Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms

Michael A. Fischbach[†,‡] and Christopher T. Walsh*[,†]

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115 and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received November 4, 2005

Contents

1. Introduction

Many thousands of small molecule natural products with interesting biological activities are polyketides (PK), nonribosomal peptides (NRP), or hybrids thereof (i.e., PK−NRP hybrids). Examples include the polyketide antibiotics erythromycin[1] and tetracycline,[2] the nonribosomal peptide antibiotics of the penicillin and cephalosporin families,[3,4] as well as the glycopeptides of the vancomycin family[5] (Figure 1a).

Christopher T. Walsh is the Hamilton Kuhn Professor of Biological Chemistry and Molecular Pharmacology (BCMP) at Harvard Medical School. He has had extensive experience in academic administration, including Chairmanship of the MIT Chemistry Department (1982–1987) and the HMS Biological Chemistry & Molecular Pharmacology Department (1987–1995) as well as serving as President and CEO of the Dana Farber Cancer Institute (1992–1995). His research has focused on enzymes and enzyme inhibitors, with recent specialization on antibiotics. He and his group have authored over 590 research papers, a text (*Enzymatic Reaction Mechanisms*), and two books (*Antibiotics: Origins, Actions, Resistance* and *Posttranslational Modification of Proteins: Expanding Nature's Inventory*). He is a member of the National Academy of Sciences, the Institute of Medicine, and the American Philosophical Society.

Michael A. Fischbach attended Harvard College, where he received his B.A. degree in 2003. During that time, he worked in the laboratory of Jeffrey Settleman at the Massachusetts General Hospital Cancer Center, where his research focused on the biochemistry of the small GTPase Ras. As a Hertz Foundation predoctoral fellow, he is currently in the Department of Chemistry and Chemical Biology at Harvard University, where he works in the laboratories of Christopher T. Walsh and David R. Liu. He is studying enzymes that tailor the nonribosomal peptide siderophore enterobactin and pursuing evolution-based strategies to engineer nonribosomal peptide synthetases.

* To whom correspondence should be addressed: christopher_walsh@hms.harvard.edu.
† Harvard Medical School.
‡ Harvard University.

10.1021/cr0503097 CCC: $59.00
Published on Web 07/19/2006

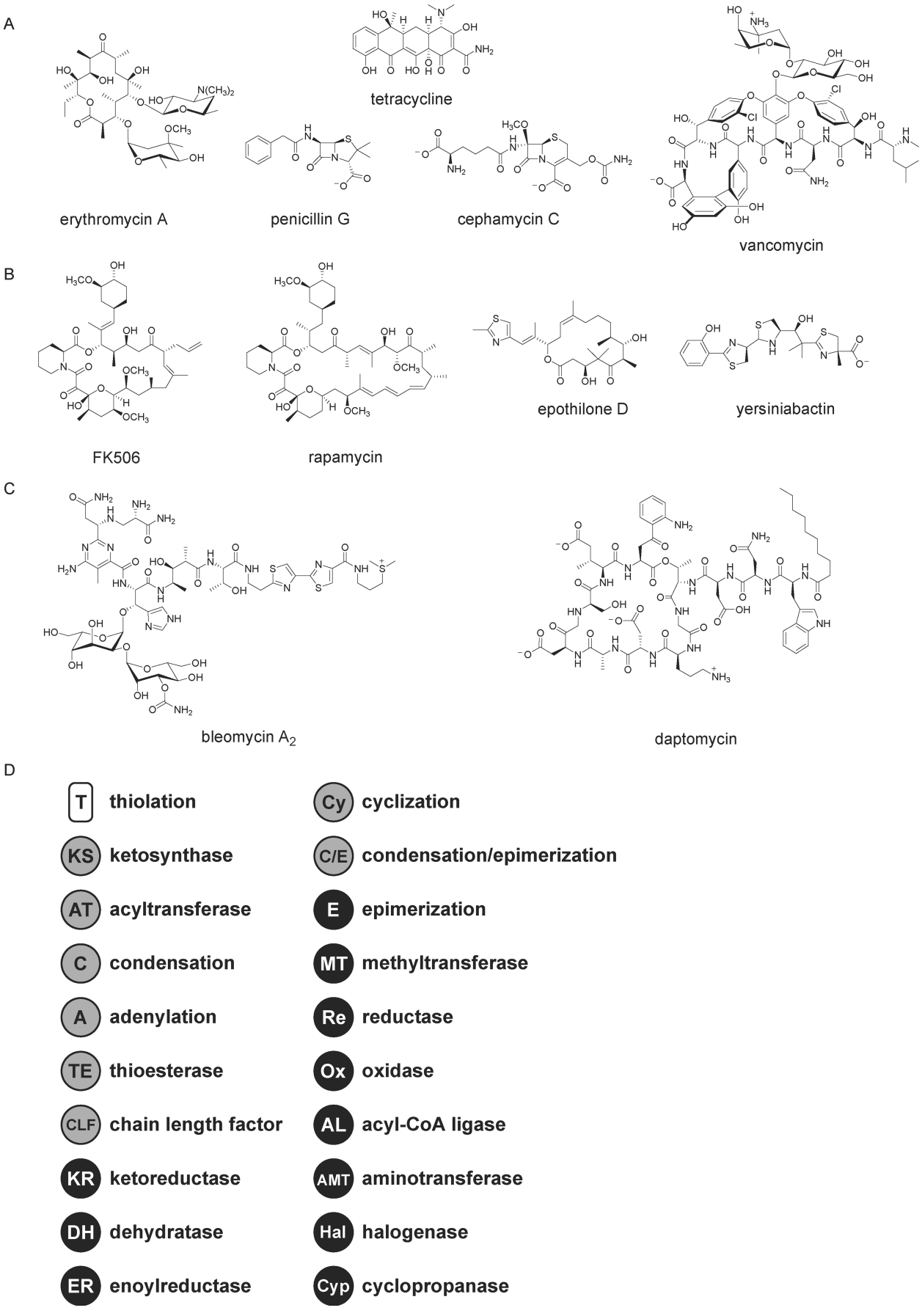

Figure 1. Therapeutically relevant PKs, NRPs, and PK−NRP hybrids. (A) The PK antibiotics erythromycin A and tetracycline and the NRP antibiotics penicillin G, cephamycin C, and vancomycin. (B) The PK−NRP hybrids FK506, rapamycin, epothilone, and yersiniabactin. (C) The *N*-acylated NRP daptomycin, and the *C*-aminated PK−NRP bleomycin. (D) Legend for the protein domains shown in subsequent figures.

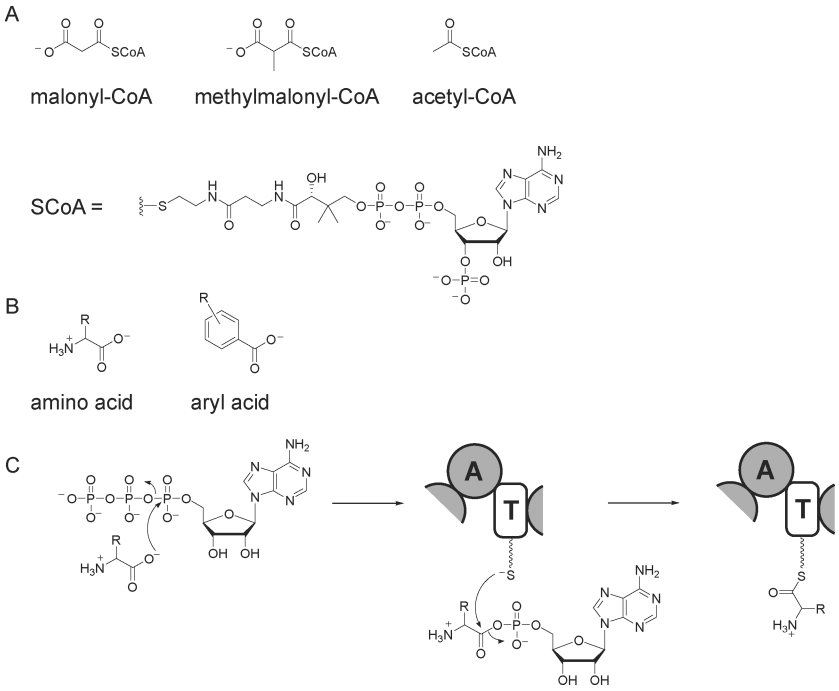

Figure 2. Use of thioester monomers by PKSs and NRPSs. (A) Common acyl-CoA thioesters used for chain initiation and elongation by PKSs. (B) Common amino and aryl acid monomers used for chain initiation and elongation by NRPSs. (C) Formation of pantetheinyl thioester from aminoacyl-AMP by NRPSs.

Medicinally relevant hybrids include the immunosuppressant drugs FK506[6,7] and rapamycin,[8,9] the antitumor agents of the bleomycin[10] and epothilone[11] families, and the *Yersinia pestis* iron chelator yersiniabactin[12] (Figure 1b).

The remarkable structural and functional diversity in polyketides arises from combinatorial utilization and template-directed elongation of only a few simple building blocks. Malonyl-CoA and methylmalonyl-CoA comprise the great bulk of monomer units incorporated during chain elongation. The chain starter units can be the thioesters of monoacyl groups such as acetyl-, propionyl-, and benzoyl-CoAs, or structural variants, such as malonamyl-CoA or methoxymalonyl-CoA.[13] For nonribosomal peptides, the 20 proteinogenic amino acids and a much wider variety of nonproteinogenic amino and aryl acids are the monomer building blocks for oligomerization and diversification during chain elongation and after chain termination.[14] NRP chains can be *N*-capped with acyl groups such as the long-chain β-OH fatty acid in daptomycin[15] (Figure 1c) or *C*-capped with amines as in bleomycin.[10]

The unique structures of thousands of PK, NRP, and NRP−PK hybrid products arise from coordinated, multistep action of enzymes organized in assembly lines. The identity and order of each protein domain in an assembly line specifies (i) the sequence of monomer units activated and incorporated, (ii) the chemistry that occurs at each way station in the assembly line, and (iii) the length and functionality of the product released from the distal end of the assembly line.

The genes for hundreds of polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), and hybrid PKS−NRPS assembly lines have been sequenced[16,17] in the 240 microbial genomes reported to date.[18] This genetic information has revealed that PKS and NRPS genes are organized in clusters,[19] perhaps reflecting coordinate regulation to activate the secondary metabolic pathway on demand and possibly suggesting horizontal gene transfer of these clusters between microbial genomes to acquire useful small molecule synthetic capabilities.[20] The clustering of biosynthetic genes facilitates the cloning of complete pathways. These genetic forays, coupled with biochemical studies on purified and/or reconstituted enzymatic components, have deciphered the logic, machinery, and mechanisms of the PKS and NRPS enzymatic assembly lines over the past 20 years. This basic knowledge of PKS and NRPS systems provides context for more efficient efforts in combinatorial biosynthesis to create collections of natural product variants with novel structure and function. At points during this review, we have taken examples and references from recent papers and reviews; for the interested reader, we intend these to serve as entry points into the founding literature.

2. Logic of Assembly-Line Enzymology

2.1. Chemical Logic of Monomer Activation and Oligomerization

The chemical logic for assembly of polyketides and nonribosomal peptides is that a small set of monomer units are incorporated into a linear oligomer by iterative chemical condensation steps. The monomers are acyl units, with the carboxyl group activated for capture by nucleophiles. The first central chemical tenet is that *thioesters* are used for activating the acyl group of each monomer. In PKS assembly lines, the monomers are acyl-CoA thioesters (e.g., acetyl-CoA, malonyl-CoA, methylmalonyl-CoA), readily conscripted from the pool of primary metabolites in microbial producer cells (Figure 2a). The monomers for NRPS assembly are proteinogenic and nonproteinogenic amino acids and other carboxylic acids (e.g., aryl acids) (Figure 2b). While some of these monomers are primary metabolites, many require a dedicated set of enzymes to divert a primary metabolite toward their production.

Comparable aminoacyl-CoAs are not known to exist in biological systems. Instead, the amino acid monomers that get incorporated by NRPS assembly lines are aminoacyl-

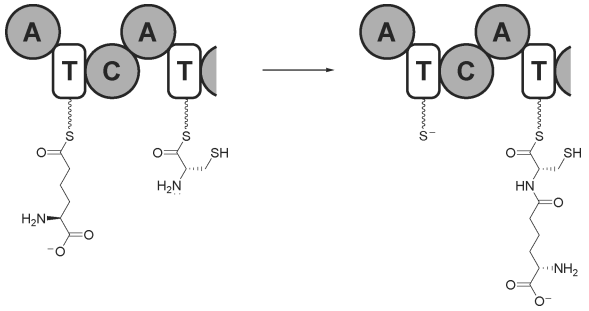

Figure 3. Chain elongation during ACV biosynthesis. The covalently tethered acyl donor, aminoadipoyl-*S*-T_1, is transferred to the covalently tethered acyl acceptor, cysteinyl-*S*-T_2, forming aminoadipoyl-cysteinyl-*S*-T_2. C = condensation, A = adenylation, T = thiolation.

AMP mixed anhydrides. This strategy mirrors the activation logic for amino acids in ribosomal protein synthesis. In every NRPS module in which an aminoacyl-AMP is selected for incorporation, thermodynamic activation is preserved by transfer of the aminoacyl group to the thiolate anion of a pantetheinyl-phospho-carrier protein domain (Figure 2c). Convergent logic is followed in PKS systems in which malonyl-CoA undergoes transthiolation to yield a comparable malonyl-pantetheinyl-phospho-carrier protein before any condensation steps occur.

The second central tenet is that chain elongation proceeds via *covalent tethering* of all the substrates, intermediates, and products.[21,22] Carrier protein or thiolation (T) domains, which bear a covalently attached phosphopantetheinyl "arm",[23,24] are the points of attachment of the activated monomers and growing acyl chains. For enzymatic assembly lines to become functional, each T domain must be post-translationally modified with a phosphopantetheinyl arm, derived from coenzyme A (see Figure 14). Since fatty acid synthases (FASs) use the same chemical logic as PKSs for monomer activation and chain elongation, these multimodular enzymes also employ tethered phosphopantetheinyl moieties and covalent attachment of the growing acyl chain from one elongation cycle to the next.

For example, in the first condensation catalyzed by the tripeptide synthetase that makes aminoadipoyl-cysteinyl-valine (ACV), the immediate precursor of the β-lactam antibiotic scaffold,[25] aminoadipoyl-*S*-T_1 is the donor and cysteinyl-*S*-T_2 the acceptor for catalyzed peptide bond formation (Figure 3) that yields the deacylated T_1-S^- and the dipeptidyl-*S*-T_2. Iterative condensations use the same chemical logic for each step of chain extension.

Whereas NRPS elongations involve C−N bond formation as an amide (peptide) link is forged in each condensation step, PKS elongations form C−C bonds via Claisen condensations. During peptide bond formation, the monomer attached to the downstream T domain acts as the nucleophilic partner in the condensation (Figure 4a). The deprotonated amine form of the aminoacyl-*S*-T_n attacks the peptidyl-*S*-T_{n-1}. In PKS modules, the polarity of condensation is maintained. The downstream (methyl)malonyl-*S*-T_n undergoes catalyzed decarboxylation to yield the stabilized C_2 carbanion (thioester enolate) (Figure 4b) as the attacking nucleophile on the acyl-*S*-T_{n-1} chain. The C−N and C−C bond formations occur during chain translocation to the downstream module.

The choice of thioester chemistry for activation of both (methyl)malonyl monomers and aminoacyl monomers provides both thermodynamic driving force and kinetically accessible nucleophiles for the condensations. During peptide bond-forming steps, the conversion of a peptidyl$_n$ thioester to the amide bond of the peptidyl$_{n+1}$ chain is thermodynamically favored. Analogously, the timed decarboxylation of the (methyl)malonyl C_3-carboxylate renders the C−C bond-forming steps irreversible. The covalent sequestration of monomers and growing PK and NRP chains as *S*-pantetheinyl-T domain thioesters in enzymatic assembly lines may be a deliberate balance between necessary thermodynamic activation of peptidyl/acyl thioesters and unwanted kinetic lability to hydrolysis. Restricting the access of water to PKS and NRPS condensation domain active sites may help prevent activated thioester intermediates from being squandered hydrolytically.

2.2. Information-Transfer Enzymology

The construction of each polyketide or nonribosomal peptide scaffold by multimodular enzymatic assembly lines

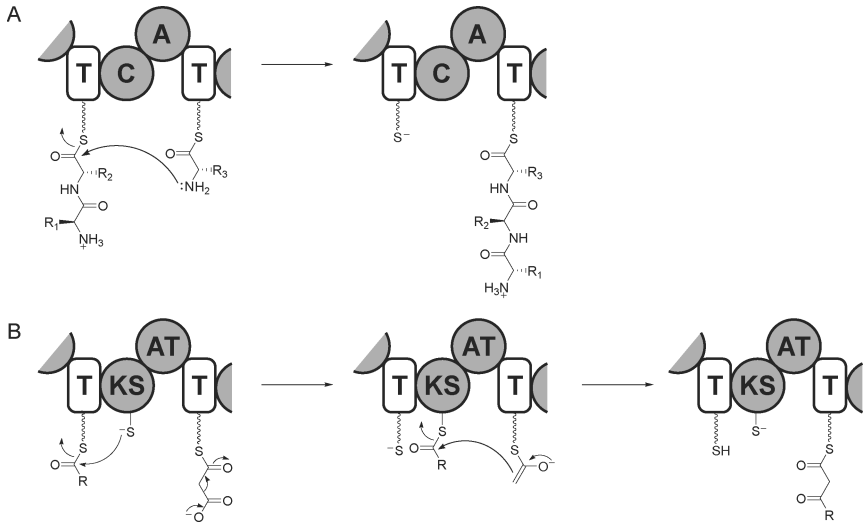

Figure 4. Catalyzed condensations during NRP and PK biosynthesis. (A) During NRP biosynthesis, nucleophilic attack of the downstream aminoacyl-*S*-T amine on the upstream peptidyl-*S*-T thioester results in amide bond formation. (B) During PK biosynthesis, decarboxylation of the downstream (methyl)malonyl-*S*-T yields a nucleophilic thioester enolate, which attacks the upstream acyl-*S*-T thioester in a Claisen condensation to form a C−C bond. KS = ketosynthase, AT = acyltransferase.

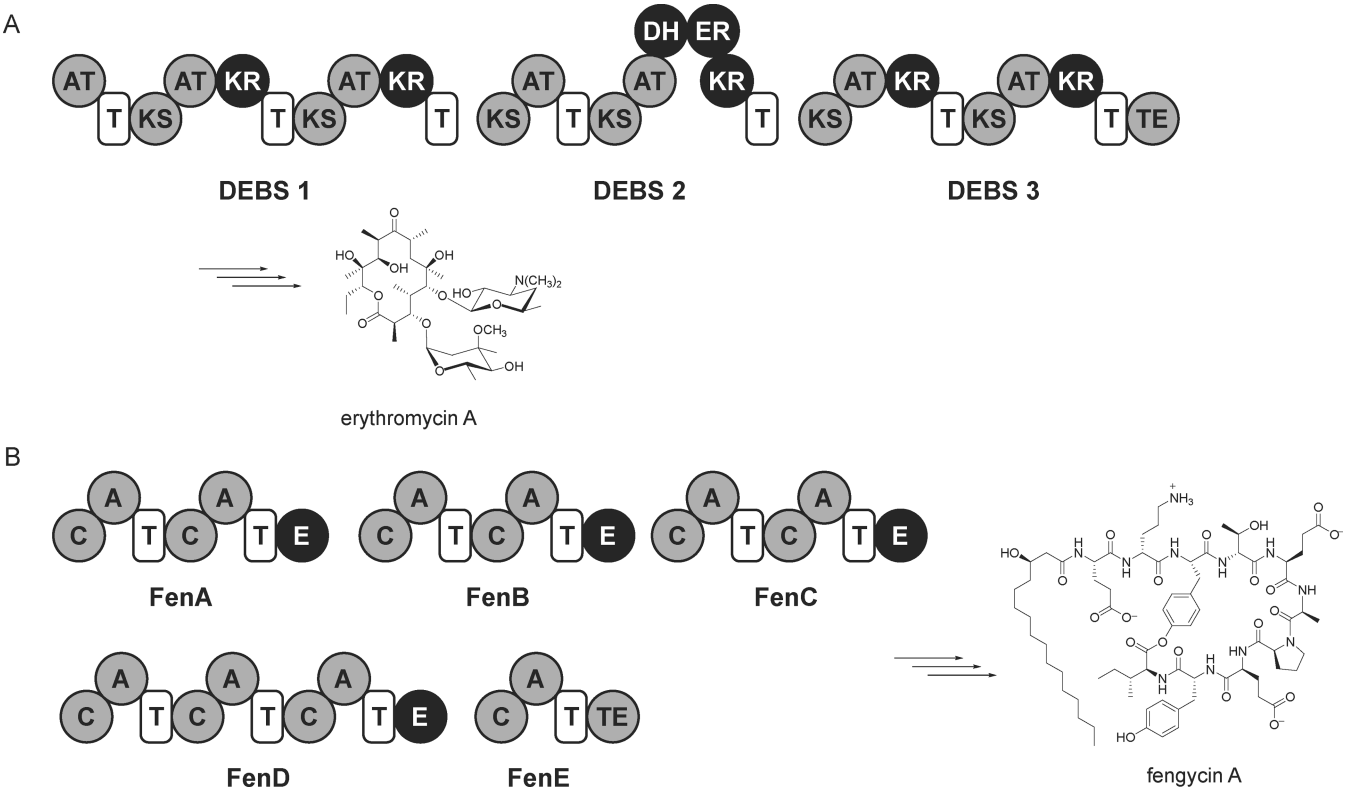

Figure 5. Representative PK and NRP assembly lines. (A) The erythromycin synthase is composed of 28 domains organized into 7 modules on 3 polypeptides.[1] (B) The fengycin synthetase comprises 35 domains organized into 10 modules on 5 polypeptides.[27] KR = ketoreductase, DH = dehydratase, ER = enoylreductase, TE = thioesterase, E = epimerase.

requires the selection and incorporation of particular monomers into the growing acyl chains in a temporally defined—and thus structurally defined—context. In a real sense, these processes comprise a system of information-transfer enzymology.

The concepts of biological information transfer are classically framed in the transcription of DNA into RNA and the translation of RNA into proteins. Both transcription and translation involve comparable processes of selecting monomer units and incorporating them into a polymer in a template-directed fashion, forming a sequence specified by the information in the template (DNA for RNA polymerization and mRNA for ribosomal protein synthesis). DNA replication involves parallel logic, where the template strand directs which monomers (in this case, dXMPs) are incorporated. Thus, DNA polymerases, RNA polymerases, and the peptidyl transferase center of the large rRNA subunit of the ribosome are information-transfer catalysts.

Each PKS and NRPS assembly line is also an information-transfer catalyst. During biosynthesis of the PK erythromycin, the seven-module 6-deoxyerythronolide B synthase performs 21 steps involving 28 domains to make the linear intermediate that is regiospecifically cyclized by the final domain of the synthase to release the 14-membered macrolactone, 6-DEB, as the aglycone scaffold of erythromycin[26] (Figure 5a). Analogously, the 35-domain fengycin synthetase performs 25 operations on 10 modules to produce the decapeptidolactone antifungal agent[27] (Figure 5b).

During ribosomal protein synthesis, mRNA is the template that specifies the amino acid monomer (recognized as an aminoacyl-tRNA) to be incorporated. The code being read by the ribosome is the hydrogen-bonding pattern between the mRNA codon in the peptidyl transferase center and the tRNA anticodon on the docked aminoacyl-tRNA.[28]

NRPS-mediated peptide elongations do not use an RNA template. Rather, the identity and ordering of protein domains (Figure 1d) in the synthetase constitutes the template (Figure 6). There are one or two gate-keeping domains in each NRPS module that specify the monomer to be selected and the chirality or modification state of the condensation partners (e.g., to generate bacitracin). The code for information transfer in polyketide synthases—comparable to NRPS logic—is that the number, identity, and order of protein modules in the assembly line determine the monomers that are incorporated and the order in which they are condensed. Other catalytic domains in each PKS module determine the chemical processing steps following each condensation to control the oxidation state at the β-carbon of each newly added C_2 or C_3 unit, as exemplified for nystatin.

Molecules with mixed PK−NRP lineage, such as bleomycin, rapamycin, and the epothilones, are constructed by hybrid assembly lines. The number and placement of PKS and NRPS modules required to make an assembly line coding for a hybrid product can be predicted from an understanding of the modular logic of PKS and NRPS systems.[29,30] The logic of PKS and NRPS systems is compatible, and growing chains can be transferred across an NRPS−PKS interface to switch from incorporating amino acid monomers to (methyl)malonyl-CoA monomers. Further downstream in a hybrid assembly line, at a PKS−NRPS interface, the process can be reversed, as in the case of the yersiniabactin or epothilone synthases. As a result, the chain connectivity of hybrid PK−NRPs switches back and forth between C−N bond connections in amide linkage to C−C bond connections in β-ketoacyl moieties. Two such switching events are shown for the first three modules of the epothilone synthase assembly line (Figure 7).

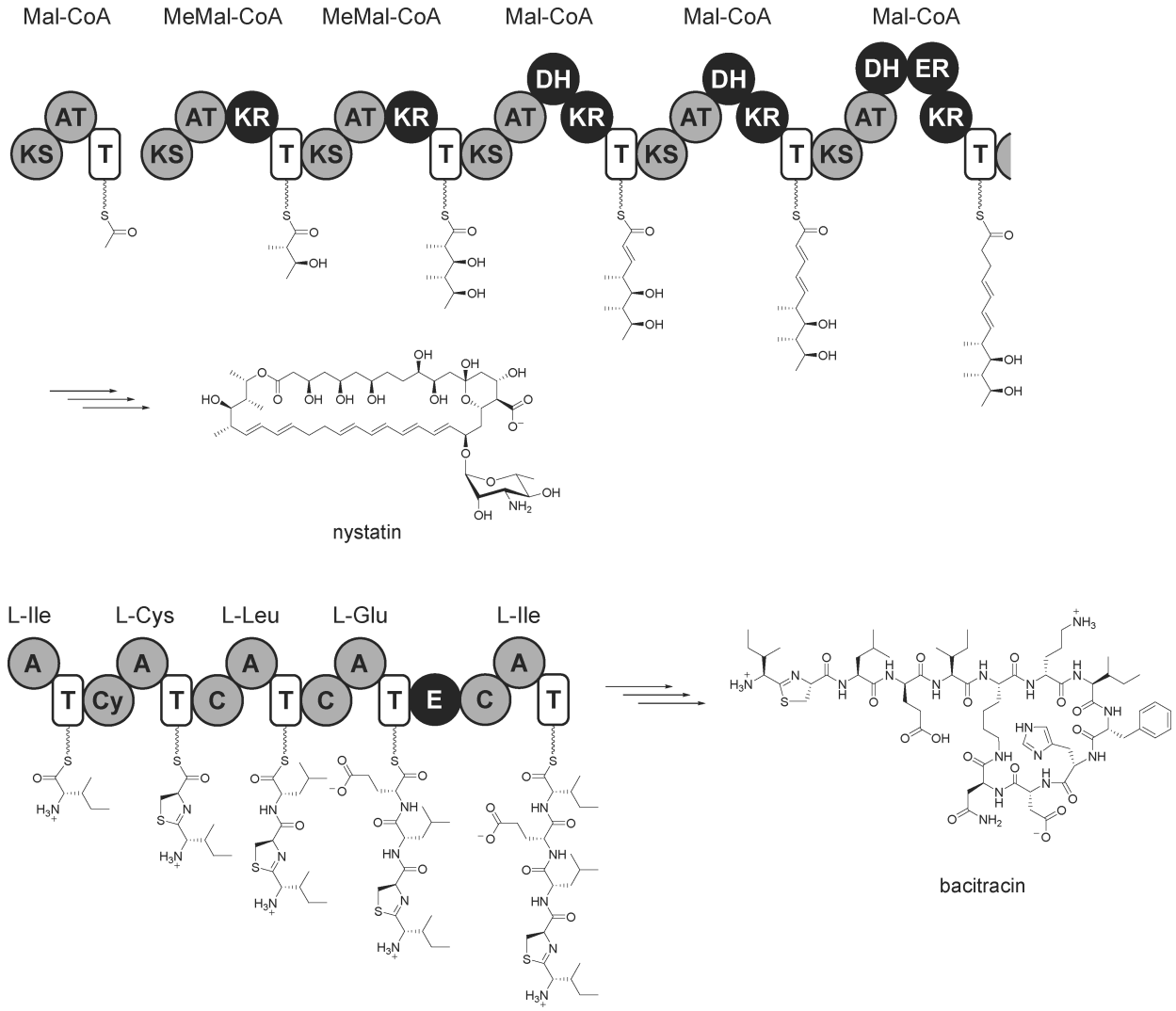

Figure 6. Product of a PK or NRP assembly line is encoded in the identity and order of its constituent domains. (A) The first six modules of the nystatin synthase[178] in which the AT domains activate either malonyl-CoA (Mal-CoA) or methylmalonyl-CoA (Me-Mal-CoA). (B) The first five modules of the bacitracin synthetase[179] in which the A domains activate the specified L-amino acids. Cy = cyclization.

2.3. Pre- and Post-Assembly-Line Enzymology

Genes that encode multimodular, megadalton PKS and NRPS assembly lines are clustered with genes that encode the production of unusual monomer units required for that particular assembly line to function. This theme is particularly well illustrated by the 30-gene biosynthetic gene cluster for chloroeremomycin, one of the vancomycin-family glycopeptide antibiotics.[5] Three of the genes are required to encode three giant proteins, CepA (three modules), CepB (three modules), and CepC (one module), to make up the seven-module NRPS assembly line for the heptapeptide scaffold of chloroeremomycin. Five of the seven residues in this scaffold are nonproteinogenic: two β-hydroxytyrosine (β-OH-Tyr), two 3,5-dihydroxyphenylglycine (3,5-Dpg), and one 4-hydroxyphenylglycine (4-Hpg) (Figure 8). Three ORFs are required for benzylic hydroxylation of tyrosine, four ORFs are needed to convert prephenate to 4-Hpg, and four ORFs encode a Type III PKS that converts four malonyl-CoA monomers to 3,5-Dpg. In total, 11 ORFs from the gene cluster are dedicated to creating nonproteinogenic amino acids as starting materials for the assembly line (Figure 9). Coordinate regulation of the clustered genes likely facilitates "just-in-time" expression of the enzymes required to divert primary metabolites down these conditional pathways.[31]

Analogously, sets of genes in PKS and hybrid PKS−NRPS biosynthetic clusters encode the production of specialized starter units, including 3,4-dihydroxy-cyclohexenecarboxylic acid for the rapamycin and FK506 synthases,[32] malonamic acid for the tetracycline synthase,[13] and 3-amino-5-hydroxybenzoic acid for the rifamycin synthase[33,34] (Figure 10).

During biosynthesis of many PKs and NRPs, the free acid or macrocyclic lactam or lactone released from the final way station of the assembly line requires further enzymatic tailoring to gain biological activity. In some cases, e.g., the macrolide antibiotics, tailoring requirements are absolute; 6-DEB does not exhibit antibiotic activity.[35] This PK scaffold requires two hydroxylations and two glycosylations to be a potent protein synthesis inhibitor.[36]

Post-assembly-line tailoring enzymes typically catalyze oxidation, glycosylation, acylation, or alkylation (usually methylation) of nascent PK and NRP scaffolds.[37] Almost all of the genes encoding these dedicated tailoring enzymes are also found in biosynthetic clusters (Figure 9) for coordinate regulation. The coordination of logic for these microbial

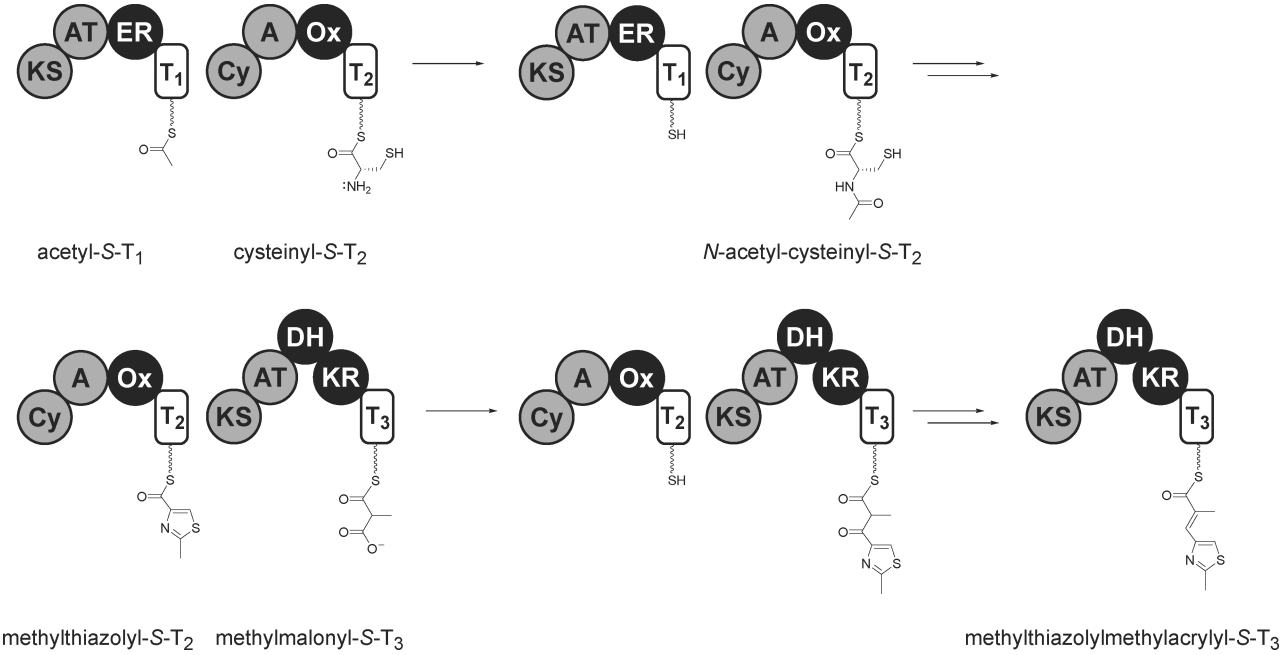

Figure 7. PKS−NRPS interfaces in the epothilone synthase.[180] Acetyl-*S*-T_1 is transferred from the PK initiation module, EpoA, to cysteinyl-*S*-T_2 on the downstream NRP module, EpoB, by its N-terminal Cy domain. Subsequent heterocyclization, dehydration, and oxidation form the new acyl donor methylthiazolyl-*S*-T_2, which is transferred to the carbanionic propionate on the downstream PK module, EpoC, by its embedded KS domain. Subsequent reduction and dehydration yield the EpoC-bound intermediate methylthiazolylmethylacrylyl-*S*-T_3. Ox = oxidase.

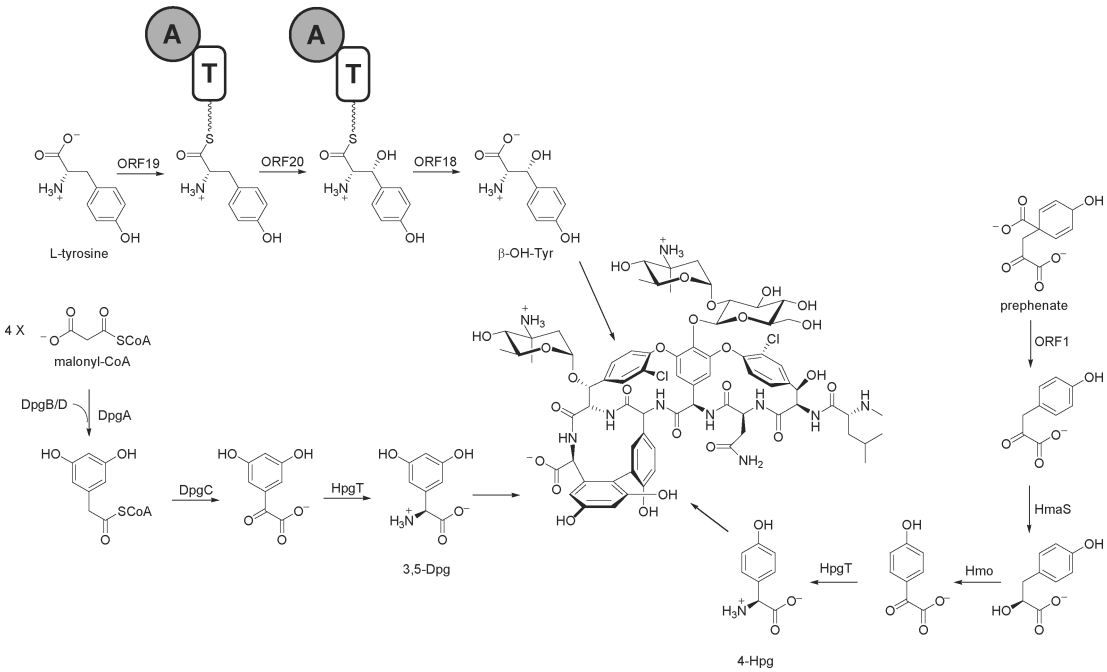

Figure 8. Pre-assembly-line protein infrastructure to create nonproteinogenic amino acids for chloroeremomycin biosynthesis: Conversion of L-tyrosine to β-hydroxy-L-tyrosine (β-OH-Tyr),[130] malonyl-CoA to 3,5-dihydroxyphenylglycine (3,5-Dpg),[181] and prephenate to 4-hydroxyphenylglycine (4-Hpg).[182]

biosynthetic factories is both elegant and admirably efficient. In addition to the assembly-line protein components, enzymes required for both chemical tailoring of PK and NRP scaffolds and diversion of primary metabolites to secondary metabolic monomers are collected into adjacent sets of operons. Inter alia, transfer of a contiguous 50−100 kb of DNA between bacteria can provide all of the genetic components required for making new small molecules in the recipient, so the biosynthetic capacity to make numerous antibiotics is portable.

3. Enzyme Machinery

3.1. Polyketide Synthase Enzymatic Machinery: Canonical Organization

PKs and fatty acids are assembled with the same logic and by equivalent enzymatic machinery (Figure 11). In some biosynthetic systems, the catalytic and carrier T domains exist as separate proteins that interact in trans to form functional

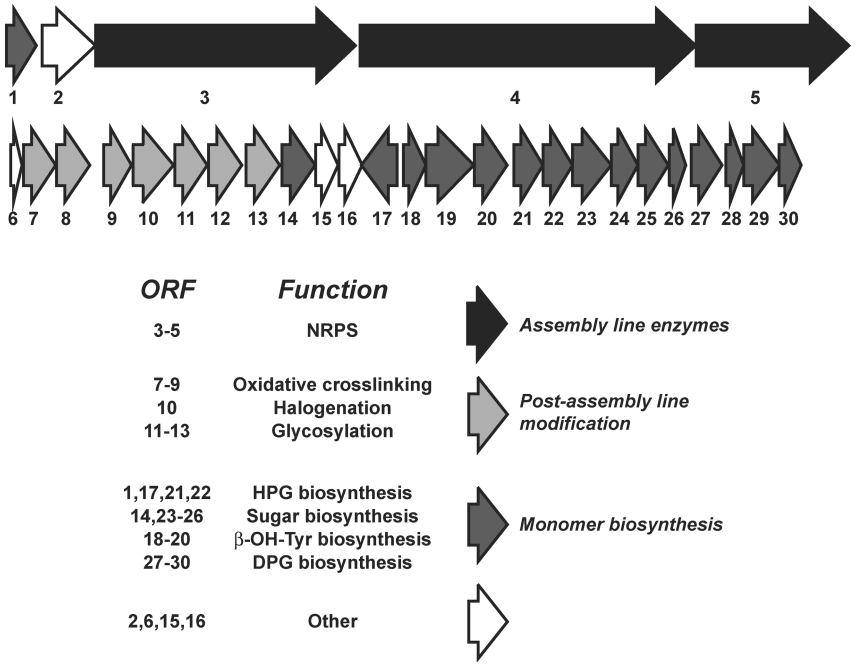

Figure 9. Chloroeremomycin gene cluster.[5,132] Genes encoding enzymes involved in assembly-line biosynthesis, post-assembly-line modifications, and provision of NDP-hexose and nonproteinogenic amino acid monomers are highlighted.

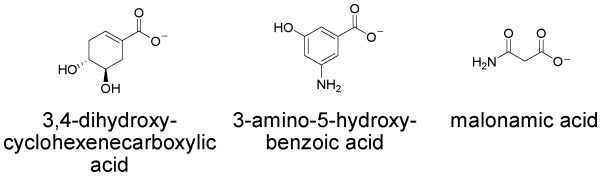

Figure 10. PK starter units requiring pre-assembly-line enzymatic activities. 3,4-Dihydroxycyclohexencarboxylic acid is the starter unit for the rapamycin and FK506 synthases,[32] 3-amino-5-hydroxy-benzoic acid is the starter unit for the rifamycin synthase,[33,34] and malonamic acid is the starter unit for the tetracycline synthase.[13]

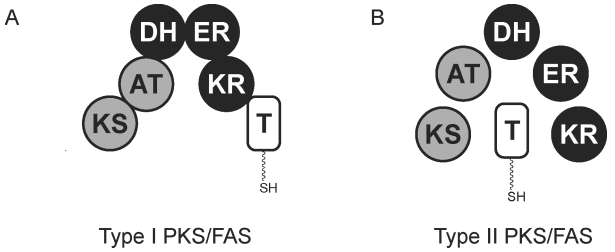

Figure 11. Comparison of Type I and Type II PKSs and FASs. (A) In Type I PKS/FAS systems, protein domains are connected in cis. (B) In Type II PKS/FAS systems, protein domains interact in trans.

complexes for acyl chain elongations. In other systems, the catalytic and the carrier T domains are fused in cis to form modules, and modules are joined together to form a multimodular assembly line. Those PKS systems with domains connected in cis have been termed Type I PKSs, while systems with domains organized in trans are known as Type II PKSs.[38,39] There exists a Type III PKS subgroup, distinguished from Types I and II by the use of malonyl-CoA rather than malonyl-*S*-pantetheinyl-T species as substrates.[40] One example of a Type III PKS is the enzyme DpgA, which uses three molecules of malonyl-CoA to form the precursor to 3,5-Dpg for chloroeremomycin assembly[41] (Figure 12).

Two catalytic domains and one T domain are the core domains present in both PKS and FAS enzymatic machinery[38] (Figure 13). The T domain is an 8−10 kDa four-helix bundle. While T domains in NRPSs and PKSs are often referred to as peptidyl carrier proteins (PCPs) and acyl carrier proteins (ACPs), respectively, the T (thiolation) domain nomenclature will be used here and denotes the function of this domain: to provide a thiol on which substrate and intermediate acyl chains are covalently tethered during chain elongations. The side chain of a specific serine in all T domains gets posttranslationally modified to install a flexible thiol-terminated pantetheinyl arm in phosphodiester linkage[42] (Figure 14). This "priming" of apo-T domains to functional, holo-T domains—a prerequisite for assembly line function—is carried out by phosphopantetheinyltransferases (PPTases), which are often encoded in FAS, PKS, and NRPS gene clusters.[23]

The two core catalytic domains are a 50 kDa acyltransferase (AT) domain and a 45 kDa ketosynthase (KS) domain. AT domains select either malonyl-CoA or methylmalonyl-CoA as substrates and thereby act as gatekeepers for specificity. They transfer the C_3 or C_4 acyl group to the thiolate terminus of the pantetheinyl arm on a holo-T domain (Figure 15a). This is a net transthiolation: an energy neutral acyl transfer step.

The KS is the C−C bond-forming catalyst. After an upstream acyl-*S*-T_{n-1} has been transthiolated onto a conserved cysteine in its active site, the KS decarboxylates the downstream malonyl-*S*-T_n (Figure 15b). The resultant β-ketoacyl-*S*-T_n is the substrate for the next cycle of elongation. Each catalytic cycle of a KS domain results in a net addition of two carbons (from a malonyl unit) or three carbons (from a methylmalonyl unit) to the growing acyl chain. During Type I PKS assembly, the growing acyl chain is translocated downstream from T_{n-1} to T_n by each action of a KS domain (Figure 16a). Each KS and T domain is used once, and the assembly-line metaphor applies. During Type II PKS assembly, there exists only one T domain, one AT domain, and one KS domain for the chain elongation steps. The chain grows by two or three carbons in each cycle, but it remains tethered to the same T domain, and KS and AT act iteratively (Figure 16b). The role of the chain length factor (CLF),

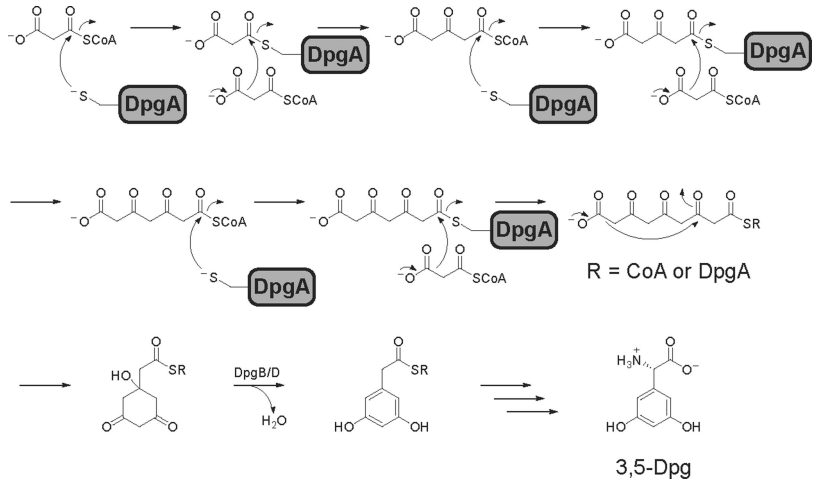

Figure 12. Formation of a 3,5-Dpg precursor by the Type III PKS DpgA.[41] DpgA uses four molecules of malonyl-CoA to make the C_8 precursor to 3,5-Dpg.

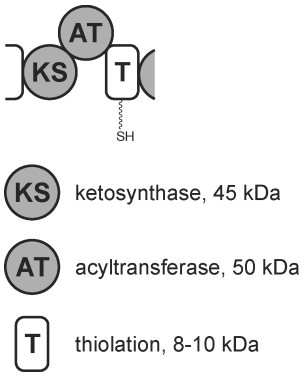

Figure 13. Core domains of PKS and FAS systems. A minimal module consists of a ketosynthase domain (45 kDa), an acyltransferase domain (50 kDa), and a thiolation domain (8−10 kDa).

which forms a heterodimer with the KS, will be discussed in section 5.4. There is no translocation of the growing chain along a cascade of T domains.

3.1.1. Type I PKS Assembly Lines

Three additional catalytic domains—invariably present in Type I and II FASs but optional in Type I and II PKSs—are ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains. Their functions are well established from classical investigation of FAS enzymology; they operate in the sequence KR → DH → ER, initially reducing the β-ketoacyl-*S*-T arising from KS-mediated condensation to the β-hydroxyacyl-*S*-T (KR action), then dehydrating to form the α,β-enoyl-*S*-T (DH action), and finally reducing the conjugated olefin to the saturated acyl-*S*-T (ER action) (Figure 17). This tandem action of three enzymes (domains) effects the four-electron reduction of the β-C═O to the β-CH_2 and completes the chain elongation process initiated by AT and KS action. In FASs, the net result is chain elongation by a CH_2-CH_2 unit.

The preceding comments on Type I PKS modules apply to all the *elongation* modules. The *initiation* and termination modules function differently. The initiation module may contain a KS domain—catalytically inactive for condensation—that effects decarboxylation of the (methyl)malonyl-*S*-T_1,[43] releasing CO_2 and yielding an acetyl- or a propionyl-*S*-T domain as the starting acyl group on the assembly line. Alternatively, the KS domain may be missing entirely from the initiation/loading module at the N-terminus of the assembly line. In that case, an acyl-CoA (other than malonyl- or methylmalonyl-CoA) is selected by the initiating AT domain and loaded on the first T domain.[44]

The last module of a Type I PKS assembly line is responsible for release of the full-length acyl chain from covalent tethering on the final T domain.[45] This process of chain termination is typically catalyzed by a 35 kDa thioesterase (TE) domain, which can catalyze hydrolysis, generating the free acid (e.g., during β-lactam biosynthesis (Figure 18a)), or intramolecular capture by one of the −OH groups in the polyketide chain, generating a cyclic lactone as shown for the proposed assembly of avermectin[46] (Figure 18b).

The hallmark of Type I and Type II PKSs is the incomplete reduction of the initial β-ketoacyl-*S*-T condensation product. While some Type I PKS modules may contain functional KR, DH, and ER domains—and therefore generate fully reduced CH_2-CH_2 units—others will preserve the intermediate α,β-enoyl, β-hydroxyacyl, or β-ketoacyl moieties (Figure 19) due to absent or nonfunctional ER, DH, or KR domains,

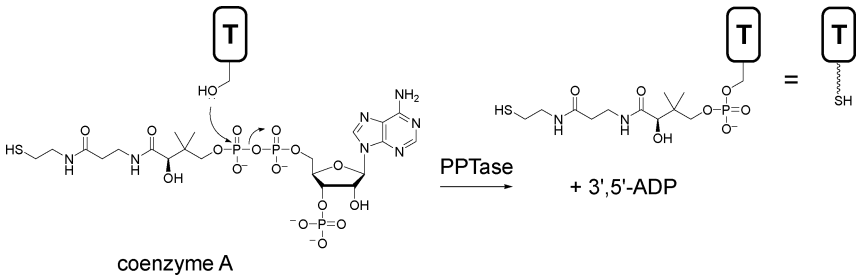

Figure 14. Posttranslational modification of T domains by phosphopantetheinyltransferases.[23] Members of this enzyme class catalyze the transfer of phosphopantetheine from coenzyme A to a conserved serine in the T domain.

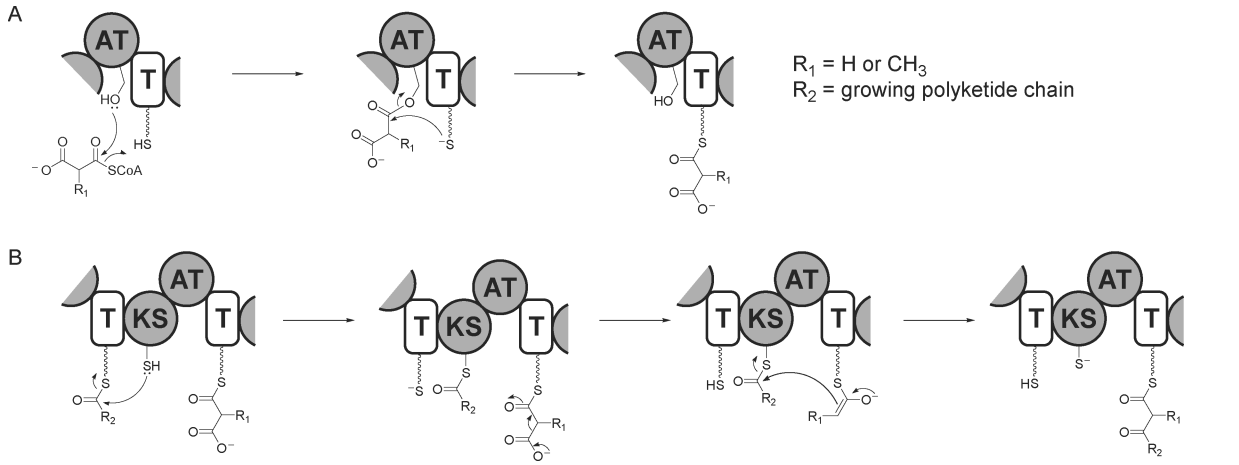

Figure 15. Activities of the AT and KS domains. (A) The AT domain selects the acyl-CoA monomer to be incorporated and catalyzes its transthiolation to the downstream T domain via an acyl-*O*-enzyme intermediate. (B) The KS domain catalyzes C−C bond formation between the upstream (transiently KS-bound) acyl thioester and the downstream carbanionic acyl acceptor resulting from decarboxylation of (methyl)-malonyl-*S*-T.

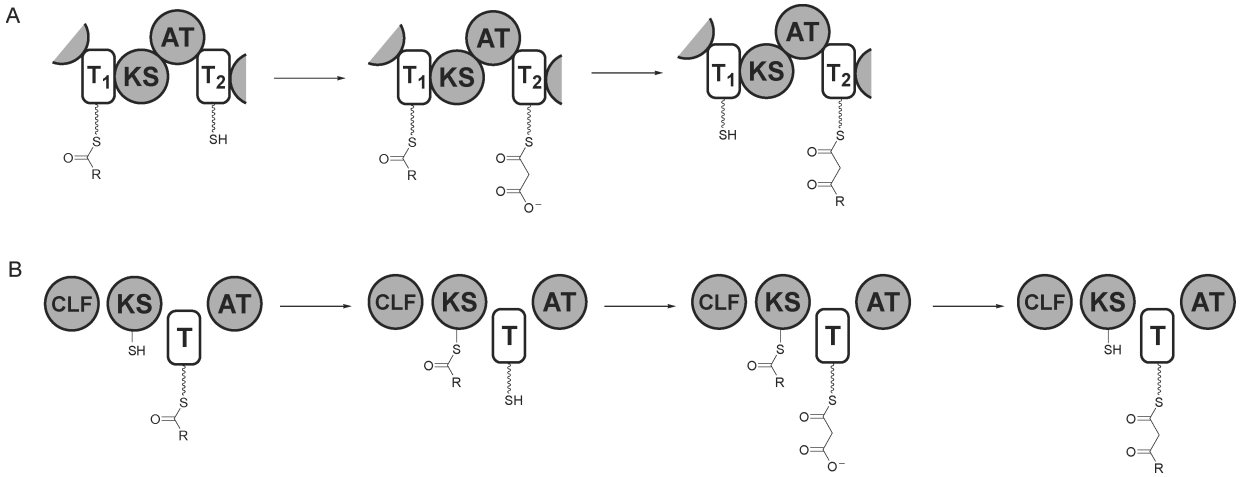

Figure 16. Regiochemistry of chain elongation in Type I and II PKS systems. (A) In Type I PKSs, each condensation results in the addition of two or three carbons to the nascent chain, which is simultaneously transferred from one T domain to the next. (B) In Type II PKSs, each condensation results in the addition of two or three carbons to the nascent chain, which remains tethered to the same T domain. CLF = chain length factor.

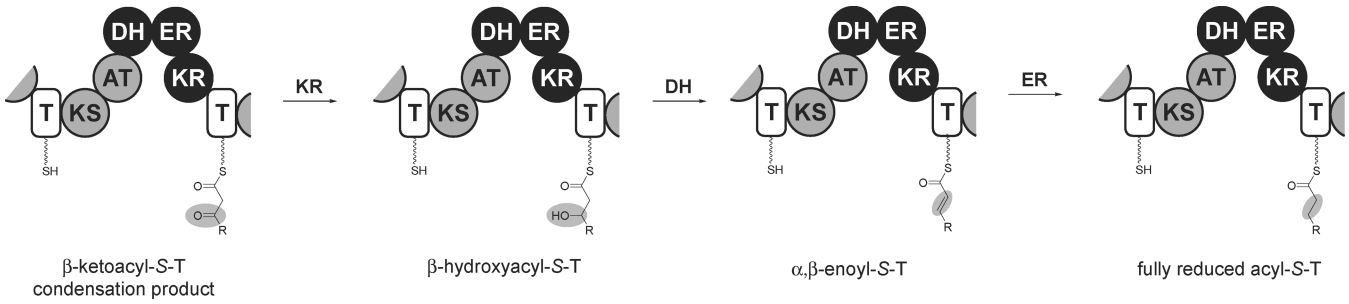

Figure 17. β-Carbon processing in PKSs. The β-ketoacyl-*S*-T condensation product is reduced by the ketoreductase (KR) domain to form β-hydroxyacyl-*S*-T, subsequent dehydration by the dehydratase (DH) domain yields an α,β-enoyl-*S*-T, and reduction of this species by the enoyl reductase (ER) domain gives the fully saturated acyl-*S*-T.

respectively. Such incomplete β-carbon processing by Type I PKSs creates the diverse chemical functionality and chemical reactivity for further elaboration of the scaffold.

The chemical diversity enabled by accumulation of polyketoacyl chains during elongation cycles—resulting from the absence of β-carbon processing—is particularly notable in Type II PKS machinery in which iterative action of core KS-AT-T subunits generates the polyketone chains for which this class of natural products is named. These highly reactive functionalities undergo enzyme-directed carbonyl chemistry as schematized for the antibiotic tetracycline and the antitumor agent doxorubicin (Figure 20). The fate of the polyketoacyl thioester chains is no doubt controlled at several stages: by the conformation of the bound polyketone chain in enzyme active sites surrounding the polyketoacyl-*S*-T, by cyclase enzymes that control regioselectivity of cyclization, and by prior regioselective reduction of one of the ketone groups in the nascent polyketoacyl-*S*-T intermediate.[47,48]

A few polyketide scaffolds generated by Type I PKSs have β-branched groups, such as jamaicamide.[49] The corre-

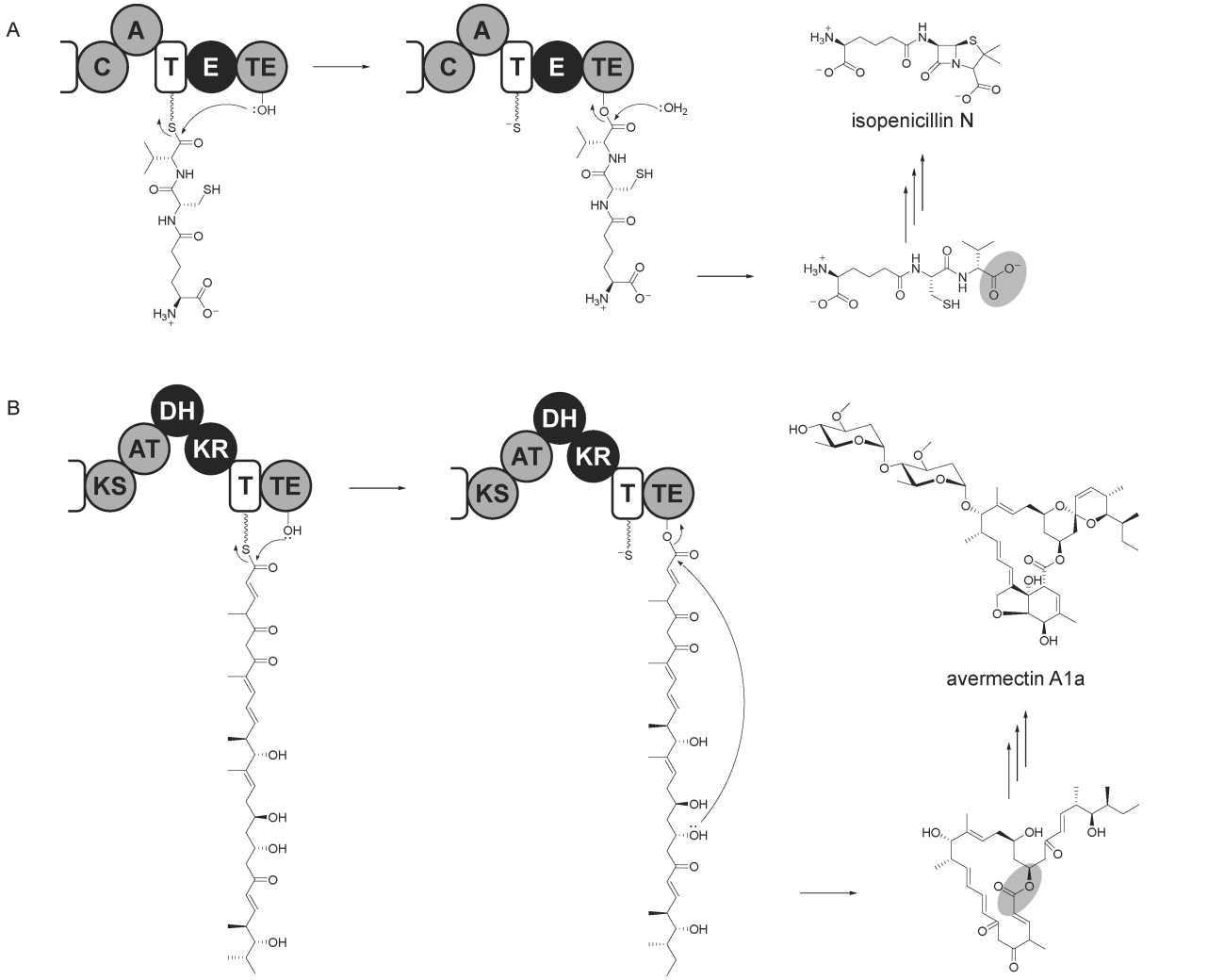

Figure 18. Chain termination by the thioesterase (TE) domain. (A) The TE domain of the ACV synthetase first acylates itself with the tripeptidyl group via a conserved serine and then catalyzes hydrolytic release. (B) Proposed chain termination for avermectin assembly. The TE domain of the avermectin synthase[46] first transfers the C_{34} linear thioester precursor to an active site serine and then catalyzes macrocyclization, forming a 20-membered ring and releasing the macrolactone from covalent linkage to the synthase.

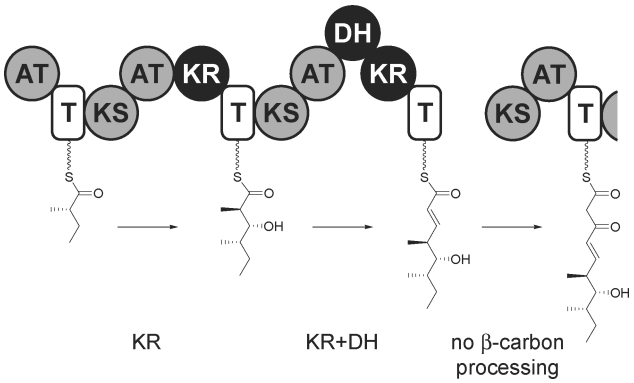

Figure 19. Correspondence between β-carbon processing domains and chemical functionality found in the polyketide product. The first three modules of the avermectin synthase[46] contain varying sets of β-carbon processing domains, which correspond to the β-carbon functionalities observed in the product.

sponding gene clusters encode homologs of hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase).[50] The HMG-CoA synthase acts in trans on the β-ketoacyl-*S*-T intermediate to add the C_2-carbanion of acetyl-*S*-T into the ketone.[184] Subsequent dehydration by ECH_1 yields a conjugated species that is primed for decarboxylation by ECH_2, forming a Δ^2-isopentenyl-*S*-T species. Elongation extends this β-branched PK chain (Figure 21). This example is probably one of many in which tailoring enzymes act in trans on growing acyl-*S*-T intermediates and therefore can be grouped with enzymes of this class, such as the *act*III KR[52] that selectively reduces one of the ketones in the octaketidyl-*S*-T intermediate that arises during actinorhodin assembly.

3.2. Nonribosomal Peptide Synthetase Machinery: Canonical Organization

Nonribosomal peptide synthetases follow the general precepts of Type I PKS logic. The biosynthetic machinery is comprised of multimodular enzymatic assembly lines that contain one module for each amino acid monomer incorporated. In analogy to the three core domains of a PKS module (KS-AT-T), there are three core domains in a minimal NRPS module that carry out an elongation step on a growing peptidyl chain tethered to a T domain in the NRPS assembly line.[53] Like PKSs, NRPSs contain a T domain and two catalytic domains, the 50 kDa adenylation (A) domain responsible for selecting the amino acid monomer, and the 50 kDa condensation (C) domain responsible for the peptide bond-forming chain elongation (Figure 22). The typical order of domains is C-A-T.

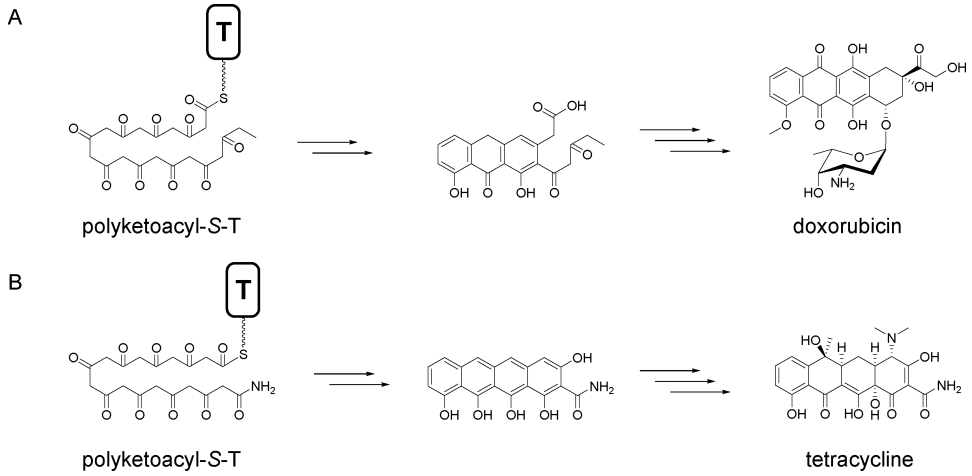

Figure 20. Biosynthesis of doxorubicin and tetracycline by Type II PKSs. In both systems, the product of the assembly line is a reactive polyketone, which is acted upon in sequence by a cyclase and an aromatase to produce the planar aromatic core of these molecules.

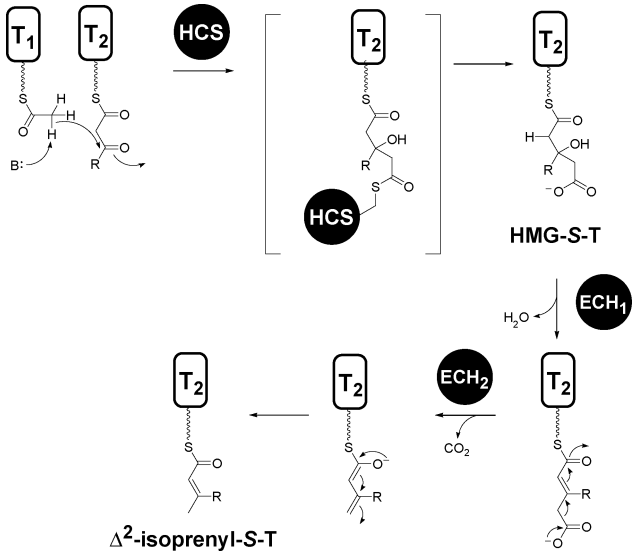

Figure 21. Formation of β-branched groups by isoprene-inserting machinery.[49,51,184] The HMG-CoA synthase adds the C_2-carbanion of acetyl-*S*-T into the β-ketone. Subsequent dehydration by ECH_1 and decarboxylation by ECH_2 yield the Δ^2-isopentenyl-*S*-T species.

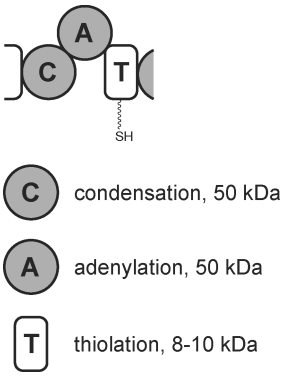

Figure 22. Core domains of NRPSs. The condensation (C) domain catalyzes amide bond formation. The adenylation (A) domain activates the amino or aryl acid monomer to be incorporated as the adenylate and catalyzes acyl transfer to the adjacent thiolation (T) domain to which monomers and nascent chains are covalently tethered.

The logic and machinery of T domain function in NRPS assembly lines is equivalent to the role of T domains in Type I PKS assembly lines. Each 8−10 kDa apo-T domain must be primed by posttranslational modification of a serine side chain with phosphopantetheine, catalyzed by dedicated PPTases.[23] During NRP assembly, aminoacyl and peptidyl chains are tethered in thioester linkage to the terminal thiolate of the phosphopantetheine prosthetic group (Figure 14).

The two catalytic domains in a minimal NRPS elongation module have functions similar to the AT and KS domains of PKS modules. The A domain, in analogy to the AT domain, selects the amino acid, activating the carboxylate with ATP to make the aminoacyl-AMP and then installing the aminoacyl group on the thiolate of the adjacent T domain. Like the KS domain, the C domain is the chain-elongating catalyst, joining an upstream peptidyl-*S*-T_{n-1} to the downstream aminoacyl-*S*-T_n. Chain elongation by one aminoacyl residue occurs concomitantly with chain translocation to T_n (Figure 23).

The *initiation* module in NRPS assembly lines is often a two-domain A-T module that selects the first amino acid and installs it covalently on T_1. The peptidyl chain then elongates, preserving the free amino group of the N-terminal residue.

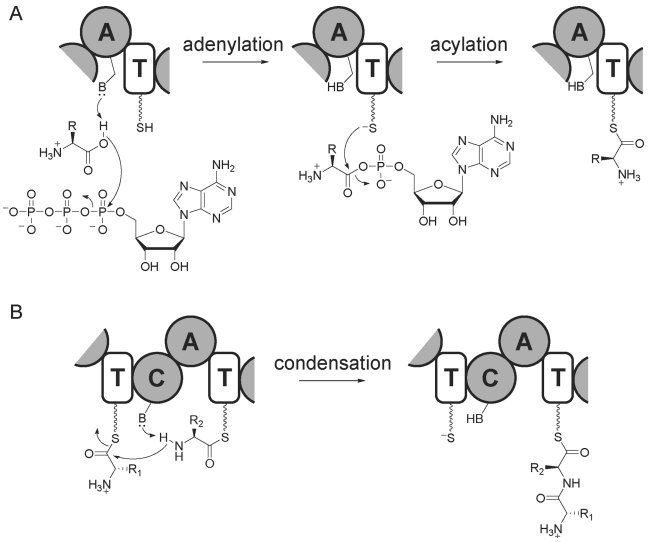

Figure 23. Activities of the A and C domains. (A) The A domain catalyzes the adenylation of the amino acid to be incorporated and its subsequent acylation to the downstream T domain. (B) The C domain catalyzes C−N bond formation between the electrophilic upstream peptidyl-*S*-T_1 and the nucleophilic downstream aminoacyl-*S*-T_2.

In some instances, the first module has a three-domain C-A-T organization; these often occur in assembly lines that make *N*-acylated peptides, such as plipastatin.[27] The first C domain has been shown in one case to catalyze *N*-acylation of the initiating amino acid while it is installed on T_1.[54]

Like Type I PKSs, the *termination* modules of NRPS assembly lines usually have a C-A-T-TE organization.[21] In analogy to PKS TE domains, NRPS TE domains can be hydrolytic or cyclizing[45,55] (Figure 18). An alternative, TE-independent route of chain release is reductase-catalyzed thioester reduction to yield an aldehyde, exemplified in myxochelin biosynthesis[56] and generation of the tetrapeptidyl aldehyde precursor of safracin[57] (Figure 24).

3.2.1. NRPS Assembly Lines: Optional Domains

3.2.1.1. Epimerization and Condensation Domains with $^{D}C_{L}$ Chirality. A pervasive feature of NRPs is the abundance of D-amino acid residues in their peptide scaffolds. Since D-amino acids (especially nonproteinogenic D-amino acids) are largely absent in microbial producer cells, L- to D-epimerization machinery is required for their production.[58] Occasionally, a separate racemase is encoded in the biosynthetic gene cluster, such as the alanine racemase that provides monomeric D-Ala for the first module of the cyclosporine synthetase.[59]

In contrast, the preponderant organization in NRPS assembly lines from Gram-positive bacteria involves a 50 kDa epimerization (E) domain in those modules where L-amino acid monomers are to be incorporated and epimerized[60] (Figure 25a). The domain organization in such modules is C_n-A_n-T_n-E_n. The C_{n+1} domain—immediately downstream of the E_n domain—is specific for the D-peptidyl-*S*-T donor.[61] Such C domains have been designated $^{D}C_{L}$ domains, for D-donor and L-acceptor. This leads to the expectation that all C domains adjacent to (and downstream of) E domains make D-,L-peptide linkages.[62] Some nonribosomal peptides, such as chloroeremomycin, contain D-,D-dipeptidyl moieties that are formed by tandem CATE-CATE modules.[5] This dimodule motif likely forms a D-,L-peptide bond and then epimerizes it to a D-,D-peptide linkage before chain transfer to the next module.

A variant of the E_n-$^{D}C_{L(n+1)}$ organization has been elucidated in assembly lines for arthrofactin,[63] syringomycin,[64] and syringopeptin[65] from Gram-negative *Pseudomonas* spp. (Figure 25b). These "dual-function" C domains are $^{D}C_{L}$ in their chiral specificity; they first generate an equilibrating mix of L-peptidyl and D-peptidyl-*S*-T intermediates and then condense only the D-peptidyl-*S*-T donor with the downstream aminoacyl-*S*-T acceptor.[66]

3.2.1.2. Condensation Domains with Heterocyclization Activity. In addition to the $^{L}C_{L}$ and $^{D}C_{L}$ domains noted above, additional C domain variants exist that catalyze amide bond formation with Cys, Ser, or Thr as the downstream acceptor. These domains then catalyze addition of the side chain thiol (Cys) or hydroxyl (Ser, Thr) to form a five-membered cyclic adduct, followed by dehydration to yield thiazoline or oxazoline rings, respectively[67] (Figure 26a). This subset of C domains are termed Cy (cyclization) domain for their cyclodehydration activity. The net formation of a five-membered heterocycle radically alters the covalent connectivity of the growing peptidyl chain. Tandem Cy domain-containing modules are found in the yersiniabactin[12] and bleomycin synthases,[10] leading to the formation of a thiazoline-thiazolidine moiety in the former and a bithiazole substituent in the latter (Figure 26b).

3.2.1.3. Additional Domains in NRPS Modules. Several other domains are found in certain NRPS modules, which presumably resulted from gene splicing at permissive sites to encode useful chemical transformations on the growing peptidyl chain. One such domain, termed oxidase (Ox), is found in modules from the bleomycin and epothilone synthases that contain Cy domains and is involved in oxidation of the thiazoline ring to a thiazole (Cy-A-T-Ox) (Figure 26c). This oxidative conversion of the hydrolytically labile dihydroheterocyclic thiazoline to the heteroaromatic

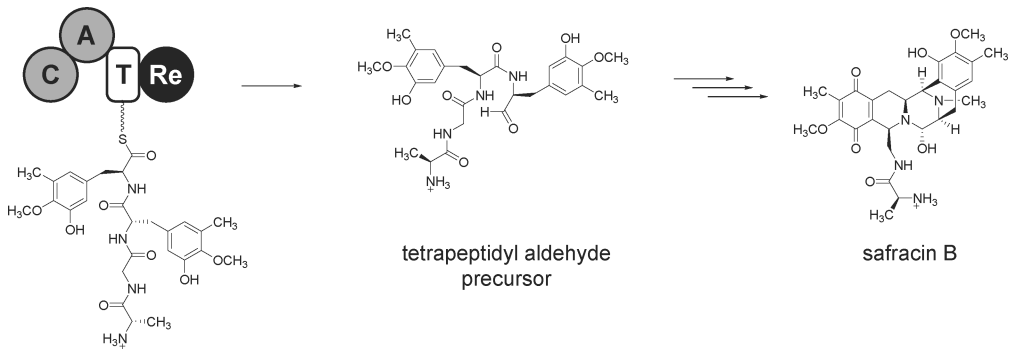

Figure 24. Proposed chain termination during safracin[57] biosynthesis. A C-terminal reductase (Re) domain likely catalyzes concomitant aldehyde formation and chain release.

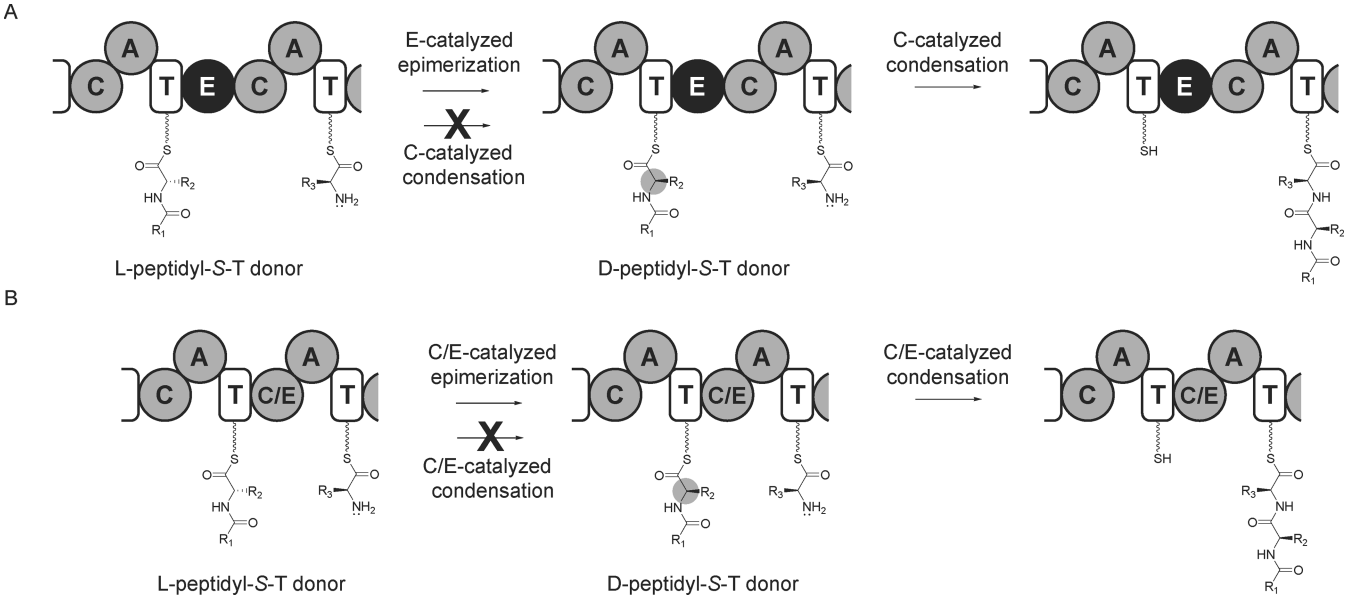

Figure 25. Two amino acid epimerization strategies in NRPSs. (A) In some NRPSs, an E-$^{D}C_{L}$ didomain functions to effect epimerization and condensation. The E domain epimerizes the upstream L-peptidyl-*S*-T acyl donor to D-peptidyl-*S*-T prior to condensation, and the $^{D}C_{L}$ domain is specific for D-peptidyl-*S*-T to ensure that condensation does not precede epimerization. (B) In other NRPSs, the E-$^{D}C_{L}$ didomain is replaced by a single bifunctional $^{D}C_{L}$/E domain that epimerizes the L-peptidyl-*S*-T to a D-peptidyl-*S*-T and then selects only the latter for subsequent condensation. C/E = condensation/epimerization.

thiazole produces a stable heterocycle.[68] In bleomycin, the bithiazole is a DNA intercalating moiety that helps target the antitumor antibiotic to DNA.[69] The Ox domain is portable in that it has been shown to function when inserted into heterologous modules[70] and might therefore be a useful tool in a combinatorial biosynthetic toolbox for reprogramming NRPS assembly lines.

NRPS assembly-line tailoring enzymes can work in trans as well as in cis. A notable case occurs during maturation of the siderophore pyochelin from *Pseudomonas aeruginosa*.[71,72] In contrast to the in cis oxidation of a thiazolinyl-*S*-T intermediate by an Ox domain in epothilone biosynthesis, pyochelin assembly involves in trans reduction of a comparable thiazolinyl-*S*-T species to the tetrahydro thiazolidinyl-*S*-T (Figure 26d) by PchG.[73] The thiazolidinyl nitrogen is now basic enough to act as a nucleophile and become *N*-methylated, stabilizing the thiazolidinyl ring from opening to the acyclic aldehyde.

Many nonribosomal peptides contain *N*-methylated amino acid residues. For example, 7 out of the 11 residues in cyclosporine are *N*-methylated.[74] *N*-Methylation is catalyzed by a methyltransferase (MT) domain that transfers the CH_3 group from *S*-adenosylmethionine (SAM) to the amino group of the aminoacyl-*S*-T intermediate in C-A-MT-T modules[75] (Figure 27a). MT domains are also found in a few modules of Type I PKS assembly lines, e.g., epothilone and yersiniabactin, and effect *C*-methylation in those contexts[76] (Figure 27b).

NRPS machinery and pyridoxal phosphate enzymology are merged in the first module of the mycosubtilin synthetase.[77] An aminotransferase domain (AMT) exists within the module presumed to generate a long chain β-ketoacyl-*S*-T intermediate. The AMT domain uses the α-amino group of glutamine in a variant of a classic transamination reaction[78] to produce α-keto-glutamine and the β-aminoacyl-*S*-T intermediate.[79] Elongation by NRPS modules ultimately leads to macrocyclization and release in which this β-amino group is the internal nucleophile during cyclization to yield the peptidolactam mycosubtilin (Figure 28).

It is clear that Nature mixes and matches assembly-line logic for both NRP and PK construction with additional modes of enzyme-mediated chemistry available in the bacterial cells. Sometimes, the auxiliary catalytic domains are accommodated in cis within a module; other times, they work as separate in trans catalysts to alter the growing chain tethered on the assembly-line way stations. Undoubtedly, the evolution of NRPS and PKS assembly lines must have involved the splicing of autonomously folded domains into and out of the synthetase, without disruption of its architecture and the function of its other domains.

3.3. PKS–NRPS Hybrid Assembly Lines

Knowing the underlying logic and protein machinery used for NRPS and Type I PKS assembly lines, one can predict the type and order of domains and modules in a hybrid assembly line for production of a mixed PK−NRP product. Many examples are known that validate these expectations.[7,9−12,80,81] Some hybrid assembly lines contain mostly PKS machinery, such as the FK520 synthase.[7] FK520 assembly involves 11 PKS modules distributed over 3 proteins, FkbB, FkbC, and FkbA. One NRPS module exists on a separate protein, FkbP. The four-domain FkbP (and analogously, RapP from the rapamycin synthase) selects and activates the six-membered imino acid pipecolate and inserts it into the PK backbone[82,83] (Figure 29a). In contrast, the bleomycin assembly line consists of 10 NRPS modules and 1 PKS module, which provides the single malonyl-CoA-derived unit in this antitumor drug[10] (Figure 29b).

The *Yersinia pestis* siderophore yersiniabactin is generated from a hybrid assembly line containing two protein subunits that function in the order HMWP2 → HMWP1.[12,84] The HMWP2 subunit contains two NRPS modules for generating the hydroxyphenylbithiazolinyl-*S*-T intermediate. The HMWP1 subunit starts with a PKS module, followed by the third NRPS module of the assembly line (Figure 29c).

One intriguing aspect of these hybrid assembly lines is how they perform either C−C or C−N bond elongations at

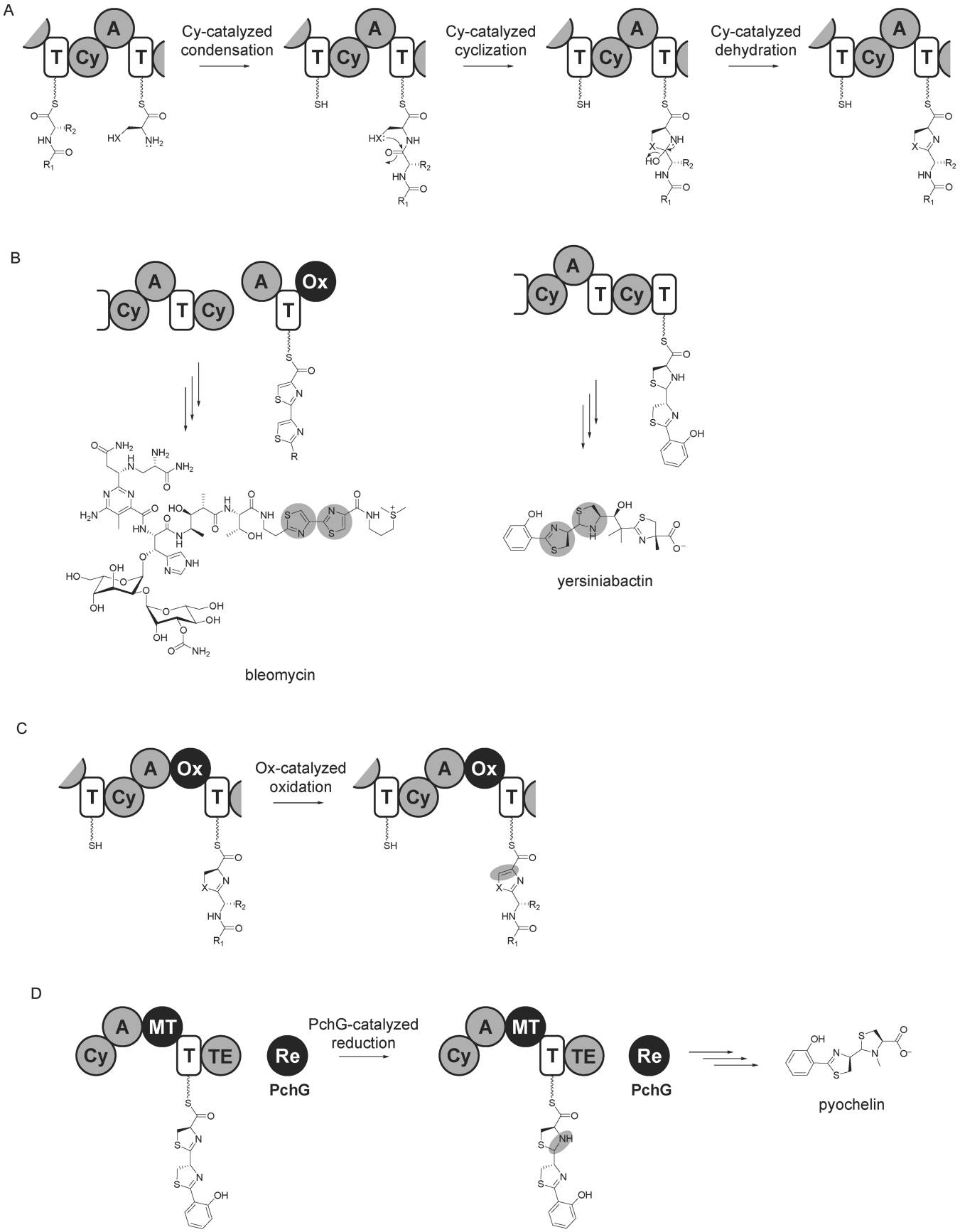

Figure 26. Formation of heterocycles in NRP backbones by Cy domains. (A) Cy domains catalyze attack of a nucleophilic Ser, Thr, or Cys side chain on the upstream carbonyl and then dehydrate the resultant adduct to yield the oxazoline or thiazoline. (B) The PK−NRP hybrids bleomycin[10] and yersiniabactin[12] contain bithiazole and thiazoline-thiazolidine moieties, respectively, that derive from adjacent Cy-containing modules in their synthetases. (C) Action of the oxidase (Ox) domain to convert the dihydroheterocyclic Cy product to the fully aromatic thiazole or oxazole. (D) In trans action of the reductase (Re) domain from the pyochelin[73] NRPS to reduce a thiazoline to a thiazolidine. MT = methyltransferase.

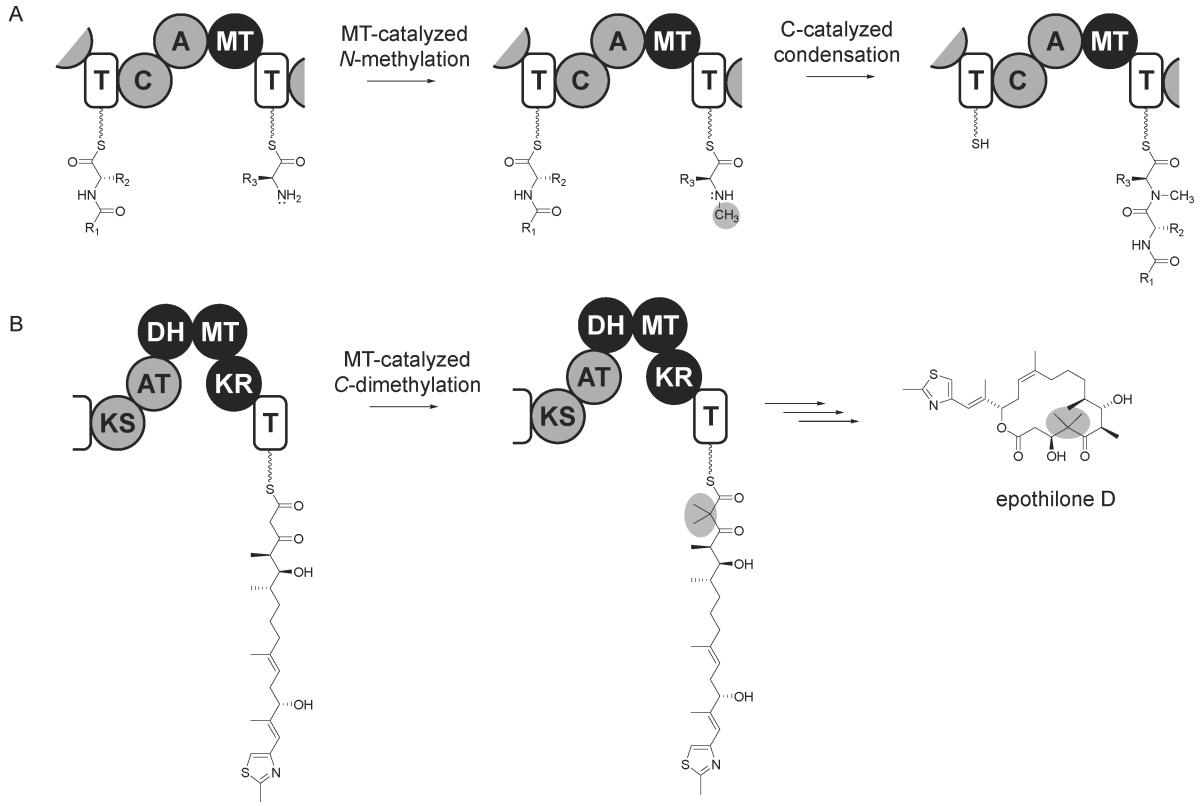

Figure 27. *N*- and *C*-Methylation of NRPs and PKs by embedded MT domains. (A) *N*-Methylation of aminoacyl-*S*-T amines in NRP assembly lines. (B) Presumed *C*-methylation of a nascent PK intermediate during epothilone biosynthesis.[11]

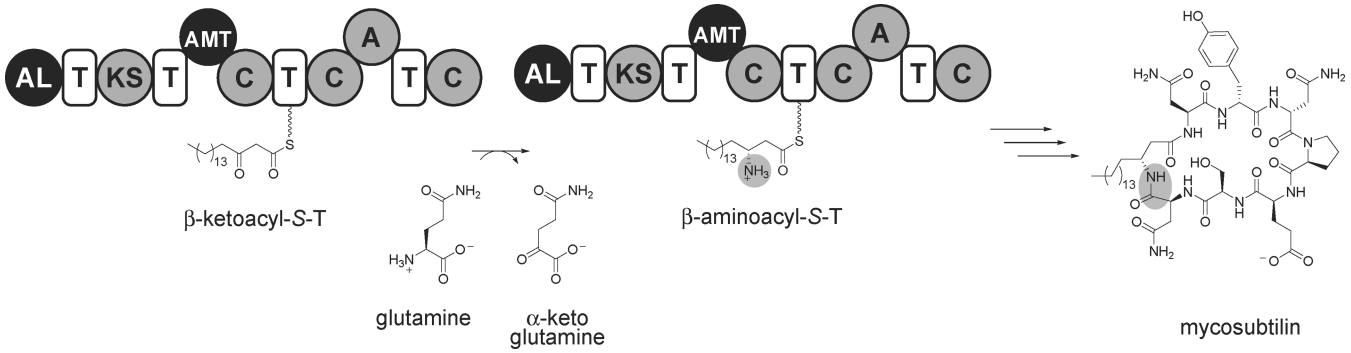

Figure 28. Action of the aminotransferase (AMT) domain in the mycosubtilin NRPS to convert a β-ketoacyl-*S*-T to a β-aminoacyl-*S*-T with concomitant conversion of glutamine to α-keto-glutamine.[79] AL = acyl-CoA ligase.

PKS/NRPS interfaces. Each time the assembly line switches from PKS to NRPS or vice versa, the growing chain must be translocated across the PKS/NRPS and NRPS/PKS interfaces. For example, at the NRPS/PKS interface of the yersiniabactin synthetase that occurs at the subunit interface between HMWP2 and HMWP1, the KS domain must accept an upstream peptidyl chain rather than ketide chain as a donor cosubstrate. In contrast, at the subsequent PKS/NRPS interface in HMWP1, the C domain must accept a dimethylmalonyl acyl donor rather than a peptidyl donor cosubstrate[84] (Figure 29d). Likewise, during the net insertion of the pipecolyl moiety into the FK520 chain, the NRPS subunit is a four-domain C_1-A-T-C_2 catalyst. The C_1 domain forms the C−N bond between the PK acyl chain and the pipecolyl nitrogen[82] (Figure 30). The C_2 domain then presumably catalyzes O−C bond formation to complete the macrocyclization, acting as a surrogate TE domain to form the macrolactone linkage.

3.4. Post-Assembly-Line Tailoring of PKs and NRPs

Following release from their assembly lines, many PKs and NRPs are chemically modified by dedicated enzymes encoded in their biosynthetic gene clusters (Figure 31). These modifications, which are often required for the biological activity of the natural product, modulate the hydrophobicity of a PK or NRP scaffold. Hydrophobic PK scaffolds such as erythromycin and fumonisin undergo hydroxylation by a heme-utilizing P450 and a non-heme iron/α-ketoglutarate-dependent dioxygenase, respectively, increasing the hydrophilicity of these molecules.[85−87] In contrast, the hydrophilic NRP aminoadipoyl-cysteinyl-valine is oxidatively cyclized to the β-lactam core by a non-heme-iron-dependent dioxygenase,[88] imparting increased rigidity and hydrophobicity on this scaffold. Glycosyltransferases can conjugate hydrophilic nucleotide-activated hexoses from primary metabolism to

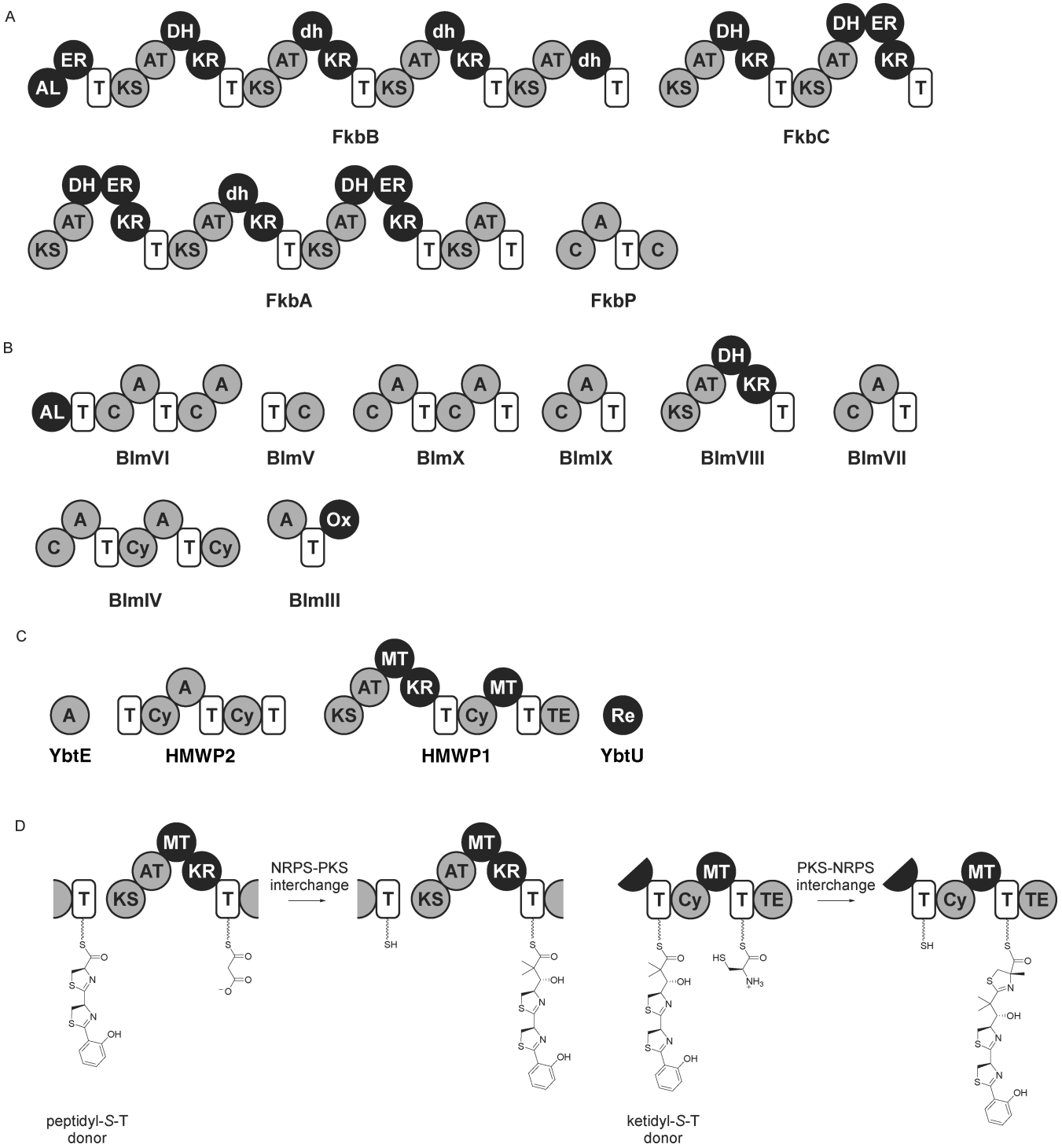

Figure 29. Representative hybrid PK−NRP assembly lines. (A) The FK520 synthase[7] contains 11 PKS modules and 1 NRPS module. (B) The bleomycin synthetase[10] comprises 10 NRPS modules and 1 PKS module. (C) The yersiniabactin synthetase[12] is composed of 3 NRPS modules and 1 PKS module. (D) Recognition of noncanonical acyl donor substrates by the KS and C domains at the PKS−NRPS interfaces of the yersiniabactin synthetase.[84] Inactive domains are indicated by lowercase lettering.

NRP scaffolds as with the glucosylation of teicoplanin.[89] Alternatively, they can transfer unusual hydrophobic deoxyhexoses—produced from primary metabolites by dedicated multistep enzymatic pathways—to PK scaffolds as with the daunosaminylation of the anthracycline daunosamine[90] and the proposed construction of the disaccharide in chartreusin.[91] These hexoses are often further tailored, as exemplified by the successive *O*-methylation and *O*-acylation of adjacent hydroxyls on the noviose sugar of the aminocoumarin novobiocin.[92]

4. Assembly-Line Operations

4.1. Size of Assembly Lines

Type I PKS assembly lines and NRPS assembly lines can require megadaltons of protein components. The 6-DEB (DEBS) assembly line contains seven modules distributed over three proteins of about 200 kDa each.[93] The tyrocidine synthetase contains 10 modules organized into 3 proteins, TycABC, comprising 30 domains in just over 1 MDa of

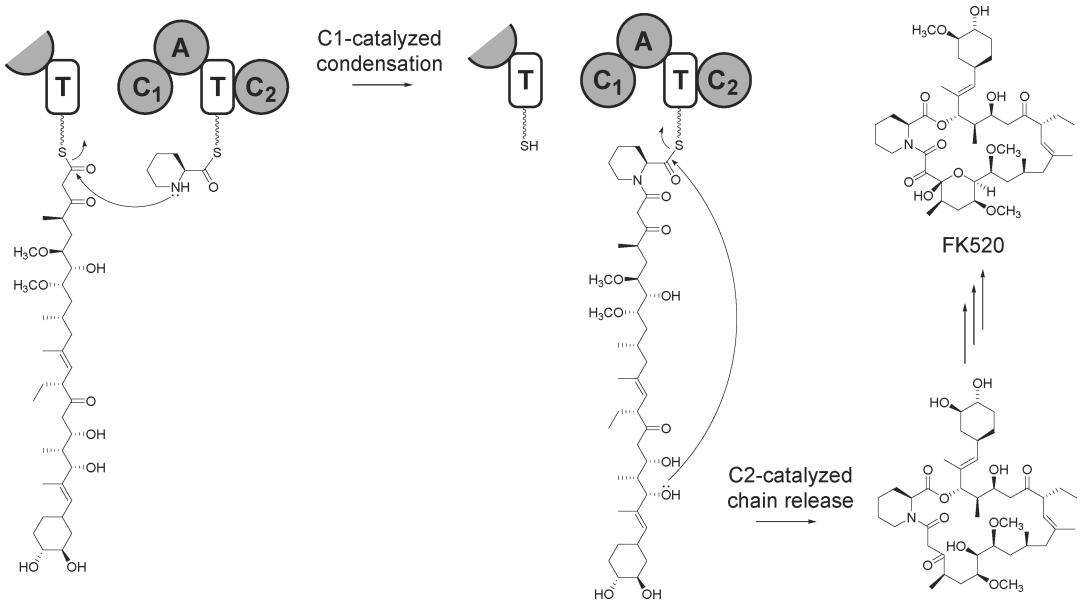

Figure 30. Action of the C1 domain[82] and proposed action of the C2 domain of the FkbP NRPS module from the FK520 assembly line.

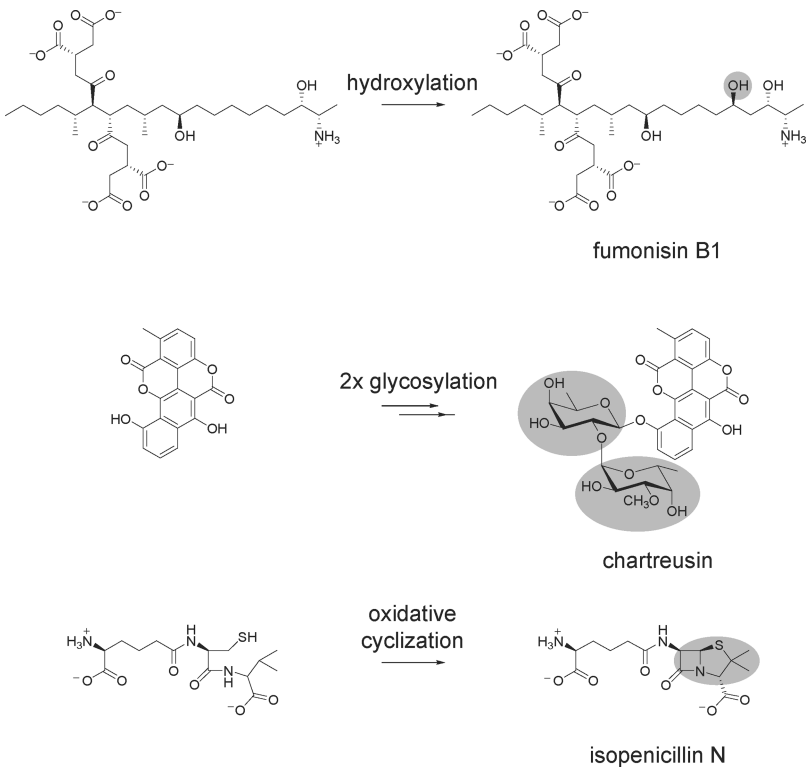

Figure 31. Post-assembly-line tailoring. Maturation of the PK fumonisin B1,[87] the PK chartreusin,[91] and the NRP isopenicillin N[88] involve post-assembly-line hydroxylation, glycosylation, and oxidative cyclization, respectively.

protein.[94] The cyclosporine synthetase has 41 domains distributed over 11 modules, encompassed in a single protein of ~15 000 amino acid residues with a mass of 1.5 MDa[95] (Figure 32a). These proportions lend themselves to a "beads on a string" analogy for the assembly lines in which each bead is a (semi)autonomously folded catalytic or carrier domain.

Presumably, proteins this large are susceptible to folding problems, vulnerable to proteolysis, and prone to mutational inactivation. The more common organizational scheme of distributing the modules over multiple protein subunits may aid in avoiding these problems and could also facilitate evolution of assembly-line components with domain additions, subtractions, or substitutions.

On the other hand, splitting an assembly line into multiple proteins creates new problems. These problems include proper domain folding at breakpoints and inter-protein recognition of high enough affinity for assembly-line components to transfer growing acyl chains across subunit interfaces. In the three-subunit DEBS assembly line, Khosla, Cane, and colleagues[96,97] pioneered detection of linker elements between the C-terminus of one subunit (e.g.,

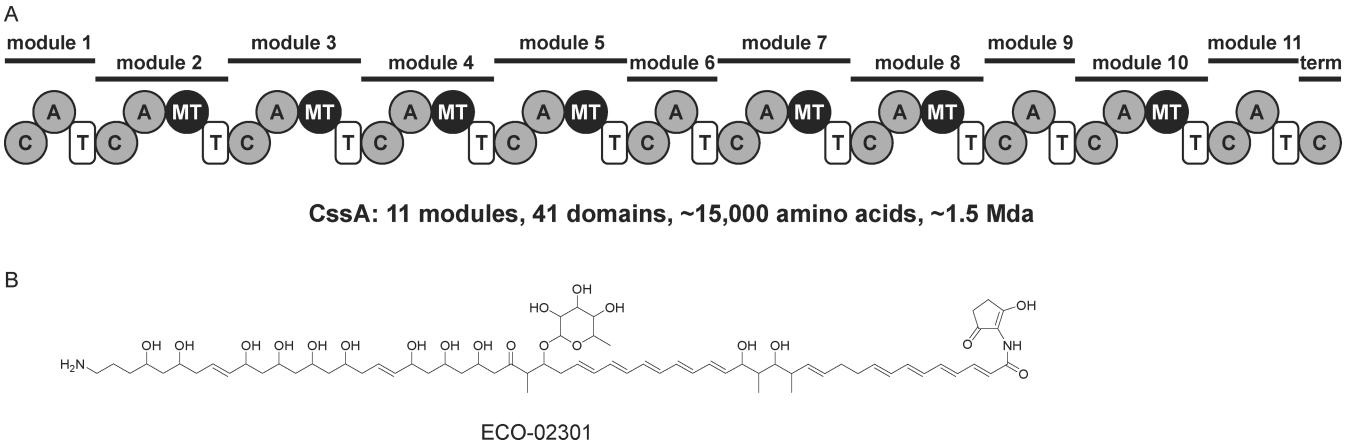

Figure 32. Large assembly lines and assembly-line products. (A) Schematic of the cyclosporine megasynthetase.[16] (B) Chemical structure of ECO-02301.[102]

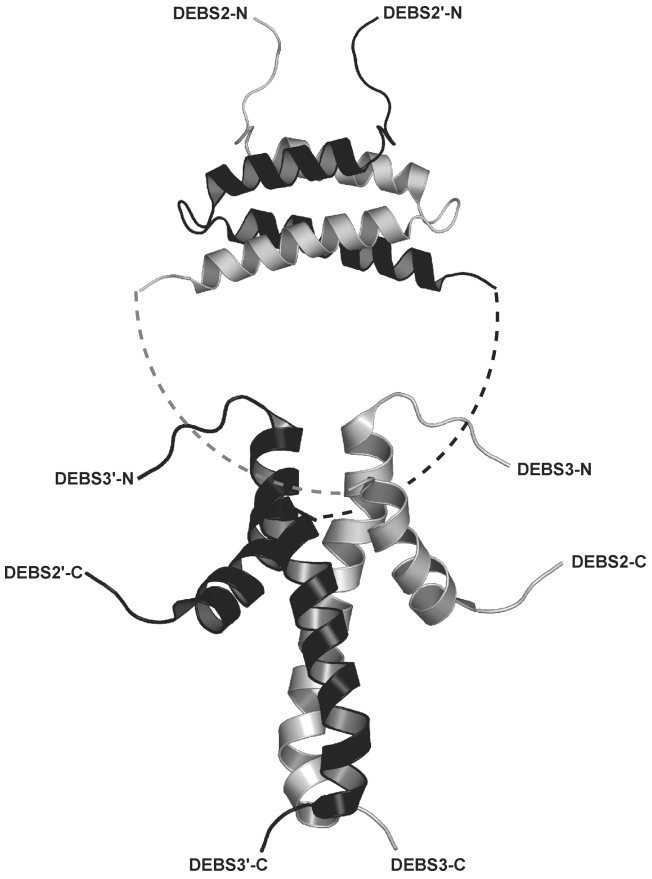

Figure 33. NMR structure of linker regions that mediate recognition between the DEBS2 and DEBS3 polypeptides of the erythromycin synthase (PDB codes 1PZQ, 1PZR). While the structure was solved as a fusion between the DEBS2 and DEBS3 fragments, part of the loop connecting these fragments is not shown in the figure in order to represent in trans docking between the DEBS2 and DEBS3 fragments. Dashed lines indicate mobile loops of poorly defined structure. Adapted from ref 98.

DEBS1) and the N-terminus of the next (DEBS2) which direct this in trans interaction. NMR analysis of the linkers has revealed the three-dimensional basis of this recognition[98] (Figure 33). There appear to be comparable intersubunit linkers between proteins in NRPS[99,100] and hybrid NRPS/PKS assembly lines,[9,101] which should facilitate re-engineering of new subunit interactions in assembly lines for novel products.

What is the upper limit for the size of a PKS or NRPS assembly line? The linear PK ECO-02301 (Figure 32b) consists of a γ-aminobutyryl starter unit and 26 C_2 or C_3 extender units; its assembly line has 122 domains spread over 9 proteins, totaling a molecular mass of 4.7 MDa for the assembly line.[102] The NRP syringopeptin contains 21 amino acid residues, and its assembly line has 68 domains spread over 3 proteins, totaling a molecular mass of 2.7 MDa for the assembly line.[65] Their producing organisms make an enormous investment of energy and amino acid mass into one machine that makes a single product. RNA-templated peptide synthesis is surely more efficient from the point of view of microbial cell economy. The importance of functions performed by PKs, NRPs, and PK−NRPs must warrant the Herculean energetic cost of making megadalton protein assembly lines.

4.2. Control of Full-Length Chain Assembly

Enzymatic assembly lines must control chain initiation, elongation, and termination in an orderly way. The mechanisms of quality control in assembly lines have been opaque but are slowly being deciphered.

Two major challenges in assembly-line action are (i) how to avoid initiation at internal modules and (ii) how to suppress premature hydrolytic termination of chain elongation. Taking NRPSs as an example, it appears that each module can stochastically select, activate, and load an aminoacyl moiety on its T domain.[103] In principle, chains could begin elongating at any internal module and generate N-terminally truncated products. The mechanism for avoiding premature hydrolytic release of incomplete chains is not understood but may arise from a combination of steric sequestration of the growing chain on each T domain and a short lifetime for growing chains on any T domain. However, the rifamycin synthetase system appears to accumulate elongating chains at every T domain way station.[104] Derailing hydrolysis is most pronounced for T domains at subunit interfaces, which suggests that these terminal motifs may be less efficient at sequestering nascent intermediates.[84]

4.3. Editing of Stalled Chains by External TEs

A complementary challenge is how to restore function to a stalled assembly line that is blocked by a partially elongated chain. Perhaps an inappropriate monomer has been misincorporated by an A domain and cannot be condensed by the downstream C domain. There exist external thioesterase

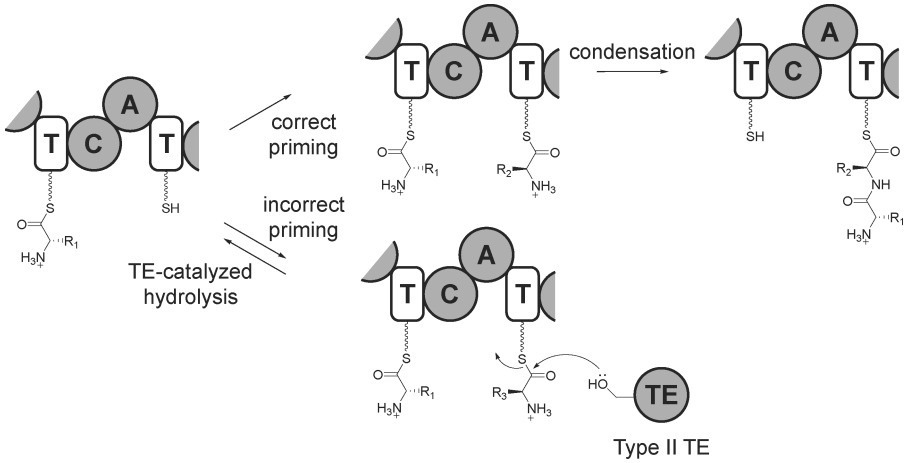

Figure 34. Type II TE domains catalyze hydrolytic editing of assembly-line intermediates. Mispriming of the T domain can be reversed by the Type II TE domain, regenerating the free pantetheinyl thiol and preventing the assembly line from stalling.

domains, often encoded in biosynthetic gene clusters, which are thought to serve as editing catalysts[105,106] (Figure 34). They are thought to patrol the assembly line in trans, and when a stalled chain is detected by its persistent accumulation at a T domain, it can be removed hydrolytically. An additional quandary is that levels of acetyl-CoA surpass those of CoASH in many bacterial cells, and the PPTases that prime apo-T domains with phosphopantetheinyl arms derived from CoASH can also use acetyl CoA and transfer acetyl-*S*-pantetheinyl phosphate.[107] The acetyl-*S*-T domains generated in this way are capped way stations, rendering the assembly line inactive. External TEs will hydrolyze the acetyl moieties, regenerating the free HS-pantetheinyl arms needed to function as the covalent way stations for elongating PK and NRP chains.[105,106]

4.4. Missing Domains, Stuttering, Skipping, and Other Collinearity Violations

Undoubtedly, microbial producers of polyketides, nonribosomal peptides, and hybrid small molecules tinker with assembly-line organizations and operations to evolve new combinations.[108] In the small sampling of assembly lines studied to date, variants have been detected in which domains normally present in cis are missing. For example, the yersiniabactin system contains three NRPS modules but only one A domain.[12] It has been shown that this A domain activates L-cysteine and supplies it in cis to one T domain and in trans to the other two T domains.[109] It is not yet clear whether Cys-AMP diffuses freely, aminoacyl transfer occurs among the T domains, or the A domain interacts directly with all three T domains.

Multiple Type I PKS clusters have now been sequenced that lack one, several, or all of the AT domains.[110−114] Instead, AT domains act in trans to supply malonyl or methylmalonyl units to the holo T domains in the AT-less modules. The free-standing AT domains may be components of Type II FAS or PKS systems in the producer organisms or stand-alone AT domains in an otherwise Type I PKS organization. There are examples of free-standing T domains, and in the jamaicamide cluster,[49] such a separate T domain follows three T domains fused in cis for a tandem run of four T domains. Any differentiating functions of the four T domains have yet to be elucidated.

There are examples in which modules appear to be skipped during assembly-line operation,[115,116] raising questions of architecture and orientation of downstream T domains in different modules. Alternatively, growing PK chains can apparently pause and be elongated twice at the same way station in a stuttering mode of assembly-line operation.[117−119] This behavior, which focuses attention on the complexities of unit operations in these protein assembly lines, will be explained by some combination of kinetics of the chemical transformations and architectural relationships of the relevant modules. It is likely that additional noncanonical arrangements and stoichiometries of various domains will continue to be uncovered as more natural product assembly lines are characterized.

Examples of iterative Type I PKSs have been described.[120] The first class, consisting of a single KS-AT-DH-(ER)-(KR)-T polypeptide, is proposed to assemble mono- and bicyclic aromatic acids including orsellinic acid[121] and naphthalinic acid[122] and polyunsaturated fatty acids such as the starter unit for myxochromide assembly.[116] The second class, which has been implicated in the assembly of the enediyne cores of C-1027 and calicheamicin, comprise single polypeptides with a KS-AT-T-KR-DH-TD domain structure in which the terminal domain (TD) has not been characterized.[123,124] It is not yet known how the mechanistic details of iterative Type I PKS elongation compare to those of Type II PKSs.

The macrotetrolide polyether nonactin is assembled by a new class of T-domain-independent Type II-like PKSs.[125,126] The nonactin synthase features five ORFs predicted to encode KSs, and genetic evidence suggests that these enzymes condense soluble acyl-CoA substrates, forming the C−C and C−O linkages found in the backbone of nonactin. A recent study in the nonactin-producing PKS[127] supports the hypothesis that the formation of tetrahydrofuran and tetrahydropyran rings embedded in the backbone of polyether PKs involves epoxidation of olefins followed by bond migration, as hypothesized by Cane and co-workers.[128]

4.5. Assembly-Line Fragments: A-T Didomains and Free-Standing A and T Domains

Bioinfomatic searches using particular domains or didomain combinations as queries reveal assembly-line fragments. The limiting case is the Type II PKS systems in which the KS, AT, and T components are classically separate proteins, as noted earlier.

Fragments of NRPS assembly lines—A-T didomains, or separate but adjacently encoded A and T domains—are found

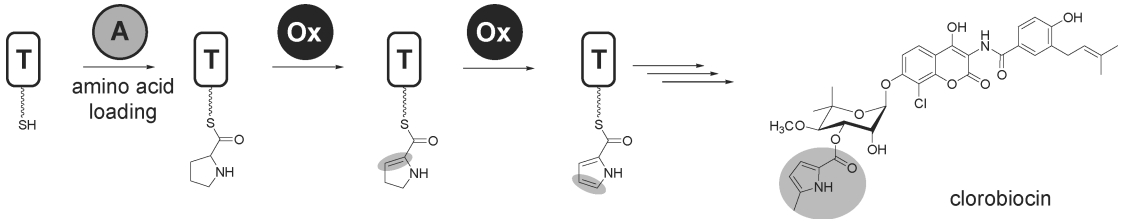

Figure 35. Conversion of A-T tethered Pro-*S*-T to pyrrole-2-*S*-T during pyoluteorin, prodigiosin, and clorobiocin biosynthesis.[135]

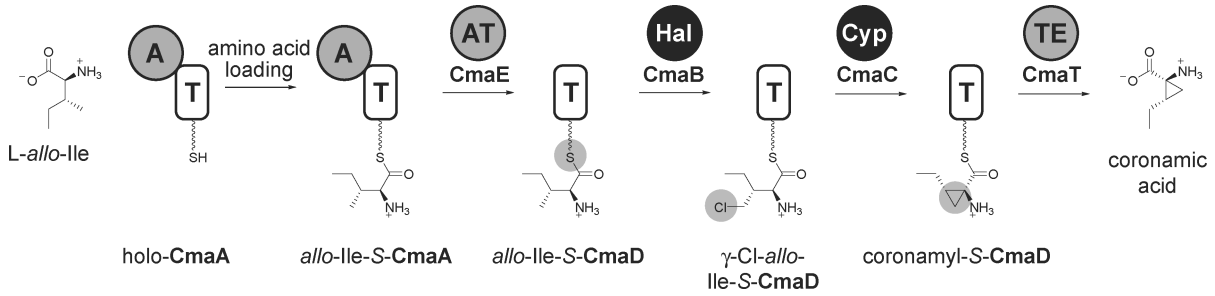

Figure 36. Formation of coronamic acid from L-*allo*-Ile by CmaABCDET.[141] CmaA loads itself with L-*allo*-Ile, which is then transferred to the free-standing T domain CmaD by the acyltransferase CmaE. L-*allo*-Ile-*S*-CmaD is the substrate for the halogenase (Hal) CmaB, and the resultant γ-Cl-L-*allo*-Ile-*S*-CmaD is converted to coronamyl-*S*-CmaD by the cyclopropanase (Cyp) CmaC. Coronamic acid is hydrolyzed from CmaD by the free-standing TE CmaT.

in the absence of any other NRPS machinery.[129] Investigation of several of these free-standing A-T (or A/T) proteins has confirmed that they activate a specific L-amino acid and install it as the aminoacyl-*S*-T. This thioester-linked aminoacyl group is then the substrate for a partner enzyme, which effects a chemical transformation at the β- or γ-carbon of the aminoacyl-*S*-T.[130]

In effect, the use of free-standing A-T or A/T motifs is a strategy to sequester a fraction of the pool of a proteinogenic amino acid and modify it to a nonproteinogenic form for use in secondary metabolic pathways. This logic is illustrated in Figure 8 for hydroxylation of a Tyr-*S*-T in vancomycin and aminocoumarin biosynthesis to yield the β-OH-Tyr-*S*-T.[131,132] In vancomycin assembly, the thioester is hydrolyzed by a thioesterase and the free β-OH-Tyr is utilized by the corresponding assembly-line modules in the vancomycin synthetase. In contrast, during biosynthesis of the aminocoumarins novobiocin, clorobiocin, and coumermycin, the β-OH-Tyr-*S*-NovH is then a substrate for the NovJ and NovK proteins, which act oxidatively to produce the β-keto-Tyr-*S*-NovH,[133] an intermediate during formation of the aminocoumarin nucleus. A comparable β-hydroxylation of His-*S*-NikP1 occurs in the nikkomycin biosynthetic pathway.[134]

Analogously, the conversion of proline to pyrrole-2-carboxylic acid occurs while proline is covalently tethered in thioester linkage to a T domain in several biosynthetic pathways, including those of pyoluteorin, prodigiosin, and clorobiocin[135,136] (Figure 35). The Pro-*S*-T is a substrate for double dehydrogenation to the pyrrolyl-*S*-T product. Activation of the carboxyl as a thioester facilitates acyl transfer to the 3′-OH of the noviosyl sugar during clorobiocin biosynthesis[137] and to the decarboxylated malonyl-*S*-T in the PKS modules for pyoluteorin[138] and prodigiosin[139] biosynthesis.

During biosynthesis of the phytotoxin coronatine by *Pseudomonas syringae*, the coronamic acid moiety (1-amino-1-carboxy-2-ethylcyclopropane) is generated from L-*allo*-Ile[129] (Figure 36). A free-standing A-T didomain forms the *allo*-Ile-*S*-T thioester,[140] which is then transferred to a distinct T domain, CmaD. The *allo*-Ile-*S*-CmaD is a substrate for γ-chlorination by CmaB,[141] which activates the γ-carbon for CmaC-catalyzed displacement by an α-carbanion to produce the coronamyl-*S*-CmaD intermediate. Coronamic acid is released from CmaD by action of a thioesterase. Overall, a six-protein ensemble converts *allo*-Ile to coronamic acid.[141] It is likely that other A-T didomains and free-standing A and T domain pairs will be discovered that carry out novel chemical traformations of proteinogenic to nonproteinogenic amino acids while tethered as aminoacyl-*S*-T species.

4.6. Assembly-Line Dynamics

PKSs and NRPSs are described as "assembly lines" because of their protein domain architecture, but do they really behave like molecular assembly lines, channeling a continuous stream of thioester-bound intermediates from one T domain to the next? Although the chemical nature of intermediates during PK and NRP assembly are known and kinetic parameters of catalysis by individual domains have been determined, the dynamics of monomer activation and chain elongation in the context of a larger assembly are still obscure. Floss and co-workers have shown that a rifamycin assembly line defective in the amide synthase that catalyzes chain termination builds up discrete intermediates at several T domains,[104] supporting the notion of assembly-line-like function. New mass spectrometric approaches,[142−144] which are able to detect T domain-bound intermediates, will be an important tool in elucidating the occupancies of multiple T domains under turnover conditions.

4.7. PKS/NRPS Structural Biology and Quaternary Structure

One notable gap in our knowledge is an understanding of the three-dimensional structure of modules, the basic unit responsible for monomer incorporation in FAS, PKS, and NRPS assembly lines. There exist X-ray or NMR structures for all of the core PKS domains[145−147] (Figure 37a), largely from Type II PKS and FAS systems, in which all of the subunits are separate proteins. There is also an X-ray structure of a PPTase for priming apo-T domains.[148] From NRPS systems, there are structures for individual C, A, T,

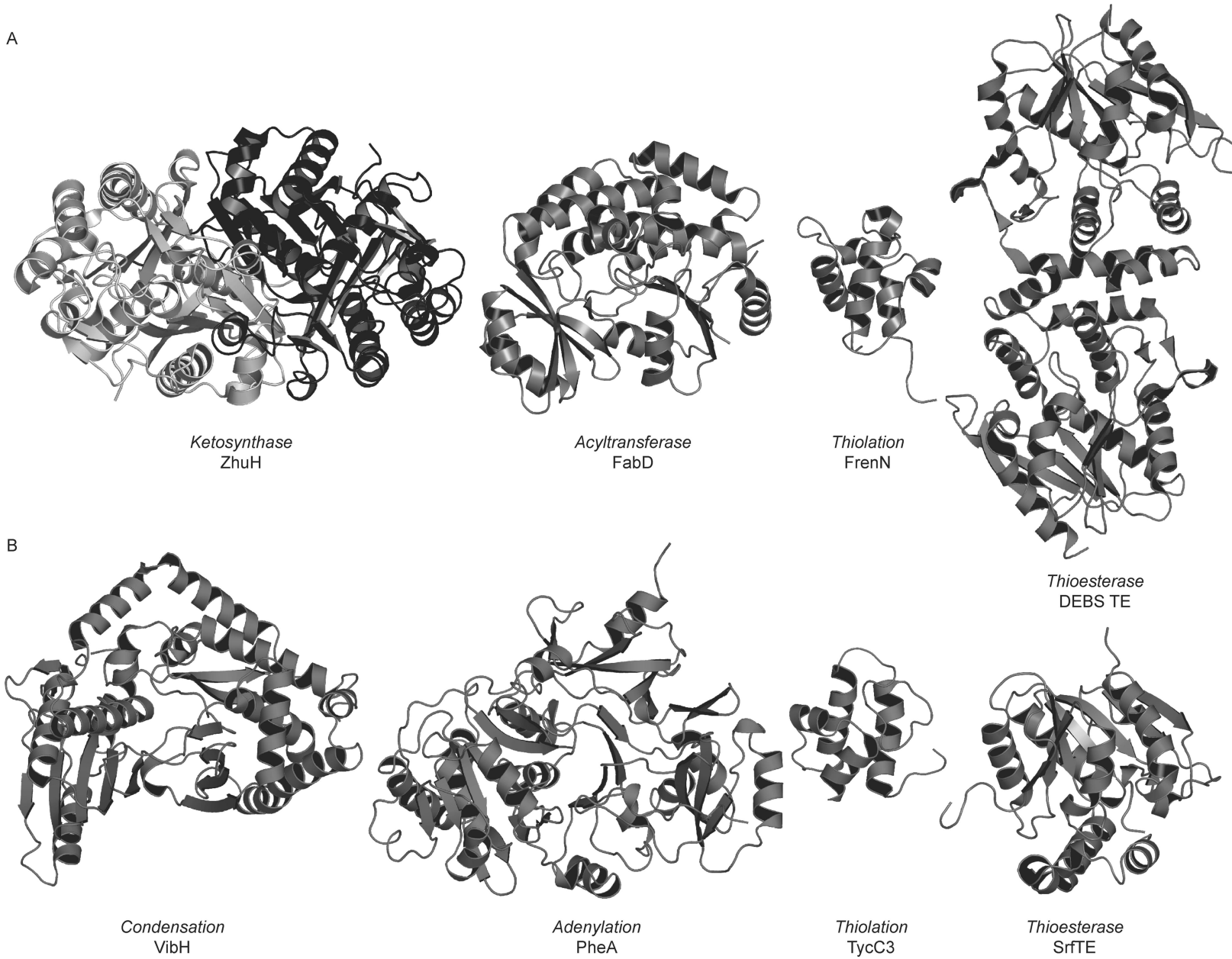

Figure 37. Structures of individual FAS, PKS, and NRPS domains. (A) Structures of the KS dimer ZhuH[146] (1MXJ) from the R1128 synthase, the *Streptomyces coelicolor* A3(2) AT domain FabD[147] (1NM2), the T domain FrenN[145] (1OR5) from the frenolicin synthase, and the TE domain DEBS TE[183] (1KEZ) from the erythromycin synthase. (B) Structures of the free-standing Cy domain VibH[149] (1L5A) from the vibriobactin synthetase, the A domain PheA[150] (1AMU) from the gramicidin synthetase, the T domain TycC3[151] (1DNY) from the tyrocidine synthetase, and the TE domain SrfTE[152] (1JMK) from the surfactin synthetase. PDB codes for each structure are given in parentheses.

and TE domains excised from NRPS assembly lines[149−152] (Figure 37b). However, no structures of multidomain fragments (e.g., C-A or A-T) or full modules have been reported, and therefore, we know very little about how domains interact with each other. Such a molecular view of the quaternary organization of assembly lines will be instrumental in understanding the organizational principles of assembly-line function. One of the central questions will be the structural role of linker regions that (i) connect protein domains in cis (within a polypeptide) and (ii) mediate the in trans interaction between modules on separate polypeptides such as those in the erythromycin PKS.[98] Another set of questions will surround the flexibility of the quaternary architecture: (i) how the bond-forming KS/C domain can interact with its two acyl-phosphopantetheinyl-*S*-T substrates, (ii) how the phosphopantetheinyl-*S*-T species can interact with upstream and downstream KS/C domains, and (iii) how optional domains such as KR/DH/ER and E/MT are accommodated within the core domain structure of a module. Taking a wider view, a recent study in the uncharacterized *B. subtilis* Pks system provides tantalizing evidence of colocalization of assembly-line components that is disrupted by the absence of an upstream multimodular assembly-line enzyme from the synthase, suggesting collinear scaffolding of the separate polypeptides comprising a single assembly line (Straight, P.; Kolter, R.; Rudner, D. Unpublished observations).

5. Mechanism

5.1. Domain and Module Architectures and Conformations

The underlying logic for the biosynthesis of PKs, NRPs, and PK−NRP hybrids is understood, and the protein machinery responsible for their assembly can be connected to the identity and arrangement of monomer units oligomerized by an assembly line. Nevertheless, there remain many mechanistic and architectural questions about synthetases and their components.

X-ray or NMR structures will be required to understand how catalytic domains act sequentially around a T domain within a module to catalyze ordered operations on the acyl/peptidyl-*S*-T domain intermediate before the chain gets transferred to the downstream module. To illustrate the myriad operations required in assembly-line function, it is instructive to focus on the biosynthetic cycle of a T domain. Within each elongation module of an NRPS, the T domain is first recognized by a dedicated PPTase for priming with the phosphopantetheinyl arm[23] (Figure 38a). Second, the 20-Å-long arm of the holo-T domain enters the active site of the A domain, in which the aminoacyl-AMP is sequestered from bulk solvent and then transferred to the pantetheinyl thiolate (Figure 38b). Third, the aminoacyl-*S*-T domain is presented to the upstream C domain (C_2 in Figure 38) for peptide bond formation (Figure 38c). Once the peptidyl chain has translocated to this T domain, it finally enters the active site of the downstream C domain (C_3 in Figure 38) to serve as the electrophilic donor for peptide bond formation (Figure 38d).

In a Type I PKS module containing KS, AT, KR, DH, and ER domains, there must be at least seven comparable recognition steps of the acyl chain and the T domain protein scaffold: one each by the PPTase, the five catalytic domains in the module, and the KS in the downstream module. For an 11-module PKS assembly line such as the FK520 synthase, >50 operations must be performed just at the T domains. It is not yet known which faces of the T domain four-helix bundle (Figure 37) are recognized by partner enzymes.[153,154] Likewise, the extent to which the pantetheinyl arm and its tethered acyl chain move between catalytic domains in a module (the classical "swinging arm" hypothesis[155]) is unclear; reciprocally, it is not known how much the conformations of catalytic domains change during a biosynthetic cycle. The flexibility required for catalytic domains to function in modules may mean that they assume a wide range of conformations; this could be one of the difficulties encountered by researchers who, to date, have unsuccessfully attempted to crystallize modules.

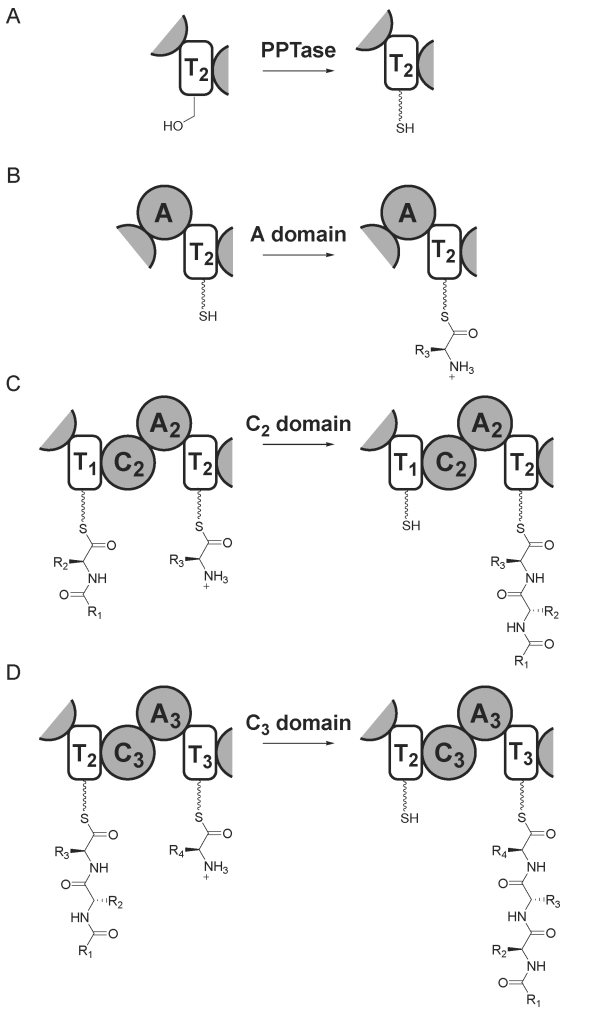

Figure 38. Interacting partners of the T domain during NRPS function. (A) Priming of the T domain by a PPTase. (B) Aminoacylation of the T domain by the upstream A domain. (C) Peptidylation of the aminoacyl-*S*-T by the upstream C domain. (D) Deacylation of the peptidyl-*S*-T by the downstream C domain.

5.2. Monomer Activation and Module Loading

Acyl thioester chemistry is used for chain elongation on both PKS and NRPS assembly lines, providing thermodynamic activation to both monomers and elongating chains covalently tethered to the T domains of assembly-line modules.

The (methyl)malonyl monomers are loaded by AT domains onto the pantetheinyl arms of the holo-T domains as (methyl)malonyl-*S*-CoAs. Malonyl-CoA is also used as the

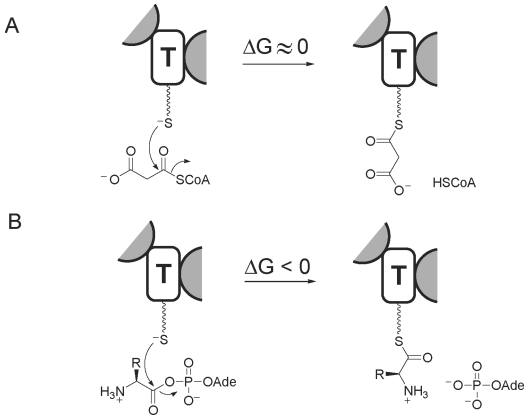

Figure 39. Thermodynamics of monomer loading in PKSs and NRPSs. (A) In PKSs, transthiolation of the (methyl)malonyl unit from CoA to the pantheteinyl thiolate is energy neutral. (B) In NRPSs, the capture of an aminoacyl-AMP mixed anhydride by the pantetheinyl thiolate is a thermodynamically favorable process.

monomeric building block in the primary metabolic pathway of fatty acid synthesis, so the use of this metabolite in PK biosynthesis merely requires its diversion down a conditional metabolic pathway. The thioesters acetyl-CoA and malonyl-CoA are finely balanced—between thermodynamic activation for group transfer of the acyl moiety as an electrophile and kinetic stability toward the weak nucleophile water—to have a useful lifetime in cells as diffusible metabolites. Thus, loading a malonyl or methylmalonyl unit from a soluble CoA thioester onto the thiolate of the pantetheinyl arm of a PKS T domain is an energy-neutral transthiolation (Figure 39a).

A convergent strategy for monomer activation is used to get to the comparable aminoacyl/peptidyl-*S*-T in NRPSs. Aminoacyl-*S*-CoA thioesters are not found in prokaryotic or eukaryotic cells. In contrast to simple acyl-CoAs such as malonyl-CoA, which have half-lives of many hours to days at neutral to slightly acidic pH, aminoacyl-CoAs are more labile due to the electronic effects of the α-amine group.[156] Their hydrolytic lability apparently precludes their use as soluble, activated aminoacyl donors. As an alternative strategy for amino acid activation, Nature employs the ATP-dependent adenylation of the carboxylate of amino acids to produce aminoacyl-AMP and PPi. The aminoacyl-AMPs are both thermodynamically activated and kinetically labile, and they are held tightly in the active sites of enzymes that generate them, which include both NRPS A domains and—more famously—aminoacyl-tRNA synthetases[157] (Figure 39b).

During aminoacyl-tRNA synthetase action, the activated aminoacyl group is attacked by the 2′- or 3′-OH of the terminal A residue of the cognate cosubstrate tRNA. Some of the thermodynamic activation in the aminoacyl-AMP is preserved in the oxoester linkage of the aminoacyl-tRNA, while enough kinetic stability to hydrolysis has been acquired (oxoester instead of thioester) that the aminoacyl-*O*-tRNAs are sufficiently stable to be used as diffusible ferries to bring amino acids to mRNA codons in the ribosome. During the action of A-T didomains of NRPS modules, the intramolecular concentration of pantetheinyl thiolate is high enough to capture the aminoacyl-AMP in the A domain active site, yielding the thioester (Figure 38b). No diffusible aminoacyl thioesters are involved, and presumably the aminoacyl-*S*-T and peptidyl-*S*-T intermediates are kinetically sequestered from hydrolysis to avoid derailing the assembly-line process.

5.3. Protein Thioester and Oxoester Intermediates

Assembly-line enzymology is, in one sense, the catalysis of an elongating cascade of acyl-*S*-protein intermediates from N-terminal T domain to C-terminal T domain. Protein acyl thioester chemistry may be quite ancient, and it is certainly useful in several modalities. There are hundreds of thiol proteases that generate peptidyl-*S*-enzyme intermediates during the course of peptide bond hydrolysis.[158] During protein splicing, thioester intermediates are generated in the first cleavage step when Cys side chains attack the upstream peptide.[159] These thioesters are not hydrolyzed, but they are subject to translocation following attack by downstream cysteine thiolates. Finally, *S*- to *N*-acyl shifts regenerate peptide bonds. The use of the pantetheinyl arm to tether acyl intermediates is also akin to the use of lipoyl arms for acyl capture in α-ketoacid oxidative decarboxylation enzymes,[155] but the pantetheinyl prosthetic group strategy may derive proximally from FASs. FASs, in turn, may have evolved first by capture of the pantetheinyl fragment of the readily available CoASH. The chemistry of the serine phosphorylation step by a serine side chain is analogous to the action of a serine protein kinase[160] (Figure 40).

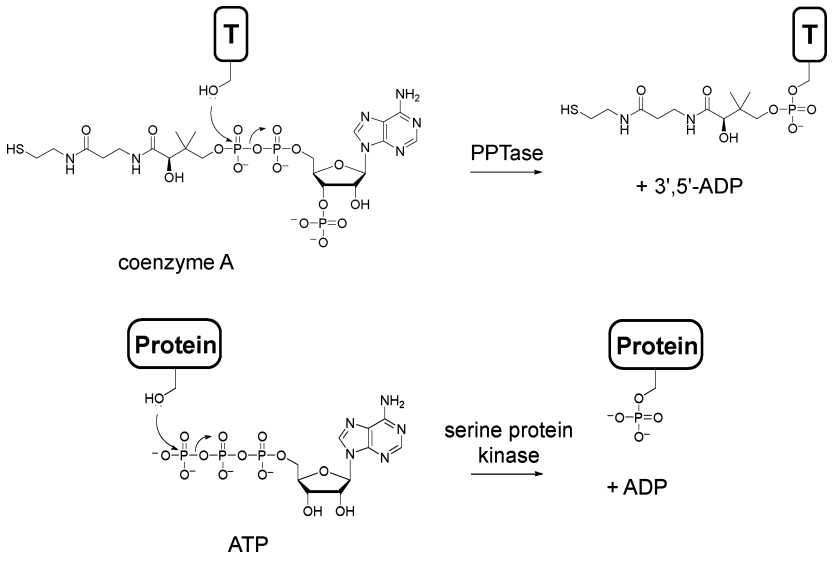

Figure 40. Comparison of protein kinase-catalyzed serine phosphorylation and PPTase-catalyzed serine phosphopantetheinylation.

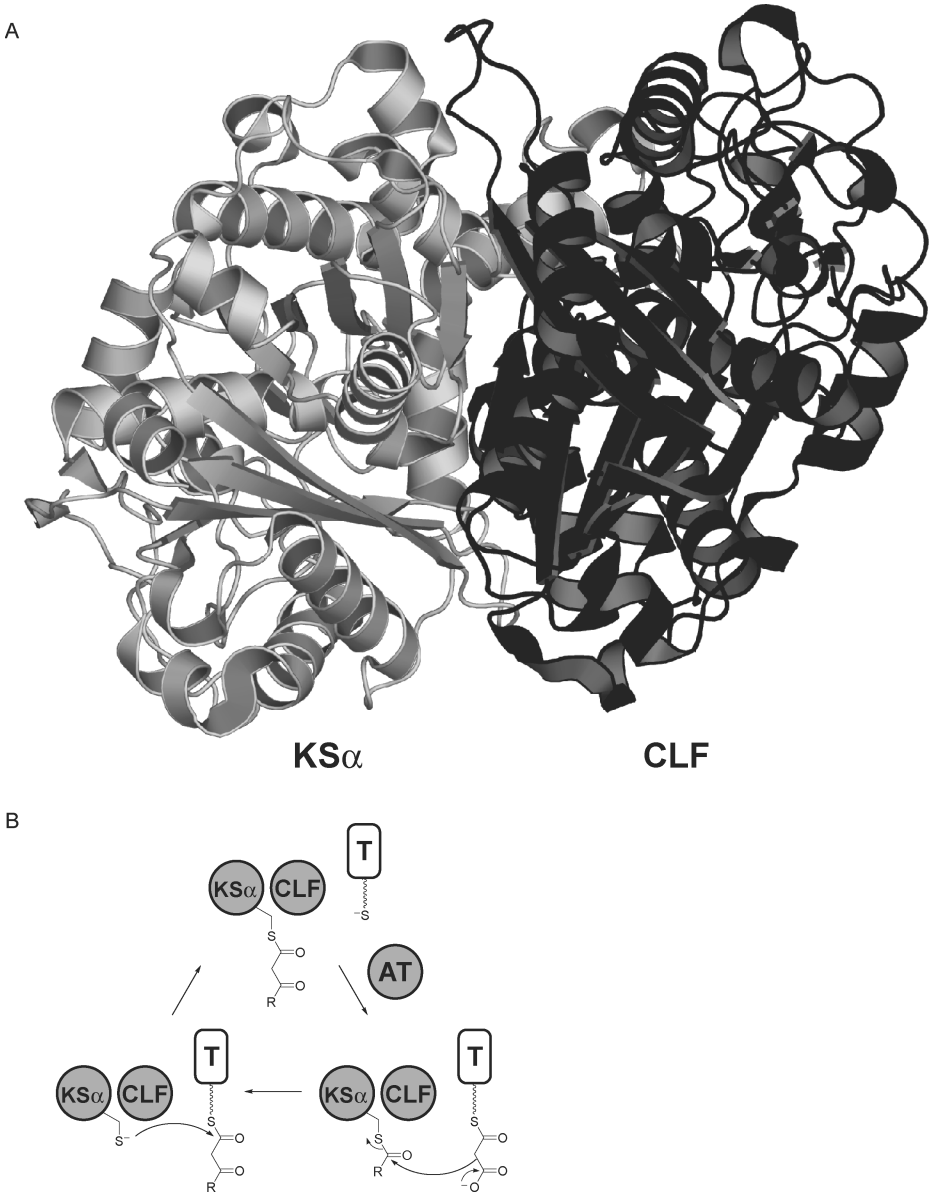

Figure 41. Structure and mechanism of KS-CLF from Type II PKSs. (A) Crystal structure of the KSα/CLF heterodimer[163] (PDB code 1TQY). (B) Mechanism of iterative chain elongation catalyzed by KSα/CLF and their associated T and AT domains.

At the end of the assembly line, to prepare for chain termination and disconnection of the covalent linkage between the full-length chain and the final T domain, the assembly line switches from acyl thioester chemistry to acyl oxoester chemistry (Figure 18). The active site serine of TE domains in both PKS and NRPS assembly lines attacks the adjacent acyl-*S*-T intermediate to create an acyl-*O*-Ser-TE as the last acyl-enzyme intermediate in the assembly line. In ribosomal protein synthesis, by analogy, the activated amino acids are ferried to the peptidyl transferase center to dock on mRNA codons as oxoesters (aminoacyl-*O*-tRNAs). TE domains are members of the α/β hydrolase superfamily of enzymes[161]—which use active site serine side chains as nucleophiles—so the evolutionary precursors to these domains might have been available as part of the protein inventory of domain building blocks in microbial cells. In Type II PKS and FAS systems, the TEs are separate proteins.

While acyl-*O*-TE intermediates are less thermodynamically activated than acyl-*S*-enzymes, their equilibria lie far in favor of hydrolysis, as evidenced by the large numbers of serine proteases found in proteomes.[162] During synthetase action, the lifetimes of acyclic polyketidyl-*O*-TEs and peptidyl-*O*-TEs must be long enough for intramolecular capture by side chain nucleophiles to proceed, yielding the macrocyclic lactone and lactam scaffolds that have conformational constraints essential to many of the biological functions of these natural products.[55]

5.4. Condensation Catalytic Domains

The catalytic domains that carry out iterative chain-elongating condensations are central to the action of multimodular assembly lines. Much is known about the chemistry and mechanism of the family of β-ketoacyl synthases from structural and mechanistic analyses of several soluble variants from Type II PKS and FAS systems[146,163−166] (Figure 41). An active site cysteine acts as a thiolate nucleophile on an acyl thioester donor to yield an acyl-*S*-Cys-KS intermediate. This species is the electrophilic donor in the subsequent condensation with a (methyl)malonyl-*S*-T. The residues in the KS domain that are required to induce decarboxylation of the (methyl)malonyl and generate the

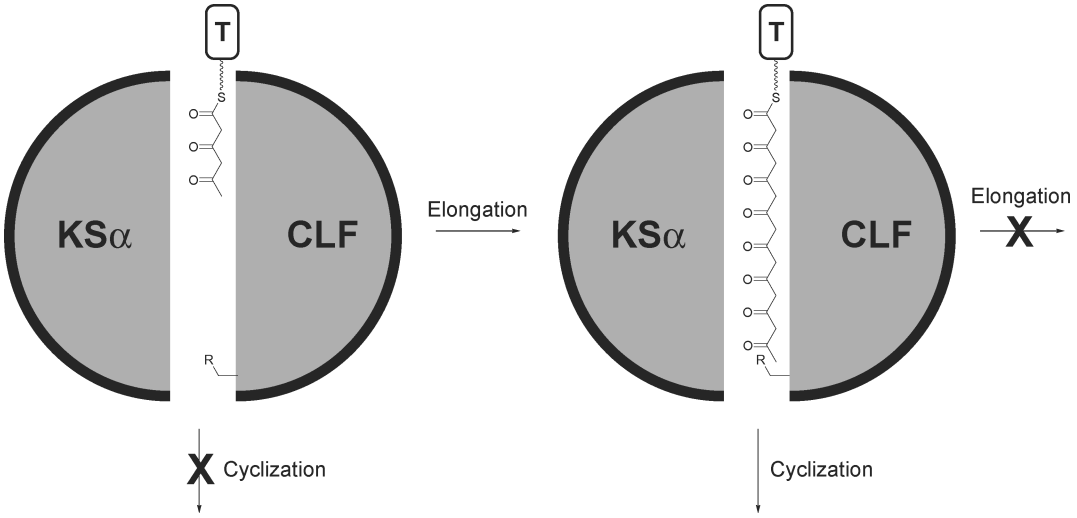

Figure 42. Proposed "volume control" by CLF active sites in determining chain length. The polyketone chain is extended until it fills the CLF tunnel, which is bounded by a gating residue indicated by R. Full-length chains are then diverted to cyclization.

C_2 carbanion of the acceptor have also been identified.[166] Studies have been conducted on donor and acceptor specificity with regard to R groups at the Cα position;[167,168] this information is relevant to the use of promiscuous KS domains in module-reprogramming efforts. In Type II PKS systems, the KS is a heterodimer of an active subunit (KS_α) and an inactive subunit termed CLF (chain length factor). Studies on the KS_α-CLF heterodimer indicate that the volume of the polyketidyl binding pocket of the CLF subunit is an important determinant of acyl chain length[163,169] (Figure 42).

Much less is known about the structure and mechanism of C domains from NRPS systems. The X-ray structure of a free-standing C domain from the vibriobactin synthetase, VibH, has been solved[149] (Figure 37). As expected from bioinformatic predictions, VibH bears homology to the chloramphenicol acetyltransferase family of enzymes.[170] In contrast to the covalent catalysis mechanism of KS domains, it is anticipated that amide/peptide bond condensation by C domains will be a direct acyl transfer, but no catalytic residues have been firmly identified.[171,172] Further structural studies of C domains—both in isolation and embedded within modules—will be needed to elucidate the mechanistic details of these NRPS bond-forming catalysts.

Some C domains have evolved the ability to catalyze epimerization (dual-function C/E domains) and cyclodehydration (Cy domains). It is not yet known how Cy domains catalyze the attack of thiol and hydroxyl side chains of Cys, Ser, and Thr residues onto the newly formed amide linkage. The attack conformations may be very similar to those in protein-splicing catalytic domains.

C domains have chiral specificity. The majority of C domains use L-donors and L-acceptors and form the corresponding L-,L-peptidyl-*S*-T products; these are $^{L}C_{L}$ catalysts.[173] We have noted that a subset of C domains introduce D-configured residues into elongating peptidyl chains. These C domains recognize D-aminoacyl groups as the residue proximal to the thioester in the upstream T domain, so they are $^{D}C_{L}$ catalysts.[61] Iterative action of modules containing E-$^{D}C_{L}$ didomains will allow generation of D-,D-linkages in elongating peptidyl chains. To date, no examples of $^{D}C_{D}$ domains have been detected.

6. Summary

Assembly lines for turning monomeric amino acids and/or acyl-CoAs into complex oligomeric natural products are remarkable protein machines. Covalently tethering elongating chains in thioester linkage for directional chain growth, they are Nature's equivalent of solid-state synthesis arrays. None of the intermediates are soluble, and—barring adventitious hydrolysis—none should escape the assembly line. The chemistry of elongation is elegantly simple and reasonably robust.

In PKS and NRPS systems, the organization of protein domains is based on *covalent* linkage within large polypeptides, and the channeling of small molecule intermediates is similarly controlled by *covalent* tethering to way stations in the assembly line. The existence of such an organizational scheme for substrate transfer begs comparison to other metabolic systems in which substrate channeling is known to operate. In the heterodimeric enzymes tryptophan synthase[174] and carbamoyl phosphate synthetase,[175] for example, the reactive intermediates indole and ammonia are channeled from one active site to another through tunnels of ~25 and 45 Å, respectively. In contrast, little is known about the cellular organization of the sets of enzymes involved in basic metabolic pathways such as glycolysis and amino acid biosynthesis. Considering the apparent efficiency of covalent linkage as an inter-enzyme substrate transfer principle, it is tempting to speculate that these fundamental metabolic systems are arranged such that noncovalent (but well-organized) protein−protein interactions replace covalent domain linkage to facilitate efficient substrate transfer. Indeed, the free-standing A and T domains of the initiation module of the enterobactin NRPS[176] may be vestiges of an earlier—noncovalent—association of protein domains in assembly-line systems.

The limitations to assembly-line biosynthesis probably come from the sheer bulk of the protein machinery needed to provide the template for the natural product. The minimal modules (KS-AT-T or C-A-T) required for selection and incorporation of a single acyl-CoA or amino acid comprise over 100 kDa (1000 amino acid residues) of protein mass. A decapeptide product such as tyrocidine requires a megadalton assembly line. The 21-mer syringopeptin—the largest

known NRP for which a biosynthetic gene cluster has been identified—requires over 2.7 MDa of protein machinery for its assembly line.[65] The energy cost to the cell to use enzymes totaling ~27 000 amino acid residues to make a cyclic 21-mer peptide may well be near the upper limit of microbial economy for secondary metabolites, no matter how efficiently the assembly line operates.

It would be less costly in protein mass to use iterative Type II PKS rather than assembly-line Type I PKS organization, and iterative use of the catalytic subunits on the one T domain may be a more efficient way to make large polyketide chains. No Type II NRPS machines have yet been reported, perhaps precluded by the absence of A domains promiscuous for iterative loading of different amino acid monomers.

The products released by TE domain action from both PKS and NRPS assembly lines may be biosynthetic end products themselves, but in many—perhaps most—pathways, assembly-line products are only late-stage intermediates. As soluble small molecules, these assembly-line products are substrates for a diverse set of tailoring enzymes that decorate the PK or NRP scaffold. These tailoring enzymes can alter polarity (acylation, methylation, glycosylation, hydroxylation) and introduce functional groups (e.g., deoxyhexoses[35,177]) that often serve as recognition elements for interaction with biological targets.

The genes that encode the great majority of natural products built by assembly-line logic are clustered together, and they include four types of functions: (i) "just in time" provision of specific monomers for use by the assembly line, (ii) assembly-line modules themselves, (iii) enzymes for modifications during assembly-line action and post-assembly-line tailoring, and (iv) efflux and other self-protection mechanisms for the microbial producer.

Continued insights into the logic, machinery, and mechanism of assembly lines will reveal how building blocks are oligomerized by microbial enzyme factories into the vast architectural and functional range of PKs and NRPs. Bacteria and fungi are clearly evolving new assembly lines to generate molecular diversity that may confer on them a selective advantage. Humans might be able to practice combinatorial biosynthesis by reprogramming, using principles deciphered thus far.

7. Acknowledgments

Work in the authors' laboratory has been supported by NIH grants GM20011, GM49338, and AI47238 and a graduate fellowship from the Hertz Foundation. We thank Chris Calderone, Nathan Magarvey, and Ellen Yeh for critical reading of the manuscript and helpful comments. We are indebted to all of our colleagues noted in the individual citations listed in this review.

8. Note Added in Proof

Recent work by Calderone et al.[184] has elucidated the roles of the enzymes involved in isoprene β-branch insertion. Two recent reports by Ban and coworkers[185,186] show the structural architecture of mammalian and fungal fatty acid synthases. A recent report by Koglin et al.[187] sheds light on the dynamics of T domain structure.

9. References

(1) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675.
(2) Kim, E. S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. *Gene* **1994**, *141*, 141.
(3) Aharonowitz, Y.; Cohen, G.; Martin, J. F. *Annu. Rev. Microbiol.* **1992**, *46*, 461.
(4) Martin, J. F. *Appl. Microbiol. Biotechnol.* **1998**, *50*, 1.
(5) van Wageningen, A. M.; Kirkpatrick, P. N.; Williams, D. H.; Harris, B. R.; Kershaw, J. K.; Lennard, N. J.; Jones, M.; Jones, S. J.; Solenberg, P. J. *Chem. Biol.* **1998**, *5*, 155.
(6) Motamedi, H.; Shafiee, A. *Eur. J. Biochem.* **1998**, *256*, 528.
(7) Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D. *Gene* **2000**, *251*, 81.
(8) Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J. B.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839.
(9) Aparicio, J. F.; Molnar, I.; Schwecke, T.; Konig, A.; Haydock, S. F.; Khaw, L. E.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 9.
(10) Du, L.; Sanchez, C.; Chen, M.; Edwards, D. J.; Shen, B. *Chem. Biol.* **2000**, *7*, 623.
(11) Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. *Science* **2000**, *287*, 640.
(12) Pelludat, C.; Rakin, A.; Jacobi, C. A.; Schubert, S.; Heesemann, J. *J. Bacteriol.* **1998**, *180*, 538.
(13) Moore, B. S.; Hertweck, C. *Nat. Prod. Rep.* **2002**, *19*, 70.
(14) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275.
(15) Miao, V.; Coeffet-Legal, M. F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M.; Silva, C. J.; Wrigley, S. K.; Baltz, R. H. *Microbiology* **2005**, *151*, 1507.
(16) Marahiel, M. A. *Chem. Biol.* **1997**, *4*, 561.
(17) McDaniel, R.; Welch, M.; Hutchinson, C. R. *Chem. Rev.* **2005**, *105*, 543.
(18) DOE JGI Integrated Microbial Genomes Database, http://img.jgi-.doe.gov/pub/main.cgi, accessed 10/24/2005.
(19) Paulsen, I. T.; Press, C. M.; Ravel, J.; Kobayashi, D. Y.; Myers, G. S.; Mavrodi, D. V.; DeBoy, R. T.; Seshadri, R.; Ren, Q.; Madupu, R.; Dodson, R. J.; Durkin, A. S.; Brinkac, L. M.; Daugherty, S. C.; Sullivan, S. A.; Rosovitz, M. J.; Gwinn, M. L.; Zhou, L.; Schneider, D. J.; Cartinhour, S. W.; Nelson, W. C.; Weidman, J.; Watkins, K.; Tran, K.; Khouri, H.; Pierson, E. A.; Pierson, L. S., III; Thomashow, L. S.; Loper, J. E. *Nat. Biotechnol.* **2005**, *23*, 873.
(20) Stinear, T. P.; Mve-Obiang, A.; Small, P. L.; Frigui, W.; Pryor, M. J.; Brosch, R.; Jenkin, G. A.; Johnson, P. D.; Davies, J. K.; Lee, R. E.; Adusumilli, S.; Garnier, T.; Haydock, S. F.; Leadlay, P. F.; Cole, S. T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1345.
(21) Keating, T. A.; Walsh, C. T. *Curr. Opin. Chem. Biol.* **1999**, *3*, 598.
(22) Stein, T.; Vater, J.; Kruft, V.; Otto, A.; Wittmann-Liebold, B.; Franke, P.; Panico, M.; McDowell, R.; Morris, H. R. *J. Biol. Chem.* **1996**, *271*, 15428.
(23) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923.
(24) Majerus, P. W.; Alberts, A. W.; Vagelos, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *53*, 410.
(25) Byford, M. F.; Baldwin, J. E.; Shiau, C. Y.; Schofield, C. J. *Chem. Rev.* **1997**, *97*, 2631.
(26) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *Nature* **1995**, *378*, 263.
(27) Steller, S.; Vollenbroich, D.; Leenders, F.; Stein, T.; Conrad, B.; Hofemeister, J.; Jacques, P.; Thonart, P.; Vater, J. *Chem. Biol.* **1999**, *6*, 31.
(28) Ogle, J. M.; Brodersen, D. E.; Clemons, W. M., Jr.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. *Science* **2001**, *292*, 897.
(29) Cane, D. E.; Walsh, C. T. *Chem. Biol.* **1999**, *6*, R319.
(30) Khosla, C.; Harbury, P. B. *Nature* **2001**, *409*, 247.
(31) Zaslaver, A.; Mayo, A. E.; Rosenberg, R.; Bashkin, P.; Sberro, H.; Tsalyuk, M.; Surette, M. G.; Alon, U. *Nat. Genet.* **2004**, *36*, 486.
(32) Lowden, P. A.; Wilkinson, B.; Bohm, G. A.; Handa, S.; Floss, H. G.; Leadlay, P. F.; Staunton, J. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 777.
(33) Admiraal, S. J.; Walsh, C. T.; Khosla, C. *Biochemistry* **2001**, *40*, 6116.
(34) Yu, T. W.; Muller, R.; Muller, M.; Zhang, X.; Draeger, G.; Kim, C. G.; Leistner, E.; Floss, H. G. *J. Biol. Chem.* **2001**, *276*, 12546.
(35) He, X. M.; Liu, H. W. *Annu. Rev. Biochem.* **2002**, *71*, 701.
(36) Staunton, J.; Wilkinson, B. *Chem. Rev.* **1997**, *97*, 2611.
(37) Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 525.
(38) Hopwood, D. A. *Chem. Rev.* **1997**, *97*, 2465.
(39) Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380.
(40) Austin, M. B.; Noel, J. P. *Nat. Prod. Rep.* **2003**, *20*, 79.
(41) Tseng, C. C.; McLoughlin, S. M.; Kelleher, N. L.; Walsh, C. T. *Biochemistry* **2004**, *43*, 970.

(42) Mootz, H. D.; Finking, R.; Marahiel, M. A. *J. Biol. Chem.* **2001**, *276*, 37289.
(43) Bisang, C.; Long, P. F.; Cortes, J.; Westcott, J.; Crosby, J.; Matharu, A. L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F. *Nature* **1999**, *401*, 502.
(44) Wilkinson, C. J.; Frost, E. J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2001**, *8*, 1197.
(45) Keating, T. A.; Ehmann, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T. *Chembiochem* **2001**, *2*, 99.
(46) Ikeda, H.; Nonomiya, T.; Usami, M.; Ohta, T.; Omura, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9509.
(47) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Hutchinson, C. R. *Chem. Biol.* **1997**, *4*, 433.
(48) Zawada, R. J.; Khosla, C. *Chem. Biol.* **1999**, *6*, 607.
(49) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* **2004**, *11*, 817.
(50) Campobasso, N.; Patel, M.; Wilding, I. E.; Kallender, H.; Rosenberg, M.; Gwynn, M. N. *J. Biol. Chem.* **2004**, *279*, 44883.
(51) Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W. H. *J. Nat. Prod.* **2004**, *67*, 1356.
(52) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. *Structure (Camb)* **2004**, *12*, 1865.
(53) Sieber, S. A.; Marahiel, M. A. *Chem. Rev.* **2005**, *105*, 715.
(54) Schmoock, G.; Pfennig, F.; Jewiarz, J.; Schlumbohm, W.; Laubinger, W.; Schauwecker, F.; Keller, U. *J. Biol. Chem.* **2005**, *280*, 4339.
(55) Kohli, R. M.; Walsh, C. T. *Chem. Commun. (Camb.)* **2003**, 297.
(56) Gaitatzis, N.; Kunze, B.; Muller, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11136.
(57) Velasco, A.; Acebo, P.; Gomez, A.; Schleissner, C.; Rodriguez, P.; Aparicio, T.; Conde, S.; Munoz, R.; de la Calle, F.; Garcia, J. L.; Sanchez-Puelles, J. M. *Mol. Microbiol.* **2005**, *56*, 144.
(58) Walsh, C. T. *J. Biol. Chem.* **1989**, *264*, 2393.
(59) Hoffmann, K.; Schneider-Scherzer, E.; Kleinkauf, H.; Zocher, R. *J. Biol. Chem.* **1994**, *269*, 12710.
(60) Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, *39*, 5775.
(61) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2003**, *42*, 12095.
(62) Linne, U.; Doekel, S.; Marahiel, M. A. *Biochemistry* **2001**, *40*, 15824.
(63) Roongsawang, N.; Hase, K.; Haruki, M.; Imanaka, T.; Morikawa, M.; Kanaya, S. *Chem. Biol.* **2003**, *10*, 869.
(64) Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D. C.; Grandi, G. *J. Biol. Chem.* **1998**, *273*, 32857.
(65) Scholz-Schroeder, B. K.; Soule, J. D.; Gross, D. C. *Mol. Plant Microbe Interact.* **2003**, *16*, 271.
(66) Balibar, C. J.; Vallaincourt, F. V.; Walsh, C. T. *Chem. Biol.* **2005**, in press.
(67) Gehring, A. M.; Mori, I.; Perry, R. D.; Walsh, C. T. *Biochemistry* **1998**, *37*, 11637.
(68) Schneider, T. L.; Shen, B.; Walsh, C. T. *Biochemistry* **2003**, *42*, 9722.
(69) Sakai, T. T.; Riordan, J. M.; Glickson, J. D. *Biochemistry* **1982**, *21*, 805.
(70) Schneider, T. L.; Walsh, C. T. *Biochemistry* **2004**, *43*, 15946.
(71) Cox, C. D.; Rinehart, K. L., Jr.; Moore, M. L.; Cook, J. C., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4256.
(72) Crosa, J. H.; Walsh, C. T. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 223.
(73) Reimmann, C.; Patel, H. M.; Serino, L.; Barone, M.; Walsh, C. T.; Haas, D. *J. Bacteriol.* **2001**, *183*, 813.
(74) Kleinkauf, H.; von Dohren, H. *Eur. J. Biochem.* **1990**, *192*, 1.
(75) Hacker, C.; Glinski, M.; Hornbogen, T.; Doller, A.; Zocher, R. *J. Biol. Chem.* **2000**, *275*, 30826.
(76) Miller, D. A.; Walsh, C. T.; Luo, L. *J. Am. Chem. Soc.* **2001**, *123*, 8434.
(77) Duitman, E. H.; Hamoen, L. W.; Rembold, M.; Venema, G.; Seitz, H.; Saenger, W.; Bernhard, F.; Reinhardt, R.; Schmidt, M.; Ullrich, C.; Stein, T.; Leenders, F.; Vater, J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13294.
(78) Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383.
(79) Aron, Z. D.; Dorrestein, P. C.; Blackhall, J. R.; Kelleher, N. L.; Walsh, C. T. *J. Am. Chem. Soc.* **2005**, in press.
(80) Tang, G. L.; Cheng, Y. Q.; Shen, B. *Chem. Biol.* **2004**, *11*, 33.
(81) Silakowski, B.; Schairer, H. U.; Ehret, H.; Kunze, B.; Weinig, S.; Nordsiek, G.; Brandt, P.; Blocker, H.; Hofle, G.; Beyer, S.; Muller, R. *J. Biol. Chem.* **1999**, *274*, 37391.
(82) Gatto, G. J., Jr.; McLoughlin, S. M.; Kelleher, N. L.; Walsh, C. T. *Biochemistry* **2005**, *44*, 5993.
(83) Konig, A.; Schwecke, T.; Molnar, I.; Bohm, G. A.; Lowden, P. A.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **1997**, *247*, 526.
(84) Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T. *Chem. Biol.* **2002**, *9*, 333.
(85) Lambalot, R. H.; Cane, D. E.; Aparicio, J. J.; Katz, L. *Biochemistry* **1995**, *34*, 1858.
(86) Weber, J. M.; Leung, J. O.; Swanson, S. J.; Idler, K. B.; McAlpine, J. B. *Science* **1991**, *252*, 114.
(87) Ding, Y.; Bojja, R. S.; Du, L. *Appl. Environ. Microbiol.* **2004**, *70*, 1931.
(88) Roach, P. L.; Clifton, I. J.; Fulop, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. *Nature* **1995**, *375*, 700.
(89) Losey, H. C.; Peczuh, M. W.; Chen, Z.; Eggert, U. S.; Dong, S. D.; Pelczer, I.; Kahne, D.; Walsh, C. T. *Biochemistry* **2001**, *40*, 4745.
(90) Olano, C.; Lomovskaya, N.; Fonstein, L.; Roll, J. T.; Hutchinson, C. R. *Chem. Biol.* **1999**, *6*, 845.
(91) Xu, Z.; Jakobi, K.; Welzel, K.; Hertweck, C. *Chem. Biol.* **2005**, *12*, 579.
(92) Freel Meyers, C. L.; Oberthur, M.; Xu, H.; Heide, L.; Kahne, D.; Walsh, C. T. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 67.
(93) Kao, C. M.; Katz, L.; Khosla, C. *Science* **1994**, *265*, 509.
(94) Mootz, H. D.; Marahiel, M. A. *J. Bacteriol.* **1997**, *179*, 6843.
(95) Kleinkauf, H.; Dittmann, J.; Lawen, A. *Biomed. Biochim. Acta* **1991**, *50*, S219.
(96) Kumar, P.; Li, Q.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2003**, *125*, 4097.
(97) Wu, N.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2001**, *123*, 6465.
(98) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723.
(99) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15585.
(100) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 275.
(101) Liu, F.; Garneau, S.; Walsh, C. T. *Chem. Biol.* **2004**, *11*, 1533.
(102) McAlpine, J. B.; Bachmann, B. O.; Piraee, M.; Tremblay, S.; Alarco, A. M.; Zazopoulos, E.; Farnet, C. M. *J. Nat. Prod.* **2005**, *68*, 493.
(103) Mootz, H. D.; Schwarzer, D.; Marahiel, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5848.
(104) Yu, T. W.; Shen, Y.; Doi-Katayama, Y.; Tang, L.; Park, C.; Moore, B. S.; Richard Hutchinson, C.; Floss, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9051.
(105) Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A. *J. Biol. Chem.* **2002**, *277*, 48028.
(106) Yeh, E.; Kohli, R. M.; Bruner, S. D.; Walsh, C. T. *Chembiochem* **2004**, *5*, 1290.
(107) Carreras, C. W.; Gehring, A. M.; Walsh, C. T.; Khosla, C. *Biochemistry* **1997**, *36*, 11757.
(108) Wenzel, S. C.; Muller, R. *Curr. Opin. Chem. Biol.* **2005**, *9*, 447.
(109) Keating, T. A.; Suo, Z.; Ehmann, D. E.; Walsh, C. T. *Biochemistry* **2000**, *39*, 2297.
(110) Cheng, Y. Q.; Tang, G. L.; Shen, B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3149.
(111) Mochizuki, S.; Hiratsu, K.; Suwa, M.; Ishii, T.; Sugino, F.; Yamada, K.; Kinashi, H. *Mol. Microbiol.* **2003**, *48*, 1501.
(112) Carvalho, R.; Reid, R.; Viswanathan, N.; Gramajo, H.; Julien, B. *Gene* **2005**, *359*, 91.
(113) Piel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14002.
(114) El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M. *Chem. Biol.* **2003**, *10*, 419.
(115) Beck, B. J.; Yoon, Y. J.; Reynolds, K. A.; Sherman, D. H. *Chem. Biol.* **2002**, *9*, 575.
(116) Wenzel, S. C.; Kunze, B.; Hofle, G.; Silakowski, B.; Scharfe, M.; Blocker, H.; Muller, R. *Chembiochem* **2005**, *6*, 375.
(117) He, J.; Hertweck, C. *Chem. Biol.* **2003**, *10*, 1225.
(118) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blocker, H.; Hofle, G.; Muller, R. *J. Biol. Chem.* **2002**, *277*, 13082.
(119) Olano, C.; Wilkinson, B.; Moss, S. J.; Brana, A. F.; Mendez, C.; Leadlay, P. F.; Salas, J. A. *Chem. Commun. (Camb.)* **2003**, 2780.
(120) Shen, B. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.
(121) Gaisser, S.; Trefzer, A.; Stockert, S.; Kirschning, A.; Bechthold, A. *J. Bacteriol.* **1997**, *179*, 6271.
(122) Sthapit, B.; Oh, T. J.; Lamichhane, R.; Liou, K.; Lee, H. C.; Kim, C. G.; Sohng, J. K. *FEBS Lett.* **2004**, *566*, 201.
(123) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S. *Science* **2002**, *297*, 1173.
(124) Liu, W.; Christenson, S. D.; Standage, S.; Shen, B. *Science* **2002**, *297*, 1170.
(125) Kwon, H. J.; Smith, W. C.; Scharon, A. J.; Hwang, S. H.; Kurth, M. J.; Shen, B. *Science* **2002**, *297*, 1327.
(126) Kwon, H. J.; Smith, W. C.; Xiang, L.; Shen, B. *J. Am. Chem. Soc.* **2001**, *123*, 3385.
(127) Bhatt, A.; Stark, C. B.; Harvey, B. M.; Gallimore, A. R.; Demydchuk, Y. A.; Spencer, J. B.; Staunton, J.; Leadlay, P. F. *Angew. Chem., Int. Ed. Engl.* **2005**.
(128) Cane, D. E.; Liang, T.; Hasler, H. *J. Am. Chem. Soc.* **1981**, *103*, 5962.
(129) Ullrich, M.; Bender, C. L. *J. Bacteriol.* **1994**, *176*, 7574.

(130) Chen, H.; Thomas, M. G.; O'Connor, S. E.; Hubbard, B. K.; Burkart, M. D.; Walsh, C. T. *Biochemistry* **2001**, *40*, 11651.
(131) Chen, H.; Walsh, C. T. *Chem. Biol.* **2001**, *8*, 301.
(132) Hubbard, B. K.; Walsh, C. T. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 730.
(133) Pacholec, M.; Hillson, N. J.; Walsh, C. T. *Biochemistry* **2005**, *44*, 12819.
(134) Chen, H.; Hubbard, B. K.; O'Connor, S. E.; Walsh, C. T. *Chem. Biol.* **2002**, *9*, 103.
(135) Garneau, S.; Dorrestein, P. C.; Kelleher, N. L.; Walsh, C. T. *Biochemistry* **2005**, *44*, 2770.
(136) Thomas, M. G.; Burkart, M. D.; Walsh, C. T. *Chem. Biol.* **2002**, *9*, 171.
(137) Pojer, F.; Li, S. M.; Heide, L. *Microbiology* **2002**, *148*, 3901.
(138) Nowak-Thompson, B.; Chaney, N.; Wing, J. S.; Gould, S. J.; Loper, J. E. *J. Bacteriol.* **1999**, *181*, 2166.
(139) Cerdeno, A. M.; Bibb, M. J.; Challis, G. L. *Chem. Biol.* **2001**, *8*, 817.
(140) Couch, R.; O'Connor, S. E.; Seidle, H.; Walsh, C. T.; Parry, R. *J. Bacteriol.* **2004**, *186*, 35.
(141) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. *Nature* **2005**, *436*, 1191.
(142) Kelleher, N. L.; Hicks, L. M. *Curr. Opin. Chem. Biol.* **2005**, *9*, 424.
(143) McLoughlin, S. M.; Kelleher, N. L. *J. Am. Chem. Soc.* **2004**, *126*, 13265.
(144) Hicks, L. M.; O'Connor, S. E.; Mazur, M. T.; Walsh, C. T.; Kelleher, N. L. *Chem. Biol.* **2004**, *11*, 327.
(145) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. *Biochemistry* **2003**, *42*, 4648.
(146) Pan, H.; Tsai, S.; Meadows, E. S.; Miercke, L. J.; Keatinge-Clay, A. T.; O'Connell, J.; Khosla, C.; Stroud, R. M. *Structure (Camb.)* **2002**, *10*, 1559.
(147) Keatinge-Clay, A. T.; Shelat, A. A.; Savage, D. F.; Tsai, S. C.; Miercke, L. J.; O'Connell, J. D., III; Khosla, C.; Stroud, R. M. *Structure (Camb.)* **2003**, *11*, 147.
(148) Reuter, K.; Mofid, M. R.; Marahiel, M. A.; Ficner, R. *EMBO J.* **1999**, *18*, 6823.
(149) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. *Nat. Struct. Biol.* **2002**, *9*, 522.
(150) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997**, *16*, 4174.
(151) Weber, T.; Baumgartner, R.; Renner, C.; Marahiel, M. A.; Holak, T. A. *Struct. Fold Des.* **2000**, *8*, 407.
(152) Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T. *Structure (Camb.)* **2002**, *10*, 301.
(153) Mofid, M. R.; Finking, R.; Essen, L. O.; Marahiel, M. A. *Biochemistry* **2004**, *43*, 4128.
(154) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. *Biochemistry* **2003**, *42*, 6588.
(155) Perham, R. N. *Annu. Rev. Biochem.* **2000**, *69*, 961.
(156) Belshaw, P. J.; Walsh, C. T.; Stachelhaus, T. *Science* **1999**, *284*, 486.
(157) Martinis, S. A.; Plateau, P.; Cavarelli, J.; Florentz, C. *Biochimie* **1999**, *81*, 683.
(158) Doran, J. D.; Tonge, P. J.; Mort, J. S.; Carey, P. R. *Biochemistry* **1996**, *35*, 12487.
(159) Perler, F. B.; Xu, M. Q.; Paulus, H. *Curr. Opin. Chem. Biol.* **1997**, *1*, 292.
(160) Matte, A.; Tari, L. W.; Delbaere, L. T. *Structure* **1998**, *6*, 413.
(161) Heikinheimo, P.; Goldman, A.; Jeffries, C.; Ollis, D. L. *Struct. Fold Des.* **1999**, *7*, R141.
(162) Puente, X. S.; Lopez-Otin, C. *Genome Res.* **2004**, *14*, 609.
(163) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. *Nat. Struct. Mol. Biol.* **2004**, *11*, 888.
(164) McGuire, K. A.; Siggaard-Andersen, M.; Bangera, M. G.; Olsen, J. G.; von Wettstein-Knowles, P. *Biochemistry* **2001**, *40*, 9836.
(165) Huang, W.; Jia, J.; Edwards, P.; Dehesh, K.; Schneider, G.; Lindqvist, Y. *EMBO J.* **1998**, *17*, 1183.
(166) Dreier, J.; Khosla, C. *Biochemistry* **2000**, *39*, 2088.
(167) Watanabe, K.; Wang, C. C.; Boddy, C. N.; Cane, D. E.; Khosla, C. *J. Biol. Chem.* **2003**, *278*, 42020.
(168) Chuck, J. A.; McPherson, M.; Huang, H.; Jacobsen, J. R.; Khosla, C.; Cane, D. E. *Chem. Biol.* **1997**, *4*, 757.
(169) Tang, Y.; Tsai, S. C.; Khosla, C. *J. Am. Chem. Soc.* **2003**, *125*, 12708.
(170) Leslie, A. G. *J. Mol. Biol.* **1990**, *213*, 167.
(171) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. *J. Biol. Chem.* **1998**, *273*, 22773.
(172) Ehmann, D. E.; Trauger, J. W.; Stachelhaus, T.; Walsh, C. T. *Chem. Biol.* **2000**, *7*, 765.
(173) Doekel, S.; Marahiel, M. A. *Chem. Biol.* **2000**, *7*, 373.
(174) Weyand, M.; Schlichting, I. *J. Biol. Chem.* **2000**, *275*, 41058.
(175) Kim, J.; Raushel, F. M. *Biochemistry* **2004**, *43*, 5334.
(176) Gehring, A. M.; Mori, I.; Walsh, C. T. *Biochemistry* **1998**, *37*, 2648.
(177) Mendez, C.; Salas, J. A. *Trends Biotechnol.* **2001**, *19*, 449.
(178) Fjaervik, E.; Zotchev, S. B. *Appl. Microbiol. Biotechnol.* **2005**, *67*, 436.
(179) Konz, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M. A. *Chem. Biol.* **1997**, *4*, 927.
(180) O'Connor, S. E.; Chen, H.; Walsh, C. T. *Biochemistry* **2002**, *41*, 5685.
(181) Chen, H.; Tseng, C. C.; Hubbard, B. K.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14901.
(182) Hubbard, B. K.; Thomas, M. G.; Walsh, C. T. *Chem. Biol.* **2000**, *7*, 931.
(183) Tsai, S. C.; Miercke, L. J.; Krucinski, J.; Gokhale, R.; Chen, J. C.; Foster, P. G.; Cane, D. E.; Khosla, C.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 14808.
(184) Calderone, C. T.; Kowtoniuk, W. E.; Kelleher, N. L.; Walsh, C. T.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, 10.1073/pnas.0603148103.
(185) Jenni, S.; Leibundgut, M.; Maier, T.; Ban, N. *Science* **2006**, *311*, 1258.
(186) Maier, T.; Jenni, S.; Ban, N. *Science* **2006**, *311*, 1263.
(187) Koglin, A.; Mofid, M. R.; Lohr, F.; Schafer, B.; Rogov, V. V.; Blum, M. M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dotsch, V. *Science* **2006**, *312*, 273.

CR0503097