. Cheng, X. Liu, S. An, C. Chang, Y. Zou, L. Huang, J. Zhong, Q. Liu, Z. Jiang, J. Zhou and L.-H. Zhang, *Mol. Plant-Microbe Interact.*, 2013, **26**, 1294–301.
181 S. L. Wenski, D. Kolbert, G. L. C. Grammbitter and H. B. Bode, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 565–572.
182 S. W. Fuchs, F. Grundmann, M. Kurz, M. Kaiser and H. B. Bode, *ChemBioChem*, 2014, **15**, 512–516.
183 C. T. Walsh and W. Zhang, *ACS Chem. Biol.*, 2011, **6**, 1000–1007.
184 W. Zhang, I. Ntai, M. L. Bolla, S. J. Malcolmson, D. Kahne, N. L. Kelleher and C. T. Walsh, *J. Am. Chem. Soc.*, 2011, **133**, 5240–5243.
185 C. Maruyama, J. Toyoda, Y. Kato, M. Izumikawa, M. Takagi, K. Shin-Ya, H. Katano, T. Utagawa and Y. Hamano, *Nat. Chem. Biol.*, 2012, **8**, 791–797.
186 H. B. Park, C. E. Perez, K. W. Barber, J. Rinehart and J. M. Crawford, *Elife*, 2017, **6**, e25229.
187 C. E. Perez, H. B. Park and J. M. Crawford, *Biochemistry*, 2018, **57**, 354–361.
188 T. Izoré, J. Tailhades, M. H. Hansen, J. A. Kaczmarski, C. J. Jackson and M. J. Cryle, *Proc. Natl. Acad. Sci.*, 2019, **116**, 2913–2918.
189 S. Hartwig, C. Dovengerds, C. Herrmann and B. T. Hovemann, *FEBS J.*, 2014, **281**, 5147–5158.
190 K. Zaleta-Rivera, C. Xu, F. Yu, R. A. E. Butchko, R. H. Proctor, M. E. Hidalgo-Lara, A. Raza, P. H. Dussault and L. Du, *Biochemistry*, 2006, **45**, 2561–2569.
191 R. A. E. Butchko, R. D. Plattner and R. H. Proctor, *J. Agric. Food Chem.*, 2006, **54**, 9398–9404.
192 S. Lin, S. G. Van Lanen and B. Shen, *Proc. Natl. Acad. Sci.*, 2009, **106**, 4183–4188.
193 C. Chang, J. R. Lohman, T. Huang, K. Michalska, L. Bigelow, J. D. Rudolf, R. Jedrzejczak, X. Yan, M. Ma, G. Babnigg, A. Joachimiak, G. N. J. Phillips and B. Shen, *Biochemistry*, 2018, **57**, 3278–3288.
194 J. R. Lohman, S.-X. Huang, G. P. Horsman, P. E. Dilfer, T. Huang, Y. Chen, E. Wendt-Pienkowski and B. Shen, *Mol. Biosyst.*, 2013, **9**, 478–491.
195 R. P. Mcglinchey, M. Nett and B. S. Moore, *J. Am. Chem. Soc.*, 2008, **130**, 2406–2407.
196 S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea, S. Zhou, V. Fülöp, N. J. Oldham, S.-C. Tsai, G. L. Challis and J. R. Lewandowski, *Nat. Chem.*, 2019, **11**, 913–923.
197 J. Masschelein, P. K. Sydor, C. Hobson, R. Howe, C. Jones, D. M. Roberts, Z. Ling Yap, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *Nat. Chem.*, 2019, **11**, 906–912.
198 Y. Shinohara, S. Takahashi, H. Osada and Y. Koyama, *Sci. Rep.*, 2016, **6**, 32865.
199 S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi and R. Müller, *Chem. Biol.*, 2014, **21**, 855–865.
200 Y. Hai, M. Jenner and Y. Tang, *J. Am. Chem. Soc.*, 2019, **141**, 16222–16226.
201 M. J. Calcott and D. F. Ackerley, *Biotechnol. Lett.*, 2014, **36**, 2407–2416.
202 H. Kries, *J. Pept. Sci.*, 2016, **22**, 564–570.
203 M. Winn, J. K. Fyans, Y. Zhuo and J. Micklefield, *Nat. Prod. Rep.*, 2016, **33**, 317–347.
204 A. S. Brown, M. J. Calcott, J. G. Owen and D. F. Ackerley, *Nat. Prod. Rep.*, 2018, **35**, 1210–1228.
205 A. Stanišić and H. Kries, *ChemBioChem*, 2019, **20**, 1347–1356.
206 M. Kaniusaite, J. Tailhades, E. A. Marschall, R. J. A. Goode, R. B. Schittenhelm and M. J. Cryle, *Chem. Sci.*, 2019, **10**, 9466–9482.
207 R. Li, R. A. Oliver and C. A. Townsend, *Cell Chem. Biol.*, 2017, **24**, 24–34.
208 S. W. Haynes, B. D. Ames, Y. Tang, X. Gao and C. T. Walsh, *Biochemistry*, 2011, **50**, 5668–5679.
209 K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert and H. B. Bode, *Nat. Chem.*, 2018, **10**, 275–281.
210 K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F. Fleischhacker, Y.-N. Shi, P.

Grün and H. B. Bode, *Nat. Chem.*, 2019, **11**, 653–661.
211 C. Steiniger, S. Hoffmann and R. D. Süssmuth, *Cell Chem. Biol.*, 2019, **26**, 1526–1534.
212 C. Steiniger, S. Hoffmann and R. D. Süssmuth, *ACS Synth. Biol.*, 2019, **8**, 661–667.
213 S. Lin, T. Huang, G. P. Horsman, S.-X. Huang, X. Guo and B. Shen, *Org. Lett.*, 2012, **14**, 2300–2303.
214 M. Strieker and M. A. Marahiel, *ChemBioChem*, 2009, **10**, 607–616.
215 R. A. Lewis, L. Nunns, J. Thirlway, K. Carroll, C. P. Smith and J. Micklefield, *Chem. Commun.*, 2011, **47**, 1860–1862.
216 Q. Liu, W. Fan, Y. Zhao, Z. Deng and Y. Feng, *Biotechnol. J.*, 2020, **15**, 1900175.
217 W. Fan, H. Liu, P. Liu, X. Deng, H. Chen, Q. Liu and Y. Feng, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 653–660.