FEATURE ARTICLE

ChemComm
www.rsc.org/chemcomm

Enzymology of acyl chain macrocyclization in natural product biosynthesis

Rahul M. Kohli and Christopher T. Walsh
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. E-mail: christopher_walsh@hms.harvard.edu

Received (in Cambridge, UK) 28th August 2002, Accepted 6th November 2002
First published as an Advance Article on the web 22nd November 2002

Polyketides and nonribosomal peptides constitute a large and diverse set of natural products with biological activity in microbial survival and pathogenesis, as well as broad pharmacological utility as antineoplastics, antibiotics or immunosuppressants. These molecules are biosynthesized by the ordered condensation of monomer building blocks, acyl-CoAs or amino acids, leading to construction of linear acyl chains. Many of these natural products are constrained to their bioactive conformations by macrocyclization whereby, in one of the terminal steps of biosynthesis, parts of the molecule distant in the constructed linear acyl chain are covalently linked to one another. Typically, macrocyclization is catalyzed by a thioesterase domain at the C-terminal end of the biosynthetic assembly line, although alternative strategies are known. The enzymology of these macrocyclization catalysts, their structure, mechanism, and catalytic versatility, is the subject of this review. The diversity of macrocyclic structures accessed by enzyme catalyzed cyclization of linear acyl chains as well as their inherent substrate tolerance suggests their potential utility in reprogramming natural product biosynthesis pathways or accessing novel macrocyclic structures.

Rahul M. Kohli graduated in 1998 with a B.S. in Biochemistry with highest honors from the University of Michigan. While there, he conducted research on flavoprotein enzymology under the supervision of Professor Vincent Massey. He is currently an MD/Ph.D. student at Harvard Medical School, carrying out his graduate work in the Department of Biological Chemistry and Molecular Pharmacology under the guidance of Professor Christopher T. Walsh. His work has focused on macrocyclizing enzymes in natural product biosynthesis.

Professor Christopher T. Walsh took his undergraduate degree in biology from Harvard and the Ph.D from Rockefeller University where his doctoral research was on enzymatic mechanisms of citrate cleavage for fatty acid biosynthesis. Following postdoctoral study at Brandeis University, in 1972 he joined the MIT faculty, both in the Chemistry and Biology Departments, serving ultimately as Chemistry Chair. In 1987 he moved to Harvard Medical School to combine the Biological Chemistry and Pharmacology Departments and serve as chair. His research has been at the interface of biology, chemistry and medicine, centered on enzymatic reaction mechanisms with attention to enzymes of therapeutic relevance. Current topics of investigation include enzymes in the biosynthesis of antibiotics such as vancomycin and tyrocidine, iron chelating siderophores such as enterobactin, pyochelin, vibriobactin and yersiniabactin, and hybrid polyketide/nonribosomal peptides such as epothilone and rifamycin. Walsh has had substantial involvement in the scientific advisory boards and/or boards of directors of several biotechnology companies, including Genzyme, Immunogen, Leukosite, Caliper, Millennium, Kosan, Versicor, Transform and Critical Therapeutics. He is a member of the National Academy of Sciences and the Institute of Medicine and has won the Lilly Award, the Repligen award, and the Bader Award, all from the American Chemical Society. He is the author of the authoritative text on enzymology, Enzymatic Reaction Mechanisms *and of a book in press on* Antibiotics: Actions, Origins, Resistance. *He is currently the Hamilton Kuhn Professor of Biological Chemistry and Molecular Pharmacology at Harvard Medical School.*

Introduction

Biologically active natural products must present the proper functionality in the precise orientation required for interaction with a molecular target. A common strategy employed by nature to achieve this aim is to constrain a molecule to a limited set of conformations by covalent linkage of distant parts of the molecule, lowering the entropic cost of populating an active conformation.[1] For many diverse natural products, including polyketides and nonribosomally synthesized peptides, covalent constraints are selectively achieved in densely functional molecules by enzymatic cyclization of linear acyl chains.

Polyketide natural products are assembled from acyl monomer units activated as thioesters. The fundamental chain elongation step is C–C bond formation mediated by attack of an enzyme-generated carbon nucleophile upon an upstream biosynthetic intermediate. In type I polyketide synthases (PKS) the elongated acyl chain is translocated from upstream to downstream carrier protein domains that contain the tethering thiol group. In type II PKS, the elongated acyl chain stays tethered to the same carrier protein while the acyl monomers are on distinct subunits and the chain elongation is iterative. Nonribosomal peptides are assembled by parallel logic to the type I PKS. Peptidyl chains grow by consecutive addition of activated aminoacyl monomer units. The fundamental chain elongation step is peptide bond formation and the elongated chain is translocated each time from upstream to downstream carrier proteins during chain elongation.

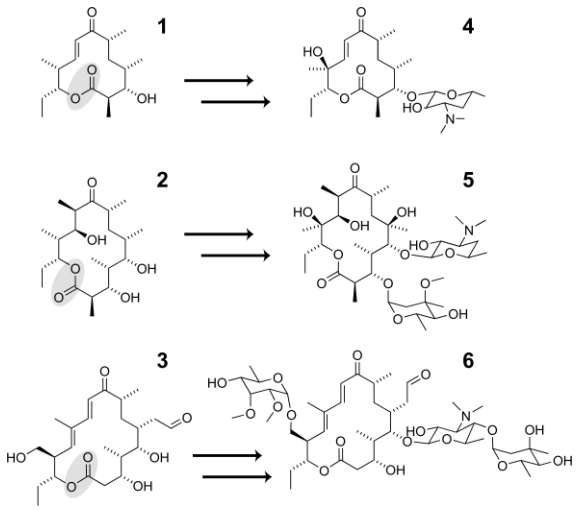
In both PKS and nonribosomal peptide synthetases (NRPS), once the acyl chain reaches its full length on the most downstream carrier protein, it has to be released from its covalent thioester tether. Typically, the most C-terminal domain in these protein assembly lines is a thioesterase (TE) domain whose role is to catalyze the chain disconnection reaction, although alternative enzymatic solutions are known. While hydrolysis is one common product of enzymatic chain termination, intramolecular macrocyclization reactions are catalyzed by several PKS and NRPS terminal domains. The nature of these macrocyclization catalysts and the reactions they catalyze are the subject of this review.

DOI: 10.1039/b208333g

Macrocyclic polyketides, peptides and proteins

Examination of the myriad structures accessed by macrocyclization of linear acyl chains makes evident the versatility of enzymatic macrocyclization catalysts. Examples can be found of almost all permutations of a host of factors—the nature of the acyl chain (polyketide, polypeptide or hybrid), the nature of the linkage formed (macrolactone, macrolactam, cyclooligomer), and the type of cyclic structure formed (closed cycles or lariat structures).

In PKS assembly lines, macrolactone formations occur from acyclic precursors with regiospecific attack by a particular hydroxyl group from a folded conformer of the linear acyl chain (Scheme 1). Cyclic lactones with 12, 14 and 16 atoms in the

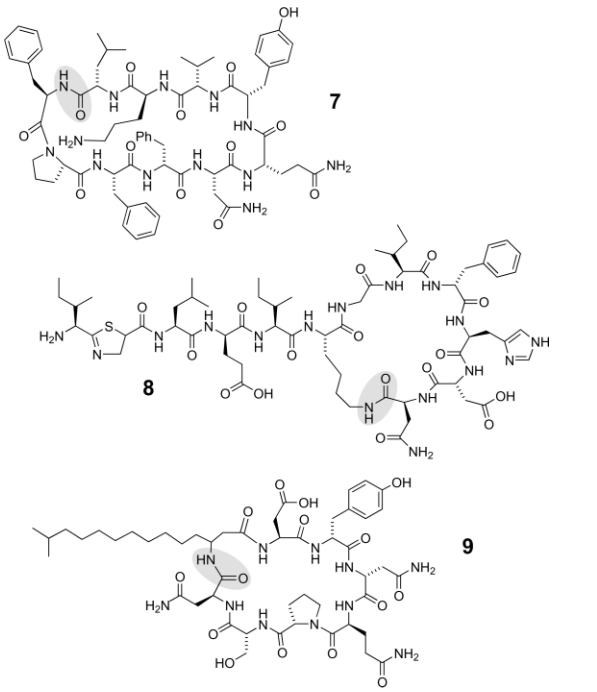


Scheme 1 Macrolactone PKS products with bonds formed during macrocyclization highlighted by shading. (**1**) 10-deoxymethynolide, (**2**) 6-deoxyerythronolide B, (**3**) tylactone, (**4**) methymycin, (**5**) erythromycin A, (**6**) tylosin.

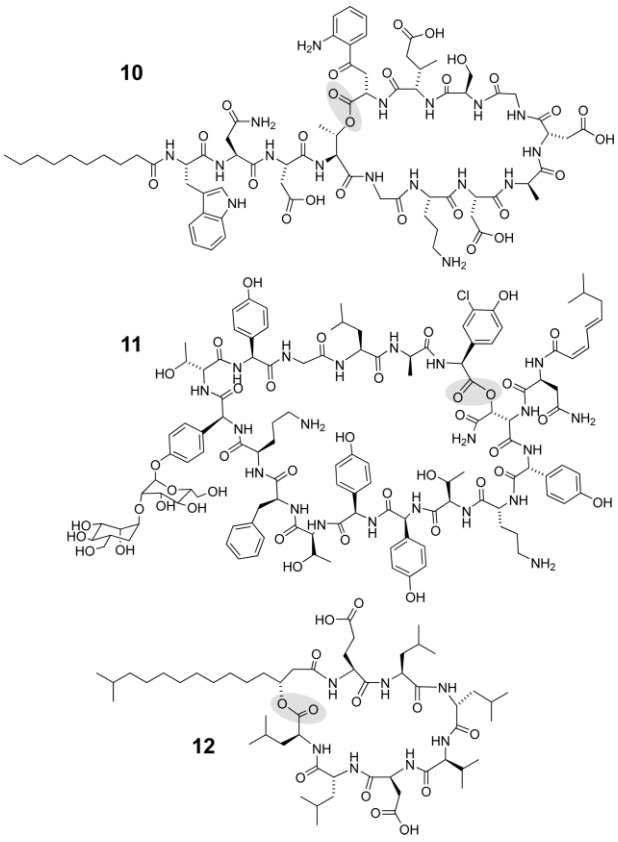
macrocycle are well known, in such scaffolds as 10-deoxymethynolide **1**, 6-deoxyerythronolide B **2** and tylactone **3**, that then become further decorated by tailoring enzymes that hydroxylate and glycosylate regio- and stereospecifically producing methymycin **4**, erythromycin A **5** and tylosin **6**, respectively.[2–5]

Several species of fungi and bacteria elaborate cyclic peptides and peptidolactones, with most examples arising by NRPS enzymatic machinery (Scheme 2). Peptide natural products offer many possible sources of nucleophiles that can be used in constraining macrocyclization reactions. Cyclic peptides include the head-to-tail cyclized decapeptide tyrocidine A **7**, that arises from linkage of the amino-terminal NH_2 of D-Phe1 with the C-terminal Leu10.[6,7] The dodecapeptide bacitracin A **8**, used as a topical antibiotic, instead has a lariat structure, with the macrocycle arising from enzyme-directed capture of the carbonyl of Asn12 by the ε-NH_2 of the Lys6 side chain.[8] Additionally, the nucleophile in the cyclization reaction may be derived from a β-amino fatty acid as in mycosubtilin **9**.[9]

Cyclic peptidolactones are biosynthesized analogously by nonribosomal depsipeptide synthetases, where the attacking nucleophile in macrocyclization is an intramolecular hydroxyl substituent (Scheme 3). This can be the side chain of a proteinogenic β-OH-amino acid such as Thr4 in the biosynthesis of daptomycin **10**, or of a nonproteinogenic β-OH amino acid, such as β-OH-Asn2 in the formation of the macrocyclic structure of the antibiotic ramoplanin **11**.[10,11] Peptidolactones such as daptomycin and ramoplanin are often enzymatically acylated or glycosylated during enzymatic assembly. The fatty acyl substituent of a β-OH-fatty acid can additionally be the macrocyclization nucleophile, as in the assembly of surfactin A **12**.[12]

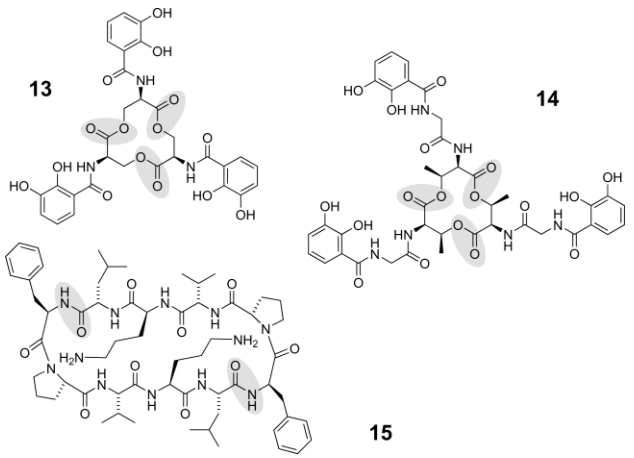


Scheme 2 Macrolactam NRPS products. (**7**) tyrocidine A, (**8**) bacitracin A, (**9**) mycosubtilin.



Scheme 3 Macrolactone NRPS products: daptomycin (**10**), ramoplanin (**11**), surfactin A (**12**).

In iron-deficient microenvironments, bacteria synthesize and secrete NRPS-derived iron chelators known as siderophores that scavenge iron by binding Fe^{3+} with picomolar affinity.[13,14] The enterobactin **13** and bacillibactin **14** siderophores from

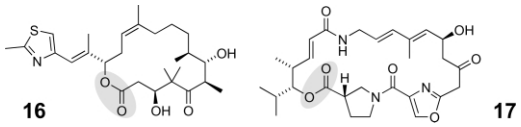


Scheme 4 Products of cyclooligomerization: enterobactin (**13**), bacillibactin (**14**), gramicidin S (**15**).

Escherichia coli and *Bacillus subtilis* (Scheme 4), are cyclic trilactones that arise, respectively, from cyclotrimerization of 2,3-dihydroxybenzoyl (DHB)-seryl thioesters or DHB-glycyl-threonyl thioester precursors yielding 12-membered trilactones.[15,16] These siderophores present three intramolecular catechol ligands for avid chelation of ferric iron. Similar oligomerization and cyclization is seen in the biosynthesis of the symmetric decapeptide antibiotic gramicidin S **15** produced by *Bacillus brevis*.[17]

There are a distinct class of cyclic peptides that are produced on ribosomes by traditional protein biosynthesis as larger precursors, and then cyclized while undergoing processing.[18] These include microcin J25, a 21-residue cyclic peptide from certain *E. coli* strains, and bacteriocin AS-48, a pore-forming 70-residue cyclic peptide from *Enterococcus faecalis*.[19,20] A family of plant insecticidal cyclotides, from 28–37 amino acids in length, are produced with three Cys–Cys disulfides imposed on the circular backbone to yield highly constrained, knotted structures.[21] Mammalian circular peptides are also known, including the rhesus θ defensins (RTD) comprised of 18 amino acids cyclized into a single beta-sheet that is further rigidified by three disulfide bonds.[22,23] Unlike the cyclotides, which are derived from a single transcript, some of the θ defensins are derived from two linear precursors. For these families of ribosomally derived cyclic peptides, however, little is known about the processing of the linear precursors by enzymatic cyclization and neither will be discussed further in this review.

We will note in the next section the parallel enzymatic assembly-line logic and modular organization for building polyketides and nonribosomal peptides and for their enzymatic cyclization.[24,25] There are natural products that are obvious polyketide/nonribosomal peptide hybrids (Scheme 5). The antineoplastic epothilones **16** and the anti-infective pristinamycins **17** stand as examples of such hybrid macrocycles.[26,27] The hybrids arise from assembly-line convergence, the mixing and matching of PKS and NRPS modules into a single hybrid assembly line.[28]



Scheme 5 Macrocyclic hybrid PKS/NRPS products: epothilone D (**15**), pristinamycin IIB (**16**).

Assembly-line enzymatic machinery: initiation, elongation and termination modules

The enzymatic catalysts that perform the macrocyclization reactions are the most downstream domains in the multidomain, modular assembly lines of polyketide synthases and nonribosomal peptide synthetases. To understand the function of these cyclases in release of the full-length polyketide and polypeptide chains we note some conserved features of the enzymatic logic and organization common to both type I PKS and NRPS assembly lines.

The protein modules of NRPS assembly lines are organized into chain initiation, chain elongation and chain termination modules, proceeding from N-terminal to C-terminal modules, respectively (Fig. 1).[29,30] The carrier protein domain within each module contains a phosphopantetheine prosthetic group that provides a terminal thiol group for attachment to the acyl chains in thioester linkages.[31,32] The most upstream, initiation module selects the first monomer and attaches it to the carrier protein, known also as a thiolation (T) domain, *via* a thioester linkage. In NRPS assembly lines, the amino acid monomer is selected and activated by an adenylation (A) domain, so the typical initiation module is a two-domain A–T module. The subsequent modules downstream also have A–T domains for activation and tethering of each amino acid added to the chain, but additionally have a peptide-bond-forming condensation (C) domain, for a typical C–A–T elongation module organization. The most downstream module in NRPS assembly lines serves as a chain termination module. Typically, a TE domain is found, in the canonical order C–A–T–TE, where the TE functions to disconnect the covalent thioester linkage between the full length peptidyl chain and the last T domain in the assembly line. Alternative chain termination strategies will be discussed later in this review, including terminal condensation domains which are proposed to act as macrocyclization catalysts.[33]

PKS assembly lines follow analogous multidomain, multimodular logic with a carrier protein T domain in each module embedded among various catalytic domains.[24,34] Since the elongation chemistry in PKS assembly lines is C–C bond formation rather than amide bond construction, the catalytic domain responsible for each chain elongation step is a

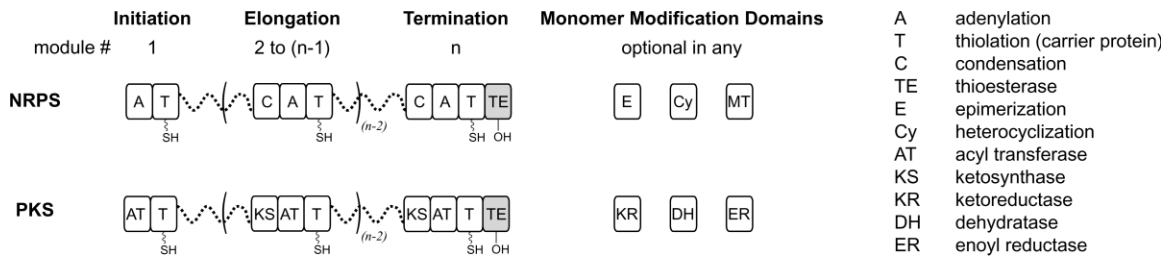


Fig. 1 Generic organizational scheme of NRPS and PKS assembly lines. The domains involved in the assembly lines are diagrammed as boxes with function denoted at right. A module is defined by the domains dedicated to the incorporation and modification of a single building block. Each module contains a thiolation domain with a phosphopantetheine tether (–SH). The dotted line indicates that the modules are often strung together into large multi-domain, multi-module synthetase subunits where the sum of several subunits constitutes a full assembly line. Within individual modules, domains may be present which function to modify the building block being incorporated into the growing peptide or polyketide.

ketosynthase (KS), creating the required carbon nucleophile for a Claisen condensation by decarboxylation of malonyl or methylmalonyl-*S*-T domains. The malonyl- and methylmalonyl-*S*-protein intermediates are installed by acyl transferase (AT) domains from soluble acyl–CoA substrates. Since polyketides can undergo post-condensation reduction, dehydration and olefin reduction, the corresponding ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains may also be present in PKS modules, for example in the order KS–AT–KR–DH–ER–T. Different PKS elongation modules can have different complements of functional domains although KS–AT–T represent the minimal three domain core. When the full-length polyketide chain has reached the most downstream T domain it must be released from its covalent tethering in order for the enzymatic assembly line to act processively to build the next acyl chain and release the next polyketide product. In full analogy to the NRPS assembly lines, most PKS assembly lines have a chain terminating thioesterase as the most downstream domain in the termination module. As with NRP synthesis, alternative chain termination strategies are known, including macrocycle formation from linear aryl-capped polyketide chains by homologues of arylamine *N*-acetyltransferase.[35,36]

Briefly, it is notable that elegant enzymatic solutions to the challenge of introducing five- or six-membered cyclic constraints into linear molecules by C–C bond formation, rather than macrolactam or macrolactone formation, have been demonstrated in other biosynthetic pathways. Type II polyketide synthases, which differ from type I synthases in that catalytic domains are used iteratively and associate transiently rather than being connected in a modular fashion and *in cis*, often synthesize products such as the tetracyclines or daunorubicin that are polycyclic or polyaromatic.[37] In these systems, the polyketidyl acyl chain is usually constructed without reduction of β-keto intermediates during chain elongation. Cyclase enzymes in the biosynthetic cluster act upon the polyketone containing linear intermediate to promote the folded conformation which is closed by aldol C–C bond formation. Alternate remarkable enzymatic polycyclization chemistry is used in the production of unfavored product of Diels–Alder cyclization in lovastatin nonaketide biosynthesis.[38]

The thioesterase domains as chain termination catalysts

The TE domains embedded at the downstream end of both type I PKS and NRPS enzymatic assembly lines are members of the α,β hydrolase superfamily of enzymes.[39] They are about 28-35 kDa in size and use an active site serine as a nucleophilic catalyst. The paired T–TE domains in the termination modules of PKS and NRPS assembly lines act in tandem when the full length acyl chain has become parked on the T domain. Chain release is initiated by transfer of the nascent peptidyl or polyketidyl acyl chain to the active site serine of the adjacent TE domain to generate an acyl-*O*-TE intermediate,[40,41] and regenerate the T domain in its HS-pantetheinyl form (Fig. 2). The acyl-*O*-TE covalent enzyme intermediate can undergo two kinds of subsequent catalytic fates. It can be captured by an external nucleophile, usually water for net hydrolysis, but occasionally by a specific amine cosubstrate for net aminolysis as in vibriobactin biosynthesis.[42,43] Alternately, it can be captured by a nucleophile (–OH, $–NH_2$) internal to the acyl chain to release a macrocyclic product, for example to yield 6-deoxyerythronolide B **2**, tyrocidine A **7** and surfactin A **12** (Fig. 2).

Hydrolysis is the fate, for example, in release of the heptapeptide precursor of vancomycin group antibiotics or the tripeptide precursor of the β-lactam antibiotics by their respective NRPS assembly lines. The macrocyclization fates must instead reflect a distinct kinetic outcome in which the acyl-*O*-TE is kinetically sequestered from water while a specific conformer of the acyl chain is populated to permit productive approach and capture of the acyl chain carbonyl by an internal nucleophile. As noted in the Introduction, different TE domains must control the orientation of the folded acyl chain since cyclizations are regiospecific, chemospecific and stereospecific as exemplified by structures **1**–**17**.

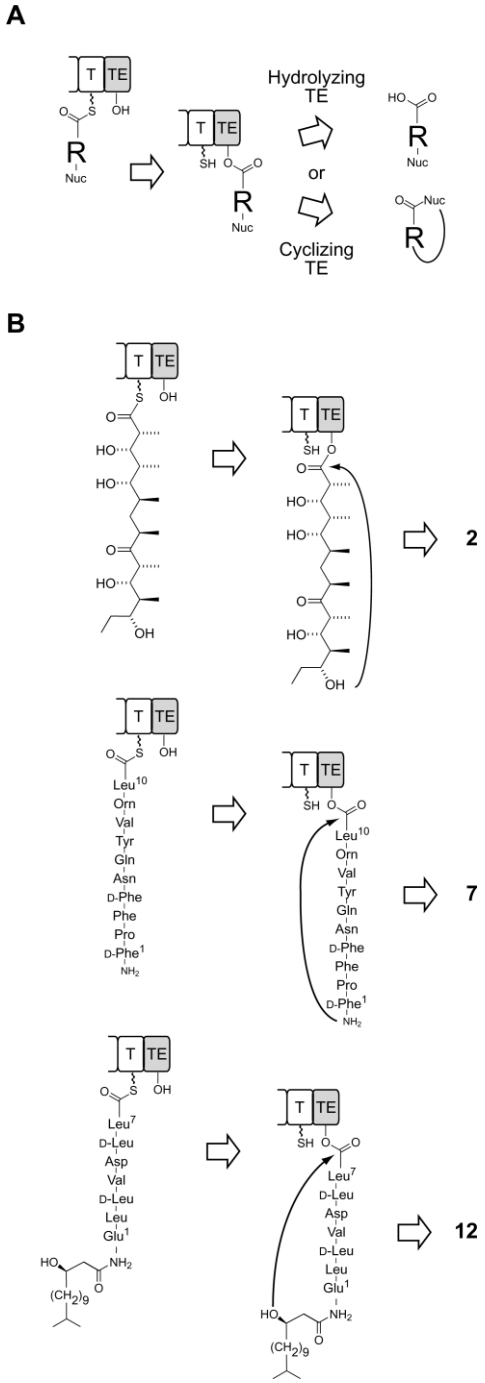


Fig. 2 The mechanism of terminal thioesterase domains. (**A**) An acyl chain loaded on the terminal carrier protein (T) of a synthetase is transferred from its the phosphopantetheine tether (–SH) to an active site serine (–OH) of the TE forming the acyl-enzyme intermediate. The intermediate may break down by the attack of water to give a linear product or by attack of an intramolecular nucleophile on the acyl-enzyme ester bond, producing a cyclic product. (**B**) Enzymatic macrocyclization of the linear erythromycin precursor, tyrocidine precursor or surfactin precursor involves the formation of an acyl-enzyme intermediate followed by regiospecific attack of a particular nucleophile to generate the macrocyclic natural product structures.

An alternative fate to strictly intramolecular cyclization by the NRPS TE domains is net cyclooligomerization effected on aminoacyl or peptidyl-*S*-T domains by TE domains. The five-module gramicidin synthetase must dimerize a pentapeptidyl-*S*-T acyl enzyme to the decapeptidyl-enzyme and then cyclize it head-to-tail (Fig. 3).[17] Similarly, the enterobactin synthetase TE

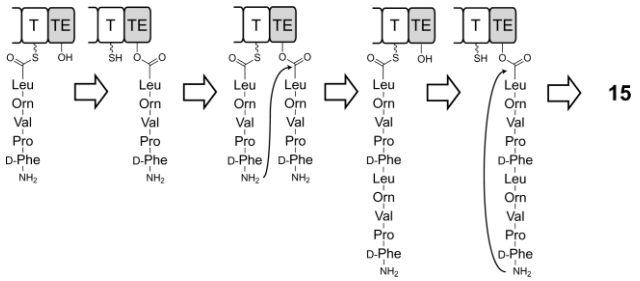


Fig. 3 The mechanism of cyclooligomerization by thioesterase domains. In the biosynthesis of the symmetric decapeptide antibiotic gramicidin S, a pentapeptide is initially transferred to the TE domain where the acyl-enzyme intermediate serves as a way-station while a second pentapeptide is constructed by the synthetase and brought to the terminal T domain. The TE domain initially catalyzes oligomerization to form the linear decapeptide tethered to the terminal T domain which subsequently serves as a substrate for macrocyclization.

accumulates 2,3-dihydroxybenzoyl-seryl-*S*-T domain intermediate lodged on the EntF constituent of the assembly line, makes the linear DHB-ser-DHB-ser-*O*-enzyme, then presumably the linear $(DHB\text{-}ser)_3$-*O*-enzyme before internal capture to generate the 12-membered trilactone siderophore.[15] Mass spectrometry has provided evidence for the dimer-*O*-TE intermediate.[41] Analogous enzymatic cyclization chemistry is intuited for formation of the DNA intercalator triostin,[44] the emetic toxin cereulide from *Bacillus cereus*,[45] as well as the vicibactin siderophore produced *via* cyclotrimerization of a D-N^5-hydroxy-N^5-(D-3-hydroxybutyryl)-enzyme intermediate by the nitrogen-fixing *Rhizobium leguminosarum*.[46]

Autonomous activity and portability of TE domains

NRPS assembly lines in general have several domains connected in a single polypeptide chain into modules and multiple modules in turn collected in large subunits of 10^5–10^6 dalton molecular weights (Fig. 4). Type I PKS and mammalian fatty acid synthase assembly lines follow this protein organizational logic.[47,48] On the other hand there are both bacterial fatty acid synthases and polyketide synthases that produce aromatic products such as tetracyclines, tetracaenomycins and anthracyclines[37,47,49] where the domains, including the TE domains, exist as separate protein subunits. This validates the view that the many domains in the multimodular enzymatic assembly lines may function as "beads on a string", where genetic mix and match or domain shuffling events can introduce diversity during evolution of natural product biosynthetic pathways.[24,34] It also indicates that the domains are autonomously folding protein units that could retain activity when excised from the assembly lines.

Separate TE proteins that hydrolyze fatty acyl-*S*-acyl carrier proteins have been purified both from mammalian and bacterial sources.[50–52] Further, in many polyketide and nonribosomal peptide biosynthetic clusters that encode the multimodular protein assembly lines that terminate in integrated TE domains, there are also genes encoding external TE proteins (also known as type II TEs), for example in tylosin, bacitracin and tyrocidine biosynthesis (TycF in Fig. 4).[8,53,54] Genetic knockouts of the external TEs reduce antibiotic production about an order of magnitude, suggesting the external thioesterases have a supplementary role to play in PK and NRP biosynthesis.[55,56] One hypothesis for the function of the external TEs is in editing or quality control, hydrolytically removing any incorrect acyl chains that have been installed but are unable to elongate, and thus freeing up stalled modules.[57] While their function has yet to be elucidated, they are clearly active as isolated proteins, suggesting that terminal TE domains could be similarly active if excised from their larger synthetase proteins.

The embedded terminal TE domains in both PKS and NRPS assembly lines have been found to be portable. When placed in frame in more upstream modules they act catalytically on shorter acyl chains from acyl-*S*–T domains. In the DEBS system, novel macrolactone products have been detected after relocation of the terminal TE domain to upstream modules (Fig. 5).[58–60] At times, action of a terminal TE domain on upstream modules can have a physiological role, as seen with pikromycin biosynthesis where the internal TE can act either on the final or

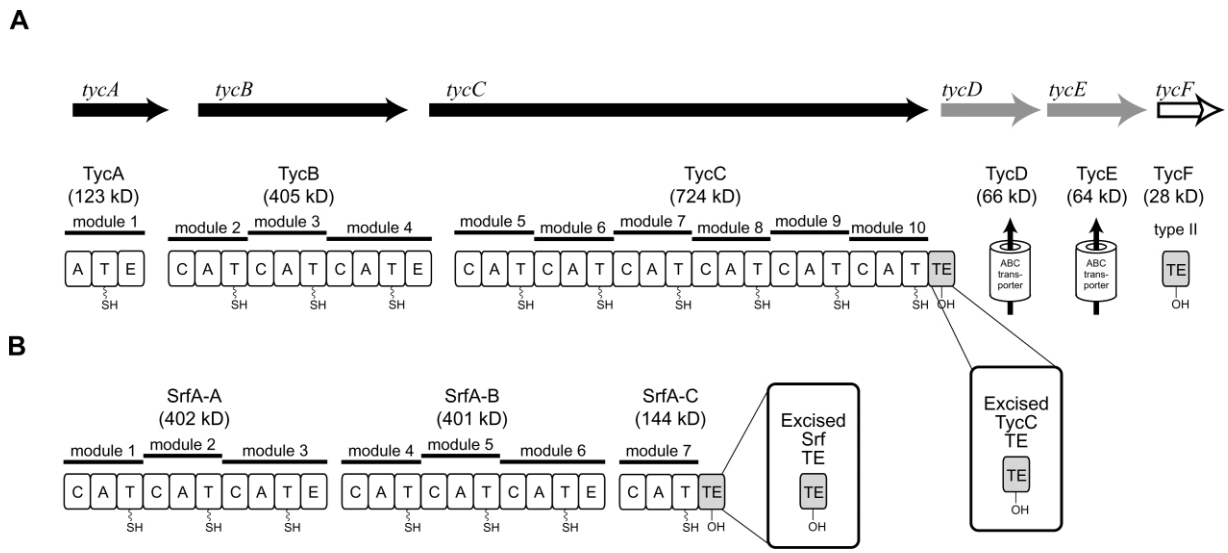


Fig. 4 Synthetases responsible for the production of tyrocidine A and surfactin A. (**A**) The gene cluster for tyrocidine biosynthesis contains six open reading frames. Three encode for the synthetase proteins TycA, TycB and TycC which contain one, three and six modules, respectively. The termination module of TycC contains the C-terminal thioesterase domain (TycC TE) which has been excised from the larger synthetase for biochemical characterization. Also contained in the cluster are two putative ABC transporters likely used in antibiotic export and self-resistance and a protein with homology to type II thioesterases from fatty acid biosynthesis with putative editing function. (**B**) The synthetase proteins resposible for the production of surfactin A consists of three subunits. The C-terminal TE domain, Srf TE, has been excised and characterized structurally and biochemically.

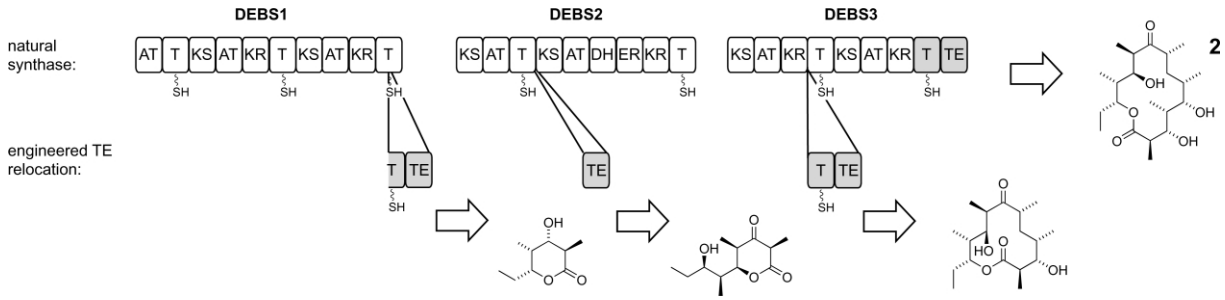


Fig. 5 The portability of the terminal TE domain in DEBS synthase. The natural synthase produces the macrocyclic erythromycin precursor 6-deoxyerythronolide. Engineered relocation of the TE domain with part or all of its upstream T domain results in the formation of novel products including the cyclic and macrocyclic products predicted to arise from the action of a TE domain on a biosynthetic intermediate in the DEBS assembly line.

penultimate carrier protein to generate 14- or 12-membered macrolactones, respectively.[61] Further, the appending of a TE domain to an internal module converts that module from a stoichiometric acylation module to one that turns over catalytically, amplifying the signal to noise in assays for detection of products and facilitating analysis of the catalytic capacity of other domains in the PKS or NRPS module. This has been of tremendous utility in the DEBS system to examine the selectivity and promiscuity of elongation modules in the DEBS assembly line by release of such acyl intermediates into solution.[62] Similarly, engineered NRPS variants constructed by unnatural fusion of modules required the presence of a terminal TE for catalytic generation of linear peptide products.[63]

In addition to shuffling of such embedded TE domains from the termination module into elongation modules, the TE domains from such assembly lines as DEBS, epothilone, tyrocidine and surfactin have been expressed as excised domains and found to retain catalytic activity.[7,64,65] The DEBS TE is active in hydrolysis of acyclic model substrates.[64,66,67] The terminal TE from the epothilone synthetase has been assayed for hydrolysis of the macrolactone ring of authentic epothilone and can catalyze macrolactonization to the 16-membered product.[68] The excised thioesterase domains from the tyrocidine synthetase, TycC TE and surfactin, Srf TE, have been biochemically analyzed for macrocyclization capacity and will be further discussed below.

Structure and mechanism of cyclizing thioesterases

The PKS and NRPS terminal thioesterase domains involved in macrocyclization of linear precursors have in common the responsibility to direct regiospecific macrocyclization while excluding water to prevent non-productive hydrolytic outcomes. The structure of the single domain integrated thioesterases excised from the PKS producing DEBS and the surfactin NRPS have been solved and allow for consideration of the architectural features used to solve the challenge of enzymatic macrocyclization (Fig. 6).[69,70]

The structural studies have revealed similarities to structures previously solved for α/β hydrolase family members with several novel aspects of note with relation to mechanism.[39] TE domains act in concert with the upstream carrier proteins that deliver the acyl chain to the active site Ser. The DEBS TE domain is built around a central seven-stranded β-sheet flanked by α helices.[69] In the DEBS TE structure, a long cleft exists spanning the β-sheet core, a feature also revealed in the recent structure of the terminal thioesterase domain from the picromycin synthase.[71,72] Centered in this cleft is the catalytic triad—an active site Ser positioned on a nucleophilic elbow at the end of a hydrogen bond network including the active site His and Asp. At one end of this cleft, an Arg rich patch could serve as the interaction site for the upstream electronegative thiolation

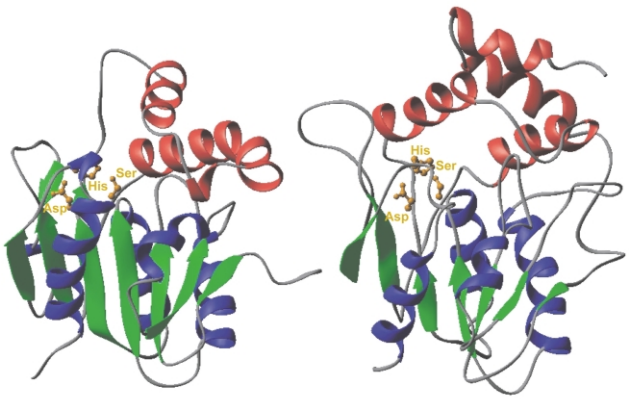


Fig. 6 The structures of macrocyclizing TE domains from surfactin synthetase[70] (left) and DEBS synthase[69] (right). Both TE domains are members of the α/β hydrolase family with the structural scaffolds (green/blue) and variable regions which define the 'lid' (red). The catalytic machinery is defined by the active site Asp-His-Ser triad highlighted (yellow).

domain based on simulated docking of a homologous carrier protein. In the Srf TE structure, a sulfate ion bound near the N-terminus of the domain defines a potential site for binding of the phosphate of the phosphopantetheine tether.[70] Docking of the structure of a peptidyl carrier protein previously solved by NMR suggests that the upstream carrier protein would be positioned appropriately to allow for the 20 Å long phosphopantetheine arm to deliver the acyl thioester to the active site Ser.

In the DEBS TE structure, surrounding the active site Ser is a largely hydrophobic cavity sufficient to accommodate the DEB product.[69] Modeling of the product in the active site suggests several hydrophobic residues conserved amongst PKS TE domains and additional nonconserved hydrophilic residues could interact to define the substrate binding pocket. The structure of the Srf TE domain was solved in the presence of a substrate analog which remarkably resulted in an asymmetric dimer with only one monomer containing substrate.[70] The Srf TE most significantly differed from the canonical fold of the α/β hydrolase family by an extended insertion of three α-helices between core β-strands. In the monomer lacking the substrate, a single helix of this 'lid' is pulled over the active site and connected to the neighboring β-strand by a disordered loop. In the monomer with substrate bound, the helix is pulled away from the active site and the loop ordered, creating an active site cavity in the form of a bowl lined by mostly aromatic and hydrophobic residues. The additional density in substrate bound monomer allowed for partial modeling of the substrate analog, showing several side-chain binding pockets for residues closest to the C-terminus of the surfactin linear substrate. It is notable that in the DEBS TE structure, a helix surrounding the active site had the highest *B* factor suggesting flexibility which could serve analogously to the 'lid' in the Srf TE domain. Addition-

ally, based on alignment, the TE domains from gramicidin and tyrocidine synthetases differ from Srf TE most significantly in the region of this lid, suggesting that the substrate specificity is encoded in this flexible and structurally dynamic region of the TE. The formation of a largely hydrophobic active site cavity in the presence of substrate mechanistically suggests that the cavity may be used to passively allow the linear acyl-*O*-TE intermediate to populate a cyclizing conformation. The non-conserved residues in the lid region may actively assist in selecting the conformer that would yield the proper regiochemical linkage.

TE domains as autonomous cyclization catalysts: scope of reaction

TycC TE

The first excised thioesterase domain to be examined in detail for retained cyclization activity was the TE domain from the tyrocidine synthetase TycC subunit.[7] The physiological substrate for this TE domain when embedded in the TycC subunit should be the linear decapeptidyl-*S*-T_{10} thioester (Fig. 7). This acyl-*S*-protein substrate is difficult to come by, as it is an acylated form of the 724 kDa TycC subunit and would be kinetically labile in any isolation scheme. Therefore a linear tyrocidine decapeptidyl-*S*–*N*-acetylcysteamine thioester (SNAC) was synthesized and utilized as substrate with the excised, purified TycC TE and found to be cyclized with a k_{cat} of 60 min^{-1} and a K_m of 3 μM (Fig. 7).

The fact that the TycC TE recognized a soluble decapeptidyl-SNAC surrogate of the normal thioester enzyme substrate enabled structure activity variation to examine substrate specificity in several manners (Fig. 8). First, a scan through all ten positions of the decapeptidyl chain revealed that D-Phe1 and L-Orn9 were required but the side chains of the other eight residues could be altered to alanine without notable consequence.[7] It was also observed that different size macrocyclic lactams, from eighteen atoms (from hexapeptidyl-SNAC) to 42 atoms (from tetradecapeptidyl-SNAC) could be efficiently cyclized by TycC TE, indicating remarkable permissivity for macrolactam ring size.[65] With a shorter substrate, a pentapeptidyl D-Phe-Pro-Val-Orn-Leu-SNAC, the TE domain could elongate it to the tandem decapeptidyl dimer and release a mixture of linear and cyclic decapeptide products, the latter authentic gramicidin S (essentially the reaction of Fig. 3).[7] This presumably reflects the logic and operation of the TE embedded at the end of the gramicidin S synthetase fifth module.[65] Depsipeptide soluble substrates gave further insight into the versatility of this cyclization catalyst, including the demonstration of macrolactone formation by alteration of the intramolecular nucleophile, the $–NH_2$ of D-Phe1, to the –OH of D-phenylactate1.[73]

In an exercise to determine the scope of substitution permitted in a decapeptidyl-SNAC by TycC TE and to move the product from infectious disease to cardiovascular pharmacology, a decapeptide was synthesized containing the RGD tripeptide motif, associated with high affinity for integrin receptors.[74] Despite the fact that the substrate **18**, contained alterations from the natural sequence of the identity of seven

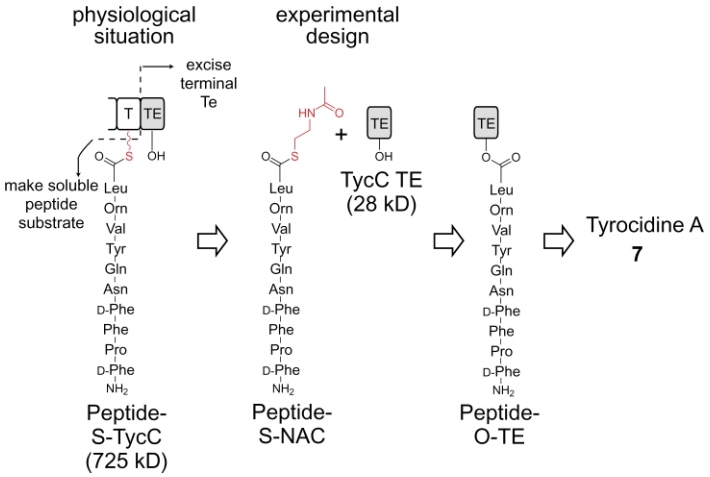


Fig. 7 The experimental design for the study of TE domains excised from their natural synthetases. A soluble peptide thioester substrate serves as a surrogate for the carrier protein tethered linear acyl intermediate. The TE domain is cloned and expressed as an isolated domain for characterization.

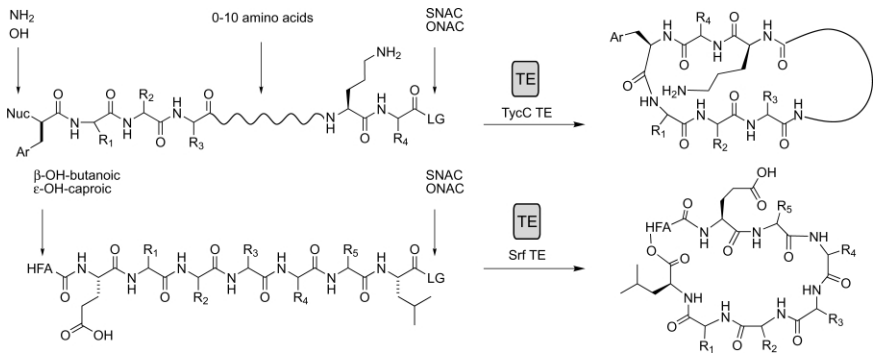
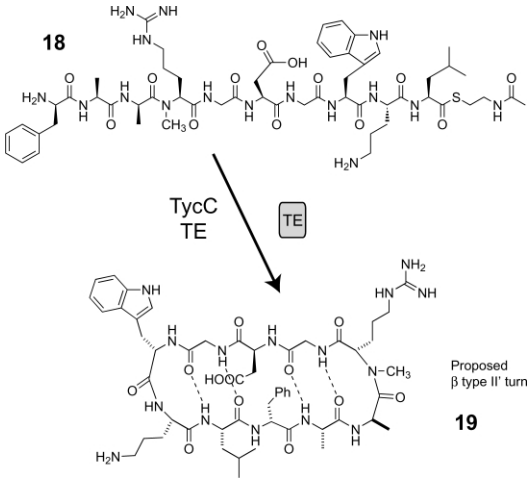


Fig. 8 The versatility of TE catalyzed macrocyclization. The substrates shown represent the predictions for minimal substrate requirements for macrocyclization based upon biochemical studies. Variations which allow for cyclization to hydrolysis ratios that are greater than 1 : 1 or that represent a less than five-fold reduction in rate of cyclization are shown. Nuc = Nucleophile, Ar = aromatic side chain, R = any natural or non-natural side chains, LG = leaving group, HFA = hydroxy fatty acid.

residues of the decapeptide, as well as several changes in stereochemistry, and *N*-methylation of a single residue, it was cyclized by TycC TE, albeit in lower yield (Scheme 6). The



Scheme 6 Cyclization of integrin binding peptides by tyrocidine TE domain.

chemoenzymatic synthesis could be scaled up to allow sufficient cyclic product, **19**, to be isolated to demonstrate nanomolar inhibition of the platelet integrin $\alpha_{IIb}\beta_{IIIa}$. The conformational constraint of cyclic peptide increased its potency relative to a linear analog. Presumably one could optimize the sequence of such chemoenzymatically generated cyclic peptides to optimize potency and integrin receptor selectivity.

Srf TE

The Srf TE is the most downstream domain in the seven module surfactin synthetase assembly line (Fig. 4) for production of the *N*-acylated lipoheptapeptide that has been macrocyclized between the β-hydroxyl substituent of the fatty acyl chain and the Leu7 carbonyl.[12,75] This TE when excised has been crystallized as noted above and retains macrolactonization activity, with stereochemical discrimination for the 3*R*-OH of 3-hydroxybutyryl-heptapeptidyl SNAC substrate.[65,70] The acyclic precursor is poorly soluble and installation of 2,3-diaminopropionate (Dap) residues impart increased water solubility, allowing studies to probe requirements in the three portions of substrate: heptapeptide, fatty acyl group and the leaving group (Fig. 8).[76] In the peptidic portion, substitution of the C-terminal most residues by Dap affected cyclization and hydrolysis, suggesting the importance of the C-terminal residues in formation of the acyl-enzyme intermediate. In contrast, alteration in the central residues of the heptapeptide is tolerable to the TE. In the fatty acid moiety, a β-amino butyryl analog was hydrolyzed but not cyclized despite the likelihood that iturins and mycosubtilin **9**, arise by such an attack.[9] Nor was an α-hydroxybutyryl acyl moiety competent to cyclize, although an ε-hydroxycaproyl heptapeptide did undergo cyclization, so lactone ring size can be expanded. The SNAC leaving group could be replaced by an ONAC, indicating the oxoester substrate was still sufficiently activated for acyl-O-TE formation. The TE would also carry out the back reaction, hydrolysis of the surfactin macrolactone scaffold.

Mutation of the Ser and His residues of the presumed Srf TE catalytic triad inactivated both cyclization and hydrolysis activities.[76] In the structurally related lipases, there is a conserved glycine in the active site that is instead a proline in the NRPS TE domains.[70] Mutation of Pro26 to Gly in Srf TE changed the partition ratio of cyclization to hydrolysis by 10-fold in favor of hydrolysis, a lipase-like product outcome, suggesting the conformation constraint imposed by Pro26 in TE domains may favor a sequestered active site and a longer life time for the peptidyl-O-TE acyl enzyme intermediates to undergo intramolecular cyclization (lactamization, lactonization) over competing hydrolysis.[76]

Applications of TycC TE to synthesis of cyclic peptide libraries

Macrocyclization by chemical means is often a formidable synthetic challenge. As seen in the storied example of synthesis of erythromycin by Woodward, with subtle variations in a linear precursor conditions that allow for cyclization, if they can be found, can vary widely and unpredictably.[77] The study of their enzymology has suggested that the thioesterase domains from NRPS and PKS sources have potential utility in the synthesis of cyclic molecule libraries, particularly in synthesis of natural product analogs.

Combinatorial synthesis has benefited greatly from the use of solid-phase chemistry, where molecules constructed on the solid phase can be split into different reaction conditions to create diversity and as it obviates the need for purification steps.[78] Nonribosomal peptide synthesis is essentially a biological equivalent of solid-phase peptide synthesis, where a carrier protein with its phosphopantetheine tether substitutes for a solid-phase resin.[32] Building upon this recognition, a solid-phase polyethylene glycol amide resin was derivatized with a mimic of the phosphopantetheine arm terminating in a free hydroxyl (Fig. 9). When the linear sequence corresponding to

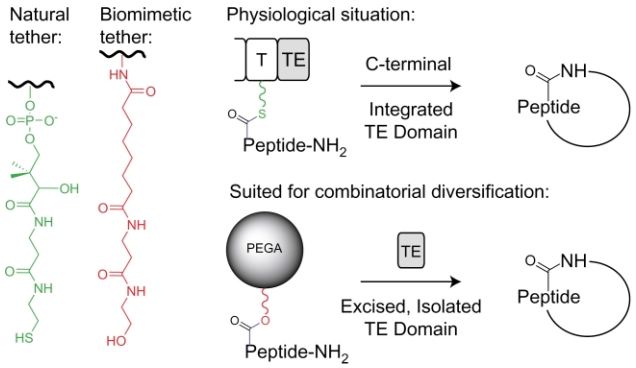
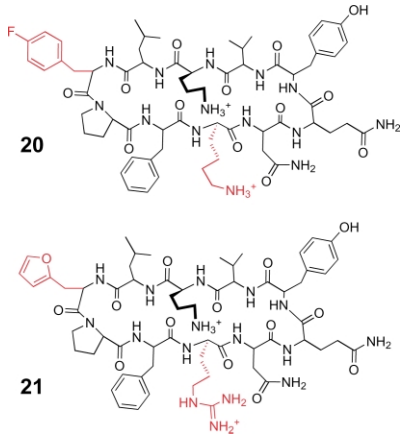


Fig. 9 The strategy of biomimetic macrocyclic synthesis. In the natural biosynthesis the acyl chain is linked to a carrier protein (T) via a phosphopantetheine tether (green). In biomimetic macrocycle synthesis, a solid-phase PEGA resin functionalized with a synthetic tether (red) substitutes for the natural carrier protein. Solid-phase peptide synthesis is used in place of nonribosomal peptide synthesis and allows for access to variants of the linear sequence by solid-phase combinatorial techniques. The isolated TE domain then catalyzes release from the solid-phase tether and macrocyclization generating the natural product and analogs.

tyrocidine was synthesized on the resin and incubated with the excised TycC TE, the enzyme could productively catalyze release from the resin and cyclization in a reaction equivalent to the physiological release from the carrier protein tether and cyclization (Fig. 9).[79] This system allowed for the synthesis of libraries of linear substrates on the solid-phase, followed by enzymatic cyclization to generate cyclic peptide libraries. By synthesis of a set of linear molecules in which the D-Phe4 position was replaced by one of 96 natural or non-natural amino acids, the method was utilized to define the substrate specificity of TycC TE.

In addition to being an enzymological tool, the method allowed for the study of the bioactivity of the library of natural product analogs. Tyrocidine A, a member of the family of cationic antimicrobial peptides, is a membranolytic antibiotic that normally shows minimal discrimination between bacterial

and eukaryotic cell membranes. In an effort to increase selectivity, a matrix of peptides with variations in two positions were synthesized and subjected to enzymatic cyclization. The products of the reaction were tested for their antibiotic activity against *Bacillus subtilis* and in the hemolysis of red blood cells. The library of products revealed that substitution of D-Phe4 in the natural product by a positively charged D-amino acid led to 30-fold selectivity for bacterial membranes.[79] The cyclizing enzyme could be used in chemoenzymatic synthesis of two of the best analogs, **20** and **21**, on preparative scale (Scheme 7).

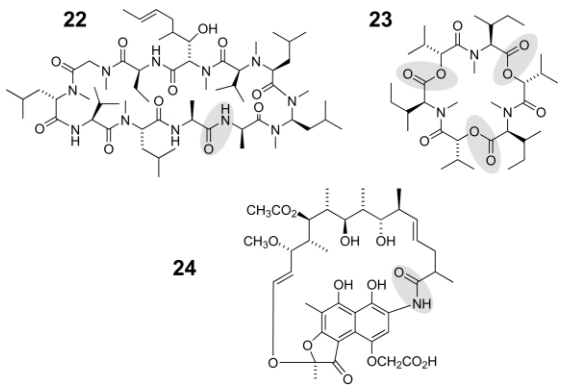


Scheme 7 Improved analogs of tyrocidine A with increased specificity for bacterial membranes synthesized by biomimetic macrocycle synthesis.

When compared to tyrocidine A, in addition to increased specificity, these analogs gained activity against gram negative organisms including *Pseudomonas aeruginosa*. This combinatorial method for generation of libraries of macrocycles with novel or improved characteristics can be translated back to biosynthetic systems as well. Once a target analog is unveiled by this method, the modular nature of synthetases makes it possible to consider reprogramming or swapping modules to generate the target analog, utilizing the terminal TE to catalyze chain termination and macrocyclization.[34,80,81]

Alternative macrocyclization strategies

While C-terminal TE domains represent the most common solution to the catalytic disconnection and macrocyclization of linear acyl chains in NRP and PK biosynthesis, alternative strategies are known (Scheme 8). In NRP synthetases, one



Scheme 8 Products of alternative macrocyclization catalysts: cyclosporin A (**22**) utilizes a terminal C domain, enniatin A (**23**) synthesized *via* a terminal T–C didomain and rifamycin B (**24**) synthesized by an arylamine *N*-acetyltransferase homologue.

alternative strategy that draws from the catalytic toolbox used in construction of the linear acyl chains is proposed to be at work in the biosynthesis of many macrocyclic peptides from fungal sources. Sequencing of the synthetase responsible for the production of cyclosporin A **22** revealed several remarkable features including the presence of a condensation domain, in place of a TE domain, at the C-terminal end of the massive synthetase.[82] The linear cyclosporin undecapeptide precursor is constructed by the eleven contiguous modules, followed by intramolecular capture of the C-terminal L-Ala11 of the peptide by the N-terminal D-Ala1, giving the macrocyclic product. The normal function of a condensation domain is to catalyze the attack of a tethered amino acid on a downstream T domain upon the thioester tethered peptide on an upstream domain. C domains are also known to use soluble amine nucleophiles as in vibriobactin biosynthesis where a soluble norspermidine is linked to a tethered dihydroxybenzoate.[42,43,83] In the synthesis of cyclosporin A, the terminal C domain is proposed to function by analogy to the latter reaction where the nucleophile, rather than coming from a soluble amine, is the N-terminal amine of D-Ala1. In the synthesis of the symmetric cyclohexadepsipeptide enniatin **23**, two modules responsible for the synthesis of D-2-hydroxyisovaleryl-*N*-Me-Val are followed by a terminal T–C didomain, giving the unusual T–T–C domain organization at the C-terminus.[84] While in cyclooligomerization catalyzed by TE domains, the single TE domain functions both as a way-station for intermediates (acyl-*O*-TE) and as the catalyst for cyclooligomerization, in enniatin biosynthesis these responsibilities are presumed to be divided, with the T domain functioning in the former capacity while the C domain functions as the cyclooligomerization catalyst. As yet, biochemical evidence for these functions is lacking, and it will be interesting to explore if, by full analogy, isolated terminal C domains will be versatile catalysts as terminal TE domains have proven to be.

Polyketide biosynthetic pathways have been known to reach outside of the usual catalytic toolbox for macrocyclization of linear precursors. In the biosynthesis of the antimycobacterial ansamycin family member rifamycin B **24**, the termination module of the synthetase lacks a TE domain.[35,36] The initiation module of the PKS loads 3-amino-5-hydroxybenzoate (AHB) and the synthase constructs the linear polyketide capped with the arylamine AHB. Immediately downstream of the gene encoding the final synthetase protein is *rifF*, which encodes a protein with homology to arylamine *N*-acetyltransferases. A genetic knockout of *rifF* resulted in the production of several linear biosynthetic intermediates along the pathway to the final linear intermediate, consistent with the suggestion that *rifF* is the macrocyclization and chain termination catalyst.[85,86] The stand-alone enzyme is proposed to form a bond between the 3-amino group derived from AHB and the carrier protein tethered end of the linear polyketide precursor. The structure of an arylamine *N*-acetyltransferase family member revealed the presence of a cysteine protease-like catalytic triad, suggesting a mechanism for catalysis analogous to TE mediated cyclization for *rifF*.[87] The genetic evidence is further supported by similar biosynthetic schemes in the cluster for the anti-tumor ansamitocin and antibiotic napthomycin, where the synthetase proteins lack terminal TE domains but have separate genes with homology to *rifF*.[88,89] While biochemical evidence for the function of the arylamine amide synthases is lacking, the recent heterologous expression and purification of the product of *rifF* may soon open the door to such characterization.[90]

Versatility of enzymatic macrocyclization reactions

In nature, many biologically active molecules are initially synthesized as linear molecules and subsequently constrained by covalent bonds bridging distant parts. We have examined macrocyclization of polyketide and nonribosomal peptide natural products in detail.

The TE domains embedded at the downstream ends of NRPS, PKS and hybrid NRPS/PKS assembly lines are versatile and efficient enzymatic catalysts. Their physiologic function is to act on the full length natural product acyl chain that has arrived at the adjacent thiolation domain (Fig. 2) to disconnect the covalent linkage between the acyl chain and the phosphopantetheinyl tether on the T domain. The catalytic disconnection can use an external nucleophile, most often a water molecule in the active site, but also amine nucleophiles can be presented on cosubstrate binding, as in vibriobactin assembly.

The catalytic disconnection can also be effected intramolecularly with a range of chemospecificity, regiospecificity and stereospecificity in the disconnection. The product-determining step occurs after transfer of the peptidyl or polyketide acyl chain from the phosphopantetheinyl thiol tether to the active site serine in the TE domains. For cyclization, the active site of the TE must exclude water molecules from achieving a distance, orientation, and activation where they would be competent nucleophiles. Additionally, they must promote the folding of acyclic conformers of the peptidyl or polyketidyl chain to allow productive approach of a specific OH or NH_2 in the acyl chain to the carbonyl group lodged on the active site serine residue. The intramolecular nucleophiles are derived from various sources, the N-terminus of a peptide, side chains of natural or non-natural amino acids, fatty acids, or ketide hydroxyls, producing diverse natural cyclic and lariat structures.

Initial exploration of TE domains indicates they are portable as long as they are maintained adjacent to T domains elsewhere in PKS and NRPS assembly lines. When integrated TE domains are excised and studied as isolated domains, they retain their catalytic capacity. For thioesterase domains studied thus far, the versatility of the cyclization catalyst is evident in several regards. The substrate may be an acyl chain tethered to a carrier protein, as in the physiological situation, may be on a biomimetic linker on a solid-phase resin or may be a soluble peptide thioester or ester. The acyl chain itself allows for variation such that the macrocycles that can be accessed include natural products, natural product analogs with optimized properties, products of cyclooligomerization or novel products which act on unrelated therapeutic targets. It is possible that as a result of having evolved with their natural substrate delivered *in cis*, these TE domains while facing evolutionary pressure to exclude water and to precisely control the regiochemistry of cyclization, have not been compelled to exacting specificity in their choice of acyl chain. With ever increasing sequence data on secondary metabolite biosynthetic clusters, the structural diversity of the reactions they catalyze and their inherent versatility makes exploration of new natural product macrocyclization catalysts an intriguing route to accessing novel structures.[80]

Acknowledgements

We thank Steven D. Bruner for assistance with figures. Financial support has been provided by the National Institutes of Health and the Medical Scientist Training Program.

Notes and references

1 J. Rizo and L. M. Gierasch, *Annu. Rev. Biochem.*, 1992, **61**, 387–418.
2 R. H. Lambalot and D. E. Cane, *J. Antibiot. (Tokyo)*, 1992, **45**, 1981–1982.
3 S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science*, 1991, **252**, 675–679.
4 J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt and P. F. Leadlay, *Nature*, 1990, **348**, 176–178.
5 E. Cundliffe, N. Bate, A. Butler, S. Fish, A. Gandecha and L. Merson–Davies, *Antonie Van Leeuwenhoek*, 2001, **79**, 229–234.
6 R. Roskoski Jr., H. Kleinkauf, W. Gevers and F. Lipmann, *Biochemistry*, 1970, **9**, 4846–4851.
7 J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh, *Nature*, 2000, **407**, 215–218.
8 D. Konz, A. Klens, K. Schorgendorfer and M. A. Marahiel, *Chem. Biol.*, 1997, **4**, 927–937.
9 E. H. Duitman, L. W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt, C. Ullrich, T. Stein, F. Leenders and J. Vater, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 13294–13299.
10 M. A. McHenney, T. J. Hosted, B. S. Dehoff, P. R. Rosteck Jr. and R. H. Baltz, *J. Bacteriol.*, 1998, **180**, 143–151.
11 F. Parenti, R. Ciabatti, B. Cavalleri and J. Kettenring, *Drugs Exp. Clin. Res.*, 1990, **16**, 451–455.
12 P. Cosmina, F. Rodriguez, F. de Ferra, G. Grandi, M. Perego, G. Venema and D. van Sinderen, *Mol. Microbiol.*, 1993, **8**, 821–831.
13 J. H. Crosa and C. T. Walsh, *Microbiol. Mol. Biol. Rev.*, 2002, **66**, 223–249.
14 L. E. Quadri, *Mol. Microbiol.*, 2000, **37**, 1–12.
15 A. M. Gehring, I. Mori and C. T. Walsh, *Biochemistry*, 1998, **37**, 2648–2659.
16 J. J. May, T. M. Wendrich and M. A. Marahiel, *J. Biol. Chem.*, 2001, **276**, 7209–7217.
17 K. Hori, Y. Yamamoto, T. Minetoki, T. Kurotsu, M. Kanda, S. Miura, K. Okamura, J. Furuyama and Y. Saito, *J. Biochem. (Tokyo)*, 1989, **106**, 639–645.
18 M. Trabi and D. J. Craik, *Trends. Biochem. Sci.*, 2002, **27**, 132–138.
19 M. Martínez–Bueno, M. Maqueda, A. Galvez, B. Samyn, J. Van Beeumen, J. Coyette and E. Valdivia, *J. Bacteriol.*, 1994, **176**, 6334–6339.
20 A. Blond, J. Peduzzi, C. Goulard, M. J. Chiuchiolo, M. Barthelemy, Y. Prigent, R. A. Salomon, R. N. Farias, F. Moreno and S. Rebuffat, *Eur. J. Biochem.*, 1999, **259**, 747–755.
21 D. J. Craik, N. L. Daly, T. Bond and C. Waine, *J. Mol. Biol.*, 1999, **294**, 1327–1336.
22 Y. Q. Tang, J. Yuan, G. Osapay, K. Osapay, D. Tran, C. J. Miller, A. J. Ouellette and M. E. Selsted, *Science*, 1999, **286**, 498–502.
23 A. M. Cole, T. Hong, L. M. Boo, T. Nguyen, C. Zhao, G. Bristol, J. A. Zack, A. J. Waring, O. O. Yang and R. I. Lehrer, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1813–1818.
24 D. E. Cane and C. T. Walsh, *Chem. Biol.*, 1999, **6**, R319–325.
25 L. Du, C. Sanchez and B. Shen, *Metab. Eng.*, 2001, **3**, 78–95.
26 I. Molnar, T. Schupp, M. Ono, R. Zirkle, M. Milnamow, B. Nowak–Thompson, N. Engel, C. Toupet, A. Stratmann, D. D. Cyr, J. Gorlach, J. M. Mayo, A. Hu, S. Goff, J. Schmid and J. M. Ligon, *Chem. Biol.*, 2000, **7**, 97–109.
27 N. Bamas–Jacques, S. Lorenzon, P. Lacroix, C. De Swetschin and J. Crouzet, *J. Appl. Microbiol.*, 1999, **87**, 939–948.
28 J. Staunton and B. Wilkinson, *Curr. Opin. Chem. Biol.*, 2001, **5**, 159–164.
29 M. A. Marahiel, T. Stachelhaus and H. D. Mootz, *Chem. Rev.*, 1997, **97**, 2651–2674.
30 T. A. Keating and C. T. Walsh, *Curr. Opin. Chem. Biol.*, 1999, **3**, 598–606.
31 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.
32 C. T. Walsh, A. M. Gehring, P. H. Weinreb, L. E. Quadri and R. S. Flugel, *Curr. Opin. Chem. Biol.*, 1997, **1**, 309–315.
33 T. A. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger and C. T. Walsh, *Chembiochem.*, 2001, **2**, 99–107.
34 D. E. Cane, C. T. Walsh and C. Khosla, *Science*, 1998, **282**, 63–68.
35 P. R. August, L. Tang, Y. J. Yoon, S. Ning, R. Muller, T. W. Yu, M. Taylor, D. Hoffmann, C. G. Kim, X. Zhang, C. R. Hutchinson and H. G. Floss, *Chem. Biol.*, 1998, **5**, 69–79.
36 H. G. Floss and T. W. Yu, *Curr. Opin. Chem. Biol.*, 1999, **3**, 592–597.
37 C. R. Hutchinson, *Chem. Rev.*, 1997, **97**, 2525–2536.
38 K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van den Heever, C. R. Hutchinson and J. Vederas, *J. Am. Chem. Soc.*, 2000, **112**, 11519–11520.
39 M. Nardini and B. W. Dijkstra, *Curr. Opin. Struct. Biol.*, 1999, **9**, 732–737.
40 J. Li, R. Szittner, Z. S. Derewenda and E. A. Meighen, *Biochemistry*, 1996, **35**, 9967–9973.

41 C. A. Shaw–Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh, *Chem. Biol.*, 1999, **6**, 385–400.
42 T. A. Keating, C. G. Marshall and C. T. Walsh, *Biochemistry*, 2000, **39**, 15513–15521.
43 T. A. Keating, C. G. Marshall and C. T. Walsh, *Biochemistry*, 2000, **39**, 15522–15530.
44 F. Takusagawa, *J. Antibiot. (Tokyo)*, 1985, **38**, 1596–1604.
45 M. Kuse, T. Franz, K. Koga, S. Suwan, M. Isobe, N. Agata and M. Ohta, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 735–739.
46 R. A. Carter, P. S. Worsley, G. Sawers, G. L. Challis, M. J. Dilworth, K. C. Carson, J. A. Lawrence, M. Wexler, A. W. Johnston and K. H. Yeoman, *Mol. Microbiol.*, 2002, **44**, 1153–1166.
47 D. A. Hopwood and D. H. Sherman, *Annu. Rev. Genet.*, 1990, **24**, 37–66.
48 S. J. Wakil, *Biochemistry*, 1989, **28**, 4523–4530.
49 I. Fujii and Y. Ebizuka, *Chem. Rev.*, 1997, **97**, 2511–2524.
50 M. Pazirandeh, S. S. Chirala, W. Y. Huang and S. J. Wakil, *J. Biol. Chem.*, 1989, **264**, 18195–18201.
51 J. S. Mattick, J. Nickless, M. Mizugaki, C. Y. Yang, S. Uchiyama and S. J. Wakil, *J. Biol. Chem.*, 1983, **258**, 15300–15304.
52 M. H. Tai, S. S. Chirala and S. J. Wakil, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 1852–1856.
53 L. A. Merson-Davies and E. Cundliffe, *Mol. Microbiol.*, 1994, **13**, 349–355.
54 H. D. Mootz and M. A. Marahiel, *J. Bacteriol.*, 1997, **179**, 6843–6850.
55 A. R. Butler, N. Bate and E. Cundliffe, *Chem. Biol.*, 1999, **6**, 287–292.
56 A. Schneider and M. A. Marahiel, *Arch. Microbiol.*, 1998, **169**, 404–410.
57 M. L. Heathcote, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 207–220.
58 J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science*, 1995, **268**, 1487–1489.
59 C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1995, **117**, 9105–9106.
60 C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1996, **118**, 9184–9185.
61 Y. Xue and D. H. Sherman, *Nature*, 2000, **403**, 571–575.
62 C. Khosla, R. S. Gokhale, J. R. Jacobsen and D. E. Cane, *Annu. Rev. Biochem.*, 1999, **68**, 219–253.
63 H. D. Mootz, D. Schwarzer and M. A. Marahiel, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 5848–5853.
64 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.
65 R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2001, **40**, 7099–7108.
66 R. Aggarwal, P. Caffrey, P. F. Leadlay, C. J. Smith and J. Staunton, *J. Chem. Soc., Chem. Commun.*, 1995, 1519–1520.
67 K. J. Weissman, C. J. Smith, U. Hanefeld, R. Aggarwal, M. Bycroft, J. Staunton and P. F. Leadlay, *Angew. Chem., Int. Ed.*, 1998, **37**, 1437–1440.
68 C. Khosla, personal communication.
69 S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 14808–14813.
70 S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs, *Structure (Camb)*, 2002, **10**, 301–310.
71 S. C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598–12606.
72 H. Lu, S. C. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590–12597.
73 J. W. Trauger, R. M. Kohli and C. T. Walsh, *Biochemistry*, 2001, **40**, 7092–7098.
74 R. M. Kohli, J. Takagi and C. T. Walsh, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1247–1252.
75 F. Peypoux, J. M. Bonmatin and J. Wallach, *Appl. Microbiol. Biotechnol.*, 1999, **51**, 553–563.
76 C. C. Tseng, S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh and S. A. Sieber, *Biochemistry*, 2002, **41**, 13350–13359.
77 R. B. Woodward, E. Logusch, K. P. Nambiar, K. Sakan, D. E. Ward, B.-W. Au-Yeung, P. Balaram, L. J. Browne, P. J. Card, C. H. Chen, R. B. Chênevert, A. Aliri, K. Froebel, H.-J. Gais, D. G. Garratt, K. Hayakawa, W. Heggie, D. P. Hesson, D. Hoppe, I. Hoppe, J. A. Hyatt, D. Ikeda, P. A. Jacobi, K. S. Kim, Y. Kobulke, K. Kojima, K. Krowicki, V. J. Lee, T. Leutert, S. Malchenko, J. Martens, R. S. Matthews, B. S. Ong, J. B. Press, T. V. Rajan Babu, G. Rosseau, H. M. Sauter, M. Suzuki, K. Tatsuta, L. M. Tolbert, E. A. Truesdale, I. Uchida, Y. Ueda, T. Uyehara, A. T. Vasella, W. C. Vladuchick, P. A. Wade, R. M. Williams and H. N.-C. Wong, *J. Am. Chem. Soc.*, 1981, **103**, 3213–3215.
78 P. Arya, R. Joseph and D. T. Chou, *Chem. Biol.*, 2002, **9**, 145–156.
79 R. M. Kohli, C. T. Walsh and M. D. Burkart, *Nature*, 2002, **418**, 658–661.
80 C. T. Walsh, *Chembiochem*, 2002, **3**, 125–134.
81 S. Doekel and M. A. Marahiel, *Metab. Eng.*, 2001, **3**, 64–77.
82 G. Weber, K. Schorgendorfer, E. Schneider-Scherzer and E. Leitner, *Curr. Genet.*, 1994, **26**, 120–125.
83 T. A. Keating, C. G. Marshall, C. T. Walsh and A. E. Keating, *Nat. Struct. Biol.*, 2002, **9**, 522–526.
84 A. Haese, M. Schubert, M. Herrmann and R. Zocher, *Mol. Microbiol.*, 1993, **7**, 905–914.
85 A. Stratmann, C. Toupet, W. Schilling, R. Traber, L. Oberer and T. Schupp, *Microbiology*, 1999, **145**, 3365–3375.
86 T. W. Yu, Y. Shen, Y. Doi-Katayama, L. Tang, C. Park, B. S. Moore, C. Richard Hutchinson and H. G. Floss, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 9051–9056.
87 J. C. Sinclair, J. Sandy, R. Delgoda, E. Sim and M. E. Noble, *Nat. Struct. Biol.*, 2000, **7**, 560–564.
88 T. W. Yu, L. Bai, D. Clade, D. Hoffmann, S. Toelzer, K. Q. Trinh, J. Xu, S. J. Moss, E. Leistner and H. G. Floss, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 7968–7973.
89 S. Chen, D. von Bamberg, V. Hale, M. Breuer, B. Hardt, R. Muller, H. G. Floss, K. A. Reynolds and E. Leistner, *Eur. J. Biochem.*, 1999, **261**, 98–107.
90 F. Pompeo, A. Mushtaq and E. Sim, *Protein Expr. Purif.*, 2002, **24**, 138–151.