

## **Author Manuscript**

**Title:** Biosynthetic Cyclization Catalysts for the Assembly of Peptide and Polyketide Natural Products

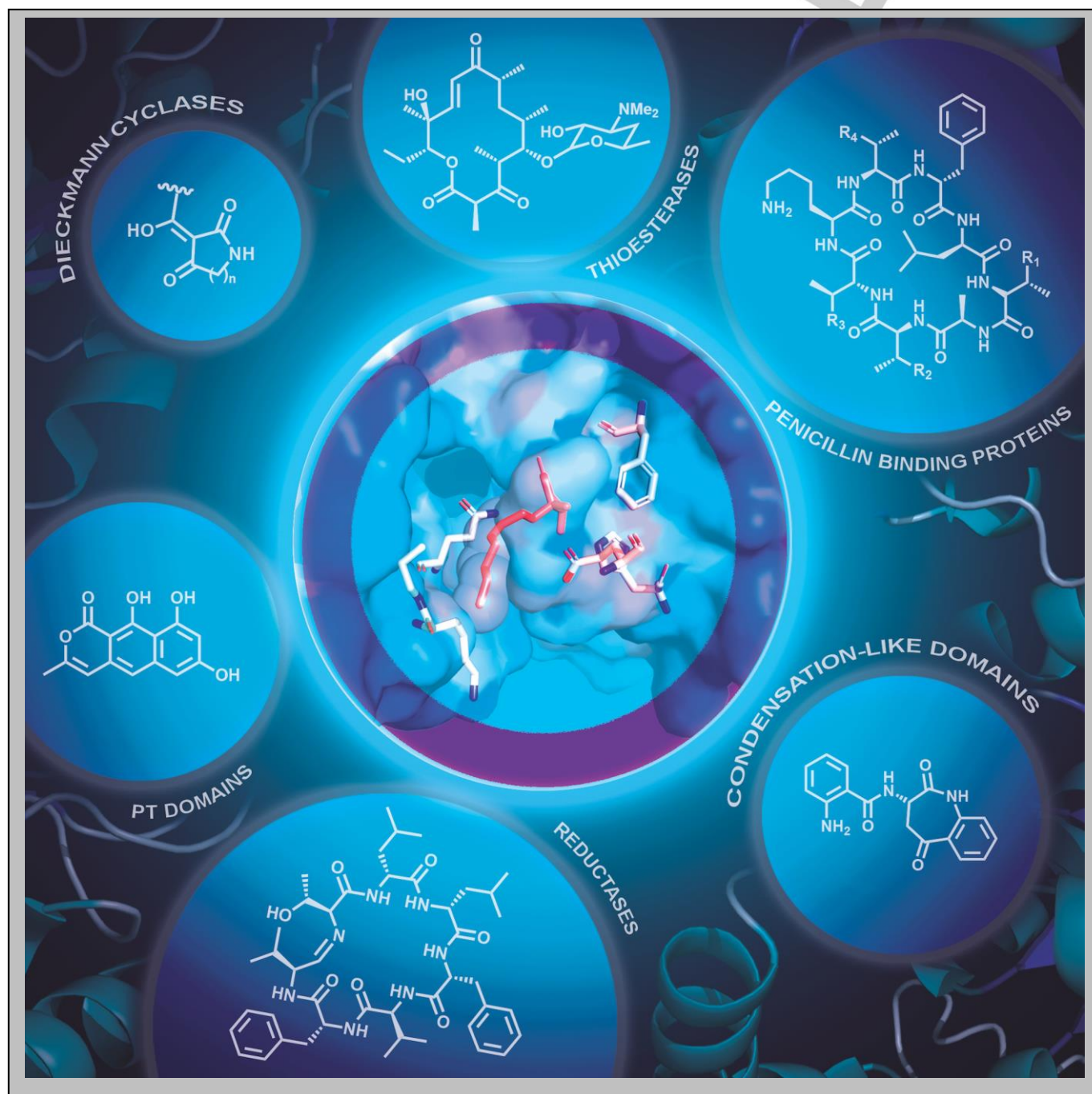
**Authors:** Maria L. Adrover-Castellano; Jennifer J. Schmidt; David H. Sherman

This is the author manuscript accepted for publication. It has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record.

**To be cited as:** 10.1002/cctc.202001886

**Link to VoR:** <https://doi.org/10.1002/cctc.202001886>

## REVIEW

**Biosynthetic Cyclization Catalysts for the Assembly of Peptide and Polyketide Natural Products**Maria L. Adrover-Castellano, Jennifer J. Schmidt, and David H. Sherman\*<sup>[a]</sup>

[a] M. L. Adrover-Castellano, Dr. J. J. Schmidt, Prof. D. H. Sherman  
Life Sciences Institute, University of Michigan  
210 Washtenaw Avenue, Ann Arbor, MI 48109-2216 (USA)

Prof. D. H. Sherman  
Life Sciences Institute, Departments of Medicinal Chemistry, Chemistry, and Microbiology &  
Immunology, University of Michigan (USA)  
E-mail: davidhs@umich.edu

**Abstract:** Many biologically active natural products are synthesized by nonribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs) and their hybrids. These megasynthetases contain modules possessing distinct catalytic domains that allow for substrate initiation, chain extension, processing and termination. At the end of a module, a terminal domain, usually a thioesterase (TE), is responsible for catalyzing the release of the NRPS or PKS as a linear or cyclized product. In this review, we address the general cyclization mechanism of the TE domain, including oligomerization and the fungal C-C bond forming Claisen-like cyclases (CLCs). Additionally, we include examples of cyclization catalysts acting within or at the end of a module. Furthermore, condensation-like ( $C_T$ ) domains, terminal reductase (R) domains, reductase-like domains that catalyze Dieckmann condensation ( $R_D$ ), thioesterase-like Dieckmann cyclases, *trans*-acting TEs from the penicillin binding protein (PBP) enzyme family, product template (PT) domains and others will also be reviewed. The studies summarized here highlight the remarkable diversity of NRPS and PKS cyclization catalysts for the production of biologically relevant, complex cyclic natural products and related compounds.

## 1. Introduction

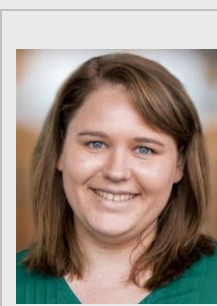
Nonribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs) and hybrids (PKS/NRPS) thereof are responsible for the synthesis of numerous natural products in bacterial and fungal species.<sup>[1]</sup> Several of these compounds are clinically approved therapeutics to treat human diseases, including cancer (e.g. calicheamicin and bleomycin) autoimmune disorders (e.g. rapamycin), parasitic infections (e.g. ivermectin) and bacterial pathogens (e.g. erythromycin, ticacumicin B and vancomycin).<sup>[2]</sup> The multifunctional machinery of these megasynthetases makes them particularly unique. Individual modules containing diverse domains incorporate specific substrates as they work together to assemble complex nonribosomal peptides or polyketides in a linear manner to generate a final product. NRPS and PKS products possess widely varied structures with linear, cyclic and branched elements.<sup>[3]</sup>

The NRPS and PKS megasynthetases contain modules that mediate substrate initiation, chain extension, processing and termination. A module contains a series of domains that catalyze diverse reactions on a peptide or polyketide chain. The NRPS core module contains an adenylation (A) domain that selects and activates amino acid monomers, a condensation (C) domain which catalyzes the formation of peptide bonds, and a thiolation domain (T) or peptidyl carrier protein (PCP) with a phosphopantetheine (Ppant) group for transferring substrates into the catalytic site. NRPS modules might also contain accessory domains including an epimerization (E) domain for the conversion

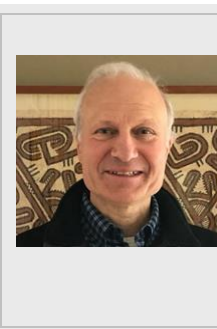
Maria L. Adrover-Castellano graduated from the University of Puerto Rico-Rio Piedras with a B.S. in chemistry in 2018. Soon after joining the Ph.D. Program in Chemical Biology at the University of Michigan-Ann Arbor, she started working under the supervision of Prof. David Sherman. Currently, she explores chemo-enzymatic synthesis of macrolactones, and macrolides utilizing polyketide synthases (PKSs) and applies enzyme bioengineering approaches for the synthesis of these complex molecules.



Jennifer J. Schmidt received her B.S. in biochemistry and molecular biology at UC Santa Cruz followed by a Ph.D. in medicinal chemistry under the supervision of Prof. David Sherman. Her work focuses on the interrogation of polyketide and nonribosomal peptide biosynthetic enzymes utilizing synthetic intermediates.



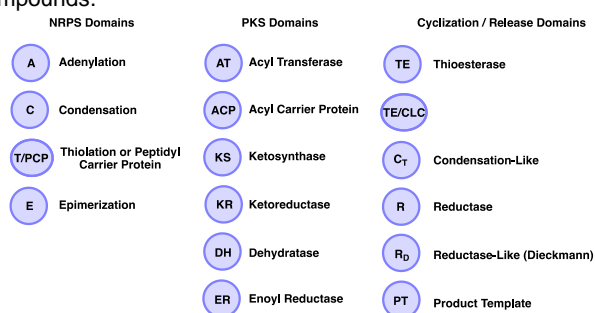
David H. Sherman received his B.A. in chemistry at UC Santa Cruz and Ph.D. at Columbia University. Following 13 years at the University of Minnesota, Prof. Sherman moved to the University of Michigan and is now the Hans W. Vahlteich Professor of Medicinal Chemistry, Professor of Chemistry, and Professor of Microbiology & Immunology. Sherman's laboratory is in the U-M Life Sciences Institute where his research focuses on the discovery and analysis of bioactive natural products and their metabolic pathways from diverse bacteria and fungi.



of L-amino acids to D-amino acids. The PKS core module contains an acyltransferase (AT) domain, which selects and transfers an extender unit, an acyl carrier protein (ACP) also with a Ppant group (similarly to the PCP in NRPS) for extender unit loading, and a ketosynthase (KS) domain for decarboxylative Claisen condensation.<sup>[4]</sup> In the case of PKS, the elective processing domains include a ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) (**Figure 1**).<sup>[1]</sup> During the biosynthesis of a fully elongated and mature linear peptide or polyketide, a terminal domain, usually a thioesterase (TE) catalyzes the release of the peptide and polyketide as a linear or cyclized product.<sup>[4]</sup> In this review, we describe examples of biosynthetic cyclization catalysts found in NRPS and PKS

## REVIEW

systems. The general cyclization mechanism of the TE domain will be discussed including oligomerization and the fungal C-C bond forming Claisen-like cyclases (CLCs). Moreover, a number of unusual or rare termination domains have been identified including terminal condensation-like ( $C_T$ ) domains, reductase (R) domains, Dieckmann cyclases, *trans*-acting TEs from the penicillin binding protein (PBP) enzyme family, product template (PT) domains and select others. Although various detailed reviews on these topics have appeared in the literature previously,<sup>[1, 5]</sup> we focus on the current literature relating to fascinating NRPS and PKS cyclization catalysts responsible for the assembly of particularly unique natural products and related compounds.



**Figure 1.** NRPS and PKS domains including the cyclization and release domains discussed throughout this review.

## 2. Thioesterases

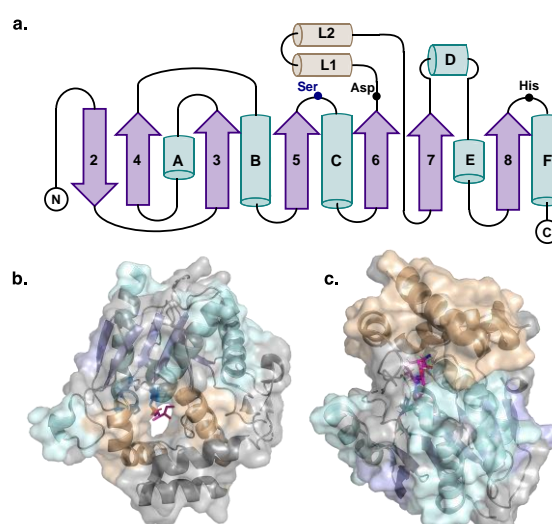
As previously described, NRPSs and PKSs are organized into a series of modules that contain distinct catalytic domains. Utilizing peptidyl or acyl chain initiation, elongation and termination mechanisms, these enzymatic assembly lines are capable of generating a mature product. At the end of a biosynthetic pathway, a terminal TE domain usually located at the C-terminus of a PKS or the NRPS module is responsible for catalyzing the cyclization step during termination.<sup>[1, 6]</sup> Due to the importance and wide range of thioesterases found in nature, this topic has been reviewed with detailed discussions on PKS TEs, NRPS TEs, comparisons of their sequence as well as insights into their structure.<sup>[5a, 7]</sup> Herein, we provide an overview of the cyclization mechanisms employed by TEs and highlight recent examples of work in this area of natural product biosynthesis. Studies on the TE domain are particularly compelling due to their ability to catalyze diverse reactions, including cyclization via oligomerization, and through carbon-carbon bond formation, among others discussed below. In addition, we will cover thioesterase-like Dieckmann cyclases later in the review. Although an important feature of PKS and NRPS systems, the TEII class of thioesterases involved in editing of polyketide intermediates during assembly will not be discussed.<sup>[8]</sup>

### 2.1. Mechanistic Overview

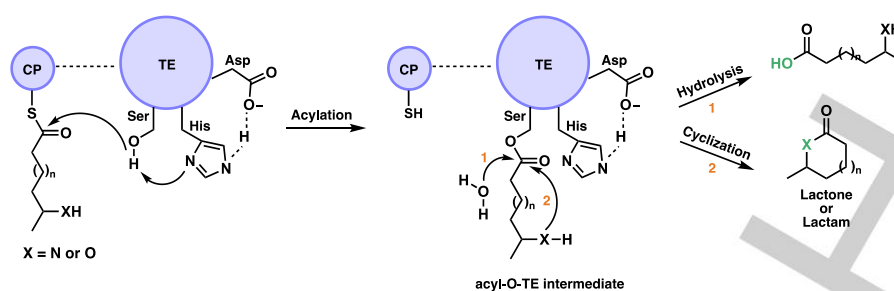
Thioesterases are prevalent in both NRPS and PKS pathways. These enzymes are broadly characterized as members of the  $\alpha/\beta$  hydrolase superfamily, which also includes serine proteases (**Figure 2**). Macrocyclizing TEs, termed type I TEs are generally found at the end of the C-terminal module of NRPS and PKS pathways. In the active site, the TE features a

conserved serine (Ser), histidine (His) and aspartate (Asp) catalytic triad (**Figure 3**). After the peptide or polyketide is fully extended and elongated, it is typically transferred from the upstream module PCP or ACP to the hydroxyl group of the Ser. The Ser residue acts as a catalytic nucleophile activated by the His and Asp dyad, which generates an acyl-O-TE covalent enzyme intermediate.<sup>[5a, 9]</sup> The Ppant group covalently attached to the ACP or PCP active site Ser residue mediates the approach and transfer of the thioester substrate to the TE. Then, an acyl hydroxyl or amino nucleophile adds into the carbonyl of the acyl-O-TE intermediate. An intermolecular nucleophile, generally water, leads to a linear hydrolyzed product, while an intramolecular nucleophile (amino or hydroxyl group) generates a lactam or lactone during cyclization, respectively (**Figure 3**).<sup>[1, 5a]</sup>

Multiple macrocyclizing TE X-ray crystal structures have been solved including for the pikromycin (**1**) and erythromycin (**2**) PKS TEs as well as tyrocidine<sup>[10]</sup> and fengycin<sup>[11]</sup> NRPS TEs. These structures provide the basis for information about the similarity of the protein fold, nature of the substrate channel and catalytic mechanism for ring closure.<sup>[12]</sup> Despite nearly identical fold architecture (**Figure 2a**), PKS TEs are dimers and contain dimerization helices at the N-terminus as well as a hydrophobic substrate channel through the protein (**Figure 2b**). In contrast NRPS TEs are monomeric and do not contain these extra helices but sometimes a primary  $\beta$ -strand at the N-terminus. NRPS TEs also have a more bowl-shaped active site that is capped by a formal lid (**Figure 2c**). These differences are born out in the current understanding of molecular mechanisms employed by TEs to effect macrocyclization. In 2006, the Fecik, Sherman and Smith groups captured a substrate mimic in the active site of the pikromycin (Pik) TE leading to a new hypothesis for macrocyclization.<sup>[13]</sup> The structure revealed a previously unrecognized water barrier in the enzyme that appears to induce a curled substrate conformation to direct macrolactone ring formation.<sup>[13]</sup> Similar studies on NRPS domains cited both key



**Figure 2.** Overall structure of PKS and NRPS thioesterases with the core  $\alpha$ -helices depicted in teal, the lid  $\alpha$ -helices depicted in yellow,  $\beta$  strands represented in purple. **a.** Topology diagram of TEs with the catalytic residues highlighted. **b.** Representation of PikTE (PDB: 2H7X) highlighting the substrate channel typically seen in PKS TEs and a triketide substrate seen in pink. **c.** Representation of VImTE (PDB: 6ECE) depicting the buried, bowl shaped active site with the lid forming the top cap.



**Figure 3.** Mechanism of the lactonization or hydrolysis of a CP-tethered peptide or polyketide intermediate. Highlighted in green throughout the review, the bond formed via the described enzyme.

interactions in the active site (especially apparent in the fengycin TE) as well as inherent preorganization of the linear peptides as the driving force behind cyclization.<sup>[9a, 11]</sup> Indeed, early studies on the tyrocidine TE demonstrated that it was capable of cyclizing substrates with widely differing amino acids compared to the native structure, lending biochemical evidence to preorganization of the linear intermediate as a driving force in the NRPS TE cyclization mechanism.<sup>[14]</sup> This new information provided insights for the design of enzymes capable of catalyzing regioselective macrocyclization of natural or synthetic substrates.<sup>[12b]</sup>

## 2.2. General Cyclization Reactions PKS/NRPS TEs

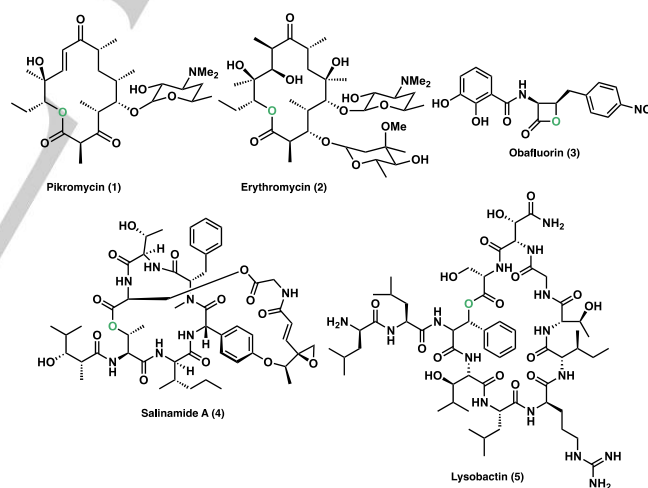
The canonical cyclization of PKSs and NRPSs refers to the TE mediated mechanism that includes the Ser-His-Asp catalytic triad and the acyl-O-TE intermediate, which facilitates the ring closing reaction.<sup>[15]</sup> A number of examples of cyclic structures can be found including lactones and lactams that are formed via canonical and non-canonical mechanisms. The different types of mechanisms will be discussed below. Moreover, chemoenzymatic applications and engineering of these systems for the synthesis of various natural product structural analogs will also be highlighted in the context of recent work involving TE domain mediated cyclization.

### 2.2.1. Lactones

Recently, the Wenczewicz group studied the NRPS *cis*-monocyclic  $\beta$ -lactone antibiotic obafluorin (**3**).<sup>[16]</sup> Obafluorin (Obi) was isolated in 1984 from *Pseudomonas fluorescens* ATCC 39502.<sup>[17]</sup> The biosynthetic gene cluster (BGC) was subsequently identified, and the ObiF TE domain shown to play a direct role in  $\beta$ -lactone formation. The authors demonstrated that the TE catalyzes cyclization of an active site-tethered  $\beta$ -hydroxythioester during product release to generate the strained ring system. Primary sequence analysis revealed that there is a cysteine active site residue (Cys1141) instead of the expected serine for the ObiF TE. This work has the potential for future NRPS engineering to produce targeted peptide  $\beta$ -lactones.

In 2016, the Moore group reported a study of salinamide A (**4**), which is a cyclic depsipeptide possessing antibacterial activity (**Figure 4**).<sup>[18]</sup> Salinamides A and B are the first depsipeptides from this class isolated from marine actinomycete *Streptomyces* sp. CNB-091.<sup>[19]</sup> They found the *sln9* gene encodes a tetradomain NRPS module with a C-terminal TE. Biochemical characterization of the Sln9 TE showed that it catalyzes an intermolecular transesterification of a serine residue of desmethylsalinamide E

with acylated glyceryl thioesters to give desmethylsalinamide C in salinamide biosynthesis. Lysobactin (**5**), is another example of a cyclic depsipeptide with high antibacterial activity against human pathogens (**Figure 4**). This molecule was originally isolated from *Lysobacter* sp. ATCC 53042 in 1988.<sup>[20]</sup> In 2011, the Marahiel group identified and characterized the gene cluster responsible for the lysobactin biosynthesis.<sup>[21]</sup> The *lyb* gene cluster codes for two NRPSs, LybA and LybB. The latter was found to contain an unusual tandem TE architecture. The authors characterized *in vitro* the two individual terminal TEs (TE1, TE2) using a thiophenol activated linear lysobactin analog. Their results demonstrated that only the penultimate thioesterase domain mediates cyclization and simultaneous release of lysobactin. TE1 exclusively catalyzes the formation of the macrocyclic structure, whereas TE2 mediates hydrolytic cleavage of the synthetic substrate.

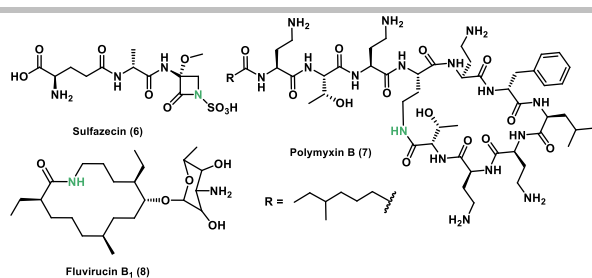


**Figure 4.** Chemical structures of the lactones pikromycin, erythromycin, obafluorin, salinamide A and lysobactin.

### 2.2.2. Lactams

In 2018, the Townsend group revealed the biosynthetic routes for monobactam synthesis in sulfazecin (**6**).<sup>[22]</sup> This  $\beta$ -lactam antibiotic is produced by NRPS enzymes and was isolated from *Pseudomonas acidophila* G-6302 in 1981.<sup>[23]</sup> In the gene cluster, they identified an unusual TE domain possessing a cysteine active site, and described possible routes for  $\beta$ -lactam formation. The proposed mechanism involves an unknown NRPS TE reaction type in which an assembled tripeptide is *N*-sulfonated

## REVIEW



**Figure 5.** Chemical structures of the lactams sulfazecin, polymyxin B and fluvirucin B<sub>1</sub>.

*in-trans* before  $\beta$ -lactam ring synthesis in the TE domain, which is distinct compared to other systems (**Figure 5**).

In 2017, the Li and Velkov groups characterized the polymyxin B (**7**) TE to understand its mode of assembly (**Figure 5**).<sup>[15a]</sup> Polymyxins remain one of the few antibiotics available for treating antibiotic resistant bacteria. In their study they found that the TE contains a catalytic cysteine in the active site similarly to the TE involved in the synthesis of sulfazecin noted above. They also demonstrated that the polymyxin B pathway TE is able to cyclize a linear *N*-acetylcysteamine (SNAC) polymyxin B analog. Additionally, homology model examinations of the TE substrate-binding cleft was found to be negatively charged, suggesting a potential mechanism for cyclization of the substrate. The authors also noted that charged interactions could play a role in the binding of the positively charged linear substrate to the negatively charged TE binding site facilitating cyclization.

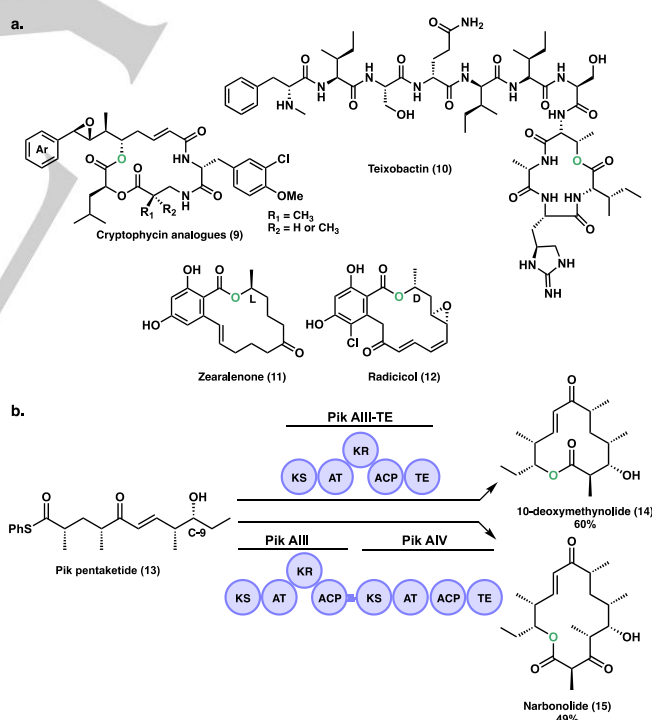
The Schnarr group in 2013 identified and characterized the fluvirucin B<sub>1</sub> (**8**) polyketide synthase (**Figure 5**).<sup>[24]</sup> This 14-membered macrolactam produced by *Actinomadura vulgaris* has shown antifungal and antiviral activity.<sup>[25]</sup> They found that the fluvirucin B<sub>1</sub> PKS consists of five extender modules, flanked by an initial *N*-terminal loading ACP that incorporates the initial  $\beta$ -alanine and a C-terminal TE, which catalyzes ring closure of the linear polyketide substrate. This macrolactam forming TE was also found to include a cysteine in the active site instead of a serine.

### 2.2.3. Chemoenzymatic syntheses

Recently, the Sherman group described the chemoenzymatic synthesis of anti-cancer cryptophycin analogs (**9**).<sup>[26]</sup> These macrocycles are a class of 16-membered ring depsipeptide natural products generated by hybrid PKS/NRPS system (**Figure 6a**). Cryptophycins were first isolated in *Nostoc* sp. ATCC 53789, and showed antifungal activity.<sup>[27]</sup> These compounds were rediscovered in *Nostoc* sp. GSV 224, as a potent antiproliferative microtubule binding agent.<sup>[28]</sup> The penultimate step in cryptophycin biosynthesis involves a macrocyclization catalyzed by the cryptophycin thioesterase (Cp TE).<sup>[29]</sup> In their work, a series of novel cryptophycin chain elongation intermediates were synthesized and the Cp TE demonstrated its utility as a biocatalyst through the formation of a library of cryptophycin analogs. Additionally, biological activity studies identified a low picomolar potency analog containing a styrene functionality that overcomes the need for a  $\beta$ -epoxide group. This work shed light into the production of complex bioactive molecules for the potential treatment of malignant diseases and may represent a promising candidate as an antibody-drug conjugate payload.<sup>[30]</sup>

In 2018, the Zhang group studied the teixobactin (**10**) TE domains.<sup>[31]</sup> Teixobactin is a depsipeptide isolated from a  $\beta$ -proteobacterium *Eleutheria terrae*, with antimicrobial activity against many notorious Gram-positive pathogens (**Figure 6a**).<sup>[32]</sup> The BGC encodes two large NRPSs, Txo1 and Txo2. In the Txo2 termination module two TE domains (TE1, TE2) were identified, similarly to the lysobactin biosynthetic system.<sup>[21]</sup> The Zhang group developed a TE based chemoenzymatic approach for the development of teixobactin analogs. Their work shows that the two TE domains operate together to generate a cyclized product and are functionally exchangeable. In contrast to lysobactin biosynthesis in which only the penultimate thioesterase domain (TE1) mediates cyclization while TE2 mediates hydrolytic cleavage, in teixobactin biosynthesis, the TE domains (TE1, TE2) appear to form a combined active site cavity that is essential for cyclization. Their work further demonstrates the impressive mechanistic diversity found in NRPS systems.

The Sherman group in 2013 reported an *in vitro*, scalable biocatalytic platform utilizing the final two modules (PikAIII and PikAIV) of the pikromycin biosynthetic pathway for the generation of 12- and 14-membered ring macrolactones (**Figure 6b**).<sup>[33]</sup> A thiophenol-activated Pik pentaketide (**13**) substrate was synthesized, and loaded onto PikAIII-TE or PikAIII-PikAIV enzyme systems to afford 10-deoxymethynolide (**14**) and narbonolide (**15**), respectively. Furthermore, these macrolactones were converted to their macrolide counterparts *in vivo* by direct appendage of *D*-desosamine and final P450-mediated C-H oxidation(s). In a follow-up study, a series of unnatural



**Figure 6a.** Chemical structure of cryptophycin analogs, teixobactin, zearalenone and radicicol. **b.** *In vitro* biocatalytic platform using the terminal two monomules (PikAIII and PikAIV) of the pikromycin pathway.

pentaketides were synthesized, which revealed the substrate tolerance of PikAIII-TE *in vitro* for production of novel 12-membered ring macrolactones.<sup>[34]</sup> However, interrogation of pentaketides with epimerized chiral centers in the C-9 hydroxyl group nucleophile revealed the inability of Pik TE to catalyze macrocycle formation. Instead, exclusive hydrolysis of the

## REVIEW

epimerized hexaketide substrates was observed, suggesting that the TE domain is a catalytic bottleneck in the processing of these chain elongation intermediates.

The Boddy group in 2014 reported work on the TE domains from the zearalenone (**11**) and radicicol (**12**) biosynthetic pathways (**Figure 6a**).<sup>[35]</sup> These natural products form part of the resorcylic acid lactone (RALs) class of macrocyclic fungal polyketides that contain a resorcyate (2,4-dihydroxybenzoate) within the 14-membered ring macrolactone. Interestingly, these compounds have opposite configuration in the lactone alcohol group, where zearalenone possesses a L (*S*) configuration, whereas radicicol contains a D (*R*) configuration. In their work, they show that the TE domains responsible for substrate cyclization are highly stereotolerant to D and L configurations contained in synthetic substrate analogs. Both zearalenone and radicicol TE have the ability to macrocyclize substrates with a different stereochemistry and low competing hydrolysis, in contrast to bacterial PKS TEs, which are highly stereoselective. They further demonstrated the ability of these RAL TEs in chemoenzymatic synthesis of 12- to 18-membered ring macrolactones and a 14-membered macrolactam.<sup>[36]</sup>

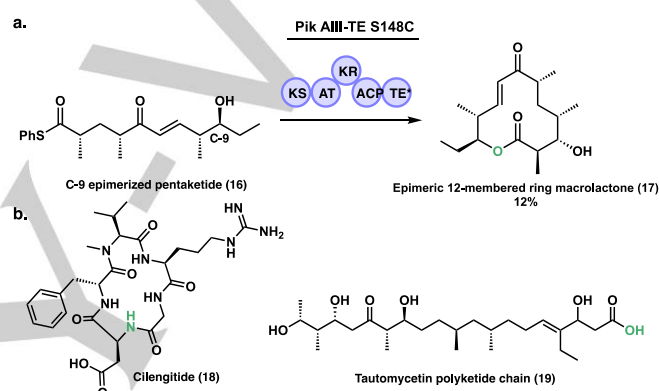
### 2.2.4. Engineering in these systems

In 2017, the Sherman group reported the engineering of the PikTE, where they mutated the serine 148 in the active site to a cysteine (**Figure 7a**).<sup>[37]</sup> The application of the engineered Pik TE<sub>S148C</sub> into a hybrid protein PikAIII-TE<sub>S148C</sub> enabled processing of the C-9 epimerized pentaketide (**16**) for the production of an epimeric 12-membered ring macrolactone (**17**). Comparison of quantum mechanics calculations for wild type Pik TE and the S148C mutant revealed that it significantly lowers the activation barriers to macrolactonization, resulting in a more efficient catalyst. In another recent study, the group created a series of hybrid PKS modules containing exchanged TE domains from heterologous pathways.<sup>[38]</sup> The modules selected were the terminal two modules of the pikromycin (Pik) and erythromycin (DEBS) pathways, and the penultimate module of the juvenimycin (Juv) system. These hybrid PKS modules were enterogated with native and non-native polyketide substrates. Their results showed that substrate TE pairing was a major driving factor in macrolactone production. These works highlight that the TE is often a critical catalytic gatekeeper for the processing of unnatural polyketide chain elongation intermediates. It also showed the substrate flexibility of PKS TE domains and their application in future engineering strategies for the production of polyketide natural product analogs.

The Zhu group in 2019, reported a chemoenzymatic method to synthesize the cyclic peptide cilengitide (**18**) based on the activity of the TE domain from *Microcystis aeruginosa* microcystin synthetase C (**Figure 7b**).<sup>[39]</sup> Moreover, they generated a single S85C mutation in McyC TE, following the strategy used by the Sherman group for the Pik TE.<sup>[37]</sup> For their chemoenzymatic method using initially the wild type McyC TE domain, they synthesized a pentapeptide-BMT substrate. Enzymatic reactions showed that the McyC TE was able to recognize the linear pentapeptide chain and cyclized it into cilengitide. This was further

improved when the McyC TE<sub>S85C</sub> mutant was employed. Kinetic analysis revealed that the McyC TE<sub>S85C</sub> mutation resulted into a superior cyclization catalyst for the pentapeptide-BMT substrate for the production of cilengitide with a 12-fold increase in the  $K_{cat}/K_m$  for macrolactonization.

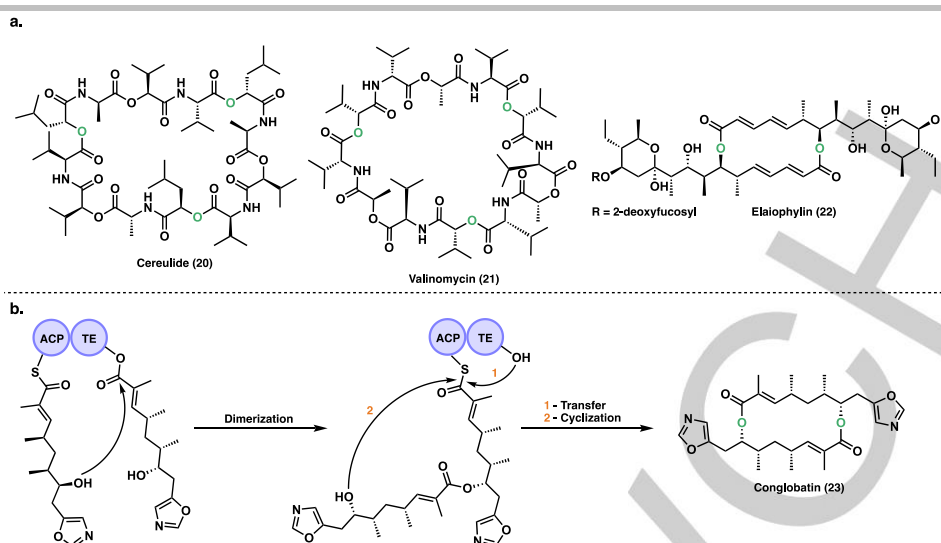
The Sherman and Kim groups in 2016 reported the polyketide tautomycin (**19**) TE exchange with the pikromycin TE in *Streptomyces* sp. CK4412 (**Figure 7b**).<sup>[40]</sup> In this work, the PKS cluster in *Streptomyces* sp. CK4412 that produces tautomycin (TMC) was engineered by replacing the TMC TE domain with the Pik TE, in order to test the ability of the constructed TMC-PikTE chimera to catalyze macrocyclization of the linear substrate. Their findings showed that indeed, the TMC-PikTE chimera produced not only the expected TMC linear product, but also a cyclized form of TMC. This work shows the potential to engineer PKS strains for *in vivo* production of new natural products by exchanging a hydrolyzing TE for a macrocyclizing TE.



**Figure 7a.** Engineering of the Pik TE for synthesis of novel macrolactones. **b.** Chemical structures of cilengitide and the tautomycin polyketide natural products.

### 2.3. TE catalyzed Oligomerization

In addition to canonical head to tail macrocyclization, TE domains are also known to catalyze oligomerization of various natural products during biosynthesis. Cereulide (**20**), valinomycin (**21**), elaiophyllin (**22**) and conglobatin (**23**) are representative examples of both PKS and NRPS derived compounds all formulated by an oligomerization mechanism (**Figure 8a**). Initially, one monomer is covalently attached to the active site serine of the thioesterase and the second to the carrier protein directly upstream (ACP or PCP). From here two modes of dimerization have been proposed including a “forward transfer” model (originally hypothesized for the enterobactin biosynthesis<sup>[41]</sup>) that involves nucleophilic attachment of the TE bound monomer on the carrier protein monomer, functionally building the oligomerized substrate on the TE. The second model, termed the “backward transfer” shifts the TE bound monomer back to a nucleophilic functional group on the carrier protein bound monomer, and building the oligomer on the carrier protein. This dimerized (or oligomerized when this cycle is repeated more than once) is then transferred back to the TE active site serine for



**Figure 8a.** Chemical structures of cereulide, valinomycin and elaiophylin. **b.** Proposed mechanism for TE oligomerization in conglobatin biosynthesis.

final cyclization. Evidence for this backward mechanism initially comes from *in vitro* work with the gramicidin S TE in which linear dimerized thioester intermediates were detected.<sup>[42]</sup> Recent structural work by the Schmeing and Chin groups on the valinomycin TE have further defined the mechanisms employed in this system.<sup>[43]</sup> Utilizing site-specific incorporation of a non-canonical amino acid 2,3-diaminopropionic acid (DAP) into recombinant TE, they demonstrated efficient capture of both the first and last acyl-enzyme intermediates. Their crystallographic data demonstrated that for this enzyme, the “backwards transfer” method is employed, further corroborating the gramicidin S TE data and indicating this is likely a global mechanism employed by this subset of TEs. Deeper investigations on the molecular mechanism of the cyclization function revealed a massive lid rearrangement associated with the dodecadepsipeptidyl bound Vlm TE. The terminal domain reorients the substrate from its position during oligomerization placing it into a pocket that entropically controls cyclization. Overall, this work provided an innovative approach to characterize and capture diverse acyl-enzyme complexes of transiently acylated proteins.

The Boddy group has also investigated the TE from the highly homologous cereulide pathway, which is also produced by two NRPS, CesA and CesB.<sup>[44]</sup> Cereulide is a depsidodecapeptide isolated in 1994 from *Bacillus cereus*.<sup>[45]</sup> Despite the similarities between the structures and BGCs of cereulide and the natural product valinomycin, it was previously suggested that cereulide was synthesized through a different mechanism.<sup>[46]</sup> However, the Boddy group demonstrated that cereulide is in fact produced in a similar manner to valinomycin where the Ces TE catalyzes the trimerization and macrocyclization of a tetradepsipeptide substrate. Interestingly, they also show the Ces TE can utilize related substrates like the valinomycin tetradepsipeptide and a cereulide-valinomycin hybrid tetradepsipeptide with the native substrate to generate natural product analogs. In 2015, the Leadlay group reported the C<sub>2</sub>-symmetrical macrodiolide elaiophylin produced by PKS enzymes.<sup>[47]</sup> They investigated the elaiophylin (Ela) TE domain mechanism of the diolide formation *in vitro* by using synthetic *N*-acetylcysteaminy thioesters (SNAc thioesters) of tetraketide and pentaketide analogs of the natural octaketide monomers. It was

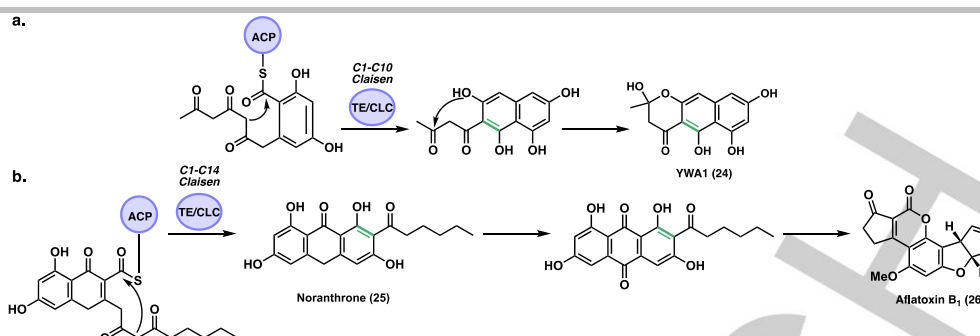
determined that the Ela TE can catalyze homodimerization of the synthetic pentaketide to a novel 16-membered decaketide diolide. Iterative use of the Ela TE active site during ligation and cyclization is suggested since a linear dimeric thioester intermediate was identified in the ring formation. The authors also proposed that Ela TE must catalyze a total of two acylation and two deacylation reactions to form the diolide. Thus, in a similar manner to that determined for the NRPS dimerizing TEs, the Ela TE first ligates two monomers and then re-loads the linear dimer in the active site for cyclization. However, further studies are required to firmly establish the Ela TE dimerization mechanism.

In the same year, the Leadlay group also studied conglobatin biosynthesis by a PKS/NRPS hybrid (**Figure 8b**).<sup>[48]</sup> This compound is an unusual C<sub>2</sub>-symmetrical macrodiolide with antitumor activity isolated from *Streptomyces conglobatus* ATCC 31005.<sup>[49]</sup> The group utilized a single step *in vitro* cloning method to obtain the entire gene cluster that contains the five genes *congA-E* resulting in conglobatin production in a heterologous host. A model substrate was also utilized to mimic the conglobatin monomer, which revealed that the Cong TE acts iteratively. The Cong TE ligates two monomers and then re-binds the dimeric product and cyclizes it similarly to the Ela TE described above. Additional work showed that Cong TE is capable of combining two different monomers into hybrid linear dimers and trimers forming additional complex polyketides.

#### 2.4. Fungal TE domains: C-C bond forming Claisen-Like Cyclases (CLCs)

Fungal nonreducing PKS (NR-PKS) are megasynthases that include the SAT (starter unit: ACP transacylase), KS ( $\beta$ -ketosynthase), MAT (malonyl CoA: ACP transacylase), PT (product template), ACP (acyl carrier protein) and TE/CLC (thioesterase/Claisen like cyclase) domains.<sup>[50]</sup> For the majority of fungal NR-PKS, the TE catalyzes the last step, which releases the product via a Claisen cyclization of a polyketide intermediate, leading to a C-C bond formation.<sup>[51]</sup> This process is accompanied by the cleavage of the thioester bond between the ACP and the polyketide to generate fused ring structures.<sup>[52]</sup> Similarly to the TE





**Figure 9a.** WA TE/CLC mechanism for the production of YWA1. **b.** PKSA TE/CLC domain catalyzes the Claisen cyclization (C-C bond formation) of noranthrone.

domains from bacterial PKSs, the TE/CLC domain also contains a Ser, His and Asp catalytic triad.<sup>[12a]</sup> In 2001, the Ebizuka group identified for the first time a TE with a Claisen cyclase function in the *wA* PKS gene cluster (**Figure 9a**) from *Aspergillus nidulans*, which produces the naphthopyrone YWA1 (**24**).<sup>[52a]</sup>

In 2008, the Kelleher and Townsend groups studied the NR-PKS PksA from the aflatoxin B<sub>1</sub> (**26**) biosynthesis in *Aspergillus parasiticus* (**Figure 9b**).<sup>[50a]</sup> With respect to the TE/CLC domain, their results suggested that it performs a Claisen cyclization on the intermediate from the PT domain (discussed further below) leading to the formation of **25**. In 2010, Townsend and Tsai groups reported a 1.7 Å crystal structure, which provided additional information related to the structure and function of the C-C bond forming TE/CLC domain.<sup>[15b]</sup> Their study revealed that this domain conducts a Claisen cyclization to release the product from the 4'-phosphopantetheine (Ppant)-tethered ACP affording noranthrone (**25**). Further biosynthetic processing steps results in formation of aflatoxin B<sub>1</sub>.

The Townsend group also reported studies on fungal PKS TE function involved in melanin biosynthesis.<sup>[53]</sup> In 1995, the Kubo group proposed that a polyketide synthase from the *PKS1* gene in *Colletotrichum lagenarium* was involved in production of this pigment.<sup>[54]</sup> During the biosynthesis of these molecules, a 1,3,6,8-tetrahydroxynaphthalene (THN) intermediate undergoes two rounds of enzymatic reduction and dehydration forming 1,8-

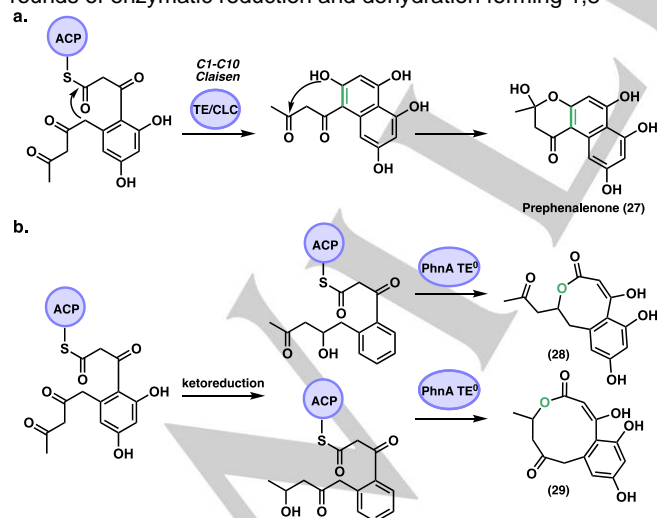
dihydroxynaphthalene (DHN), which ultimately affords melanin. THN is the major product from the NR-PKS Pks1 obtained from the heterologous expression in *Aspergillus oryzae*.<sup>[55]</sup> In their work, they performed an *in vitro* characterization of Pks1 from *Colletotrichum lagenarium*. The Townsend group showed that Pks1 TE is both a Claisen cyclase and deacetylase of an enzyme-bound monocyclic hexaketide precursor that generates THN. This rare dual function demonstrates the notable versatility of the  $\alpha/\beta$  hydrolase fold family of enzymes.

Our final example is the TE/CLC domain found in the herqueinone biosynthesis from *Penicillium herquei*. In 2016, the Houk and Tang groups studied the NR-PKS PhnA that is responsible for biosynthesis of a heptaketide backbone as well as for its cyclization into a hemiketal-containing naphtho- $\gamma$ -pyrone prephenalenone (**27**).<sup>[50b]</sup> They showed that the PhnA TE/CLC catalyzes a C1-C10 Claisen cyclization generating the corresponding naphthopyrone products (**Figure 10a**). Remarkably, when the TE/CLC domain was deleted in PhnA (TE<sup>0</sup>) shunt products **28** and **29** appeared, which are 8- and 10-membered ring macrolactones, respectively (**Figure 10b**). Their study revealed the basis for new and unexpected metabolites derived from polyketide cyclization.

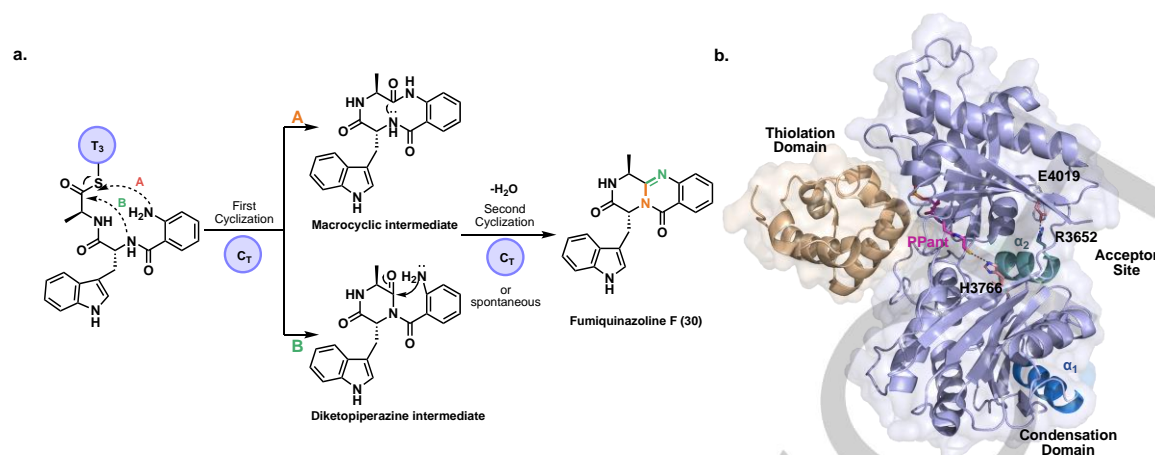
### 3. Condensation Domains

#### 3.1. Condensation-like (C<sub>T</sub>) domains

The minimal NRPS module is typically composed of a condensation (C), adenylation (A) and a carrier protein (CP) domain, also termed a thiolation (T) or a peptidyl-carrier-protein (PCP). In bacterial NRPSs, the final off-loading reaction is frequently catalyzed by a thioesterase (TE) domain responsible for the hydrolysis or cyclization of the resulting peptide.<sup>[56]</sup> However, fungal NRPSs have evolved to employ a different mechanism for macrocyclization, a C-terminal condensation-like (C<sub>T</sub>) domain, reserving TEs only for hydrolytic offloading.<sup>[1, 56b, 57]</sup> Genome sequencing data has shown that 60-90% of the fungal NRPSs have a terminal C<sub>T</sub> domains demonstrating the nearly universal adoption of this offloading mechanism in fungi.<sup>[56b]</sup> Understanding the basis of the C<sub>T</sub> domain catalysis in fungal NRPSs offers interesting avenues for biocatalysis and chemoenzymatic approaches for synthesis of numerous cyclic nonribosomal peptides.



**Figure 10a.** PhnA TE/CLC C1-C10 Claisen cyclization producing the naphthopyrone prephenalenone. **b.** Shunt products from the PhnA TE deletion (TE<sup>0</sup>).



**Figure 11a.**  $C_T$  cyclization routes. A and B represent two possible cyclization routes by the  $C_T$  domain for the formation of Fumiquinazoline F. **b.** Crystal structure of the *holo* T domain (gold) in conjunction with the  $C_T$  domain (purple). The Ppart arm and the conserved histidine are drawn in pink. The salt bridge between E4019 (pink) of the  $\beta 12$  -  $\beta 13$  loop and R3652 (teal) of helix  $\alpha 2$  is highlighted to show the closure of the solvent channel at the acceptor site.

### 3.2. Structural and Mechanistic Overview of $C_T$ domains

The macrocyclization mechanism of fungal NRPSs differs significantly from bacterial TE domains. The activity of  $C_T$  domains was originally established by the Walsh and Tang groups through the *in vitro* reconstitution of the trimodular NRPS genes (TqaA) involved in formation of fumiquinazoline F (**30**), a 6,10 fused heterocyclic ring isolated from *Aspergillus fumigatus*.<sup>[56b]</sup> They confirmed that the  $C_T$  domain was responsible for catalyzing the intramolecular attack of an amine nucleophile on the CP domain thioester linkage, and proposed the mechanistic basis for cyclization. Analysis of the fused ring structure generated two initial hypotheses regarding how this core is formed. First, the  $C_T$  domain could catalyze an initial 10-membered ring and subsequent transannular closure followed by dehydration to generate the fused system (**Figure 11a**, pathway A). Second, the  $C_T$  domain could form an initial diketopiperazine like intermediate, followed by nucleophilic attack of the aniline amino group (**Figure 11a**, pathway B) to produce the same scaffold. In order to distinguish between these two mechanisms, an *N*-methylated tripeptide analog that would preclude amine attack was synthesized and used as a mechanistic probe. The experiment showed buildup of the 10-membered ring product, suggesting initial macrocyclization is favored over DKP formation. Moreover, this mechanism is comparable to that proposed for the  $C_T$  domains from the 6,11 bicycle in the asperlicin systems,<sup>[58]</sup> indicating the larger rings are likely formed first, followed by the intramolecular rearrangement to generate the smaller fused heterocyclic systems.

Further efforts have highlighted the structural and mechanistic differences between  $C_T$  domains, canonical NRPS condensation domains and bacterial TEs.<sup>[59]</sup> As previously discussed, bacterial NRPSs terminate by first transferring the growing peptide to the TE domain and then activating it for cyclization through a His-Asp-Ser catalytic triad.<sup>[59]</sup> In contrast,  $C_T$  domains, along with canonical C domains, complete their chemistry on T domain tethered substrates. Biochemical investigations have demonstrated that the  $C_T$  domain is only active on substrates bound to the correctly paired T-domain partner and that peptidyl-SNac or peptidyl-CoA substrates are not

recognized by the  $C_T$  domain (as they would be for a canonical bacterial TE domain).<sup>[56b]</sup> Both types of condensation domains also utilize a histidine in the active site that functions as a general base.<sup>[56b, 60]</sup> The catalytic histidine deprotonates the amine nucleophile, which then promotes attack on the thioester carbonyl for cyclization and further release in a single step. Recently, a crystal structure of the TqaA  $C_T$  domain has demonstrated that despite low sequence homology (less than 20%), the overall  $C_T$  domain fold compared to the canonical condensation domain VibH and SrfA-C is well conserved.<sup>[59]</sup> The most notable differences are that the N-terminal loop in the C domains is replaced with a helix ( $\alpha_1$ ). This alpha helix causes a shift in the truncated  $\alpha_2$  helix toward the acceptor site. A notable salt bridge between Glu4019 and Arg3652, effectively closes off the position where a downstream T domain would interact with the acceptor site.<sup>[59]</sup> Furthermore, this  $\alpha_2$  helix is capable of making extensive contacts with the  $\beta 12$  -  $\beta 13$  loop, which obstructs the solvent channel generally found in C domains, preventing water and other nucleophiles from competing with the cyclization reaction (**Figure 11b**).

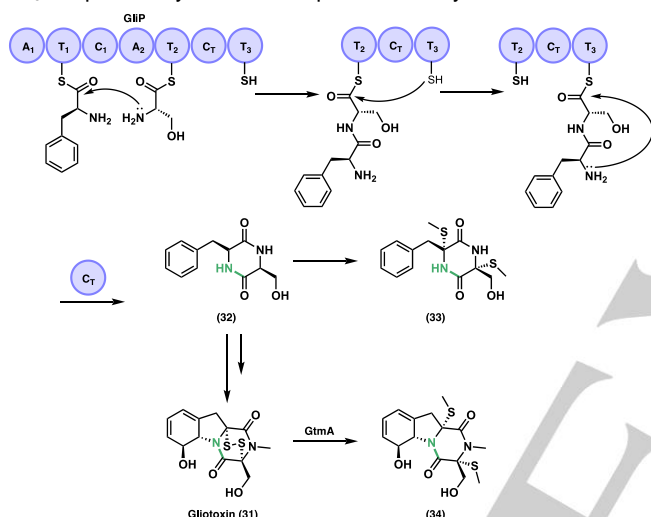
### 3.3. Examples of $C_T$ domains

#### 3.3.1. Gliotoxin

Gliotoxin (**31**) is one of the most notable members of the epipolythiodiketopiperazines (ETPs), a family of toxic NRPS-derived diketopiperazines (DKPs) (**Figure 12**).<sup>[61]</sup> The biosynthesis of gliotoxin is mediated by the *gli* BGC in *Aspergillus fumigatus* as well as other related fungi. The gene cluster contains a core bimodular NRPS *gliP* gene, that has been characterized to convert L-Phe and L-Ser into the DKP cyclo(L-Phe-L-Ser).<sup>[62]</sup> Due to the lack of an apparent offloading domain, initial reports speculated that the core DKP was formed by a spontaneous cyclization. However, the presence of a conserved histidine within an SHXXDXXS/T sequence suggested the presence of a  $C_T$  domain. Notably, although GliP specifies attachment of two amino acids, it contains a third, terminal T domain that may play a role in product offloading in this pathway, a feature not identified in the aforementioned  $C_T$  domain containing pathways. This prompted the Keller and Schroeder groups to investigate the formation of

## REVIEW

DKP by these specialized condensation domains and to elucidate its association with the second and third T domains.<sup>[62c]</sup> They first investigated the function of the GliP-C<sub>T</sub> and the GliP-T<sub>3</sub> domains *in vivo* by generating two truncation mutants, GliP-ΔC<sub>T</sub>T<sub>3</sub> and GliP-ΔT<sub>3</sub> in *A. fumigatus*. Their results show a complete loss of **31-34** for the GliP-ΔC<sub>T</sub>T<sub>3</sub> strain. Next, GliP-ΔT<sub>3</sub> strain alone was investigated and again no **31** was detected or the shunt metabolites **32** or **33** for the GliP-ΔT<sub>3</sub>, only trace quantities of **34**, the detoxification product of **31**. The experiment was repeated by growing cultures supplemented with **31** since it has been shown that it serves as a positive feedback for its own production through regulation of *gli* cluster expression.<sup>[63]</sup> Again, similar results were obtained using LC-HRMS comparison of the extracts. No production of DKP containing metabolites were observed with the GliP-ΔC<sub>T</sub>T<sub>3</sub> strain, whereas the GliP-ΔT<sub>3</sub> strain produced small amounts of cyclized materials **32** and **33**. These results demonstrate that for normal biosynthesis of **31**, the T<sub>3</sub> domain of GliP is required.<sup>[62c]</sup> Trace products observed with only the GliP-ΔT<sub>3</sub> are potentially formed via spontaneous cyclization.



**Figure 12.** Function of GliP and abbreviated biosynthesis of gliotoxin showing the most abundant intermediates or shunt metabolites (**32**) and (**33**), as well as the detoxification product (**34**).

Finally, heterologous expression of a GliP mutant protein was conducted with point mutations designed to inactivate the first and second T domains (GliP-ΔT<sub>1</sub>T<sub>2</sub>). In order to correct for potential cyclization caused by non-T<sub>3</sub> tethered substrate, a second mutant protein with all three T domains inactivated was engineered (GliP-ΔT<sub>1</sub>T<sub>2</sub>T<sub>3</sub>). Utilizing a synthetic L-Phe- L-Ser-*N*-acetylcysteamine (L-Phe-L-Ser-SNAC) in conjunction with the heterologous expressed GliP-ΔT<sub>1</sub>T<sub>2</sub> and GliP-ΔT<sub>1</sub>T<sub>2</sub>T<sub>3</sub>, *in vitro* cyclization was confirmed to be greatly enhanced by the presence of the activated T<sub>3</sub> domain. This study, in conjunction with previous *in vivo* studies demonstrated that L-Phe-L-Ser dipeptide cyclization is not spontaneous as previously proposed.<sup>[62a, 64]</sup> Rather, it requires two additional GliP domains, the second condensation-like domain (C<sub>T</sub>) and the terminal (third) thiolation domain (T<sub>3</sub>) in which the nascent dipeptide appears to be transferred from the T<sub>2</sub> domain to the T<sub>3</sub> domain prior to C<sub>T</sub> mediated cyclization (**Figure 12**).<sup>[62c]</sup> This is in contrast to the previously discussed C<sub>T</sub> domains for tripeptide substrates in the fumiquinazoline pathway.

### 3.3.2. Nanangelenin A

Nanangelenin A (**36**) contains a 1-benzazepine core, which is rare among natural products (**Figure 13a**).<sup>[65]</sup> The Piggott and Chooi groups explored the biosynthesis of this alkaloid isolated from a novel Australian fungus, *Aspergillus nanangensis*.<sup>[66]</sup> Nanangelenin A contains a 3,4-dihydro-1-benzazepine-2,5-dione-*N*-prenyl-*N*-acetoxy-anthranilamide scaffold and its biosynthesis involves the NRPS *nan* gene cluster, showing a similar T<sub>3</sub>/C<sub>T</sub> domain as discussed above for gliotoxin.<sup>[66a]</sup>

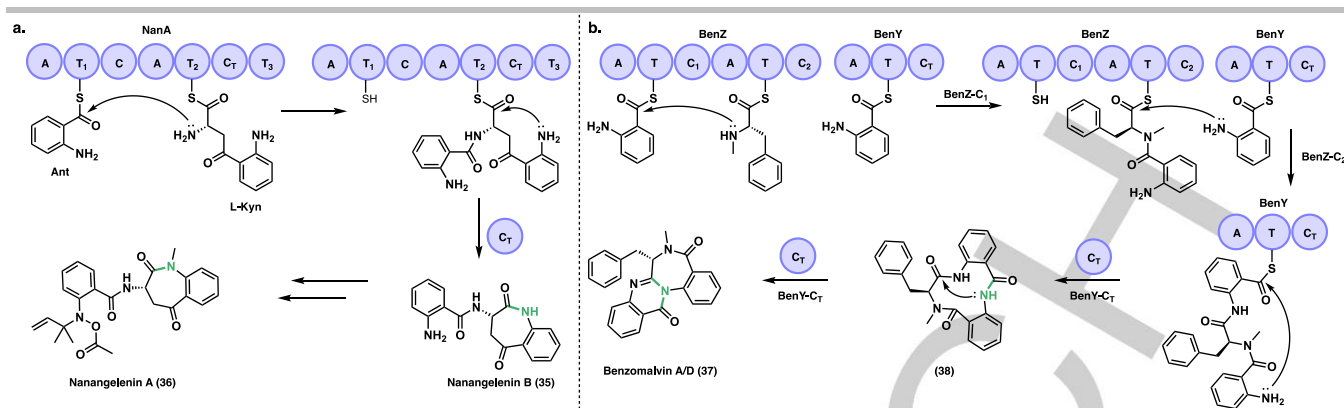
Utilizing bioinformatic analysis, the authors identified a putative biosynthetic cluster associated with the formation of this scaffold. Through sequential activation of the reconstituted gene cluster they found five biosynthetic intermediates, which established the role of each encoded enzyme towards assembly of the final product. Importantly, the NRPS *nanA* gene was found to incorporate anthranilic acid (Ant) and L-kynurenine (L-Kyn) leading to a dipeptide containing two potential cyclizing amines (**Figure 13a**). Heterologous reconstitution of *nanA* produced only a single product indicating the terminal C<sub>T</sub>/T domain pair effects regioselective cyclization to form nanangelenin B (**35**).<sup>4</sup> This was unequivocally established by constructing *NanA* carrying mutations in the C<sub>T</sub> and T<sub>3</sub> active sites. Inactivation of the *nanA* C<sub>T</sub> domain (generating the *NanA*<sub>H2106A</sub> mutant protein) resulted in the loss of (**35**) production and instead accumulated small amounts of the linear Ant-L-Kyn dipeptide. In a similar manner, the *NanA*-T<sub>3</sub> mutant (S2401A) was tested and resulted in low level production (<3% of WT) of (**35**) and even lower amounts of the linear Ant-L-Kyn. Expression of both *NanA* C<sub>T</sub> and T<sub>3</sub> mutants together with *NanC* also yielded a small amount of the linear dipeptide.<sup>[66a]</sup>

Additionally, to further probe C<sub>T</sub> domain regioselectivity, the investigators synthesized the Ant-L-Kyn dipeptide SNAc thioester (Ant-L-Kyn-SNAC) to mimic the T domain P<sub>ant</sub>-tethered Ant-L-Kyn, and expressed the stand alone C<sub>T</sub> domain in *E. coli*. The native reaction demonstrated that this C<sub>T</sub> domain was able to catalyze cyclization to **35** despite the lack of a carrier protein. Inactivation of the C<sub>T</sub> domain produced a new product. NMR studies revealed this to be a spontaneous cyclization event from the anthranilic acid amine, demonstrating the clear divergence of the enzymatically catalyzed cyclization from the kynurenine amine to the more favorable anthranilic amine.<sup>[66a]</sup>

### 3.3.3. Benzomalvin A/D

Benzomalvin A and D (**37**) are benzodiazepines, which were first isolated from a *Penicillium* species (**Figure 13b**). These compounds have been shown to antagonize the human NK1 receptor, inhibiting the effects of substance P.<sup>[67]</sup> The Wu, Keller and Kelleher groups described for the first time the BGC for benzomalvin A and D (*ben*).<sup>[68]</sup> The system was identified using fungal artificial chromosomes with a metabolomic scoring (FAC-MS) platform in *Aspergillus terreus* (ATCC 20542). The FAC-MS approach utilizes high-throughput capture of intact fungal BGCs from genomic DNA in an *Escherichia coli-Aspergillus nidulans* shuttle vector called a FAC.<sup>[69]</sup> This is followed by gene cluster heterologous expression in *A. nidulans* and detection and analysis of small molecule products by metabolomics and cheminformatics using high resolution mass spectrometry (HRMS).<sup>[69]</sup>

## REVIEW



**Figure 13a.** Initial steps in the biosynthesis of nanangelinen A. **b.** Biosynthesis of benzomalvins proceeds through formation of the Anth-NmPhe dipeptide covalently bound via a thioester bond to the second T-domain of BenZ. BenZ-C<sub>2</sub> subsequently catalyzes formation of the Anth-NmPhe-Anth tripeptide covalently bound to BenY-T. BenY-C<sub>T</sub> then catalyzes cyclization and cleavage of the thioester bond, leading to *trans*-annulation and production of benzomalvin A/D.

The *ben* BGC is composed of a *benX* gene that encodes a putative SAM-binding methyltransferase and two NRPS genes *benY* and *benZ*. Benzomalvin A and D biosynthesis proceeds through formation of an NRPS-tethered linear tripeptide (Anth-NmPhe-Anth) containing two anthranilic acid (Anth) residues and one *N*-methylphenylalanine (NmPhe) residue, which is later cyclized to form an 11-membered macrocyclic intermediate (**38**). Recently, these investigators deleted C-domains of *benY* and *benZ* from a FAC containing the entire benzomalvin gene cluster.<sup>[70]</sup> The results from the deletions were analyzed by heterologous expression in *A. nidulans* and targeted metabolomics was employed to determine the biosynthesis of benzomalvin in fungi and the role of the C domains. The results show the production of **37** and **38** by AtFAC9J20- $\Delta$ *benY-C*, although at a lower level compared to AtFAC9J20 (without any deletions), suggesting that *benY-C* encodes the C<sub>T</sub> domain. The biosynthetic pathway for benzomalvin was also deciphered from these findings (**Figure 13b**). Further data indicated that BenZ-C<sub>2</sub> acts as the second internal C-domain while BenY-C functions as the C<sub>T</sub> domain.<sup>[70]</sup>

This work suggests that BenY-C<sub>T</sub> could also mediate formation of the 7-membered ring, representing the first reported benzodiazepine synthase. This process might be occurring through a *trans*-annulation reaction using **38** as the substrate. This hypothesis was tested, and the group obtained results suggesting that BenY-C<sub>T</sub> could directly catalyze *trans*-annulation. However, the possibility that another enzyme was involved from the *A. terreus* strain or the *A. nidulans* heterologous host could not be ruled out. In general, the FAC-MS strategy was beneficial to identify the roles of the C domains in the biosynthesis of benzomalvin A and D, showing that there are two internal C-domains (BenZ-C<sub>1</sub> and BenZ-C<sub>2</sub>) and the terminal cyclization domain BenY-C<sub>T</sub>.<sup>[70]</sup>

## 4. Reductive Cyclization

Typical TE domains catalyze offloading in PKS, NRPS or PKS/NRPS hybrids to generate carboxylic acids or cyclic macrocycles. Another release mechanism sometimes utilized to terminate biosynthetic assembly involves a terminal reductase (R) domain. Originally reported for lysine primary metabolism,<sup>[1, 5b, 71]</sup> these domains are part of the short-chain

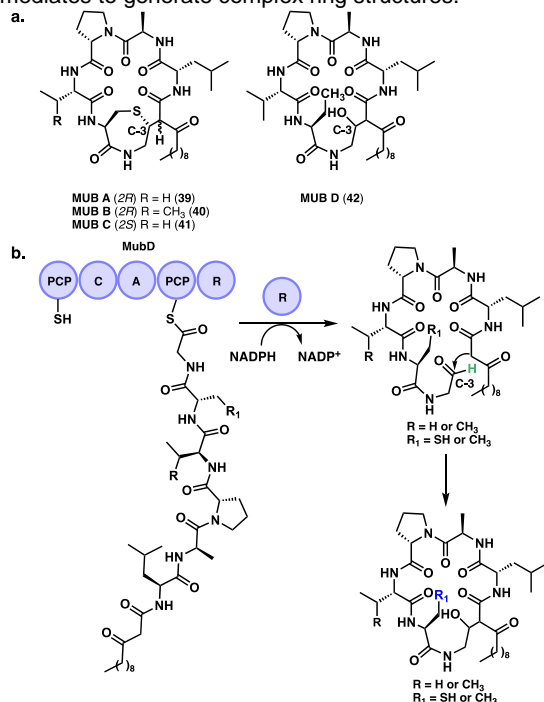
dehydrogenase/reductase (SDR) superfamily catalyzing NAD(P)H-dependent reductive release of acyl-S-polyketides or peptidyl-S-nonribosomal peptides. One of the most conserved features of the SDR enzymes is an  $\alpha/\beta$  folding pattern with a central beta sheet flanked by two - three  $\alpha$ -helices from each side, which is characteristic of a Rossmann-fold.<sup>[72]</sup> This key motif is characteristic of a NAD(P)H nucleotide binding site, which then facilitates recruiting of a NAD(P)H/NAD(P)<sup>+</sup> redox cofactor.<sup>[73]</sup> The mechanism relies on a hydride and proton transfer that involves nicotinamide in conjunction with a tyrosine residue in the active site of the enzyme.<sup>[73a]</sup> R domains enable product release via a catalyzed two-electron or four-electron reduction, resulting in an aldehyde or alcohol functional group, respectively. Interestingly, after a two-electron reduction and subsequent aldehyde formation, the product could undergo a spontaneous cyclization process.<sup>[1, 73a]</sup> Even though the formal cyclization is generally nonenzymatic, the fact that a reductase catalyzes the release of the product as a reactive aldehyde, which then spontaneously cyclizes, makes these mechanism noteworthy. Many examples of these linear and cyclized compounds are also found in a recent review from the Kelleher and Thomson groups where they discussed compounds produced from reductase terminated BGCs.<sup>[5b]</sup> Herein, we will focus on the two-electron reduction and spontaneous cyclization catalyzed by these domains.

### 4.1. Mutanobactins

Recently, the Chen group studied the macrocyclic lipohexapeptides mutanobactins A-D (**39-42**) (MUBs) derived from a PKS/NRPS hybrid pathway (**Figure 14a**).<sup>[74]</sup> These compounds were isolated from the human oral bacterium *Streptococcus mutans* UA159.<sup>[75]</sup> The *mub* BGC found in many *S. mutans* strains is the primary causative agent of tooth decay.<sup>[76]</sup> The authors found that the C-terminal reductase domain of MubD initially releases the linear peptide as an aldehyde (**Figure 14b**). From there, a three-step spontaneous process can occur depending on the reactive functional groups leading to either macrocycles or linear derivatives. Initially, an intramolecular aldol addition mediates the C-C macrocyclization event. This intermediate can be further modified spontaneously via the presence of a reactive thiol group that can attack the C-3 position generating the 1,4-thiazepane rings of select MUBs. Finally, in the presence of a hemithioacetal, the macrocyclized  $\beta$ -thiodiketones can be converted to linear lipopeptides through a retro-aldol

## REVIEW

reaction. These molecules represent an interesting route used by bacteria that takes advantage of biosynthetically formed reactive intermediates to generate complex ring structures.



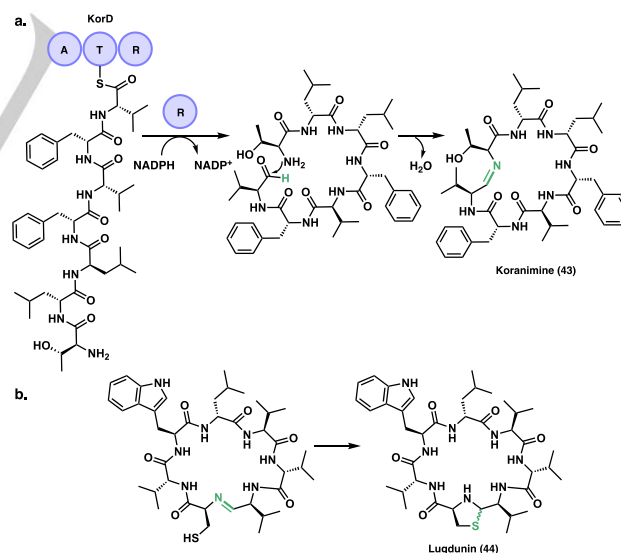
**Figure 14a.** Mutanobactins A – D. **b.** Reductase (R) domain mediated release of aldehyde precursor in the biosynthesis of mutanobactins.

#### 4.2. Macrocyclic iminopeptides

Macrocyclic iminopeptides represent another great example of aldehyde offloading mediated by an R domain in which the linear product undergoes spontaneous cyclization.<sup>[73a]</sup> In 2001, the Moore group discovered the nostocyclopeptides, the first cyclic iminopeptides from the terrestrial cyanobacterium *Nostoc* sp. ATCC 53789.<sup>[77]</sup> Despite the biosynthetic cluster terminating in a reductase domain, no linear products were observed. Further studies into the macrocyclization of these compounds, demonstrated that they are offloaded as an aldehyde by the reductase domain (NcpB) and then spontaneously cyclize to form the imine. This was further supported by NMR studies in which the solution phase structure demonstrated the importance of preorganization for cyclization over polymerization and other side reactions.<sup>[78]</sup>

More recently, other members of this family have been discovered including the scytonemides (whose biosynthesis is currently unknown but is proposed to form via a similar aldehyde amine addition) and the koranimines.<sup>[79]</sup> In 2011, the Kelleher group discovered the cyclic heptapeptide koranimine (**43**), from a *Bacillus* sp. collected in Koran, Louisiana (**Figure 15a**).<sup>[79b]</sup> They utilized a scanning microbial proteome mining approach to identify expressed gene clusters. The koranimine structure was determined using tandem mass spectrometry, feeding studies with stable isotope tracers, NMR spectroscopy and *in vitro* enzyme reconstitution. The gene cluster encodes four NRPS proteins and five adenylation (A) domains. Interestingly, the natural product was determined to be a heptapeptide with an NRPS bearing only five A domains (suggesting it should incorporate only five amino acids) due to their absence from KorA and KorC. It was proposed that a single adenylation domain in

KorA serves two carrier protein sites although the exact mechanism remains unknown. Another unusual finding is a domain structure containing a T-T-C tridomain and no linked A domain in KorC. Utilizing a Ppat ejection assay, the investigators determined that the thiolation domains in KorC are both charged by the A domain of KorD. Moreover, the cyclization process is initiated by the R domain on KorD, which performs a two-electron NADPH-dependent reduction of the C-terminal acyl-S-thioester acid to an aldehyde. This is followed by a nonenzymatic macrocyclization event that yields the imine product **43**. In 2017, the Baran group confirmed the spontaneous macrocyclization producing koranimine by the synthesis of the aldehyde precursor.<sup>[80]</sup> An additional example is the novel peptide antibiotic lugdunin (**44**) that was reported in 2016 as the first compound of a new class of macrocyclic thiazolidine peptides produced by a NRPS system (**Figure 15b**).<sup>[81]</sup> This compound has potent antibacterial activity against major human pathogens, which have been analyzed in animal models. Lugdunin was isolated from *Staphylococcus lugdunensis* IVK28, which prevented *S. aureus* colonization in the human nasal cavity. Similar to koranimine, lugdunin is comprised of a heptapeptide derived from a five A domain containing NRPS. Their study suggests that in this pathway, the single adenylation domain of LugC is responsible for the iterative activation of three consecutive valine units with alternating L- and D-configurations. The R domain in LugC cleaves the thioester-bound peptide using a NAD(P)H cofactor. The heptapeptide is released from the enzyme complex with a C-terminal aldehyde (L-Val) and is subsequently cyclized by the N-terminal amine (L-Cys) forming a macrocyclic imine, similar to that of koranimine. However, in this example, further nucleophilic attack of the cysteine thiol group generates the unusual 5-membered thiazolidine heterocycle found in **44**.



**Figure 15a.** Termination step mediated by the reductase (R) domain in the biosynthesis of koranimine B. **b.** Final step in the biosynthesis of lugdunin and nucleophilic attack of the cysteine thiol group forming the 5-membered thiazolidine heterocycle.

#### 4.3. Pyrrolobenzodiazepines

Pyrrolobenzodiazepines containing natural products feature a tricyclic ring system formed by an anthranilate (A ring), a diazepine (B ring) and a hydropyrrole (C-ring).<sup>[5b, 82]</sup> Anthramycin

## REVIEW

(45) is the first known bacterially NRPS derived benzodiazepine alkaloid with potent antitumor and antibiotic activity (Figure 16). This compound is a pyrrolo [1,4]-benzodiazepine from *Streptomyces refuineus* sbsp. *thermotolerans* discovered during the 1960s.<sup>[83]</sup> During assembly, the dipeptide anthramycin precursor is released by a terminal R domain. Intramolecular addition of the anthranilate-derived arylamine then occurs by adding into the dehydropyrroline acrylamide aldehyde resulting in the formation of the 7-membered hemiaminal benzodiazepine skeleton of anthramycin.<sup>[84]</sup> Moreover, tomaymycin (46) is another pyrrolobenzodiazepine structurally similar to anthramycin. It is also released as an aldehyde by an R domain and spontaneously cyclized (Figure 16). This compound is an antitumor antibiotic produced by *Streptomyces achromogenes*.<sup>[85]</sup> In 2009, the Gerratana group identified and sequenced the BGC providing insights into A ring formation in tomaymycin.<sup>[86]</sup> Later, in 2017, the Müller group reported an *in vitro* reconstitution of the NRPS system including intact protein mass spectrometry analysis, which dissected every step in tomaymycin biogenesis.<sup>[87]</sup> Their method could be broadly applicable to diverse systems for dissection of the reaction steps in secondary metabolic pathways.

Our last example of this class is tilivalline (47) produced by an NRPS was isolated from *Klebsiella oxytoca* found in the human gut microbiota, and represents a marker for diagnosis and treatment of colitis.<sup>[88]</sup> The Breinbauer and Zechner groups studied the biosynthesis of tilivalline, whose gene cluster did not appear to contain any genes associated with the characteristic indole functionality.<sup>[89]</sup> Genome analysis revealed a trypticase gene *tnaA*, which was subsequently inactivated. Cultures of this  $\Delta$ *tnaA* *K. oxytoca* produced a new product, tilimycin (48), containing an alcohol functionality in place of the indole moiety. Further analysis of the biosynthesis of tilimycin shows that the dipeptide is reductively released as an aldehyde via an R domain followed by a spontaneous ring closure. From there, tilimycin was shown to react nonenzymatically with free indole (generated from tryptophan) to generate tilivalline (Figure 16).

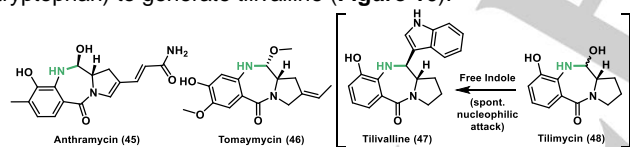


Figure 16. Structures of anthramycin, tomaymycin, tilivalline and tilimycin.

## 5. Dieckmann Cyclases

The family of Dieckmann cyclases were recently found in the biosynthesis of natural products containing tetramate or pyridone scaffolds.<sup>[5b, 90]</sup> The Dieckmann condensation is an intramolecular Claisen-type reaction of diesters or similar functionalities leading to stable 5- and 6-membered heterocyclic rings.<sup>[1]</sup> In fungal PKS/NRPS hybrids a reductase-type domain performs a non-redox, NAD(P)H independent Dieckmann condensation carrying out a cyclization. Similarly, in actinomycete-derived tetramic acid and pyridone natural products from bacterial PKS/NRPS hybrids, stand-alone Dieckmann cyclases have been identified as cyclization catalysts for the biosynthesis of these compounds.<sup>[90]</sup>

### 5.1. Reductase domain catalyzed Dieckmann Condensation ( $R_D$ )

Reductase ( $R$ ) domains have been found to catalyze a non-redox Dieckmann condensation instead of a reductive release of a polyketide or a peptide, as shown previously. In this review, we will refer to these non-redox  $R$  domains as  $R_D$ . As noted previously for NRPS terminal  $R$  domains, they enable product release in an aldehyde form, which could then undergo a spontaneous cyclization. However, for these  $R_D$  domains, the release of the polyketide or peptide occurs directly via a Dieckmann condensation.<sup>[1, 5b]</sup>

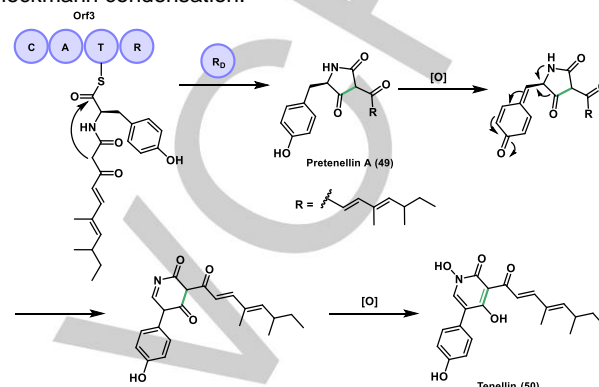
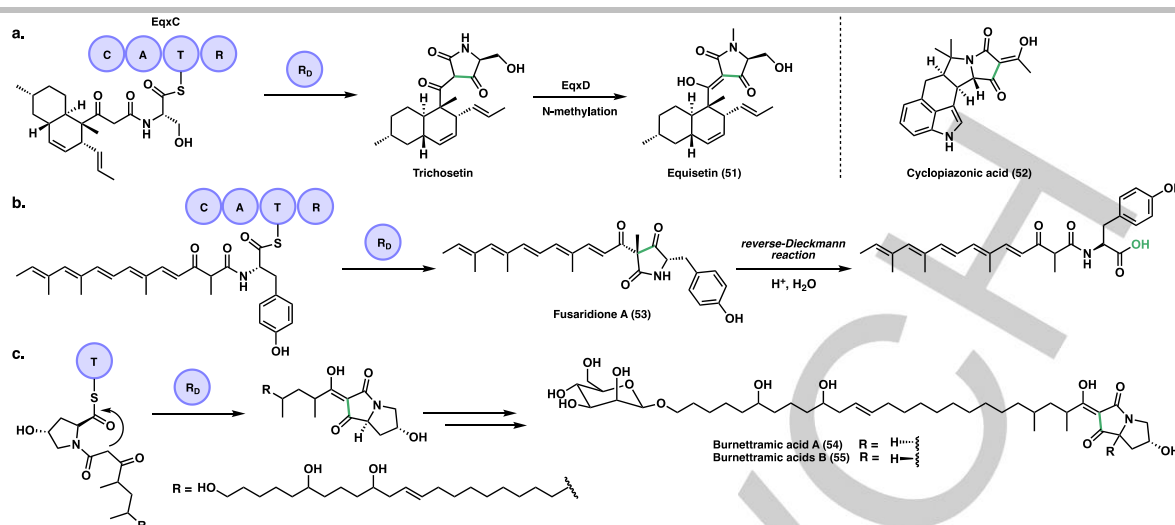


Figure 17. Dieckmann cyclization mediated by a reductase domain ( $R_D$ ) in the biosynthesis of tenellin.

Tenellin (50) is an example of this category, isolated in 1977 from the fungus *Beauveria bassiana* and produced by a PKS/NRPS hybrid.<sup>[91]</sup> This compound contains a 2-pyridone ring that derives from a tetramic acid (pyrrolidine-2,4-dione) via a ring expansion. In 2007, it was proposed that the tetramate ring was produced through an  $R$  domain catalyzed reductive release to give an aldehyde, which then cyclized non-enzymatically generating pretenellin A (49). Then, this tetramate ring would be further oxidized by tailoring enzymes to afford the 2-pyridone in tenellin.<sup>[92]</sup> A year later, the *tenS* gene that encodes the PKS/NRPS tenellin synthetase (TENS) was expressed in the heterologous host *Aspergillus oryzae* M-2-3.<sup>[93]</sup> It was determined that the terminal  $R$  domain in the tenellin biosynthesis catalyzes a Dieckmann-type cyclization to generate the tetramic acid moiety in pretenellin A, as opposed to a spontaneous cyclization via an off-loaded aldehyde as originally proposed. This intermediate requires two additional oxidations by two BGC-encoded cytochrome P450s to be converted to tenellin through the 2-pyridone (Figure 17).<sup>[94]</sup>

Additional tetramic acid containing natural products can be formed from an  $R_D$  domain such as equisetin (51) and cyclopiazonic acid (52) (Figure 18a). Equisetin was isolated from the fungus *Fusarium heterosporum*.<sup>[95]</sup> Later, it was found to be an HIV-1 integrase inhibitor and it also has shown antibacterial properties.<sup>[96]</sup> Recently, equisetin was shown to inhibit bacterial acetyl-CoA carboxylase (ACC), the first step in fatty acid synthesis.<sup>[97]</sup> This compound derives from the PKS/NRPS hybrid equisetin synthetase (EqiS) and involves an  $R_D$  domain mediated cyclization.<sup>[98]</sup> It also bears a structural similarity with the cholesterol lowering agent lovastatin.<sup>[99]</sup> In 2008, the Schmidt group overexpressed the EqiS  $R_D$  domain and studied its reactivity utilizing synthetic substrate analogs. Their results show that the EqiS  $R_D$  domain catalyzes a Dieckmann condensation rather than a reductive release.<sup>[98]</sup> Subsequently, they showed that EqiS does not produce equisetin, but instead it produces a 2,4-pyrrolidinedione, fusaridione A (53). This newly identified

## REVIEW



**Figure 18a.**  $R_D$  domain in the biosynthesis of equisetin and fusicurone A. The activity of the  $R_D$  domain results in release of the intermediate as the tetramate, trichosetin. Then, *N*-methylation is carried out to give equisetin. **b.** For the case of fusicurone A, the unstable pyrrolidinedione ring is opened through a reverse-Dieckmann reaction. **c.** Proposed biosynthetic pathway for burnettramic acids A and B.

compound leads to a spontaneous reverse-Dieckmann reaction due to the methylated 3-position of the compound scaffold (**Figure 18b**). A new gene cluster, *eqx* was described as responsible for the biosynthesis of equisetin.<sup>[100]</sup> Another example is cyclopiazonic acid, produced by the fungus *Aspergillus flavus*, which possesses nanomolar inhibitory activity against  $Ca^{2+}$ -ATPase.<sup>[101]</sup> Similarly, cyclopiazonic acid follows a similar mechanism to equisetin where the CpaS  $R_D$  domain catalyzes a Dieckmann condensation producing the tetramate ring. Interestingly, the Walsh group determined that the CpaS  $R_D$  domain lacks the Ser-Tyr-Lys catalytic triad found in the SDR superfamily.<sup>[102]</sup> Their results show a conserved aspartic acid residue (D3803), which may play an important role in the  $R_D$  domain since mutations abolished the activity.

Burnettramic acids A (**54**) and B (**55**) are a new class of antibiotics derived from a Dieckmann cyclization during their biosynthesis (**Figure 18c**). Recently, the Chooi and Piggott groups studied rare PKS/NRPS derived bolaamphiphilic pyrrolizidinediones, isolated from an Australian fungus *Aspergillus burnettii*.<sup>[103]</sup> They reported the discovery, biological investigation and biosynthetic studies of the burnettramic acids. In the proposed pathway, an  $R_D$  domain releases the polyketide-peptide via a Dieckmann condensation to produce a tetramic acid fused to the hydroxyproline, generating the bicyclic pyrrolidinedione moiety. The C-terminal domain in BuaA was confirmed in their study as a Dieckmann condensation domain, containing a phenylalanine in place of a tyrosine found in the typical catalytic triad of the short chain SDR superfamily.<sup>[102, 104]</sup>

## 5.2. Thioesterase-like Dieckmann Cyclases

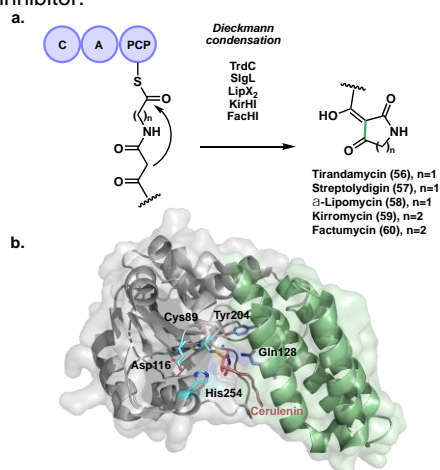
In 2015, the Ju group discovered a new family of Dieckmann cyclases related to actinomycete-derived tetramic acid and pyridone natural products (**Figure 19a**).<sup>[90a]</sup> The compounds in this study included tirandamycin B (**56**), streptolydigin (**57**),  $\alpha$ -lipomycin (**58**), kirromycin (**59**), and factumycin (**60**). Although their biosynthetic assembly lines have been identified, none have been found to contain TE or  $R$  domains presumed to produce the characteristic tetramic acid or pyridone moiety. Intrigued by this, the Ju group studied the TrdC, SlgL, LipX<sub>2</sub>, KirHI, and FacHI

group of homologous proteins. Their results showed that this highly conserved new family of enzymes catalyzes formation of tetramic acid and pyridone moieties via a Dieckmann cyclization mechanism. Further investigation into the closest homologs of TrdC revealed that it is most related to FAS, PKS and NRPS TEs. The conserved catalytic triad is also present in TrdC with the typical serine being replaced by a cysteine. Surprisingly, mutation of this catalytic cysteine to serine (C88S) retained cyclase activity. Finally, an analysis of GenBank found that these TE like Dieckmann cyclases are prevalent in bacterial secondary metabolism and may serve as effective genetic probes for identifying bacterially derived tetramic acids and 2-pyridone based natural products.

More recently, the Nair group identified a Dieckmann cyclase (NcmC) that installs the tetramate headgroup in the PKS/NRPS hybrid derived nocamycin peptide natural product.<sup>[90b]</sup> They identified a putative biosynthetic pathway for nocamycin I in *Saccharothrix syringae* NRRL B-16468. By testing a PCP-linked, and a SNAc mimic substrate, they measured NcmC activity *in vitro*, which revealed that the enzyme catalyzes a Dieckmann condensation reaction. They also reported a 1.6 Å resolution crystal structure of NcmC that provided mechanistic insights about this family of cyclases. Thus, NcmC was found to possess two subdomains, with the first consisting of the characteristic  $\alpha/\beta$  hydrolase fold found in TEs and other hydrolases (**Figure 19b**, grey). In contrast, the second subdomain consists of a four-helix bundle inserted between strands  $\beta_5$  and  $\beta_6$ , which is unique to these Dieckmann cyclases (**Figure 19b**, green). The  $\alpha/\beta$  hydrolase subdomain contains the common active site catalytic triad with a cysteine residue present instead of a serine (Cys89, Asp116 and His254), a substitution observed in previously discussed TE containing PKS pathways. Although they were unable to capture a native substrate, cerulenin was employed and formed a covalent adduct with the active site, yielding insights into other important active site residues. Identification of Tyr204 and Gln128 located on the unique four helix bundle in conjunction with mutagenesis studies demonstrated that these residues play a crucial role in catalysis. Mutations at both positions showed a marked increase in hydrolytic activity and a corresponding decrease in cyclization. These results coupled with the

## REVIEW

crystallographic analysis led the authors to hypothesize that Gln128 (positioned next to the thioester bond) is responsible for shielding the thioester intermediate from water mediated hydrolysis. The Tyr204 residue appears to be highly conserved among similar proteins, and likely participates in stabilizing the enol tautomer necessary for nucleophilic attack and cyclization. This is supported by the hydrogen bonding network observed between the tyrosine side chain, a water molecule, and the cerulenin inhibitor.



**Figure 19.** Activity and structure of Dieckmann cyclases. **a.** Dieckmann cyclization reaction in the biosynthesis of five actinomycete-derived tetramic acid and pyridone natural products. **b.** NcmC-cerulenin complex with the  $\alpha/\beta$  hydrolase subdomain highlighted in grey and the four-helix bundle in green. Catalytic residues important for function are depicted in blue and purple.

## 5. Cyclization by Other Mechanisms

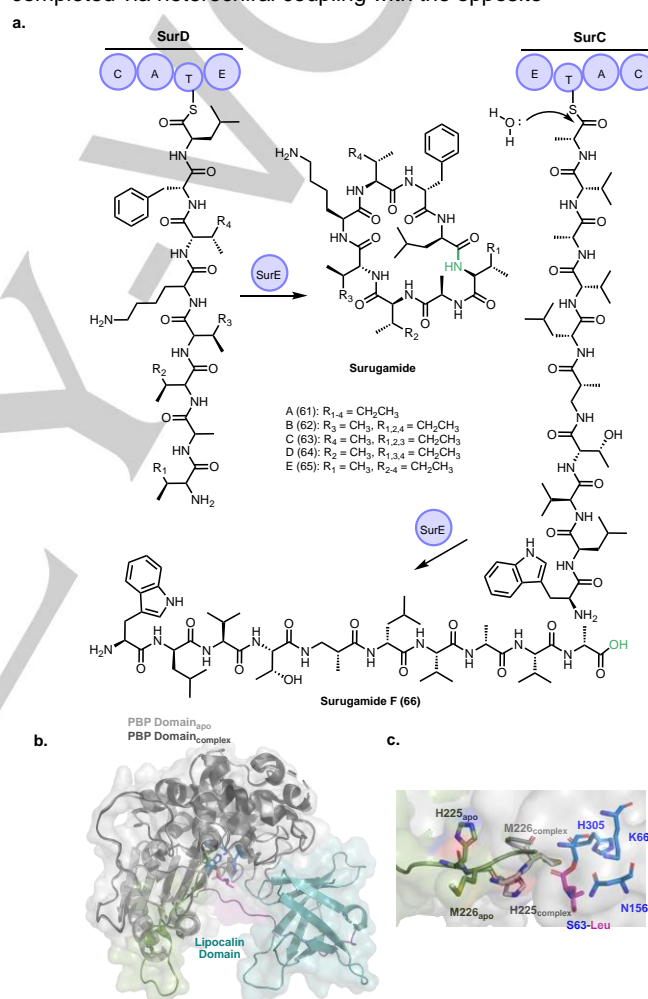
### 5.1. Penicillin Binding Protein (PBP) Cyclases

Surugamides A-E (**61-65**) are cyclic octapeptides with four D-amino acids originally isolated from the marine-derived actinomycete *Streptomyces* sp. JAMM992. These compounds have been shown to have cathepsin B inhibitory bioactivity.<sup>[105]</sup> Sequencing and genome mining identified the biosynthetic gene cluster responsible for the production of surugamides.<sup>[106]</sup> The authors reported a putative gene cluster containing four successive NRPS genes, *surA*, *surB*, *surC* and *surD*, with the *surA* and *surD* being responsible for the assembly of surugamides A-E (**61-65**) as well as *surB* and *surC* required to produce surugamide F (**66**), an unrelated linear decapeptide (**Figure 20a**).<sup>[106]</sup> Interestingly, these biosynthetic studies demonstrated a lack of chain termination domain such as a TE, R, or  $C_T$  encoded by the *surD* or *surC* genes. Hence, the enzyme responsible for the chain termination remained unknown.

In 2018, the Wakimoto group showed the total synthesis of surugamide B (**62**) and identified the chain termination enzyme.<sup>[107]</sup> Their work describes SurE as the candidate offloading enzyme, which is encoded just upstream of the NRPS genes and was originally annotated as a penicillin-binding protein (PBP).<sup>[108]</sup> Utilizing a recombinant SurE protein and SNAc chain elongation intermediate to mimicking the PCP-bound octapeptide substrate, the authors demonstrated not only that SurE catalyzes peptide cyclization *in vitro*, but also that it appeared to exhibit a preference for substrates bearing a C-terminal D-amino acid, contrary to the L orientation usually preferred by TEs. Further

investigation including *in vivo* knockouts of the *surE* gene and testing with alternative substrates revealed that SurE is catalyzes offloading of the *surA/surD* specified products **61-65**, and also the *surB/surC* product **66**.<sup>[101, 102]</sup> Notably, unlike TEs that are appended to the terminal module of NRPS modules, these PBP like TEs are stand-alone proteins that act *in-trans* to offload as either cyclic or hydrolyzed products.

Recently, the Abe and Wakimoto groups reported interesting results about the substrate scope of SurE using synthetic substrate derivatives.<sup>[109]</sup> Their work confirmed a strict requirement for heterochirality between N- and C-terminal residues with L- and D-amino acids, respectively. They also revealed that macrolactamization by canonical *cis*-acting offloading domains such as type I TE and  $C_T$  domains is completed via heterochiral coupling with the opposite



**Figure 20a.** The fully elongated peptide chains tethered to the last NRPS modules are presumed to be released by SurE to generate cyclopeptides **61-65** and the linear peptide **66**.<sup>[110]</sup> **b.** Overall Structure of SurE with PBP domain depicted in light grey for the apo form and dark grey for the complexed form, lipocalin like domain in cyan and the linker between the two in purple. Active site residues are depicted in blue with the density for the visible leucine in pink. The loop rearrangement is highlighted in green for the ordered apo form. **c.** Active site of SurE including comparison of loop region His225 and Met226 in apo (green) and complex structure with D-Leu (grey).

configuration (D-amino acid at the N-terminus and L-amino acid at the C-terminus). Based on these findings, PBP-type TEs could be employed as a strategy for creating alternative heterochirality in the biosynthesis of cyclopeptides.



## REVIEW

In order to gain mechanistic insights, the authors also disclosed a 2.2 Å resolution X-ray crystal structure for this enzyme.<sup>[109]</sup> This work revealed two subdomains linked by a long disordered loop region: an N-terminal PBP domain containing the typical  $\alpha/\beta$  hydrolase fold seen in TEs and a C-terminal lipocalin like domain comprised of  $\beta$  strands. This C-terminal domain acts as a lid on the PBP subdomain, forming a cleft that houses the active site tetrad Ser63, Lys66, Asn156 and His305. To understand the basis for the unusual C-terminal selectivity, a model of the peptide-O-SurE complex using the residues located within 8 Å of the active-site serine (Ser63) was calculated. Their data suggests that in SurE the D-stereoselectivity in the C-terminus is governed by the oxyanion hole, a side chain-recognizing hydrophobic pocket (Leu231, Ala234, Gly235 and Val309), and the carbonyl-recognizing Arg446. In addition, substrate recognition studies were conducted using the crystal of *apo*-SurE protein soaked with a substrate that cannot be cyclized, which enabled capture of a peptide-O-SurE complex. Density for a single amino acid was well defined (with the rest weakly visible) attached to the active site serine. Analysis of the complex vs the *apo* structures revealed a conformational rearrangement of a flexible loop (**Figure 20b**, green loop) induced by the formation of the peptidyl-O-SurE complex due to interactions with the tethered peptide. This rearrangement however, caused a shift in His225 and Met226 (**Figure 20c**), which now occupy the presumed hydrophobic pocket that was identified in the simulations, contradicting that original hypothesis. Although the exact mechanism for the C-terminal recognition remains elusive, these studies overall show the potential of SurE has a compelling candidate for synthetic biology applications and developing new methodology towards the biocatalytic production of NRPS based macrocycles.<sup>[109]</sup>

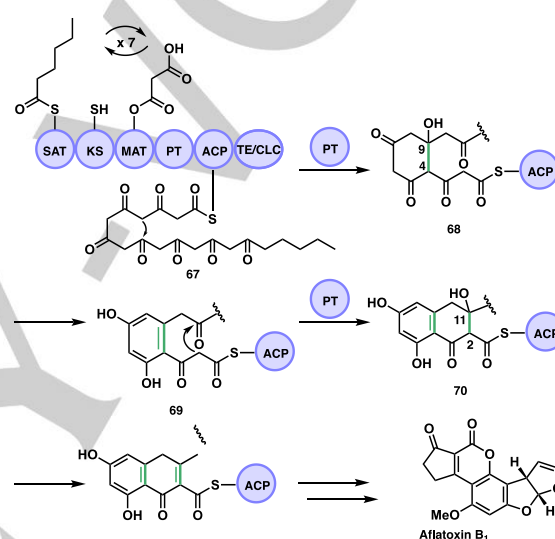
## 5.2. Product Template Domains (PTs) in Fungi

The PT domain in fungal NR-PKS controls the aldol cyclization of poly- $\beta$ -ketone intermediates. These oligoketides are assembled upstream in the  $\beta$ -ketoacyl synthase (KS) catalyzing intramolecular closure to cyclic and fused polycyclic compounds.<sup>[111]</sup> In 2008, the Kelleher and Townsend groups showed the PT domain present in the NR-PKS PksA from the aflatoxin B<sub>1</sub> biosynthesis in *Aspergillus parasiticus*. Their results demonstrated that a PT domain unites with the KS and TE in the iterative PKS to assemble precisely seven malonyl-derived building blocks to a hexanoyl starter unit and mediates a specific cyclization cascade (**Figure 21**).<sup>[50a]</sup>

In 2009, the Tsai and Townsend groups solved the crystal structure of the PT domain at 1.8 Å resolution and also conducted mutational studies in the PT domain from PksA.<sup>[112]</sup> They found that the PT structure displays a "double hot dog" (DHD) fold and differs from those observed in the dehydratase (DH) domains in animal FASs and bacterial PKSs.<sup>[113]</sup> The DHD fold controls cyclization specificity, which is one of the key programmed steps in aromatic-polyketide biosynthesis.<sup>[50a, 114]</sup> Additional data provided some insights on the possible origins of the PT domains. Interestingly, native PAGE analysis of the oligomerization states coupled with previous phylogenetic results<sup>[115]</sup> suggested that the dimeric PT domains evolved from ancient DH domains in reducing PKSs and would be positioned at the DH locus in animal FAS.

Furthermore, docking and mutagenesis studies in the PT domain allowed identification of important residues for substrate

binding and catalysis. It also identified the presence of a phosphopantetheine localization channel and a deep two-part interior binding pocket and reaction chamber. The proposed mechanism in this study suggested the ACP-bound linear substrate **67** binds to the PT substrate pocket in an extended conformation. His1345 acting together with Asp1543 (hydrogen bonded to Gln1547) functions as the key catalytic base deprotonating C4, leading to the enolate intermediate stabilized by the backbone N-H of Asn1568. The enolate subsequently collapses followed by C4 aldol addition to carbonyl C9 (C4-C9 cyclization event) leading to the first ring cyclization product **68** with a network of water molecules bound to Ser1356, Asp1543, Asn1568 and Thr1546 stabilizing the oxyanion. The second ring is formed in the same manner as the first to achieve C2-C11 closure in **70** (**Figure 21**).

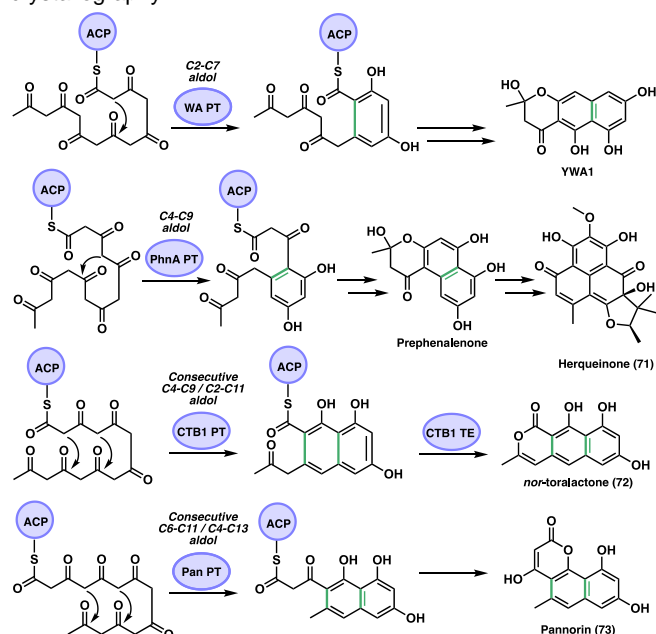


**Figure 21.** PT domain in the NR-PKS PksA for the biosynthesis of aflatoxin B<sub>1</sub>.

In 2017, Townsend, Burkart and Tsai provided new insights about the PT domain function through crystallographic applications of "atom replacement" mimetics.<sup>[111]</sup> Isoxazole rings linked by thioethers were employed to mimic the alternating sites of carbonyls in the poly- $\beta$ -ketone intermediates. In their work, they utilized 4'-phosphorylated and unphosphorylated atom-replaced mimetics, which were selected to represent substrate chain lengths of 8-16 carbon units of the linear and proposed monocyclic intermediates from aflatoxin biosynthesis. A 1.8 Å resolution co-crystal structure of the PksA PT domain was presented with a 4'-phosphopantetheinylated linear heptaketide mimetic. In summary, using structural biology, structure-based mutagenesis, and enzymatic assays the collaborative team identified key residues involved in substrate recognition that have been difficult to identify due to oligoketide substrate instability and inherent reactivity. They observed that the heptaketide C4 carbon is positioned proximal to the catalytic His1345, providing further evidence to the previously proposed mechanism of polyketide cyclization mediated by the PT domain. Also, the role of the protein-coordinated water network to selectively activate the C9 carbonyl for nucleophilic addition was shown as well as the importance of the 4'-phosphate distal to the pantetheine arm to deliver and anchor the heptaketide mimetic into the PT active site. Moreover, atom-replacement mimetics proved to be an alternative way to probe other PKS enzymes (e.g. bacterial type

## REVIEW

II PKS) together with highly reactive substrates/intermediates, which are challenging or impossible to capture using X-ray crystallography.



**Figure 22.** PT domain cyclization regioselectivities in NR-PKSs.

Another interesting study of the fungal PKS PT domains was reported by the Houk and Tang groups.<sup>[50b]</sup> They investigated the biosynthesis of herqueinone (**71**) isolated from *Penicillium herquei*. This compound forms part of the phenalenone family of natural products that contains a peri-fused tricyclic ring core system cyclized from a linear polyketide precursor. **71** is formed after the phenalenone is heavily derivatized and oxidized. The identified gene cluster is *phn* from *P. herquei* containing an encoded NR-PKS (PhnA), that contains SAT-KS-MAT-PT-ACP-ACP-TE/CLC domains. Additional genes encode a flavin-dependent monooxygenase (FMO, PhnB and PhnG), an O-methyltransferase (O-MT, PhnC), and a prenyltransferase (PrT, PhnF). In their work, the Houk and Tang team demonstrated that the NR-PKS PhnA is responsible for the synthesis of a heptaketide backbone as well as for its cyclization into a hemiketal-containing naphtho- $\gamma$ -pyrone prephenalenone. They also found that the FMO (PhnB) catalyzes a C2 aromatic hydroxylation of the prephenalenone NR-PKS product converting it into phenalenone followed by a simultaneous ring opening of the  $\gamma$ -pyrone ring. Density functional theory (DFT) calculations were presented that supported their mechanistic hypothesis. In addition, they compared prephenalenone with different NR-PKS shown to synthesize the common heptaketide precursor that is cyclized into diverse final products due to PT domain regioselectivity (**Figure 22**). The NR-PKS derived naphthopyrone products in the analysis include 1) YWA1 from *Aspergillus nidulans* *wA*<sup>[52a, 116]</sup> (WA, NR-PKS) 2) *nor*-toralactone (**72**) from *Cercospora nicotianae*<sup>[117]</sup> (CTB1, NR-PKS) and 3) pannorin (**73**) from *Chrysosporium pannorum*<sup>[118]</sup> M10539 (Pan, NR-PKS). PT catalyzed cyclization of oligoketide substrates has been demonstrated to perform consecutive C4-C9 and C2-C11 intramolecular aldol condensations.<sup>[50a, 117a]</sup> In addition, they showed that the PT domain in PhnA catalyzes only the C4-C9 aldol condensation, while the TE/CLC catalyzes a C1-C10

cyclization. Both PhnA and WA produce a heptaketide compound and catalyze an intramolecular aldol condensation and a C1-C10 Claisen cyclization generating the corresponding naphthopyrone products. A difference between them is the PT regioselectivity, which for the case of WA a C2-C7 cyclization leads to YWA1 and for PhnA a C4-C9 cyclization results in prephenalenone. In the CTB1 system, the PT domain catalyzes tandem C4-C9 and C2-C11 aldol cyclization reactions that produce *nor*-toralactone (**72**),<sup>[117a]</sup> which is similar to the PT domain in PksA described above.<sup>[112]</sup> Lastly, the NR-PKS for pannorin (**73**) has not been identified, however, the PT domain has been suggested to catalyze consecutive C6-C11 and C4-C13 cyclization reactions.<sup>[119]</sup> In general, the PT domain in the NR-PKS PhnA is functionally distinct since it catalyzes only a C4-C9 cyclization compared to the corresponding PT homologs that catalyze C4-C9 or C2-C11 cyclizations.

## 6. Summary and Outlook

In this review we describe the fascinating work reported on peptide and polyketide biosynthetic cyclization catalysts during the past decade. We presented several examples of new cyclic structures with interesting biological activities and summarized the recent findings on the diversity of enzymes that catalyze cyclization of linear peptides or polyketide chains. These enzymes possess structural and mechanistic diversity including cyclization mediated by thioesterase (TE) domains, oxidative, reductive, condensation (C<sub>T</sub> domains and Dieckmann cyclases) and other mechanisms.

Initial attention was focused on thioesterases (TEs), which have been studied extensively throughout the past two - three decades. Even though TEs are relatively well understood in comparison with other cyclization catalysts, additional NRPS and PKS BGCs were recently characterized, showing the production of cyclic lactone and lactam structures mediated by a TE domain.<sup>[15a, 16, 18, 21-22, 24]</sup> Examples of these TE cyclases are derived from obafuorin, salinamide A, lysobactin, sulfazecin, polymyxin B and fluvirucin B<sub>1</sub>. In some cases, a cysteine instead of a serine active site residue was found<sup>[15a, 16, 22, 24]</sup>; in addition to the unusual tandem TE architecture employed by lysobactin and teixobactin biosynthesis.<sup>[21, 31]</sup> Furthermore, chemoenzymatic syntheses<sup>[26, 31, 33-36]</sup> and pathway engineering in these systems<sup>[37-40]</sup> represent leading examples for the synthesis of novel compounds. TE domain oligomerization<sup>[43-44, 47-48, 120]</sup> and C-C bond forming Claisen-Like Cyclases<sup>[15b, 50, 53]</sup> (CLCs) were also described, which reflect the mechanistic diversity of these domains. The remarkable achievements toward understanding these systems show the potential of these cyclization biocatalysts. In the future, mechanistic, structural insights (X-ray crystallography and cryo-electron microscopy) and protein engineering will guide the path for deeper exploration of the capabilities of the variant TEs for the synthesis of numerous natural products and bioactive, high value analogs.

In addition to canonical TE domains, we also discussed functionally similar enzymes that perform a cyclization via a condensation mechanism (C<sub>T</sub> domains and R domain catalyzed Dieckmann condensation) or the release of a linear chain substrate as an aldehyde (terminal R domains), followed by spontaneous cyclization. For the fungal NRPS C<sub>T</sub> domains, phylogenetic analysis, mechanistic and structural investigations

## REVIEW

have provided significant information about these terminal domains.<sup>[56b, 59, 62c, 66, 70]</sup> However, only a few reports are available on C<sub>T</sub> domains, suggesting that further studies are needed on fungal NRPSs to explore their full biocatalytic potential. For the case of terminal R domains, remarkable work has been reported on these enzymes. However, it would be interesting to understand mechanistically how the initial product is compelled to cyclize upon reductive release,<sup>[5b, 74, 79b, 81, 87, 89]</sup> and similarly the initial product of a non-redox Dieckmann condensation<sup>[90a, 94, 98, 102-103]</sup> leading to cyclization. Structural elucidation of NRPS R domains could further elucidate linear off-loading or cyclization routes for these homologous SDR enzymes. Lastly, cyclization mediated by other mechanisms were also described. The PBP cyclases and product template (P<sub>T</sub>) domains are a pronounced example of additional enzymes in NRPS and PKS systems with cyclization chemistry. Recent crystal structures and mechanistic studies have demonstrated the impressive strategies resulting in cyclization of secondary metabolites.<sup>[50b, 109, 111-112]</sup> Future approaches using genome mining for the discovery of new biocatalysts, as well as protein engineering efforts on NRPS and PKS mega-enzyme systems are ripe with potential for the production of pharmaceutically relevant peptide and polyketide derived natural products.

## Acknowledgements

We gratefully acknowledge NIH grant R35 GM118101, and the Hans W. Vahlteich Professorship (D.H.S.). We also thank Rajani Arora for the design of our frontispiece.

**Keywords:** Enzymes • Natural Products • Biocatalysis • Polyketide Synthase • Nonribosomal Peptide Synthetase

- [1] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255-278.
- [2] G. J. Williams, *Curr. Opin. Struct. Biol.* **2013**, *23*, 603-612.
- [3] S. Hwang, N. Lee, S. Cho, B. Palsson, B. K. Cho, *Front. Mol. Biosci.* **2020**, *7*.
- [4] M. Z. Ansari, G. Yadav, R. S. Gokhale, D. Mohanty, *Nucleic Acids Res.* **2004**, *32*, W405-W413.
- [5] a) M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Nat. Prod. Rep.* **2016**, *33*, 183-202; b) M. W. Mallowney, R. A. McClure, M. T. Robey, N. L. Kelleher, R. J. Thomson, *Nat. Prod. Rep.* **2018**, *35*, 847-878; c) M. C. Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang, *Chem. Rev.* **2017**, *117*, 5226-5333.
- [6] K. J. Weissman, *Nat. Prod. Rep.* **2016**, *33*, 203-230.
- [7] D. C. Cantu, Y. Chen, P. J. Reilly, *Protein Sci* **2010**, *19*, 1281-1295.
- [8] M. Kotowska, K. Pawlik, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7735-7746.
- [9] a) S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure* **2002**, *10*, 301-310; b) B. R. Miller, A. M. Gulick, in *Nonribosomal Peptide and Polyketide Biosynthesis: Methods in Molecular Biology*, Vol. 1401 (Ed.: B. Evans), Humana Press, New York, **2016**, pp. 3-29.
- [10] P. J. Loll, E. C. Upton, V. Nahoum, N. J. Economou, S. Cocklin, *Biochim Biophys Acta* **2014**, *1838*, 1199-1207.
- [11] S. A. Samel, B. Wagner, M. A. Marahiel, L. O. Essen, *J Mol Biol* **2006**, *359*, 876-889.
- [12] a) S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 14808-14813; b) S. C. Tsai, H. Lu, D. E. Cane, C. Khosla, R. M. Stroud, *Biochemistry* **2002**, *41*, 12598-12606.
- [13] D. L. Akey, J. D. Kittendorf, J. W. Giraldez, R. A. Fecik, D. H. Sherman, J. L. Smith, *Nat. Chem. Biol.* **2006**, *2*, 537-542.
- [14] J. W. Trauger, R. M. Kohli, C. T. Walsh, *Biochemistry* **2001**, *40*, 7092-7098.
- [15] a) C. A. Galea, K. D. Roberts, Y. Zhu, P. E. Thompson, J. Li, T. Velkov, *Biochemistry* **2017**, *56*, 657-668; b) T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 6246-6251.
- [16] J. E. Schaffer, M. R. Reck, N. K. Prasad, T. A. Wenciewicz, *Nat. Chem. Biol.* **2017**, *13*, 737-744.
- [17] a) J. S. Wells, W. H. Trejo, P. A. Principe, R. B. Sykes, *J. Antibiot. (Tokyo)* **1984**, *37*, 802-803; b) A. A. Tymiak, C. A. Culver, M. F. Malley, J. Z. Gougoutas, *J. Org. Chem.* **1985**, *50*, 5491-5495.
- [18] L. Ray, K. Yamanaka, B. S. Moore, *Angew. Chem. Int. Ed.* **2016**, *55*, 364-367.
- [19] J. A. Trischman, D. M. Tapiolas, P. R. Jensen, R. Dwight, W. Fenical, T. C. Mckee, C. M. Ireland, T. J. Stout, J. Clardy, *J. Am. Chem. Soc.* **1994**, *116*, 757-758.
- [20] a) D. P. Bonner, J. Osullivan, S. K. Tanaka, J. M. Clark, R. R. Whitney, *J. Antibiot. (Tokyo)* **1988**, *41*, 1745-1751; b) J. Osullivan, J. E. McCullough, A. A. Tymiak, D. R. Kirsch, W. H. Trejo, P. A. Principe, *J. Antibiot. (Tokyo)* **1988**, *41*, 1740-1744.
- [21] J. Hou, L. Robbel, M. A. Marahiel, *Chem. Biol.* **2011**, *18*, 655-664.
- [22] R. A. Oliver, R. F. Li, C. A. Townsend, *Nat. Chem. Biol.* **2018**, *14*, 5-7.
- [23] M. Asai, K. Haibara, M. Muroi, K. Kintaka, T. Kishi, *J. Antibiot. (Tokyo)* **1981**, *34*, 621-627.
- [24] T. Y. Lin, L. S. Borkekey, G. Prasad, S. A. Waters, N. A. Schnarr, *ACS Synth. Biol.* **2013**, *2*, 635-642.
- [25] a) N. Naruse, M. Konishi, T. Oki, Y. Inouye, H. Kakisawa, *J. Antibiot. (Tokyo)* **1991**, *44*, 756-761; b) N. Naruse, O. Tenmyo, K. Kawano, K. Tomita, N. Ohgusa, T. Miyaki, M. Konishi, T. Oki, *J. Antibiot. (Tokyo)* **1991**, *44*, 733-740; c) N. Naruse, T. Tsuno, Y. Sawada, M. Konishi, T. Oki, *J. Antibiot. (Tokyo)* **1991**, *44*, 741-755.
- [26] J. J. Schmidt, Y. Khatri, S. I. Brody, C. Zhu, H. Pietraszkiewicz, F. A. Valeriote, D. H. Sherman, *ACS Chem. Biol.* **2020**, *15*, 524-532.
- [27] R. E. Schwartz, C. F. Hirsch, D. F. Sesin, J. E. Flor, M. Chartrain, R. E. Fromtling, G. H. Harris, M. J. Salvatore, J. M. Liesch, K. Yudin, *J. Ind. Microbiol.* **1990**, *5*, 113-123.
- [28] C. D. Smith, X. Q. Zhang, S. L. Mooberry, G. M. L. Patterson, R. E. Moore, *Cancer Res.* **1994**, *54*, 3779-3784.
- [29] N. A. Magarvey, Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore, D. H. Sherman, *ACS Chem. Biol.* **2006**, *1*, 766-779.
- [30] V. A. Verma, T. H. Pillow, L. DePalatis, G. Li, G. L. Phillips, A. G. Polson, H. E. Raab, S. Spencer, B. Zheng, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 864-868.
- [31] D. Mandalapu, X. Ji, J. Chen, C. Guo, W. Q. Liu, W. Ding, J. Zhou, Q. Zhang, *J. Org. Chem.* **2018**, *83*, 7271-7275.
- [32] L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schaberle, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. A. Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen, K. Lewis, *Nature* **2015**, *517*, 455-459.
- [33] D. A. Hansen, C. M. Rath, E. B. Eisman, A. R. Narayan, J. D. Kittendorf, J. D. Morrison, Y. J. Yoon, D. H. Sherman, *J. Am. Chem. Soc.* **2013**, *135*, 11232-11238.
- [34] D. A. Hansen, A. A. Koch, D. H. Sherman, *J. Am. Chem. Soc.* **2017**, *139*, 13450-13455.
- [35] G. W. Heberlig, M. Wirz, M. Wang, C. N. Boddy, *Org. Lett.* **2014**, *16*, 5858-5861.
- [36] G. W. Heberlig, J. T. C. Brown, R. D. Simard, M. Wirz, W. Zhang, M. Wang, L. I. Susser, M. E. Horsman, C. N. Boddy, *Org. Biomol. Chem.* **2018**, *16*, 5771-5779.

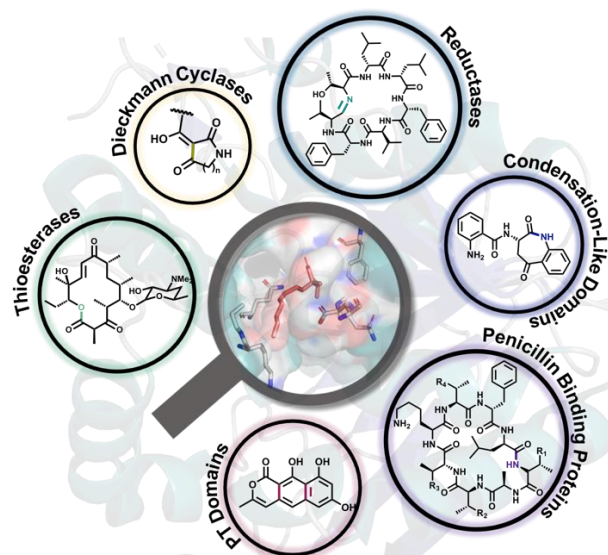
## REVIEW

- [37] A. A. Koch, D. A. Hansen, V. V. Shende, L. R. Furan, K. N. Houk, G. Jimenez-Oses, D. H. Sherman, *J. Am. Chem. Soc.* **2017**, *139*, 13456-13465.
- [38] A. A. Koch, J. J. Schmidt, A. N. Lowell, D. A. Hansen, K. M. Coburn, J. A. Chemler, D. H. Sherman, *Angew. Chem. Int. Ed.* **2020**, *59*, 13575-13580.
- [39] L. Qiao, J. Fang, P. Zhu, H. Huang, C. Dang, J. Pang, W. Gao, X. Qiu, L. Huang, Y. Li, *Protein J.* **2019**, *38*, 658-666.
- [40] A. Tripathi, S. S. Choi, D. H. Sherman, E. S. Kim, *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 1189-1193.
- [41] A. M. Gehring, I. Mori, C. T. Walsh, *Biochemistry* **1998**, *37*, 2648-2659.
- [42] K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem Biol* **2007**, *14*, 13-22.
- [43] N. Huguenin-Dezot, D. A. Alonzo, G. W. Heberlig, M. Mahesh, D. P. Nguyen, M. H. Doman, C. N. Boddy, T. M. Schmeing, J. W. Chin, *Nature* **2019**, *565*, 112-117.
- [44] G. W. Heberlig, C. N. Boddy, *J. Nat. Prod.* **2020**, *83*, 1990-1997.
- [45] a) N. Agata, M. Mori, M. Ohta, S. Suwan, I. Ohtani, M. Isobe, *FEMS Microbiol. Lett.* **1994**, *121*, 31-34; b) N. Agata, M. Ohta, M. Mori, M. Isobe, *FEMS Microbiol. Lett.* **1995**, *129*, 17-19.
- [46] S. Marxen, T. D. Stark, A. Rutschle, G. Lucking, E. Frenzel, S. Scherer, M. Ehling-Schulz, T. Hofmann, *Sci. Rep.* **2015**, *5*.
- [47] Y. J. Zhou, P. Prediger, L. C. Dias, A. C. Murphy, P. F. Leadlay, *Angew. Chem.* **2015**, *54*, 5232-5235.
- [48] Y. J. Zhou, A. C. Murphy, M. Samborsky, P. Prediger, L. C. Dias, P. F. Leadlay, *Chem. Biol.* **2015**, *22*, 745-754.
- [49] J. W. Westley, C. M. Liu, R. H. Evans, J. F. Blount, *J. Antibiot. (Tokyo)* **1979**, *32*, 874-877.
- [50] a) J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246; b) S. S. Gao, A. Duan, W. Xu, P. Yu, L. Hang, K. N. Houk, Y. Tang, *J. Am. Chem. Soc.* **2016**, *138*, 4249-4259.
- [51] H. Zhou, Y. Li, Y. Tang, *Nat. Prod. Rep.* **2010**, *27*, 839-868.
- [52] a) I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem. Biol.* **2001**, *8*, 189-197; b) T. Awakawa, K. Yokota, N. Funo, F. Doi, N. Mori, H. Watanabe, S. Horinouchi, *Chem. Biol.* **2009**, *16*, 613-623.
- [53] A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem. Biol.* **2012**, *19*, 1525-1534.
- [54] Y. Takano, Y. Kubo, K. Shimizu, K. Mise, T. Okuno, I. Furusawa, *Mol. Gen. Genet.* **1995**, *249*, 162-167.
- [55] a) I. Fujii, Y. Mori, A. Watanabe, Y. Kubo, G. Tsuji, Y. Ebizuka, *Biochemistry* **2000**, *39*, 8853-8858; b) I. Fujii, Y. Mori, A. Watanabe, Y. Kubo, G. Tsuji, Y. Ebizuka, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1445-1452.
- [56] a) M. McErlean, J. Overbay, S. Van Lanen, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 493-513; b) X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh, Y. Tang, *Nat. Chem. Biol.* **2012**, *8*, 823-830; c) S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715-738; d) M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468-3496.
- [57] a) D. Berry, W. Mace, K. Grage, F. Wesche, S. Gore, C. L. Schardl, C. A. Young, P. P. Dijkwel, A. Leuchtmann, H. B. Bode, B. Scott, *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 25614-25623; b) F. Kopp, M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735-749; c) H. von Döhren, *Fungal Genet. Biol.* **2009**, *46 Suppl 1*, S45-S52.
- [58] S. W. Haynes, X. Gao, Y. Tang, C. T. Walsh, *J. Am. Chem. Soc.* **2012**, *134*, 17444-17447.
- [59] J. Zhang, N. Liu, R. A. Cacho, Z. Gong, Z. Liu, W. Qin, C. Tang, Y. Tang, J. Zhou, *Nat. Chem. Biol.* **2016**, *12*, 1001-1003.
- [60] a) V. Bergendahl, U. Linne, M. A. Marahiel, *Eur. J. Biochem.* **2002**, *269*, 620-629; b) T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nat. Struct. Biol.* **2002**, *9*, 522-526.
- [61] a) C. B. Cui, H. Kakeya, G. Okada, R. Onose, H. Osada, *J. Antibiot. (Tokyo)* **1996**, *49*, 527-533; b) Y. Wang, Z. L. Li, J. Bai, L. M. Zhang, X. Wu, L. Zhang, Y. H. Pei, Y. K. Jing, H. M. Hua, *Chem. Biodivers.* **2012**, *9*, 385-393; c) D. H. Scharf, A. A. Brakhage, P. K. Mukherjee, *Environ. Microbiol.* **2016**, *18*, 1096-1109; d) X. Wang, Y. Li, X. Zhang, D. Lai, L. Zhou, *Molecules* **2017**, *22*, 2026; e) D. K. Nilov, K. I. Yashina, I. V. Gushchina, A. L. Zakharenko, M. V. Sukhanova, O. I. Lavrik, V. K. Svedas, *Biochemistry (Moscow)* **2018**, *83*, 152-158.
- [62] a) D. H. Scharf, T. Heinekamp, N. Remme, P. Hortschansky, A. A. Brakhage, C. Hertweck, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 467-472; b) C. J. Balibar, C. T. Walsh, *Biochemistry* **2006**, *45*, 15029-15038; c) J. A. Baccile, H. H. Le, B. T. Pfannenstiel, J. W. Bok, C. Gomez, E. Brandenburger, D. Hoffmeister, N. P. Keller, F. C. Schroeder, *Angew. Chem.* **2019**, *58*, 14589-14593.
- [63] R. A. Cramer, Jr., M. P. Gamcsik, R. M. Brooking, L. K. Najvar, W. R. Kirkpatrick, T. F. Patterson, C. J. Balibar, J. R. Graybill, J. R. Perfect, S. N. Abraham, W. J. Steinbach, *Eukaryot. Cell* **2006**, *5*, 972-980.
- [64] a) D. H. Scharf, N. Remme, A. Habel, P. Chankhamjon, K. Scherlach, T. Heinekamp, P. Hortschansky, A. A. Brakhage, C. Hertweck, *J. Am. Chem. Soc.* **2011**, *133*, 12322-12325; b) D. H. Scharf, P. Chankhamjon, K. Scherlach, T. Heinekamp, K. Willing, A. A. Brakhage, C. Hertweck, *Angew. Chem. Int. Ed.* **2013**, *52*, 11092-11095; c) S. K. Dolan, G. O'Keefe, G. W. Jones, S. Doyle, *Trends Microbiol.* **2015**, *23*, 419-428; d) S. L. Chang, Y. M. Chiang, H. H. Yeh, T. K. Wu, C. C. Wang, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2155-2157.
- [65] a) A. Chiaroni, L. Randriambola, C. Riche, H. P. Husson, *J. Am. Chem. Soc.* **1980**, *102*, 5920-5921; b) X. L. Hu, X. Q. Bian, X. Wu, J. Y. Li, H. M. Hua, Y. H. Pei, A. H. Han, J. Bai, *Tetrahedron Lett.* **2014**, *55*, 3864-3867; c) M. Kitajima, M. Nakazawa, Y. Q. Wu, N. Kogure, R. P. Zhang, H. Takayama, *Tetrahedron* **2016**, *72*, 6692-6696.
- [66] a) H. Li, C. L. M. Gilchrist, C. S. Phan, H. J. Lacey, D. Vuong, S. A. Moggach, E. Lacey, A. M. Piggott, Y. H. Chooi, *J. Am. Chem. Soc.* **2020**, *142*, 7145-7152; b) H. J. Lacey, C. L. M. Gilchrist, A. Crombie, J. A. Kalaitzis, D. Vuong, P. J. Rutledge, P. Turner, J. I. Pitt, E. Lacey, Y. H. Chooi, A. M. Piggott, *Beilstein J. Org. Chem.* **2019**, *15*, 2631-2643.
- [67] H. H. Sun, C. J. Barrow, D. M. Sedlock, A. M. Gillum, R. Cooper, *J. Antibiot. (Tokyo)* **1994**, *47*, 515-522.
- [68] K. D. Clevenger, J. W. Bok, R. Ye, G. P. Miley, M. H. Verdant, T. Velk, C. Chen, K. Yang, M. T. Robey, P. Gao, M. Lamprecht, P. M. Thomas, M. N. Islam, J. M. Palmer, C. C. Wu, N. P. Keller, N. L. Kelleher, *Nat. Chem. Biol.* **2017**, *13*, 895-901.
- [69] J. W. Bok, R. Ye, K. D. Clevenger, D. Mead, M. Wagner, A. Krerowicz, J. C. Albright, A. W. Goering, P. M. Thomas, N. L. Kelleher, N. P. Keller, C. C. Wu, *BMC Genomics* **2015**, *16*.
- [70] K. D. Clevenger, R. Ye, J. W. Bok, P. M. Thomas, M. N. Islam, G. P. Miley, M. T. Robey, C. Chen, K. Yang, M. Swyers, E. Wu, P. Gao, C. C. Wu, N. P. Keller, N. L. Kelleher, *Biochemistry* **2018**, *57*, 3237-3243.
- [71] D. E. Ehmann, A. M. Gehring, C. T. Walsh, *Biochemistry* **1999**, *38*, 6171-6177.
- [72] K. L. Kavanagh, H. Jornvall, B. Persson, U. Oppermann, *Cell. Mol. Life Sci.* **2008**, *65*, 3895-3906.
- [73] a) K. Kavanagh, H. Jornvall, B. Persson, U. Oppermann, *Cell. Mol. Life Sci.* **2008**, *65*, 3895-3906; b) L. Selles Vidal, C. L. Kelly, P. M. Mordaka, J. T. Heap, *Biochim. Biophys. Acta Proteins Proteomics* **2018**, *1866*, 327-347.
- [74] M. Wang, Z. Xie, S. Tang, E. L. Chang, Y. Tang, Z. Guo, Y. Cui, B. Wu, T. Ye, Y. Chen, *Org. Lett.* **2020**, *22*, 960-964.
- [75] a) P. M. Joyner, J. Liu, Z. Zhang, J. Merritt, F. Qi, R. H. Cichewicz, *Org. Biomol. Chem.* **2010**, *8*, 5486-5489; b) X. Wang, L. Du, J. You, J. B. King, R. H. Cichewicz, *Org. Biomol. Chem.* **2012**, *10*, 2044-2050.
- [76] a) C. Wu, R. Cichewicz, Y. Li, J. Liu, B. Roe, J. Ferretti, J. Merritt, F. Qi, *Appl. Environ. Microbiol.* **2010**, *76*, 5815-

## REVIEW

- 5826; b) W. K. Mousa, B. Athar, N. J. Merwin, N. A. Magarvey, *Nat. Prod. Rep.* **2017**, *34*, 1302-1331.
- [77] T. Golakoti, W. Y. Yoshida, S. Chaganty, R. E. Moore, *J. Nat. Prod.* **2001**, *64*, 54-59.
- [78] a) J. E. Becker, R. E. Moore, B. S. Moore, *Gene* **2004**, *325*, 35-42; b) F. Kopp, C. Mahler, J. Grunewald, M. A. Marahiel, *J. Am. Chem. Soc.* **2006**, *128*, 16478-16479; c) S. Enck, F. Kopp, M. A. Marahiel, A. Geyer, *ChemBioChem* **2008**, *9*, 2597-2601.
- [79] a) A. Kronic, A. Vallat, S. Mo, D. D. Lantvit, S. M. Swanson, J. Orjala, *J. Nat. Prod.* **2010**, *73*, 1927-1932; b) B. S. Evans, I. Ntai, Y. Chen, S. J. Robinson, N. L. Kelleher, *J. Am. Chem. Soc.* **2011**, *133*, 7316-7319.
- [80] L. R. Malins, J. N. deGruyter, K. J. Robbins, P. M. Scola, M. D. Eastgate, M. R. Ghadiri, P. S. Baran, *J. Am. Chem. Soc.* **2017**, *139*, 5233-5241.
- [81] A. Zipperer, M. C. Konnerth, C. Laux, A. Berscheid, D. Janek, C. Weidenmaier, M. Burian, N. A. Schilling, C. Slavetinsky, M. Marschal, M. Willmann, H. Kalbacher, B. Schitteck, H. Brotz-Oesterhelt, S. Grond, A. Peschel, B. Krismer, *Nature* **2016**, *535*, 511-516.
- [82] T. Huang, X. Wang, W. Guo, S. Lin, in *Comprehensive Natural Products III* (Eds.: H. W. Liu, T. P. Begley), Elsevier, **2020**, pp. 393-445.
- [83] a) M. D. Tendler, S. Korman, *Nature* **1963**, *199*, 501; b) W. Leimgruber, V. Stefanovic, F. Schenker, A. Karr, J. Berger, *J. Am. Chem. Soc.* **1965**, *87*, 5791-5793; c) W. Leimgruber, A. D. Batcho, F. Schenker, *J. Am. Chem. Soc.* **1965**, *87*, 5793-5795.
- [84] Y. Hu, V. Phelan, I. Ntai, C. M. Farnet, E. Zazopoulos, B. O. Bachmann, *Chem. Biol.* **2007**, *14*, 691-701.
- [85] K. Arima, M. Kosaka, G. Tamura, H. Imanaka, H. Sakai, *J. Antibiot. (Tokyo)* **1972**, *25*, 437-444.
- [86] W. Li, S. Chou, A. Khullar, B. Gerratana, *Appl. Environ. Microbiol.* **2009**, *75*, 2958-2963.
- [87] A. von Tesmar, M. Hoffmann, J. Pippel, A. A. Fayad, S. Dausend-Werner, A. Bauer, W. Blankenfeldt, R. Muller, *Cell Chem. Biol.* **2017**, *24*, 1216-1227 e1218.
- [88] G. Schneditz, J. Rentner, S. Roier, J. Pletz, K. A. Herzog, R. Buckner, H. Troeger, S. Schild, H. Weber, R. Breinbauer, G. Gorkiewicz, C. Hogenauer, E. L. Zechner, *Proc. Natl. Acad. Sci. U.S.A* **2014**, *111*, 13181-13186.
- [89] E. Dornisch, J. Pletz, R. A. Glabonjat, F. Martin, C. Lembacher-Fadum, M. Neger, C. Hogenauer, K. Francesconi, W. Kroutil, K. Zangger, R. Breinbauer, E. L. Zechner, *Angew. Chem. Int. Ed.* **2017**, *56*, 14753-14757.
- [90] a) C. Gui, Q. Li, X. Mo, X. Qin, J. Ma, J. Ju, *Org. Lett.* **2015**, *17*, 628-631; b) D. P. Cogan, J. Ly, S. K. Nair, *ACS Chem. Biol.* **2020**, *15*, 2783-2791.
- [91] C. K. Wat, A. G. McInnes, D. G. Smith, J. L. C. Wright, L. C. Vining, *Can. J. Chem.* **1977**, *55*, 4090-4098.
- [92] K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus, T. J. Simpson, *ChemBioChem* **2007**, *8*, 289-297.
- [93] L. M. Halo, J. W. Marshall, A. A. Yakasai, Z. Song, C. P. Butts, M. P. Crump, M. Heneghan, A. M. Bailey, T. J. Simpson, C. M. Lazarus, R. J. Cox, *ChemBioChem* **2008**, *9*, 585-594.
- [94] L. M. Halo, M. N. Heneghan, A. A. Yakasai, Z. Song, K. Williams, A. M. Bailey, R. J. Cox, C. M. Lazarus, T. J. Simpson, *J. Am. Chem. Soc.* **2008**, *130*, 17988-17996.
- [95] a) R. F. Vesonder, L. W. Tjarks, W. K. Rohwedder, H. R. Burmeister, J. A. Laugal, *J. Antibiot. (Tokyo)* **1979**, *32*, 759-761; b) S. B. Singh, D. L. Zink, M. A. Goetz, A. W. Dombrowski, J. D. Polishook, D. J. Hazuda, *Tetrahedron Lett.* **1998**, *39*, 2243-2246.
- [96] a) N. J. Phillips, J. T. Goodwin, A. Fraiman, R. J. Cole, D. G. Lynn, *J. Am. Chem. Soc.* **1989**, *111*, 8223-8231; b) Y. C. Jeong, M. G. Moloney, *Future Med. Chem.* **2015**, *7*, 1861-1877.
- [97] E. C. Larson, A. L. Lim, C. D. Pond, M. Craft, M. Cavuzic, G. L. Waldrop, E. W. Schmidt, L. R. Barrows, *PLoS One* **2020**, *15*, e0233485.
- [98] J. W. Sims, E. W. Schmidt, *J. Am. Chem. Soc.* **2008**, *130*, 11149-11155.
- [99] C. D. Campbell, J. C. Vederas, *Biopolymers* **2010**, *93*, 755-763.
- [100] T. B. Kakule, D. Sardar, Z. Lin, E. W. Schmidt, *ACS Chem. Biol.* **2013**, *8*, 1549-1557.
- [101] a) P. K. Chang, K. C. Ehrlich, I. Fujii, *Toxins* **2009**, *1*, 74-99; b) X. Liu, C. T. Walsh, *Biochemistry* **2009**, *48*, 8746-8757; c) N. W. Seidler, I. Jona, M. Vegh, A. Martonosi, *J. Biol. Chem.* **1989**, *264*, 17816-17823.
- [102] X. Liu, C. T. Walsh, *Biochemistry* **2009**, *48*, 8746-8757.
- [103] H. Li, C. L. M. Gilchrist, H. J. Lacey, A. Crombie, D. Vuong, J. I. Pitt, E. Lacey, Y. H. Chooi, A. M. Piggott, *Org. Lett.* **2019**, *21*, 1287-1291.
- [104] R. Ladenstein, J. O. Winberg, J. Benach, *Cell. Mol. Life Sci.* **2008**, *65*, 3918-3935.
- [105] K. Takada, A. Ninomiya, M. Naruse, Y. Sun, M. Miyazaki, Y. Nogi, S. Okada, S. Matsunaga, *J. Org. Chem.* **2013**, *78*, 6746-6750.
- [106] A. Ninomiya, Y. Katsuyama, T. Kuranaga, M. Miyazaki, Y. Nogi, S. Okada, T. Wakimoto, Y. Ohnishi, S. Matsunaga, K. Takada, *ChemBioChem* **2016**, *17*, 1709-1712.
- [107] T. Kuranaga, K. Matsuda, A. Sano, M. Kobayashi, A. Ninomiya, K. Takada, S. Matsunaga, T. Wakimoto, *Angew. Chem. Int. Ed.* **2018**, *57*, 9447-9451.
- [108] R. F. Pratt, *Cell. Mol. Life Sci.* **2008**, *65*, 2138-2155.
- [109] K. Matsuda, R. Zhai, T. Mori, M. Kobayashi, A. Sano, I. Abe, T. Wakimoto, *Nat. Catal.* **2020**, 507-515.
- [110] Y. J. Zhou, X. Lin, C. M. Xu, Y. Y. Shen, S. P. Wang, H. Z. Liao, L. Li, H. Deng, H. W. Lin, *Cell Chem. Biol.* **2019**, *26*, 737-744.
- [111] J. F. Barajas, G. Shakya, G. Moreno, H. Rivera, D. R. Jackson, C. L. Topper, A. L. Vagstad, J. J. La Clair, C. A. Townsend, M. D. Burkart, S. C. Tsai, *Proc. Natl. Acad. Sci. U.S.A* **2017**, *114*, E4142-E4148.
- [112] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorphe, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-1143.
- [113] a) T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315-1322; b) A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941-953.
- [114] J. M. Crawford, C. A. Townsend, *Nat. Rev. Microbiol.* **2010**, *8*, 879-889.
- [115] S. Kroken, N. L. Glass, J. W. Taylor, O. C. Yoder, B. G. Turgeon, *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 15670-15675.
- [116] a) M. E. Mayorga, W. E. Timberlake, *Genetics* **1990**, *126*, 73-79; b) M. E. Mayorga, W. E. Timberlake, *Mol. Gen. Genet.* **1992**, *235*, 205-212; c) P. Rugbjerg, M. Naesby, U. H. Mortensen, R. J. Frandsen, *Microb. Cell Fact.* **2013**, *12*, 31.
- [117] a) A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem. Commun.* **2012**, *48*, 11772-11774; b) A. G. Newman, C. A. Townsend, *J. Am. Chem. Soc.* **2016**, *138*, 4219-4228.
- [118] H. Ogawa, K. Hasumi, K. Sakai, S. Murakawa, A. Endo, *J. Antibiot. (Tokyo)* **1991**, *44*, 762-767.
- [119] A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford, C. A. Townsend, *Angew. Chem. Int. Ed.* **2013**, *52*, 1718-1721.
- [120] F. Kudo, Y. Asou, M. Watanabe, T. Kitayama, T. Eguchi, *Synlett* **2012**, 1843-1846.

## Entry for the Table of Contents



**Biosynthetic cyclization catalysts:** In this review, we discuss the biosynthetic enzymes involved in formation of cyclic natural products from polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs). Recent developments and the remarkable diversity of these cyclization catalysts to produce biologically relevant compounds are highlighted.