1-1-2005

Nature's Assembly Line Logic for Natural Products

Christopher T. Walsh
Harvard University

Ryan G. Kruger
Harvard University

David A. Vosburg
Harvey Mudd College

FEATURE ARTICLE

Nature's Assembly Line Logic for Natural Products

Christopher T. Walsh, Ryan G. Kruger, *and* ***David A. Vosburg,*** *Harvard Medical School*

Nature fashions a very large number of diverse products (molecular weights ranging from ca. 200 to 2000 daltons) from simple monomeric metabolites used in primary metabolism. These include polyketide scaffolds generated from the simple C_3 malonyl CoA and C_4 methylmalonyl CoA monomers and nonribosomal peptides made from both the 20 proteinogenic amino acids and dozens of nonpoteinogenic amino acids.[1] A large family of terpenoid skeletons are built from the Δ^2- and Δ^3-alkene isomers of the biological isoprene monomer, isoprenyl-pyrophosphate.[2]

Iterative Condensation of Monomeric Building Blocks by Enzymes That Assemble Isoprenoid, Polyketide, and Nonribosomal Peptide Natural Products

A common underlying strategy in all three of these natural product classes is the *iterative condensation* of monomer units onto growing chains under the aegis of enzymes acting as regiospecific and stereospecific elongation catalysts. The iterative aspects of the monomer oligomerizations comprise a simple blueprint for building the characteristic scaffolds of the three natural product classes. However, the assembly logic can be quite distinct.

Two of the three enzyme-mediated iterative elongations produce C–C bonds as the links connecting monomers, while the third generates amide (peptide) linkages as the backbone connectivity in the growing chain. Polyketides arise by iterative Claisen-type thioester condensations from malonyl and

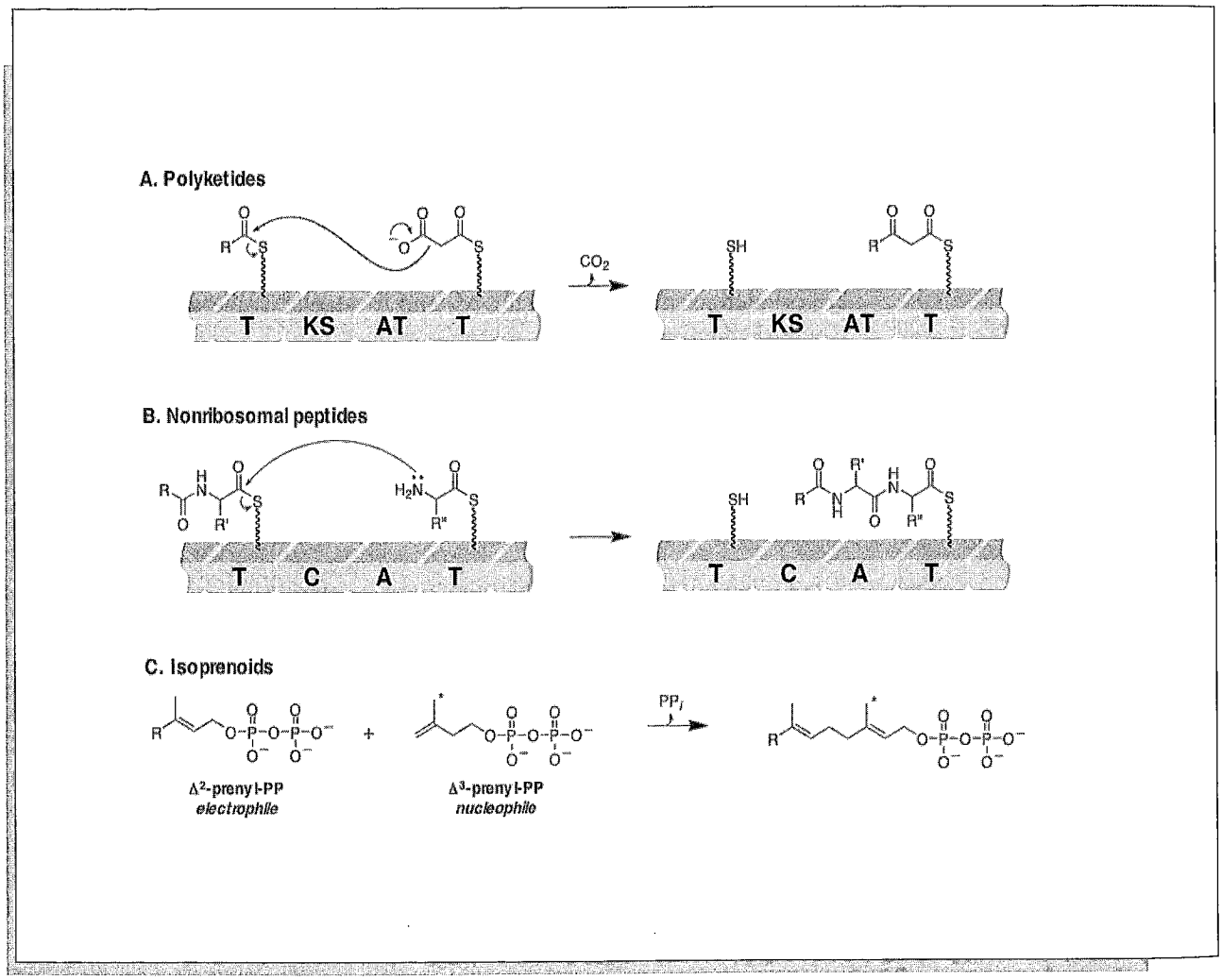
Address correspondence and reprint requests to: Christopher T. Walsh, Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; e-mail: christopher_walsh@hms.harvard.edu.

CHEMTRACTS—ORGANIC CHEMISTRY **18:** 307–320 (2005)

CCC 1431-9268

methylmalonyl thioesters, either as diffusible coenzyme A (CoA) thioesters or as protein-bound thioesters. The CoA thioesters are kinetically stable, but thermodynamically activated, acylation monomers used in every organism for fatty acid biosynthesis and also serve for elaboration of polyketides. Thiol interchange reactions transfer the malonyl and methylmalonyl moieties to thiols tethered on the ends of pantetheinyl arms in turn covalently attached to carrier protein domains in the enzymes. Controlled decarboxylation of malonyl or methylmalonyl-*S*-pantetheinyl species yields kinetically accessible C_2 and C_3 carbanion equivalents for head-to-tail chain growth as shown in Scheme 1.

Terpenoid natural products arise instead not by carbanion-dependent reaction of acyl thioesters but rather by allylic carbocation capture from S_N1-type ionization of Δ^2-isopentenyl-PP as the elongating partner. Capture of



Scheme 1. Chain elongation on assembly lines. **A**, Polyketide Claisen-type C–C bond formation via malonyl thioester decarboxylation. **B**, Peptide bond formation through acyl chain transfer from sulfur to nitrogen. **C**, Isoprenoid Δ^2-allyl carbocation capture by the π e^- of the Δ^3-partner (PP = pyrophosphate).

the incipient allyl carbocation at C_1 of the allyl system in an enzyme active site can occur with the π electrons of a Δ^3-isopentenyl-PP monomer. In the first instance the product is a C_{10}-compound, geranyl-PP. Further iterations of C_5-unit addition build C_{10}, C_{15}, and C_{20} chains most commonly, but can be extended all the way up to C_{100} and longer polyisoprenoids. In addition to elaboration of linear isoprenoid chains, enzyme-directed intramolecular capture and subsequent cationic rearrangements send the scaffolds down many cyclization pathways (e.g., to steroid or taxane scaffolds).

In enzyme-mediated alkylative elongation of the polyisoprenoid chains by five carbons at a time the growing chain is typically the electrophilic partner and the Δ^3-isoprenyl-PP monomer is the nucleophile. This is the same polarity as the iterative condensations in polyketide chain growth where the elongating ketide chain is the electrophile and the C_2 or C_3 monomer is the nucleophile.

The fundamental chain elongation step in peptide natural products is the amide bond-forming condensation. As in the polyketide and polyisoprenoid chain elongation steps, the growing peptide chain acts as the electrophilic partner in each condensation with the deprotonated amino group of the amino acid monomer as nucleophile (Scheme 1). Thus, the polarity of all three of these chain elongations is equivalent. However, the isoprenoid elongation steps occur on soluble substrates while the polyketide and nonribosomal peptide chains are elongated while covalently tethered to multimodular protein way stations.

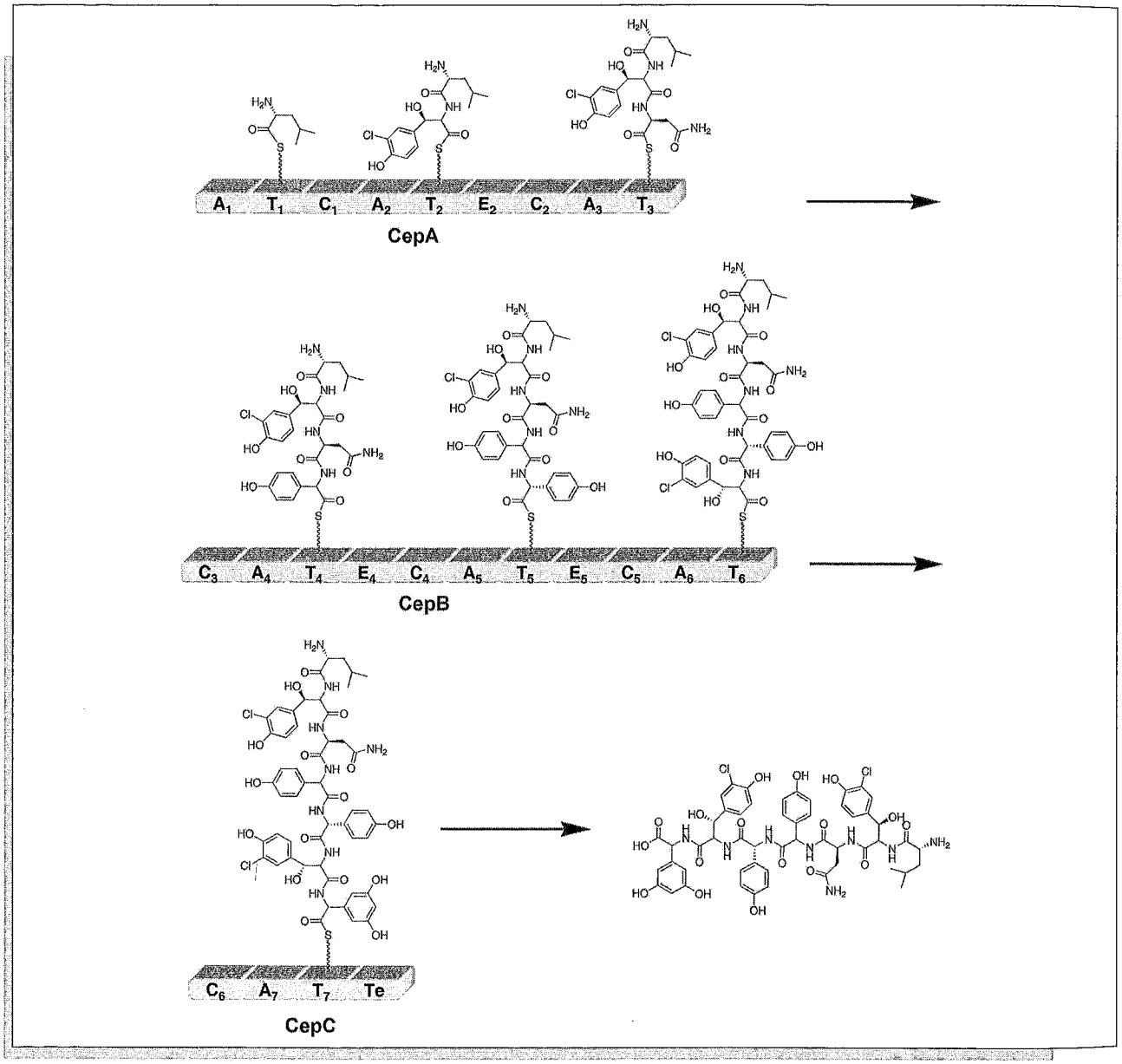
Protein Templating to Control the Identity of Polyketide and Nonribosomal Peptide Natural Products

These two routes of building up carbon skeletons in natural products (i.e., C_5 alkylations vs. C_2 or C_3 Claisen thioester condensations) differ fundamentally in terms of templating. The growth of prenyl and terpenyl chains is controlled by the active sites of the elongation enzymes but occur without specific templating. The same Δ^3-isopentenyl-PP monomers are used in every cycle. In contrast, the Claisen-type chain growth of polyketides is generally templated by specific protein modules (for type I and II polyketides, but not type III),[3] in common with chain growth of nonribosomal peptides. Templating requires distinct enzymatic logic and machinery from selection of two substrates from solution. In principle, templating also offers coding and information storage possibilities. The informational content of biologically derived oligomers and polymers made by templating is most apparent in DNA and RNA molecules, where the parent strands are copied by DNA or RNA polymerases. The *translation* of messenger RNA triplet codons into proteins is the classic RNA templating of protein biosynthesis, converting oligonucleotide language to oligopeptide code.

Nonribosomal peptides and polyketides also are built in specific sequences, again driven by templating. In these cases, the templates are not RNA triplet codons but rather the specific order and identity of protein modules. In a typical polyketide synthase (PKS) or nonribosomal peptide synthetase (NRPS) assembly line, one monomer is activated and incorporated by one module.

Thus, erythromycin with seven two-carbon units, derived from malonyl CoAs, is built on an assembly line of seven modules.[4] Analogously the heptapeptide scaffold of vancomycin is constructed on a seven-module NRPS assembly line (Scheme 2). In both cases the modules are spread over three protein subunits, in a 2-2-2 distribution for 6-deoxyerythronolide B synthase and a 3-3-1 distribution for the heptapeptide synthetase.[4,5]

Each module of a PKS or NRPS assembly line is composed of one thiolation (T) domain and up to several catalytic domains. The T domains, also



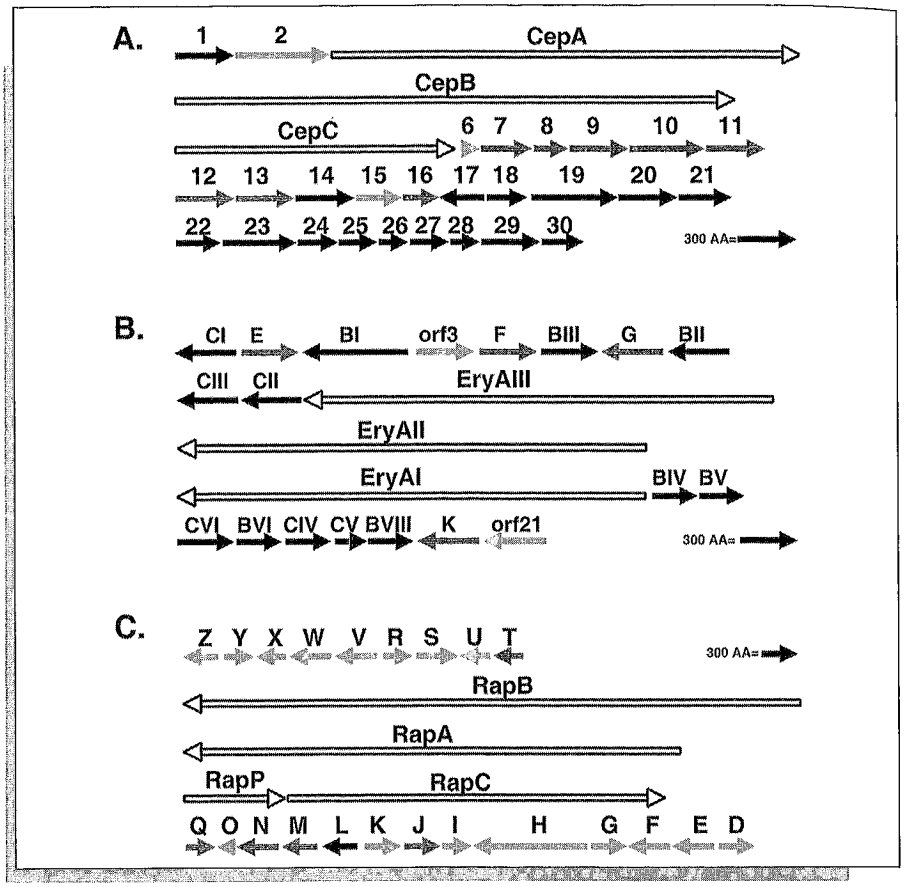
Scheme 2. NRPS assembly line for the production of the vancomycin heptapeptide. Three NRPS proteins catalyze the formation of the vancomycin heptapeptide. Elongation is terminated by the thioesterase (Te) domain, which hydrolyzes the thioester bond holding the completed peptide to domain T_7.

called acyl carrier proteins (ACP) in PKS assembly lines and peptidyl carrier proteins (PCP) in NRPS assembly lines, are 80- to 100-residue autonomously folding units that must be posttranslationally phosphopantetheinylated on a conserved serine side chain to install the pantetheinyl arm with its terminal thiol that will be the docking site for the growing acyl chain in each module. In each module there are core catalytic domains and there may be optional ones. Core catalytic domains include those that select the monomers, malonyl and methylmalonyl CoAs for PKS, and amino acids for NRPS, and catalyze the chain elongation/condensation step. In a prototypic PKS module these are acyltransferase (AT) and ketosynthase (KS) domains so a minimal module has the three domains KS-AT-ACP, at a molecular weight of 100–120 kDa. For a minimal NRPS module, the amino acid adenylation (A) and condensation (C) domains are paired with a PCP domain to yield a C-A-PCP tridomain protein, also in the 100–120 kDa size range (Scheme 2).

The assembly lines operate in a linear fashion, with the sequence of domains within modules and modules within the assembly line acting as the protein-based templates to specify the identity and the number of monomers incorporated into the growing chain, as well as the connectivity. There is an a *initiation* module, followed by a series of *elongation* modules acting in tandem. The most C-terminal module acts as *termination* module. Often, the most C-terminal domain in the C-terminal module has thioesterase activity, releasing the full-length acyl/peptidyl chain as the free carboxylic acid. In many other assembly lines, a nucleophile internal to the full-length acyl/peptidyl chain serves as the kinetically competent nucleophile to create macrocyclic lactones or lactams, with the exquisite regioselectivity anticipated for enzymatic catalysis.

Biosynthetic Gene Clusters

Another distinction between terpenoid/isoprenoid biosyntheses and the polyketide and nonribosomal peptide biosynthetic pathways is the clustering of biosynthetic genes (Scheme 3). The latter two classes of natural products are the paradigms for collocation of large numbers of genes required to convert central metabolites of primary metabolism into complex natural products. The genes are thought to be clustered to provide coordinated regulation to turn on the genes and translate the required enzymes to carry out all the operations in an efficient time frame. We note that biological construction of both polyketides and nonribosomal peptides can be divided into three phases: the production of any unusual monomers, the utilization of the monomers in chain elongation steps by the protein assembly lines, and post-assembly line tailoring steps (Table 1). All three types of genes are located together for coordinated regulation. While PKS assembly lines do not usually employ exotic malonyl CoA derivatives as starting monomers, when they do (e.g., methoxymalonyl CoA), the genes required to generate the building blocks are collocated with genes encoding the assembly line modules and any tailoring enzymes.



Scheme 3. Biosynthetic gene clusters from representative NRPS and PKS assembly lines. Gene function is identified as follows: white arrows depict NRPS/PKS enzymes; black arrows indicate enzymes required for the biosynthesis of various monomer units; dark gray arrows show enzymes responsible for scaffold tailoring; and light arrows indicate proteins of other or unknown function. Arrow length corresponds to protein size as indicated. **A**, Chloroeremomycin; **B**, erythromycin; **C**, rapamycin.

The genes required for actinomycetes to make the vancomycin family antibiotic chloroeremomycin illustrate Nature's logic explicitly.[5] Some 30 genes are clustered. One subset encodes 10 enzymes required to make 4-OH-phenylglycine, 3,5-$(OH)_2$-phenylglycine, and 3-OH-tyrosine. These nonproteinogenic amino acids make up five of the seven residues in the heptapeptide scaffold of the antibiotic. These monomers are made by the

Table 1. Three kinds of functions encoded in biosynthetic gene clusters

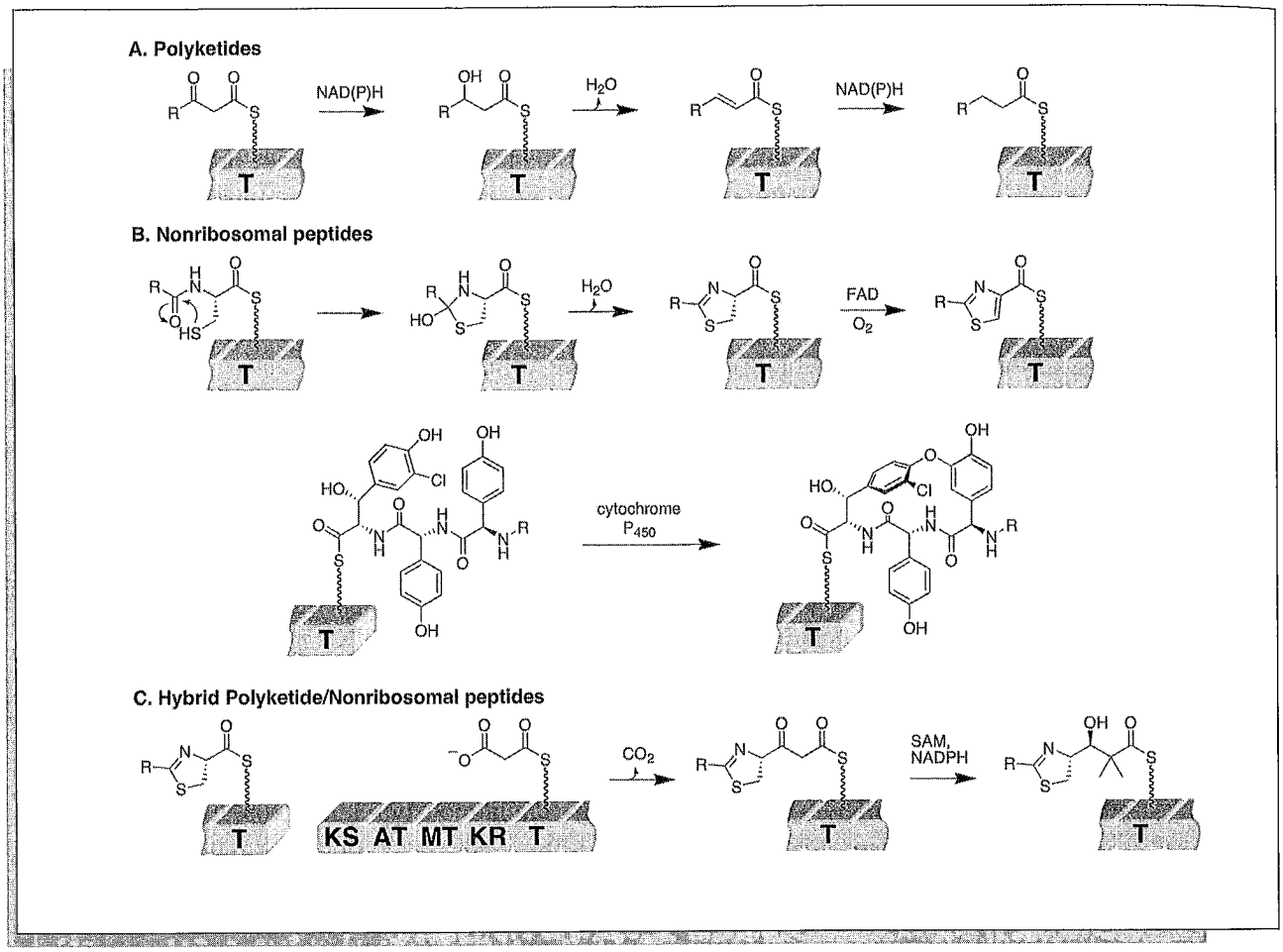
1.	Enzymes to produce dedicated monomers as building blocks (nonproteinogenic amino acids, dTDP-deoxyhexoses, methoxymalonyl CoA)
2.	Assembly line modules: one module per monomer incorporated
3.	Post-assembly line tailoring enzymes for natural product maturation

antibiotic-producing cell with "just in time" inventory control. A second subset of genes encodes the seven modules of the NRPS assembly line that releases the completed heptapeptide after a series of 24 tethered intermediates on the seven PCP docking sites of the assembly line. A third subset of genes encodes the post-assembly line tailoring enzymes, detailed in a subsequent section, required for maturation of the nascent product of assembly line action. These tailoring enzymes are likewise made only when the assembly line is running. Indeed, regulation of assembly line production and natural product resistance are additional functions often encoded in biosynthetic gene clusters. The genetic economy of production of polyketide and nonribosomal peptides, and hybrids thereof, are high examples of the evolution of biological engineering principles in Nature.

Tailoring Reactions During Assembly Line Operations: *in cis* and *in trans*

In addition to chain elongation steps, the protein modules of PKS and NRPS assembly lines can also carry out tailoring reactions on the growing chain (Scheme 4). The operation of two minimal PKS modules in tandem would generate a 3-keto-acyl chain tethered to the pantetheinyl arm of the second module. To convert the 3-keto group to a 3-CH_2 moiety—as found in fatty acids—requires three additional catalytic domains: a ketoreductase (KR), a dehydratase (DH), and an enoylreductase (ER) domain. Thus, a full-service PKS module would have six autonomously folding protein domains, KS-AT-KR-DH-ER-ACP, and perform the three additional operations of ketoreduction, dehydration, and enoyl reduction on the tethered acyl chain before allowing its transfer to the next downstream module of the assembly line. The absence of an ER domain or a defect in its catalytic activity would leave the 2,3-enoyl chain, while a defect or absence of the DH domain would allow the 3-OH-acyl group to persist. It is clear that four different functionalizations of the acyl chain can exist for transfer to the next downstream module. The incomplete processing of the initial 3-ketoacyl-*S*-ACP chains at each way station of PKS assembly lines with different functional domains in each module allows dramatic diversity generation. Nature creates remarkable variation in polyketide scaffolds from combinatorial use of domains and modules in such assembly lines, as evidenced in such therapeutically useful molecules as FK506 and rapamycin.

For NRPS chain growth, the most common optional domains to be found in modules are *N*-methyltransferase (MT) and epimerase (E) domains. The MT domains produce *N*-methylated peptides that lower energy barriers for distortion of the peptide bonds out of planarity (e.g., 7 of 11 residues in cyclosporin), while E domains allow interconversion of L-aminoacyl moieties to D-aminoacyl moieties, before or after a condensation step during NRPS assembly line action.[6] Most remarkable of the domains in NRPS assembly lines is the occasional replacement of the C domains, catalyzing peptide bond condensation as the fundamental step in chain elongation, with a variant Cy domain that catalyzes both condensation, cyclization, and dehydration of Cys, Ser, or Thr residues to generate thiazoline or oxazoline rings, respectively. Cy domain action converts a transiently formed peptide bond



Scheme 4. Tailoring reactions that occur during chain elongation on assembly lines. **A**, Reduction of the 3-keto group showing different intermediates that are used as diversity elements in natural products. **B**, Cysteines can be converted to thiazoles through thiazoline intermediates by the action of *in cis* tailoring domains. *In trans* tailoring is also seen as in the case of oxidative cross-linking of vancomycin family heptapeptides. **C**, In hybrid systems, features of both NRPS and PKS systems are combined to allow the elongation of amino acid chains by acyl groups and vice versa.

into the five-membered heterocycles, dramatically altering backbone connectivity of the peptide scaffold.[7] These are found in several iron-chelating peptides, siderophores, secreted by bacteria to chelate Fe^{3+} for reuptake into the bacterial cell.

The tailoring steps of the initial 3-ketoacyl-*S*-ACP chains or the peptidyl-*S*-PCP chains are effected by tailoring domains embedded *in cis*, as integral parts of the multimodular protein assembly lines. There are dedicated tailoring enzymes that act *in trans*, as separate proteins, operating on a particular chain docked on a particular carrier protein in the assembly line. This strategy is exemplified with thiazoline reductases that convert thiazolinyl-*S*-PCP to thiazolidinyl-*S*-PCP intermediates. The tetrahydroaromatic rings may then undergo *N*-methylation or *C*-methylation, as in the siderophores pyo-

chelin and yersiniabactin from the bacterial pathogens *Pseudomonas aeruginosa* and *Yersinia pestis*.[8,9] On the other hand, catalytic oxidase domains can be embedded *in cis* in modules containing Cy domains to create thiazole or oxazole rings.[10] The tandem action of two such modules can generate the bithiazole moiety that is a DNA intercalator in the antitumor drug bleomycin. Most recently, it appears that the oxidative cross-linking of the electron-rich aromatic side chains at five of the seven residues in the heptapeptide scaffold of vancomycin family members may reflect *in trans* tailoring by heme protein oxidases.[11]

It is likely that evolutionary experiments are continually at play in the microbial world, moving domains in and out of assembly lines, to function *in cis* or *in trans*, as part of the creation of new assembly lines with altered product capabilities. Any selective advantage conferred to such evolved assembly lines from the biological action of the products will propagate the microbes with those biosynthetic capacities.

Hybrid Assembly Lines

Several of the molecules noted in passing above (including rapamycin, FK506, bleomycin, and yersiniabactin) are hybrid PK/NRP natural products, fashioned from both malonyl CoA/methylmalonyl CoA and amino acid monomers. Not surprisingly, the responsible protein assembly lines are hybrid, with both PKS and NRPS modules. The linear assembly of PKS and NRPS-derived products, with chain growth from the amino terminus to the carboxy terminus of the protein assembly lines, is templated by the number, order, and identity of the modules in any assembly line. Nature has clearly learned how to mix and match NRPS and PKS modules such that protein–protein recognition between domains and modules of a hybrid assembly line are maintained. Also the recognition of the growing acyl chain, whether polyketidyl or peptidyl, for condensation with an immediately downstream malonyl-*S*-ACP or an aminoacyl-*S*-PCP must be maintained for hybrid natural product elongations.

Inspection of the identity and order of domains/modules in an assembly line allows prediction of how Nature assembles complex products. For example, the rapamycin and FK506 assembly lines are mostly PKS modules with only a single participating NRPS module. The NRPS module, activating and inserting a pipecolyl residue, acts at the end of the line, effectively inserting the pipecolyl moiety, essential for therapeutic activity, into the polyketide macrolactone scaffold. The bleomycin assembly line is biased in the other direction: it is mostly an NRPS assembly line with one PKS module, placed near the middle of the assembly process. The causative agent of plague, *Y. pestis*, fashions yersiniabactin via one NRPS module, then a PKS module, then an NRPS module, with optional catalytic domains in each of the three modules. Understanding the recognition features required for NRPS/PKS and PKS/NRPS junctions to function efficiently is a prerequisite to high-throughput engineering of novel hybrid assembly lines for combinatorial biosynthesis.

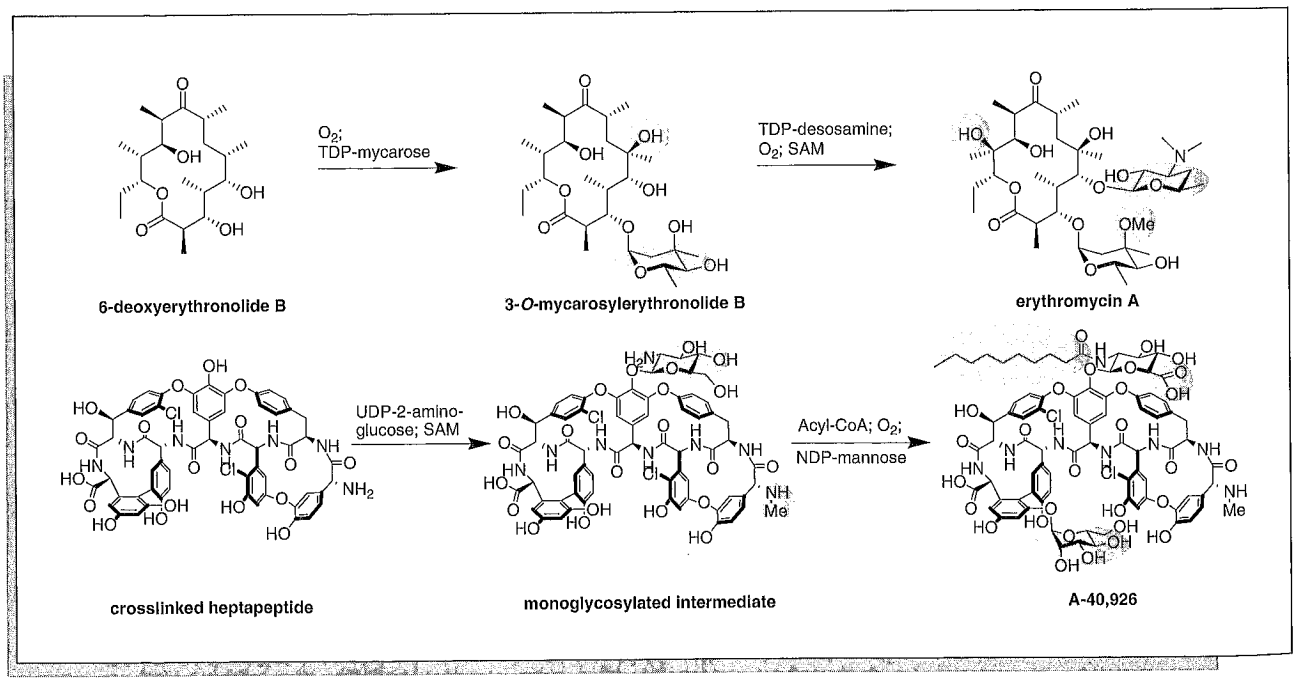
There is reason to believe one could import other types of catalytic domains into natural product assembly lines to increase the kinds of chemistry prac-

ticed during iterative elongations. For example, pyridoxal-P-dependent catalytic domains are embedded *in cis* in the mycosubtilin synthetase and in the prodiginine assembly lines,[12,13] opening up the full range of PLP-mediated amino acid carbanion enzymology to go along with the redox reactions (dehydrogenase/reductase) dependent on FAD and NADH and the methylations dependent on *S*-adenosylmethionine already known in assembly lines.

Post-Assembly Line Tailoring Steps

In many instances the products released from the most C-terminal thiolation domain of PKS and NRPS assembly lines are not biologically active until further enzymatic processing by a suite of dedicated tailoring enzymes has occurred. As noted above, such tailoring enzymes are encoded by sets of genes embedded in the biosynthetic clusters and coordinately regulated to be available when the nascent products start to roll off the assembly lines. Post-assembly line modifications most often include oxidation and glycosylation enzymatic action but acylations and methylations can occur as well (Scheme 5).

Typically the redox enzymes are hemeproteins, catalyzing oxidative modifications at carbon centers to generate alcohol functionalities on the natural product scaffolds, such as the hydroxylations at C_6 and C_{12} of the deoxyerythronolide scaffold on the path to erythromycin. Late stage enzymatic hydroxylations at the periphery of a natural product scaffold also occur in terpenoid natural products, with eight hydroxylations catalyzed by a suite of hemeprotein hydroxylases to convert the taxane skeleton into the octa-oxygenated taxol scaffold.[14]



Scheme 5. Post-assembly line tailoring such as methylation, hydroxylation, glycosylation, and acylation occurs in both PKS and NRPS systems.

O-Glycosylations occur most often on hydroxyl groups on the mature scaffolds as late-stage tailoring steps,[15] but *N*-glycosylations are known (e.g., in the indolocarbazole examples of rebeccamycin and staurosporines[16,17]) and *C*-glycosylations are found in anguicyclines among other polyketides.[18] The glycosyltransferases (Gtfs) use nucleoside diphospho (NDP) sugars as glycosyl donors to attach a broad range of D- and L-hexoses regio- and stereospecifically to nucleophilic sites on the aglycones. The NDP-hexoses found in many natural products are not those used in primary metabolism (TDP-D-glucose, GDP-D-mannose) but rather a series of deoxy and aminodeoxy sugars. Desosamine attached to C_5 of the erythromycin scaffold (Scheme 5) is such an example as is vancosamine in vancomycin and daunosamine in the antitumor polyketide daunomycin. These modified sugars are often important pharmacophores for interaction with biological targets and strike a balance of hydrophilic and hydrophobic surfaces for optimized interaction as well as electrostatic interaction with the amino groups. In novobiocin, the L-noviosyl sugar is further decorated enzymatically by *O*-methylation at the 4'-position and then *O*-carbamoylation at the 3' locus. It is this final *O*-carbamoylation that provides a 200-fold increase in antibiotic potency by specific interaction with the B subunit of bacterial DNA gyrase.[19] Consistent with the logic of collocation of all biosynthetic genes necessary for production of a biologically active natural product, the genes that encode enzymes that convert dTDP-D-glucose to dTDP-D-desosamine, or to dTDP-L-vancosamine, are embedded in the biosynthetic gene clusters for erythromycin and chloroeremomycin so that all enzymes, including the enzymes needed to make the dedicated NDP-hexose monomers, are produced at the same time when the gene clusters are activated for transcription.

Multiple glycosylations can occur, typically with the logic of one sugar per glycosyltransferase. Thus, three glycosyltransferases are encoded in the chloroeremomycin cluster.[5,20] The first transfers D-glucose to the phenolic-OH of 4-OH-PheGly$_4$ in the cross-linked aglycone. The other two use dTDP-L-epivancosamine as the hexosyl donor. One catalyst transfers this L-trideoxyhexose to the 2-OH of the newly introduced glucose to create the D-glucosyl-2,1-L-epivancosamine disaccharide moiety. The third glycosyltransferase from this cluster transfers epivancosamine to the benzylic OH of the β-OH-Tyr$_6$ residue in the cross-linked heptapeptide scaffold. The tailoring Gtfs have a common two-domain architecture, making them good candidates for chimerization by molecular biology-mediated domain swaps to vary either NDP-hexose donor or aglycone acceptor specificity for combinatorial biosynthesis purposes.[21]

Acylations of natural products are also purposeful modifications, enacted by acyltransferases similarly located in the biosynthetic gene clusters. *O*-acylation of the terminal sugar of the macrolide carbomycin and *N*-acylation of the glucosamine moiety of teicoplanin are important maturation steps for gain of biological activity and constitute significant pharmacophore adjustment.[22,23]

The post-assembly line tailoring steps offer promise for combinatorial biosynthetic manipulations both enzymatically and by a mixture of chemoenzymatic manipulations. They occur after the scaffolds have been fashioned by the assembly lines and may occur after the hydroxylation steps as well.

Regiospecific acylation and alkylation of the deoxysugar substituents also permits modulation of activities by late-stage decorations on the periphery of the natural product scaffold.

Concluding Remarks

This article has not addressed the rationale for why microbes make such a rich array of natural products but focuses only on how they take simple molecules of primary metabolism and elaborate them into a rich diversity of scaffolds and functional groups to create the architectures that enable high-affinity recognition by particular biological targets, including proteins, DNA, peptidoglycans, lipid phosphates, and doubtless many other cellular components. Such ligand–receptor interactions (with selectivity, affinity that allows occupancy under physiologic conditions, and useful on/off rates) are prerequisites for the natural product ligands to function as signaling molecules of many types and in many contexts.

We have focused on polyketides, nonribosomal peptides, and hybrids thereof to illustrate two broad principles used by producer organisms: templating without RNA and the consequent assembly line logic of many protein domains in a "beads on a string" array. The products of newly discovered assembly lines can be predicted just from the number, identity, and placement of protein domains in such assembly lines, including crossover points between PKS and NRPS logic.

The strategy of tethering the two workhorse malonyl and methylmalonyl acyl groups to make polyketides and more than 100 amino acids to make nonribosomal peptides on carrier protein way stations, each containing an HS-pantetheinyl arm, has several useful consequences. First the acyl/peptidyl thioester chemistry is likely to be ancient and perhaps even prebiotic and provides the thermodynamic activation to drive each chain elongation step and to stabilize the carbanion nucleophiles from controlled decarboxylation of the malonyl thioesters. Second, each protein module is a miniature chemical factory, containing the thiolation domain and two to six catalytic domains. There are several examples of single modules excised from assembly line contexts that act on their own, to activate proteinogenic amino acids, carry out chemical modifications on the aminoacyl-*S*-carrier proteins, and then release the now nonproteinogenic amino acid monomer. Third, assembly lines could evolve from fusion of modules into megaproteins. The shuffling of domains in and out of modules and modules in and out of assembly lines is a constant flux in the microbial world that is Nature's practice of combinatorial biosynthesis and surely has lessons for new, man-made combinatorial efforts.[24]

The operation of natural product multimodular protein assembly lines is indeed a marvel of microbial chemical engineering. There is an integrated strategy for inventory control of monomer availability, and chain initiation starts only at the N-terminus of the multimodular, often multiprotein, assembly lines. Chain elongation involves completion of all the chemical steps in each module before the temporally controlled passage of the growing chain to the next module, loaded with its nucleophilic monomer loaded on the next

way station. In the assembly line for cyclosporine synthetase, the 11 modules are in a single protein of 14,000 amino acids, a molecular weight of 1.5 megadaltons, with 47 recognizable domains (47 beads on the string).[25] The covalent tethering of the growing chain avoids its loss by diffusion, at any time in the 47-step elongation process, and ensures high yield. Of course the sheer size of such protein assembly lines limits the ultimate size of the number of iterative elongations that can be built by this machinery. The largest size of a nonribosmal peptide is the 21-residue syringopeptin of *Pseudomonas syringae*.[26]

Chain termination occurs when the full-length chain reaches the most downstream pantetheinyl way station in the protein assembly line. Almost always there is one more domain C-terminal to the last T domain that executes the covalent disconnection of the full-length acyl/peptidyl chain from the assembly line. The disconnection can be hydrolytic or reductive, but is most intriguing when catalysis of specific intramolecular macrocyclization is enacted.[27]

We have noted that while generation of the aglycone scaffolds by such assembly lines may be monumental chemical achievements, the released scaffold may be only halfway through its biosynthetic journey. During formation of the antibiotic teicoplanin, for example, there are eight additional tailoring enzymes to methylate, acylate, oxidize, and glycosylate the scaffold to its mature, biologically active form. There are 11 oxidative tailoring steps in the conversion of the nascent dodecaketide to the napthoquinone and spiroketal functionalities found in griseorhodin.[28]

The assembly line strategy is chemically economic for creating complex scaffolds from simple 2- to 5-carbon building blocks, executed through a cascade of elongating acyl thioesters. Diversification of the scaffolds by sets of tailoring enzymes, again using variants of simple building blocks found in or diverted from primary metabolism, is also simple conceptually but well executed. Perhaps the most remarkable economy of microbial producers is in their genomic organization. The collection of two to three dozen genes in adjacency for coordinate regulation allows a turnkey approach to turn on synthesis of all the enzymes needed to generate the substrates, the assembly line, and the catalysts to convert intermediates into finished products. Much of this strategy probably derives from the ancient ability to activate acyl groups as thiol esters. No such catalytic devices for preserving the activity of tethered monomers and simultaneously creating a favorable equilibrium for iterative condensation appears to have been built up for other classes of natural products.

Acknowledgments

This work was supported in part by NIH grants GM20011, GM49338, AI 42738 (C.T.W.), and postdoctoral fellowships from the NIH (R.G.K.) and the Jane Coffin Childs Memorial Fund (D.A.V.).

References

1. Walsh, C.T. *Science* **2004**, *303*, 1805.
2. Cane, D.E. In *Comprehensive Natural Products Chemistry*, Barton, D.H.B., Nakanishi, K., Meth-Cohn, O., eds., Elsevier: New York, 1999, vol. 2, ch. 1.
3. Austin, M.B., Noel, J.P. *Nat. Prod. Rep.* **2003**, *20*, 79.
4. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L. *Science* **1991**, *252*, 675.
5. Van Wageningen, A.M., Kirkpatrick, P.N., Williams, D.H., et al. *Chem. Biol.* **1998**, *5*, 155.
6. Clugston, S.L., Sieber, S.A., Marahiel, M.A., Walsh, C.T. *Biochemistry* **2003**, *42*, 12095.
7. Marshall, C.G., Burkart, M.D., Keating, T.A., Walsh, C.T. *Biochemistry* **2001**, *40*, 10655.
8. Patel, H.M., Walsh, C.T. *Biochemistry* **2001**, *40*, 9023.
9. Miller, D.A., Walsh, C.T., Luo, L. *J. Am. Chem. Soc.* **2001**, *123*, 8434.
10. Schneider, T.L., Shen, B., Walsh, C.T. *Biochemistry* **2003**, *42*, 9722.
11. Bischoff, D., Bister, B., Bertazzo, M., et al. *ChemBioChem* **2005**, *6*, 267.
12. Duitman, E.H., Hamoen, L.W., Rembold, M., et al. J. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13294.
13. Cerdeno, A.M., Bibb, M.J., Challis, G.L. *Chem. Biol.* **2001**, *8*, 817.
14. Chau, M., Croteau, R. *Arch. Biochem. Biophys.* **2004**, *427*, 48.
15. (a) Walsh, C.T., Freel Meyers, C.L., Losey, H.C. *J. Med. Chem.* **2003**, *46*, 3425; (b) Bechthold, A., Weitnauer, G., Luzhetskyy, A., et al. *Ernst Schering Res. Found. Workshop* **2005**, *51*, 147.
16. Sanchez, C., Butovich, I.A., Brana, A.F., Rohr, J., Mendez, C., Salas, J.A. *Chem. Biol.* **2002**, *9*, 519.
17. Onaka, H., Taniguchi, S., Igarashi, Y., Furumai, T. *J. Antibiot.* **2002**, *55*, 1063.
18. Trefzer, A., Hoffmeister, D., Kunzel, E., et al. *Chem. Biol.* **2000**, *7*, 133.
19. Hooper, D.C., Wolfson, J.S., McHugh, G.L., Winters, M.B., Swartz, M.N. *Antimicrob. Agents Chemother.* **1982**, *22*, 662.
20. Lu, W., Oberthur, M., Leimkuhler, C., Tao, J., Kahne, D., Walsh, C.T. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4390.
21. Thorson, J.S., Barton, W.A., Hoffmeister, D., Albermann, C., Nokolov, D.B. *ChemBioChem* **2004**, *5*, 16.
22. Epp, J.K., Huber, M.L., Turner, J.R., Goodson, T., Schoner, B.E. *Gene* **1989**, *85*, 293.
23. (a) Li, T.L., Huang, F., Haydock, S.F., Mironenko, T., Leadlay, P.F., Spencer, J.B. *Chem. Biol.* **2004**, *11*, 107; (b) Kruger, R.G., Lu, W., Oberthur, M., Tao, J., Kahne, D., Walsh, C.T. *Chem. Biol.* **2005**, *12*, 131.
24. Vosburg, D.A., Walsh, C.T. *Ernst Schering Res. Found. Workshop* **2005**, *51*, 261.
25. Weber, G., Schorgendorfer, K., Schneider-Scherzer, E., Leitner, E. *Curr. Genet.* **1994**, *26*, 120.
26. Scholz-Schroeder, B.K., Soule, J.D., Lu, S.E., Grgurina, I., Gross, D.C. *Mol. Plant Microbe Interact.* **2001**, *14*, 1426.
27. Kohli, R.M., Walsh, C.T. *Chem. Commun.* **2003**, 297.
28. Li, A., Piel, J. *Chem. Biol.* **2002**, *9*, 1017.