REVIEW: BIOCHEMISTRY

Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations

David E. Cane, Christopher T. Walsh, Chaitan Khosla*

Polyketides and non-ribosomal peptides are two large families of complex natural products that are built from simple carboxylic acid or amino acid monomers, respectively, and that have important medicinal or agrochemical properties. Despite the substantial differences between these two classes of natural products, each is synthesized biologically under the control of exceptionally large, multifunctional proteins termed polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) that contain repeated, coordinated groups of active sites called modules, in which each module is responsible for catalysis of one complete cycle of polyketide or polypeptide chain elongation and associated functional group modifications. It has recently become possible to use molecular genetic methodology to alter the number, content, and order of such modules and, in so doing, to alter rationally the structure of the resultant products. This review considers the promise and challenges inherent in the combinatorial manipulation of PKS and NRPS structure in order to generate entirely "unnatural" products.

ver the past decade striking advances in microbial genetics have propelled a revolution in our ability to deduce, analyze, and manipulate the biosynthesis of structurally complex and biologically important families of natural products, most notably the classes known as polyketides and as non-ribosomal peptides. An explosion of discoveries and technological innovations is rapidly enhancing our capacity to "mutate" the structure of natural products using heuristics and procedures that are analogous to those routinely used in the generation of structurally altered nucleic acids and proteins. It is tempting to speculate that before too long, it will be possible to modify the structures of natural products such as the immunosuppressants FK506 or cyclosporin with the same degree of precision and convenience with which the primary amino acid sequences of their corresponding binding proteins, FKBP and cyclophilin, can be altered. What is the state-of-art in the manipulation of microbial natural products? Where are the existing hurdles to an expansion of the biologist's synthetic capabilities? How might these barriers be overcome? What would be the biological, biochemical, and medical implications as the

structural manipulation of a target natural product becomes an increasingly straightforward task? This review addresses these questions.

What Are Polyketides and Non-ribosomal Peptides?

Polyketides and non-ribosomal peptides are two large families of natural products, the first built from acyl-coenzyme A (CoA) monomers and the second from amino acids (1). These metabolites include many important pharmaceuticals, veterinary agents, and agrochemicals, a few of which are illustrated in Fig. 1. The enormous structural diversity and complexity of these biomolecules is impressive. Although the actual biological roles of each of these metabolites in the native producing organisms (primarily actinomycetes, bacilli, and filamentous fungi) are unclear, an extraordinary variety of pharmacological properties have been associated with naturally occurring polyketides and non-ribosomal peptides. Other widely used polyketides include rifamycin (antibacterial, inhibitor of bacterial RNA polymerase), FK506 (immunosuppressant, binds FK506 binding protein and calcineurin), and lovastatin (anti-cholesterolemic, hydroxymethylglutaryl-CoA reductase inhibitor), while the antibacterial vancomycin (cell wall biosynthesis inhibitor) is an important non-ribosomal peptide (2).

How Are They Made?

Non-ribosomal peptide synthetases (NRPSs) and a class of polyketide synthases (PKSs) called modular PKSs use a strikingly similar

strategy for assembly of these distinct classes of natural products (1) (Fig. 2). Both groups of metabolites are biosynthesized by exceptionally large, multifunctional proteins that are organized into coordinated groups of active sites termed modules, in which each module is responsible for catalysis of one cycle of polyketide or polypeptide chain elongation and associated functional group modifications (1, 3, 4). Within each module is a carrier protein domain to which the growing polyketide or polypeptide chain is covalently tethered (1, 3). The order of biosynthetic modules from NH2- to COOH-terminus on each PKS and NRPS polypeptide and the number and type of catalytic domains within each determine the order of structural and functional elements in the resulting natural product. The size and complexity of the ultimately formed polyketide or polypeptide are controlled by the number of repeated acyl chain extension steps, which are in turn a function of the number and placement of carrier protein domains in these multimodular enzymes. In analysis of the molecular logic of these megasynthases, four phases can be delineated: priming, chain initiation, chain elongation, and termination.

Posttranslational priming of the apo-synthases. Modular PKS enzymes contain multiple copies of 75- to 90-amino acid acyl carrier protein (ACP) domains, one to each module, whereas the NRPSs have analogous peptidyl carrier protein (PCP) domains, each of which carries a 20 Å-long phosphopantetheinyl group derived from CoA covalently attached to a conserved serine residue. The phosphopantetheinyl prosthetic group serves as a flexible tether terminating in a cysteamine thiol group that becomes the site of covalent attachment for both the monomer units and the growing acyl chain intermediates of PKS and NPRS catalysts. This posttranslational priming of each megasynthase apoACP and apoPCP is mediated by ACPand PCP-specific variants of phosphopantetheinyl transferases (PPTases) (5), ensuring pathway-specific regulation and partner protein specificity. In the assembly line strategy for chain growth by way of covalent acyl-S-ACP/PCP-enzyme intermediates, every ACP or PCP (ACP/PCP) domain must carry its pantetheinyl residue, or chain elongation will be prematurely terminated. For example, 6-deoxyerythronolide B synthase (DEBS),

D. E. Cane, Department of Chemistry, Box H, Brown University, Providence, RI 02912–9108, USA. C. T. Walsh, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. C. Khosla, Departments of Chemical Engineering, Chemistry, and Biochemistry, Stanford University, Stanford, CA 94305–5025, USA.

^{*}On sabbatical at Kosan Biosciences, Burlingame, CA 94010, USA.

the PKS that is responsible for the biosynthesis of the parent macrolactone precursor of the erythromycin antibiotics, has seven such ACP sites (Fig. 3A), one used in the initiation phase of polyketide chain building and six in the elongation phase (3). The trimodular ACV (aminoadipoyl-cysteinyl-valine) synthetase of penicillin biosynthesis has one PCP domain for each of its three amino acidactivating modules (Fig. 3B) (δ) whereas the undecapeptide immunosuppressive drug cyclosporin A is assembled on a megasyn-

thetase containing 11 primed PCP domains. After each PKS or NRPS has been converted from inactive apo- to phosphopantetheinylcontaining holo-ACP/PCP forms, it is able to support initiation, elongation, and termination of its characteristic polyketide or polypeptide product. In principle, these polymodular proteins should be capable of simultaneous, sequential processing of as many polyketide or polypeptide chains as there are ACP/PCP domains, analogous to the functioning of a typical industrial assembly line,

SCIENCE'S COMPASS



Fig. 1. Examples of polyketide and non-ribosomal peptide natural products.

although definitive evidence for such multiple loading is still not available.

Initiation of polyketide and polypeptide chain growth. The biosynthesis of polyketide and polypeptide chains is initiated by covalent loading of activated monomer units onto the first holo-ACP/PCP domains from which the resultant acylthio-enzyme intermediates act as donors in the first elongation cycle. In PKSs, as in fatty acid synthases, the starter units are acyl-CoA thioesters, usually acetyl-CoA or propionyl-CoA, with alternate starters such as isobutyryl-CoA (avermectin), cyclohexenoyl-CoA (FK506), and 3-amino-5hydroxybenzoyl-CoA (rifamycin) also being used. The corresponding NRPS loading monomers are not the analogous aminoacyl-CoAs, which are hydrolytically unstable. Instead, each amino acid is activated on the spot by dedicated adenylation domains (7), analogous to those in aminoacyl tRNA synthetases, that are placed upstream of each holo-PCP domain so as to produce a transient aminoacyl-AMP species that is captured by the PCP thiol to yield the aminoacyl-S-PCP as donor for chain initiation. In fact, such paired adenylation and PCP domains are the hallmarks of NRPSs, not just at the initiation site but also at every condensation and elongation way station, as illustrated for ACV synthetase (8) (Fig. 3B).

Acyl chain growth and translocation: Elongation cycles. The PKS and NRPS enzymes use equivalent elongation strategies in which the growing chain, docked as an acyl-S-enzyme at the upstream ACP/PCP domain, acts as an electrophile and is attacked by a nucleophilic species associated with the proximal downstream module, thereby enforcing



2 OCTOBER 1998 VOL 282 SCIENCE www.sciencemag.org

directionality of chain transfer (Fig. 2A). In the case of a PKS, the polyketide acyl chain is transferred from the ACP to the active site cysteine of the ketosynthase (KS) domain in the downstream module, from whence it is condensed with a malonyl- or methylmalonyl-S-ACP by a decarboxylative acylation, resulting in formation of a β -ketoacyl-S-ACP. In an analogous manner, the peptidyl-S-PCP_{*n*-1} of an NRPS is attacked by the amino group of the downstream aminoacyl-S-PCP_{*n*} to generate the characteristic peptide bond.

Within the same module as the KS-ACP domain pair in any PKS are up to three additional fatty acid-like catalytic domains that carry out ketone reduction, dehydration, and conjugate double-bond reduction, depending on the eventual extent of functional group modification that takes place within each chain elongation cycle (Fig. 2B). The resultant acyl-S-ACP intermediate then acts as the electrophilic donor for the next chain elongation cycle in which it is transferred downstream to the next module. The intermediate oxidation states of the polyketide acyl chain generated in each cycle persist once the polyketide chain is transferred out of that biosynthetic way station and accumulate in the eventually formed polyketide product (9). Analogously the elongation modules of an NRPS may contain one or more optional domains clustered around a core condensation-PCP domain pair. Thus, epimerization domains (L- to D-), methyl transferase domains (NH2- and COOHmethylation), and heterocyclization domains (serine to oxazoline; cysteine to thiazoline) can functionalize the growing peptide chain before transfer to the downstream aminoacyl-S-PCP (Fig. 2B).

SCIENCE'S COMPASS

Chain termination and release of the fulllength acyl chains. When the elongation and chain translocation cycles have brought the growing PK or NRP acyl chain to the furthest downstream ACP/PCP domain, chain termination occurs with release of the full-length acyl or peptidyl chain and regeneration of the free holo forms of the megasynthases. The final acyl-S-ACP/PCP chains are released from their respective synthases either by intermolecular attack by water, resulting in net hydrolysis, or by intramolecular capture by a hydroxyl or amino group of the acyl chain itself, giving rise to a lactone or cyclic peptide product. The first termination route yields the free acid, as in the release of aminoadipoyl-cysteinyl-valine from ACV synthetase (Fig. 3B) during penicillin biosynthesis. Alternatively, DEBS-catalyzed capture by a polyketide chain hydroxyl group results in the formation of the 14-membered macrolactone 6-deoxyerythronolide B (Fig. 3A) (10).

The common catalytic element used by the megasynthases for chain termination is a 25- to 35-kD polypeptide segment located at the very COOH-terminus of these polyfunctional enzymes that has been termed a thioesterase (TE) domain (11). TE domains have an active-site serine to which the fulllength acyl chain is transferred from the last ACP/PCP site to generate a transient acyl-O-TE intermediate. This final covalent acyl-enzyme species is then cleaved by hydrolysis or cyclization, depending on the nature of the recruited nucleophile. The TE domain of DEBS is clearly portable; when fused downstream of other modules in this PKS, the TE directs lactonization and release of the resultant intermediate PK chains (12). In the Escherichia coli NRPS that mediates the formation of the ironchelating siderophore enterobactin (tris-*N*dihydroxybenzoylserine trilactone), the TE domain may perform both elongation and cyclotrimerization functions (*13*).

Once the mature polyketide or polypeptide chains have been released from the bucket brigade of pantetheinyl tethers on the respective PKS and NRPS complexes, they frequently undergo further enzymatic tailoring by ancillary enzymes. Indeed, such late-stage modification is often required for the final product to be biologically active. Most such tailoring enzymes are dedicated to the biosynthetic pathway itself and are encoded by genes that are clustered with the core PKS and NRPS genes. These modifying activities can include hydroxylases, glycosyl transferases, and methyl transferases, a combination of which, for example, is required for the conversion of 6-DEB to erythromycin A (14). Analogous dedicated hydroxylases and glycosyl transferases are encoded in biosynthetic operons for the vancomycin class of glycopeptide antibiotics along with genes for peroxidative enzymes responsible both for chlorination and the oxidative cross-linking of the heptapeptide skeleton to create the rigid scaffolding required for the high-affinity recognition of bacterial peptidoglycan termini characteristic of these antibiotics (15).

Combinatorial Manipulation of the Structures of Polyketides and Non-ribosomal Peptides

Despite the enormous number of known polyketides and non-ribosomal peptides, it has recently become evident that combinatorial genetics and chemistry can be used to expand vastly the molecular diversity of these pharmacologically important metabolites, tak-



Fig. 3. Examples of modular polyketide synthases and non-ribosomal peptide synthetases. (**A**) Deoxyerythronolide B synthase (DEBS), a PKS generating the precursor of erythromycin; AT, acyl transferase, TE, thioesterase (abbreviations as in Fig. 2). (**B**) Aminoadipyl-cysteinyl-valine

synthetase (ACVS), a NRPS biosynthesizing the precursor of penicillins and cephalosporin. A_{Aad}, aminoadipoyl adenylase; A_{Cys}, cysteine adenylase; A_{Val}, valine adenylase; C, peptide-condensing enzyme.

ing advantage of the natural modularity of polyketide and non-ribosomal peptide biosynthesis (16). For example, there are four degrees of freedom in polyketide biosynthesis that can be independently manipulated by genetic engineering: (i) the length of the polyketide chain, which is determined by the number of modules that comprise the polyketide synthase (3, 12); (ii) the choice of primer and extender units, each controlled by gatekeeper acyl transferase (AT) domains (17); (iii) the degree of reduction of the polyketide backbone, which is determined by the set of enzyme domains present in each module (18); and (iv) the stereochemistry at centers carrying alkyl and hydroxyl substituents, which is locally controlled by enzyme domains that are responsible for generating the stereocenter in question (12, 19). Likewise, in non-ribosomal peptide biosynthesis, the degrees of freedom include the number, type, and configuration of each amino acid building block, as well as the specific combination of modifications controlled by each NRPS module, including N-methylations, epimerizations, and cyclizations. Even for a modest-size enzyme complex, the theoretical number of possible products is extremely high. Practical realization of the combinatorial potential for generating new natural products will depend on the extent of molecular recognition by individual PKS and NRPS modules and the development of an optimal toolbox for exploiting this modularity.

Because modules 1, 2, 5, and 6 of the erythromycin PKS each contain only a ketoreductase domain in addition to the core ketosynthase. AT, and ACP domains, the polyketide chain elongation intermediates generated by each of these modules carry hydroxyl functions at the relevant positions of the chain. Likewise, the presence of methyl transferase domains in modules 2, 3, 4, 5, 7, 8, and 10 of the cyclosporin synthetase results in N-methylation of the corresponding amide bonds. In practice, the actual synthetic potential is limited to a certain extent by the intrinsic chemistry of each module. Thus, dehydration can occur only if ketoreduction has taken place; similarly, enoylreduction requires prior dehydration (20).

A potentially far more serious limitation to the manipulation of the modularity of the biosynthetic apparatus arises from the possibility that downstream modules in a PKS or NRPS may not accept or process efficiently the anomalous product of an engineered upstream module. Indeed, the factors controlling transfer of growing polyketide or polypeptide chains to downstream modules are poorly understood, especially when those modules are located in distinct protein subunits. Two extreme models may be considered. In one case, intermodular chain transfer would be dominated primarily by proteinprotein recognition, with more or less nonselective transfer of the growing chain from the donor module. Contrasted with this fixed menu model is an à la carte model, in which acceptor modules actively recognize key structural features of their natural substrates and discriminate to varying degrees against functionally modified analogs. That small molecule-protein recognition plays at least some, if not the dominant, role in the efficient acylation of KS domains in target modules is evident from the success of precursordirected biosynthesis experiments in which exogenously administered N-acetyl cysteamine (NAC) thioester derivatives of natural polyketide chain elongation intermediates and their analogs (21) are recognized and processed by the target DEBS modules to give 6-dEB and its structural analogs.

Individual modules are considerably tolerant toward loss-of-function, change-of-specificity, or gain-of-function mutagenesis (22, 23). A major requirement for the practical exploitation of combinatorial biosynthesis is the development of an efficient and flexible genetic toolbox for rapid, convenient, and high-yield manipulation of PKS and NRPS structural genes. Polyketides and non-ribosomal peptides are biosynthesized by microorganisms spanning Gram-positive bacteria, Gram-negative bacteria, and filamentous fungi, thus complicating the choice of a surrogate host, given the serious issues of promoter compatibility and major differences in codon usage. In addition, the large size of PKS or NRPS gene clusters (10 to 100 kb) makes them particularly challenging targets for protein engineering. Moreover, every acyl carrier protein (ACP or PCP) domain within these systems must be posttranslationally phosphopantetheinvlated in order for the PKS to be functional. Furthermore, precursors for polyketide and non-ribosomal peptide biosynthesis, including unusual amino acids and even simple precursors such as methylmalonyl-CoA, may require specialized biosynthetic pathways not present in all organisms. Finally, formation of pharmacologically active products from the initially generated polyketide or polypeptide chains almost always requires the further action of dedicated tailoring enzymes responsible for a variety of biochemical modifications, including cyclization, glycosylation, methylation, and oxygenation. The balance between intrinsic substrate specificity and catalytic plasticity of these tailoring enzymes could be critical to the successful modification of polyketide and non-ribosomal peptide libraries.

Two general strategies have been developed for genetic engineering of PKSs and NRPSs. In one approach, homologous recombination has been used to replace or delete segments of genomic DNA in the native producing organism, ranging from individual base pairs through discrete domains up to entire modules (1, 3, 17, 18). This method has the advantage that all the native host enzymes and relevant regulatory elements are in principle present. Moreover, enzymes controlling the supply of both starter and chain extension units as well as post-PKS tailoring enzymes should be fully active, thereby providing an adequate and well-regulated supply of substrates and allowing the appropriate modification of the initially generated polyketide or polypeptide product. Homologous recombination through use of the native producer suffers from several drawbacks, however, the most serious being that it is often tedious and technically difficult, especially when applied to slow-growing organisms that have poorly developed transformation systems or that lack other important genetic tools.

The major alternative to the use of homologous recombination has been to use appropriate vectors to transfer the entire PKS or NRPS to a heterologous host or hosts, in which the background genetic methodology is already more highly developed. The most successful example to date of such a strategy has been the utilization of a Streptomyces coelicolor-based host-vector system that has been extensively used for the heterologous expression of bacterial and fungal PKSs (24). Eventually, a model host such as E. coli might prove to be the organism of choice for polyketide and non-ribosomal peptide biosynthesis, given its rapid growth rate, extensive genetic toolbox, and proven capability as a host for protein overexpression. Until very recently, a major drawback to the use of any heterologous host has been the potential absence of enzymes essential for the requisite posttranslational modification of the PKS or NRPS gene products and for the biosynthesis of required substrates. Cloning of a number of phosphopantetheinvl transferases (5) has made possible the successful overexpression of posttranslationally modified PKS and NRPS modules in E. coli and yeast (25). Production of polyketides and non-ribosomal peptides in E. coli is still limited by the restricted availability of suitable metabolic precursors, a problem that can in principle be addressed by established metabolic engineering approaches. Similar strategies could also be exploited to introduce the necessary tailoring enzymes into the E. coli host. Analogous approaches may also be contemplated with other well-developed host-vector systems, including those for yeast, bacillus, and ultimately mammalian and plant hosts.

The power of genetic engineering for harnessing the biosynthetic potential of PKSs and NRPSs can be considerably amplified when coupled with complementary chemical approaches (26). A hybrid chemical-biological strategy for the generation of libraries of "unnatural" natural products would involve in vitro enzymatic synthesis of polyketides and nonribosomal peptides. The feasibility of the latter process has already been demonstrated by the

SCIENCE'S COMPASS

successful cell-free synthesis of complex polyketides such as 6-dEB (27) and non-ribosomal peptides such as enterobactin (28). The successful overexpression of posttranslationally modified modules in E. coli has made possible an abundant supply of such multifunctional catalysts (25). In the case of non-ribosomal peptides, a wide range of natural and unnatural amino acid precursors can be cheaply prepared, and methods for in situ regeneration of ATP are well established. The prohibitively high cost of CoA thioesters has recently been addressed through the development of a facile method for the synthesis of malonyl- and methylmalonylthioesters of N-acetylcysteamine, and the demonstration that these simplified substrates can replace the more commonly used CoA thioesters as substrates for polyketide chain elongation (29).

Future Directions of Combinatorial Biosynthesis

The impact of combinatorial biosynthesis is most likely to be felt in the near future in connection with promising new natural product leads, such as the antibacterial ketolides, the neurotrophic agent FK506, and the anticancer epothilones (30). In addition to providing methods for the optimization of existing pharmacological leads and for the dissection of novel modes of action, combinatorial biosynthesis promises to be a new resource for the discovery of cell-permeable bioactive ligands. The primary bottleneck for semisynthetic modification and biological evaluation of both natural and unnatural polyketide and non-ribosomal peptides is often their limited supply (31).

Increases in the size and diversity of combinatorial biosynthetic libraries will also require the development of advanced strategies for combinatorial manipulation of PKS and NRPS gene clusters. The development of sets of compatible expression vectors carrying individual open reading frames derived from the parent multigene clusters could significantly reduce the special difficulties associated with in vitro manipulation of large DNA fragments while simplifying access to multiple mutants of large PKS or NRPS genes based on plasmid cotransformation. Furthermore, the enhancement of transformation, transfection, and conjugation methods, or the development of improved techniques for in vivo genetic engineering, could be expected to increase the accessible size of combinatorial biosynthetic libraries. Finally, state-ofthe-art combinatorial mutagenesis and DNAshuffling methods (32) are likely to have a major impact both on the generation of new molecules and on the improvement of product yields.

Technological advances are also needed for the rapid and convenient development of new pathways amenable to combinatorial biosynthesis. At present, this is an intrinsically labor-intensive process that requires cloning, verification, and sequencing of gene clusters that can span as many as 100 kb. Complicating these problems is the fact that these genes often reside in genetically poorly characterized, experimentally hard-to-manipulate microorganisms, for which new transformation or recombination systems must often be developed de novo. The design of robust plasmids capable of carrying large DNA fragments (for example, >100-kb inserts can be maintained in bacterial artificial chromosomes) could facilitate shotgun cloning of entire biosynthetic gene clusters (including regulatory genes) from producing organisms. On the other hand, the development of model host organisms within genera of prolific natural product producers such as the actinomycetes, myxobacteria, and filamentous fungi, when coupled with high throughput micro-analytical technologies, will undoubtedly facilitate functional evaluation of shotgun-cloned gene clusters.

The incorporation of unnatural precursors into polyketide and non-ribosomal peptides is likely to expand dramatically the size and diversity of resulting libraries. Two prerequisites must be satisfied: access to suitable precursor libraries, and the development of "gatekeeper" ATs or adenylating domains with broad substrate specificities that have orthogonal specificity to most naturally occurring domains (equivalent in a sense to the role of amber suppressor codons) (33). Perhaps the most intriguing feature of combinatorial biosynthesis is that it provides a selfreplicating, biological approach to the synthetic generation of small-molecule diversity. One exciting prospect will be to couple library generation to lead selection and amplification through suitable in vivo or in vitro assays. Given the explosive growth in the use of direct, cell-based assays optimized for molecular targets of interest, it may soon be possible to engineer and exploit contrived "autocrine" systems in hosts that produce "unnatural natural product" libraries. By coupling the marginal survival of a bacterial clone to the biological efficacy of its polyketide or non-ribosomal peptide product, one could conceivably mimic in the laboratory the evolutionary processes that may have been the very source of the already vast repertoire of bioactive natural products.

References and Notes

- For a thematic issue reviewing the recent literature on the molecular genetics and enzymology of polyketide and non-ribosomal peptide biosynthesis, see *Chem. Rev.* 97 (November 1997).
- 2. It has often been argued that the observed biological potency and selectivity of these microbial metabolites is the result of evolutionary pressures. While this simple argument may be true for some natural products, an increasing number of examples suggest that the complete picture is far more complex. For exam-

ple, erythromycin A, an antibiotic produced by the soil organism Saccharopolyspora erythraea and which inhibits ribosomal protein biosynthesis in bacteria, readily undergoes an intramolecular cyclization under acidic conditions such as found in the stomach so as to generate a rearranged metabolite that is inactive as an antibacterial but is now a potent agonist of the motilin receptor [S. Omura et al., J. Med. Chem. 30, 1943 (1987); P. A. Lartey et al., ibid. 38, 1793 (1995)]. Clearly these two properties associated with the same fundamental natural product skeleton are mechanistically unrelated, and it is unlikely that evolutionary pressures led to the simultaneous optimization of both properties. What is undisputed about these families of natural products are their intrinsic pharmacological properties.

- J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **348**, 176 (1990); S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **252**, 675 (1991); P. Caffrey, D. J. Bevitt, J. Staunton, P. F. Leadlay, *FEBS Lett.* **304**, 225 (1992).
- 4. Many polyketides, particularly aromatic polyketides such as tetracyline, are biosynthesized by structurally simpler polyketide synthases called iterative or type II synthases. Although these PKSs are also subject to combinatorial manipulation, they are not discussed in this review. For recent comprehensive reviews, see (1).
- 5. R. H. Lambalot et al., Chem. Biol. 3, 923 (1996).
- 6. G. Weber, K. Schorgendorfer, E. Schneider-Scherzer, E. Leitner, *Curr. Genet.* **26**, 120 (1994).
- M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* 97, 2651 (1997); H. von Döhren, U. Keller, J. Vater, R. Zocher, *ibid.*, p. 2675.
- D. J. Smith, J. E. Alison, G. Turner, *EMBO J.* 9, 2743 (1990); S. Guttierez, B. Diez, E. Montenegro, J. F. Martin, J. Bacteriol. 173, 2354 (1991); J. R. Coque, J. F. Martin, P. Liras, *Mol. Microbiol.* 5, 1125 (1991); A. P. MacCabe *et al.*, *J. Biol. Chem.* 266, 12646 (1991).
- D. E. Cane and C.-C. Yang, J. Am. Chem. Soc. 109, 1255 (1987); S. Yue, J. S. Duncan, Y. Yamamoto, C. R. Hutchinson, *ibid.*, p. 1253.
- Internal capture by the amino group of the mature polypeptide chain is the release mechanism in the formation of the characteristic macrolactam of such cyclic peptides as cyclosporin A, gramicidin S, and bacitracin.
- 11. K. J. Weissman et al., Angew. Chem. Int. Ed. Engl. 37, 1437 (1998).
- J. Cortes et al., Science 268, 1487 (1995); C. M. Kao, G. Luo, L. Katz, D. E. Cane, C. Khosla, J. Am. Chem. Soc. 117, 9105 (1995); *ibid.* 118, 9184 (1996); I. Böhm et al., Chem. Biol. 5, 407 (1998).
- A. M. Gehring, I. M. Mori, C. T. Walsh, *Biochemistry* 37, 2648 (1998).
- 14. R. G. Summers et al., Microbiology 143, 3251 (1997).
- A. M. A. van Wageningen et al., Chem. Biol. 5, 155 (1998).
- C. J. Tsoi and C. Khosla, *ibid.* 2, 355 (1995); J. Rohr, Angew. Chem. Int. Ed. Engl. 34, 881 (1995); P. F. Leadlay, Curr. Opin. Chem. Biol. 1, 162 (1997).
- M. Oliynyk, M. J. B. Brown, J. Cortes, J. Staunton, P. F. Leadlay, *Chem. Biol.* **3**, 833 (1996); L. Liu, A. Thamchaipenet, H. Fu, M. Betlach, G. Ashley, *J. Am. Chem. Soc.* **119**, 10553 (1997); X. Ruan *et al.*, *J. Bacteriol.* **170**, 6416 (1997).
- S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **252**, 675 (1991); S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson, L. Katz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7119 (1993); D. Bedford, J. R. Jacobsen, G. Luo, D. E. Cane, C. Khosla, *Chem. Biol.* **3**, 827 (1996); R. McDaniel et *al.*, *J. Am. Chem. Soc.* **119**, 4309 (1997); C. M. Kao et *al.*, *ibid.*, p. 11339.
- C. M. Kao et al., J. Am. Chem. Soc. **120**, 2478 (1998);
 K. J. Weissman et al., Biochemistry **36**, 13849 (1997).
- 20. In some cases, presumptive domains recognized on the basis of their characteristic consensus sequences may be functionally silent. For example, although various modules of the 14-module rapamycin synthase possess auxiliary domains with inferred ketoreductase, dehydratase, or enoylreductase properties, for reasons that are still obscure these latent activities are not always manifest in the structure of the eventually formed natural product []. F. Aparicio et and the structure of the sequences of the sequence of the sequence

SCIENCE'S COMPASS

al., *Gene* **169**, 9 (1996); T. Schwecke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7839 (1995)].

- J. R. Jacobsen, C. R. Hutchinson, D. E. Cane, C. Khosla, Science 277, 367 (1997); J. R. Jacobsen, D. E. Cane, C. Khosla, J. Am. Chem. Soc. 120, 9096 (1998); J. R. Jacobsen, A. Keatinge-Clay, D. E. Cane, C. Khosla, Bioorg. Med. Chem. 6, 117 (1998); A. F. A. Marsden, B. Wilkinson, J. Cortes, N. J. Dunster, J. Staunton, P. F. Leadlay, Science 279, 199 (1998).
- Summarized in C. Khosla, R. Gokhale, J. R. Jacobsen, D. E. Cane, Annu. Rev. Biochem., in press.
- 23. In some cases transfer of the resultant structurally modified intermediate to downstream modules may limit the overall processivity of the recombinant multienzyme system. Although the exact magnitude and causes of these kinetic bottlenecks remain to be assessed, reductions in in vivo productivity of modified polyketides of up to three orders of magnitude compared with the titer of the natural product produced by the parent PKS have been observed. For example, replacement of the methylmalonyl-specific AT domain in module 2 of the erythromycin PKS by a suitable malonyl-CoA transferase from the rapamycin PKS results in biosynthesis of the expected normethyl-triketide by a truncated, two-module variant of DEBS in yields that approximate natural triketide production, implying that the activities of the native enzymes in the hybrid module are not significantly eroded by the introduction of a heterologous AT domain. When the same AT substitution is performed in the full, 6-module DEBS PKS, however, the expected normethyl-macrolide is produced in considerably reduced yields compared with those of the natural product 6-dEB (J. Lau, H. Fu, D. E. Cane, C. Khosla, unpublished results). The reasons for this reduction in yield are not clear. Developing strategies to overcome these deleterious effects on chain transfer represents a major challenge.
- R. McDaniel, S. Ebert-Khosla, D. Á. Hopwood, C. Khosla, J. Am. Chem. Soc. **115**, 11671 (1993); C. M. Kao, L. Katz, C. Khosla, *Science* **265**, 509 (1994); D. J. Bedford, E. Schweizer, D. A. Hopwood, C. Khosla, J. Bacteriol. **177**, 4544 (1995).
- J. Kealey, L. Liu, D. V. Santi, M. Betlach, P. J. Barr, Proc. Natl. Acad. Sci. U.S.A. 95, 505 (1998); R. S. Gokhale and C. Khosla, unpublished results; A. Gehring, I. Mori, R. Perry, C. Walsh, Biochemistry 37, 11637 (1998).
- 26. For example, genetic engineering can be used to

introduce novel functionality into or delete existing functional groups from a polyketide or non-ribosomal peptide backbone, thereby facilitating subsequent semisynthetic modification. In such cases genetics can be viewed as yet another tool in the medicinal chemist's armamentarium. Alternatively, chemical synthesis can precede engineered biosynthesis, whereby synthetic mimics of biosynthetic starting materials or chain elongation intermediates are fed to a recombinant strain carrying a lesion in an early step in the biosynthetic pathway. Processing of these synthetic analogs by the appropriate domains of the multi-enzyme synthase results in the formation of analogs of the natural product containing functional groups or other chemical properties that are ordinarily unavailable from primary metabolism. (21).

- R. Pieper, G. Luo, D. E. Cane, C. Khosla, *Nature* **378**, 263 (1995).
- A. M. Gehring, K. A. Bradley, C. T. Walsh, *Biochemistry*, 36, 8495 (1997).
- N. L. Pohl, R. S. Gokhale, D. E. Cane, C. Khosla, J. Am. Chem. Soc., in press.
- 30. Ketolides are a class of semisynthetic derivatives of erythromycin that possess excellent activity against a wide range of drug-resistant bacteria [C. Agouridas, Y. Benedetti, A. Denis, O. Le Martret, J. F. Chantot, 35th Interscience Conference On Antimicrobial Agents and Chemotherapy, San Francisco, 17 to 20 September 1995 (abstract F157); G. Griesgraber et al., J. Antibiot. 49, 465 (1996)]. Exploration of the pharmacological properties of ketolides has been limited by the difficulty of semisynthesis from naturally occurring macrolides. Combinatorial biosynthesis is ideally suited to the selective de novo construction of several of the required functional elements in the target ketolide molecule. Recent reports that FK506 can stimulate nerve regeneration [B. G. Gold, Mol. Neurobiol. 15, 285 (1997)] have stimulated interest in the identification of the relevant structural features responsible for the newly discovered activity of this polyketide. The epothilones are a class of polyketide natural products from Sorangium cellulosum, which are potent Taxol-like cytotoxic agents that are considerably more water-soluble than Taxol and active against Taxol-resistant tumors [K. C. Nicolaou et al., Angew. Chem. Int. Ed. Engl. 36, 2097 (1997); D.-S. Su et al., ibid., p. 2093.
- 31. For example, although the erythromycin PKS genes

are expressed in the heterologous S. coelicolor host at levels that are considerably higher than are observed in the natural erythromycin producer, Sac. erythraea, actual production of the resultant polyketides is an order of magnitude higher in the native producing strain. This difference is most likely due to the limited supply of the requisite polyketide building blocks, propionyl-CoA and methylmalonyl-CoA, in S. coelicolor. Traditionally, problems of low metabolite production have been overcome by classical strain improvement based on routine but time-consuming empirical application of random chemical mutagenesis, coupled with iterative screening of producing microorganisms, eventually leading in some cases to improvement in metabolite titers of more than three to four orders of magnitude over those of the wild-type strains. The application of functional genomic technologies in synergism with emerging approaches for kinetic analysis of complex metabolic networks could provide powerful new tools for the rapid and efficient increases in metabolite production.

- A. Crameri, S.-A. Raillard, E. Bermudez, W. P. C. Stemmer, *Nature* **391**, 288 (1998); P. A. Patten, R. J. Howard, W. P. Stemmer, *Curr. Opin. Biotechnol.* **8**, 724 (1997).
- 33. The structural determinants of chain extension unit specificity have been narrowed to a sequence of 20 to 30 amino acids (23), suggesting that it should be possible to modify the substrate specificity of AT domains. Likewise, the recent determination of the structure of the adenylation domain of the phenylalanine-activating module of the gramicidin NRPS opens the door to a structurebased approach toward the engineered biosynthetic incorporation of novel amino acids into nonribosomal peptides (7).
- 34. Research on many of the topics covered in this review has been supported by grants from the NIH (GM22172 to D.E.C., CA66736 to C.K., GM20011 to C.T.W.) and by a David and Lucile Packard Foundation Grant (to C.K.). The authors are also members of the Scientific Advisory Board of Kosan Biosciences. We thank P. Belshaw for assistance in design of the figures.

So instant, you don't need water...

NEW! Science Online's Content Alert Service

There's only one source for instant updates on breaking science news and research findings: *Science*'s Content Alert Service. This free enhancement to your *Science* Online subscription delivers e-mail summaries of the latest research articles published each Friday in *Science* – **instantly**. To sign up for the Content Alert service, go to *Science* Online – and save the water for your coffee.



For more information about Content Alerts go to www.sciencemag.org. Click on Subscription button, then click on Content Alert button.