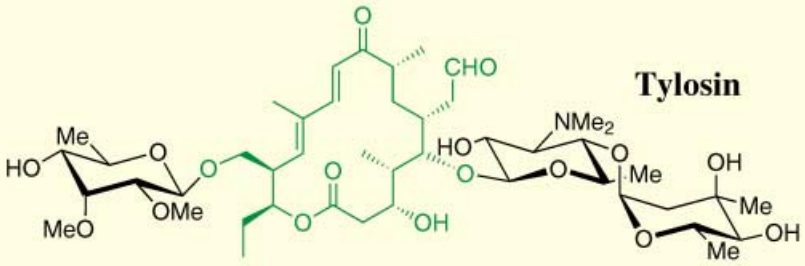
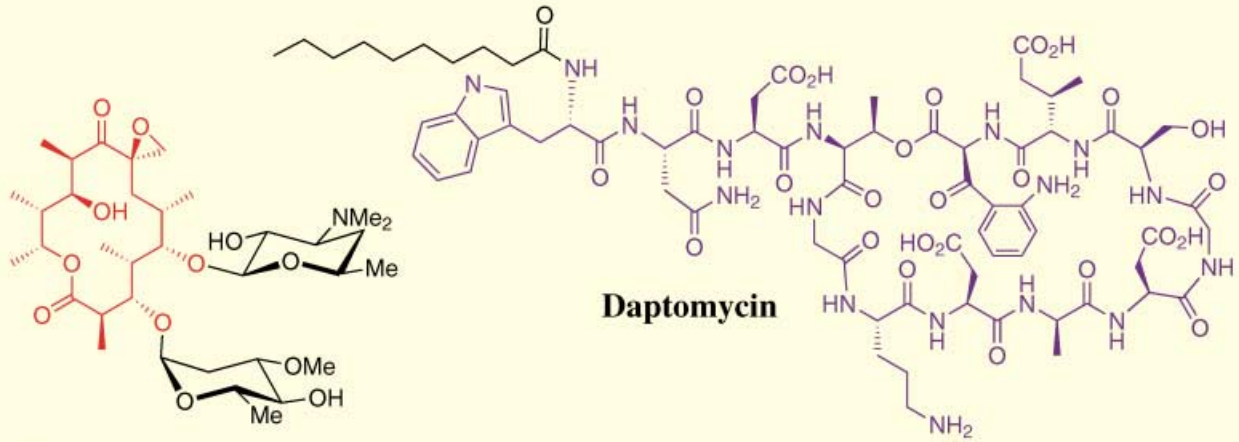
Nonribosomal Peptide and Polyketide Antibiotics

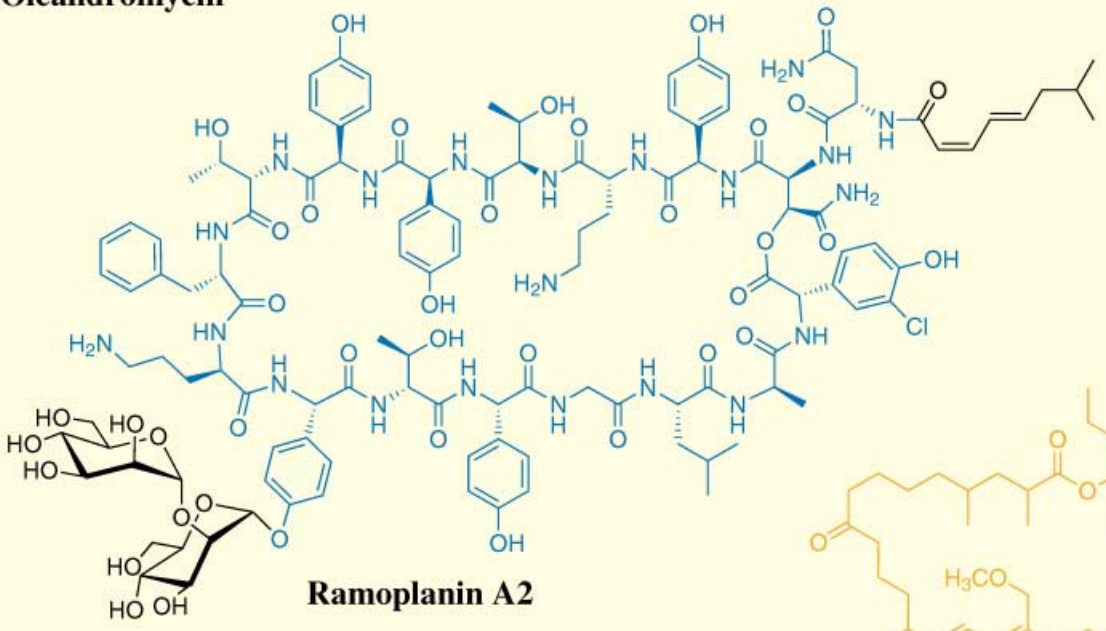


Tylosin



Daptomycin

Oleandromycin



Ramoplanin A2

Antibiotic TA

  1439-4227/02/03/02-03 $ 17.50+.50/0 ChemBioChem **2002**, 3, 124 – 134

Combinatorial Biosynthesis of Antibiotics: Challenges and Opportunities

Christopher T. Walsh*[a]

Natural products with antibiotic activity have been central agents in human therapeutics over the past fifty years. They are likely to remain crucial in the decades to come. These molecules, often termed secondary metabolites because they are the end products of dedicated metabolic pathways that are turned on when microbes are stressed by environmental factors such as starvation, can acheive considerable architectural and functional group complexity that allows specific targeting. The programmed manipulation of the genes that encode the enzymes in the biosynthetic pathways offers promise for redesign of antibiotic structures to create new activities and overcome bacterial resistance to existing antibiotics.

KEYWORDS:

antibiotics · biosynthesis · combinatorial chemistry

1. Introduction

Dramatic increases in the number of antibiotic-resistant pathogenic bacteria in the past decade[1] have focused attention on the need for new antibiotics. With the genomic sequences of several antibiotic-producing microbes available, there has been much interest in combinatorial biosynthetic approaches[2] to manipulate the genes of antibiotic biosynthesis pathways to generate hybrid or variant antibiotics that may have novel properties and efficacies against problematic and resistant pathogens. Some classes of antibiotics are judged to be more appropriate than others for combinatorial manipulations of biosynthetic pathways. When the genes relevant for the biosynthesis are tightly clustered in the bacterial chromosomes, both gene replacements and coregulation of all the genes necessary for formation of an antibiotic are facilitated.

2. Polyketide and Nonribosomal Peptide Antibiotics

The polyketide (PK) antibiotics, both the aromatic family[3] represented by tetracyclines and the macrolactone family[4] represented by erythromycins, are in the category mentioned above, while the aminoglycosides are not. The genes of the latter are not coregulated and are dispersed throughout the genomes of the producer organisms. The peptide-based antibiotics—the vancomycins, bacitracins, and β-lactam antibiotics—are made on nonribosomal peptide synthetase (NRPS) assembly lines[5, 6] with logic paralleling the polyketide biosynthesis, and they have been considered candidates for combinatorial biosynthesis approaches. In the macrolide antibiotics, the first-generation erythromycin (**1**, see Scheme 1a) has been substantially supplanted by the second-generation derivatives azithromycin and clarithromycin, and a third generation, the ketolides, are in late-stage clinical trials. On the vancomycin front, the dramatic rise in vancomycin-resistant enterococci (VRE) and the resulting high mortality rates in hospitals and clinics have spurred efforts towards the development of second-generation semisynthetic glycopeptide antibiotics;[7, 8] however, total synthesis is likely to remain impractical due to the structural complexity of the vancomycins.

The three-subunit protein assembly lines for the aglycon of erythromycin (**1**)[9, 10] and of a vancomycin analogue, chloroeremomycin (**2**),[11] in which the peptide scaffold is identical to that of vancomycin, are schematized in Schemes 1a and b, respectively. The three-subunit deoxyerythronolide B synthase (DEBS) has one loading module and six extender modules (modules 1–6) that utilize one propionyl-CoA and six methylmalonyl-CoA monomers. The polyketide chain grows as a cascade of translocating acyl-S-enzyme intermediates, which are tethered to the acyl carrier protein (ACP) domain of each module by a phosphopantetheinyl prosthetic group. Release involves intramolecular attack by the C13 hydroxyl oxygen on the thioester carbonyl of the full-length acyl-S-ACP to give the 14-membered macrolactone erythronolide. Analogously, the heptapeptide backbone of the vancomycin family is assembled by the three-subunit synthetase Cep, which has seven modules,[12] one for each amino acid to be activated, and is tethered as an aminoacyl-S-carrier protein. Peptidyl chain growth again occurs by an N-to-C translocation of an elongating series of peptidyl-S-enzyme intermediates, the last of which is hydrolyzed to release the heptapeptide acid.

[a] *Prof. C. T. Walsh*
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, MA 02115 (USA)
Fax: (+1) 617-432-0438
E-mail: Christopher_walsh@hms.harvard.edu

Each aglycon then undergoes a series of post-assembly-line transformations carried out by dedicated tailoring enzymes also encoded within the antibiotic biosynthetic clusters. The maturation of both the aglycons shown in Scheme 1 into the active antibiotics involves regiospecific oxidation/oxygenation reactions and glycosylations, typically with deoxysugars, as will be discussed below.

Since the sequencing of the DEBS genes about a decade ago,[9, 10] that assembly line has been the main focus of efforts directed towards combinatorial biosynthesis. Several milestones have been met, including the development of host strains of streptomycetes for genetic manipulation and protein expression.[13–15] To implement combinatorial biosynthesis for the polyketide macrolides and more generally for the peptide antibiotics as well, three challenges have been identified. One is the availability of monomers to serve as substrates for acyl chain growth in the biosynthesis as shown in Scheme 1. A second requirement is the reengineering of the assembly lines to reprogram chain initiation, chain elongation, and chain termination steps to carry out combinatorial strategies. A third issue is altering or lowering the specificity of the tailoring enzymes that are committed to the maturation of the antibiotics.

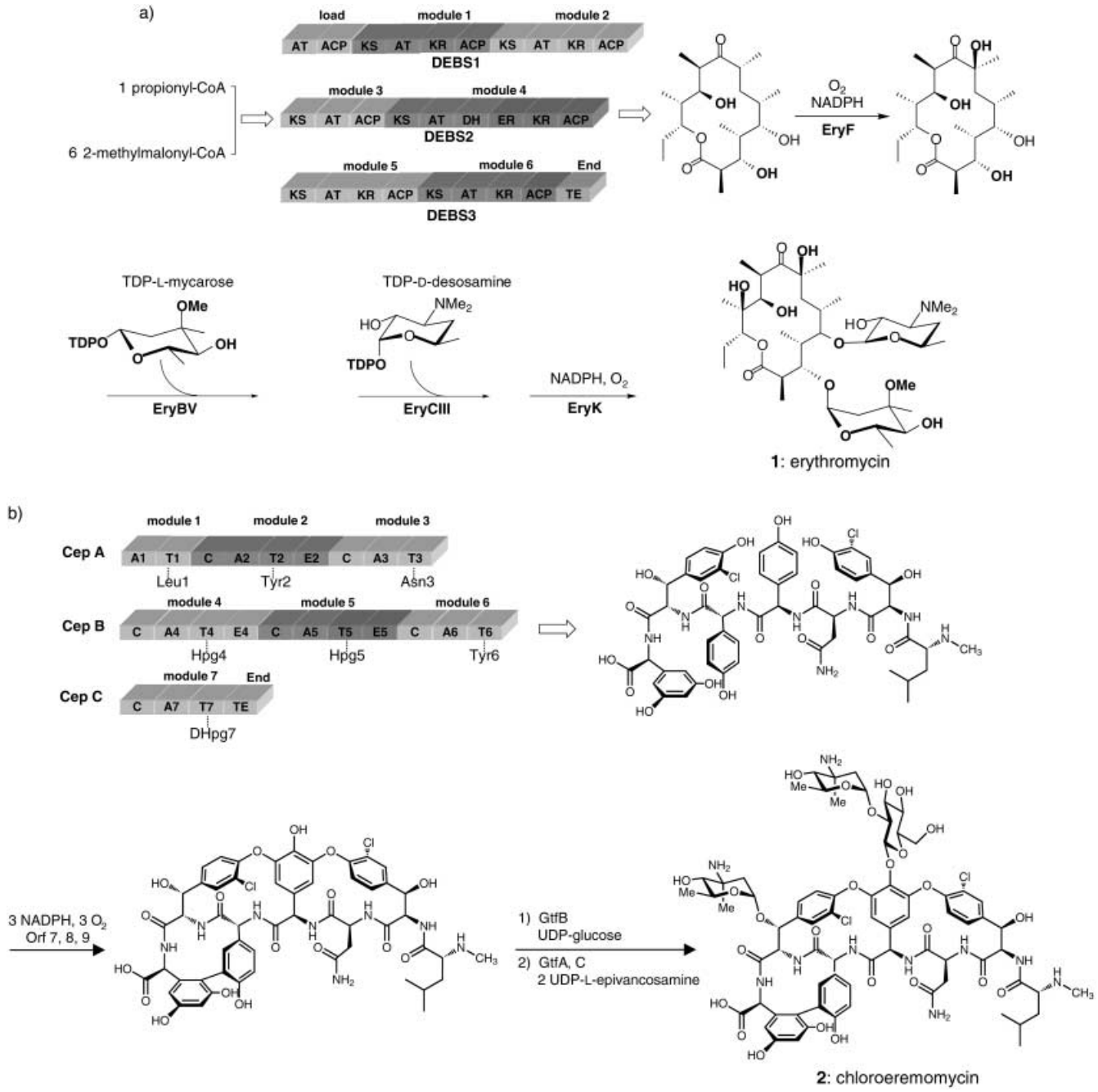
Christopher T. Walsh, born in 1944, majored in biology at Harvard College and then obtained a PhD in biochemistry in the laboratory of F. Lipmann at the Rockefeller University, New York. After postdoctoral research with R. H. Abeles from 1970 – 1972, he joined the faculty of the Massachusetts Institute of Technology (MIT), in both the chemistry and biology departments, to study enzymatic reaction mechanisms, with an emphasis on inhibitor design in both the mechanism-based and slow-binding classes. His group also studied enzymatic catalysts involved in Baeyer – Villiger oxidations, cyclopropane fragmentations, light-induced repair of thymine dimers, nickel-containing enzymes involved in methanogenesis, and the enzymatic detoxification of mercurials. After fifteen years at MIT he left to chair a combined department of biological chemistry and molecular pharmacology at Harvard Medical School. At Harvard his group has studied enzymes of therapeutic relevance, including the peptidyl-prolyl isomerase that is the binding protein for cyclosporin A. Most of the recent focus has been on the mechanism of action of antibiotics, including the antibacterial drug of last resort, vancomycin. He and his group have determined the function of five enzymes that are necessary and sufficient to cause high-level clinical resistance to vancomycin by pathogenic enterococci. The interest in vancomycin, a nonribosomally produced heptapeptide, has extended to studies on the biosynthesis of other peptide antibiotics and bacterial iron-chelating siderophores by multimodular enzyme assembly lines with the intent of decoding the mechanisms and utilizing them for combinatorial biosynthesis.

3. Substrate Monomer Availability: Initiator and Extender Molecules

For polyketide antibiotics the monomers are acyl-CoAs while for nonribosomal peptide antibiotics, amino acids, many of them nonproteinogenic, are activated as aminoacyl-AMP species in each module. For glycosylated antibiotics the relevant sugar monomers are the NDP-deoxyhexoses (NDP = nucleoside diphospho).[16, 17]

The chain-initiating acyl-CoA is generally distinct from chain-extending acyl-CoAs, which are almost always either malonyl-CoA or methylmalonyl-CoA. The decarboxylation of malonyl-S-carrier proteins provides the driving force for C–C bond formation in each elongation, translocation step of chain growth. The chain-initiating monomers can be diverse: propionyl- and phenylacetyl-CoA for erythromycin (**1**) and phenyerythromycin (**3**), respectively,[4] cyclohexanecarboxyl-CoA in the antiparasitic drug dormectin (**4**),[18] and (5-amino-3-hydroxybenzoyl)-CoA in rifamycin B (**5**)[19] (Scheme 2). Even if a chain-initiating loading domain can be mutated to relaxed specificity or swapped onto the start of an assembly line[20] to broaden specificity, there is still the real problem of providing the specific starter acid in useful levels. While one can assay the selectivity of the loading domain in vitro,[21] or add the free acid exogenously in precursor-directed biosynthesis,[22, 23] providing the molecule in vivo is likely to require transfer of the genes that code for monomer biosynthesis. This has been achieved recently for propionyl-CoA in *Escherichia coli* containing the three DEBS subunits (see Scheme 1 a),[24] and also in the production of dormectin (**4**), where five genes for cyclohexane carboxylate biosynthesis were transferred from *Streptomyces collinus* to the producer strain of *S. avermitilis*.[18] Fine-tuning the regulation of gene activity and levels of the activated acyl-CoAs may be challenging for each initiator unit. There is much less choice and versatility in the carboxyacyl-CoA extender units, where malonyl and methylmalonyl are essentially the exclusive two-carbon or three-carbon building blocks recognized by the acyl transferase catalytic domains in each elongation cycle.

In the fermentations to provide nonribosomal peptide antibiotics the 20 proteinogenic amino acids are likely to be available; however, the nonproteinogenic amino acids will probably be biosynthesized by enzymes specifically encoded in the antibiotic biosynthetic cluster. For example, 4-hydroxyphenylglycine is found at residues 4 and 5 of the vancomycin scaffold (Scheme 3) and is present in complestatin, ramiplanin, and nocardicins.[25, 26] In the cluster for the vancomycin family member chloroeremomycin (**2**) there is a four-enzyme cassette that diverts carbon flux from the shikimate pathway to convert *p*-OH-phenylpyruvate into *p*-OH-mandelate and then into *p*-OH-benzoylformate, which is transaminated to give *p*-OH-phenylglycine.[27, 28] At position 7 of the vancomycin heptapeptide is 3,5-dihydroxyphenylglycine, and its biosynthetic enzymes also appear to be committed members of the biosynthetic open reading frames (Orfs) in the Cep gene cluster (see Scheme 1 b).[12] The directed approach of providing precursors has been utilized in generating rapamycin analogues where a *RapL* knockout gene blocks the enzymatic cyclization of lysine to L-pipecolate,

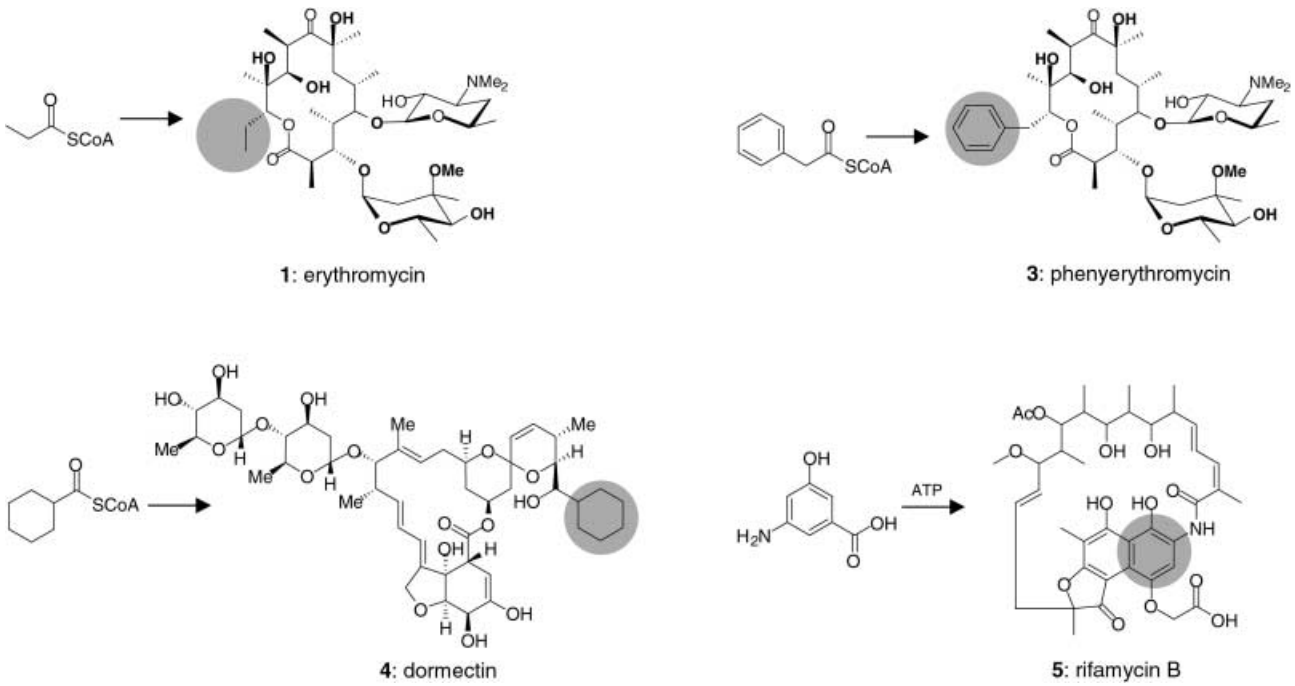


Scheme 1. *a) Biosynthesis of the polyketide erythromycin (**1**): the three-subunit DEBS assembly line and the tailoring hydroxylases and glycosyl transferases. b) Biosynthesis of chloroeremomycin (**2**), a vancomycin family member: the three-subunit Cep NRPS assembly line, oxidative cyclization, and glycosylation give the mature glycopeptide antibiotic. For abbreviations, please refer to the text.*

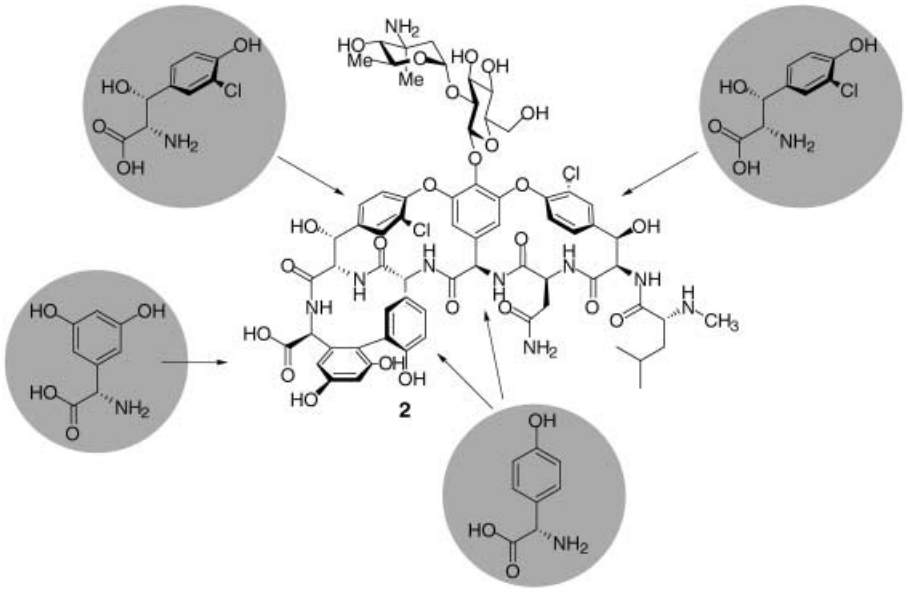
allowing analogues to compete for incorporation by the rapP enzyme.[29]

In glycosylated antibiotics, the sugars are occasionally primary metabolites such as glucose or GlcNAc, but most often they are deoxy sugars, deoxygenated at carbon atoms 2, 3, 4, or 6. For example in vancomycin, L-vancosamine is a 2,3,6-trideoxyhexose that has been methylated and aminated at C3, while D-desosamine is a 3-dimethylamino-4,6-dideoxyhexose (see compound **7** in Scheme 4). The activated donor forms of sugar substrates are NDP sugars (specifically TDP-sugars in Scheme 4; T = thymidine), and there are dedicated pathways to such NDP deoxyhexoses that involve four to six enzymes, again typically encoded within the biosynthetic clusters. The chloroeremomycin cluster has at least 29 contiguous Orfs, several of which are dedicated to providing the nonproteinogenic amino acid monomers, dTDP-epivancosamine, the NRPS assembly line subunits, the crosslinking oxidases, the glycosyl transferases, hydroxylases, and methylases.[12]

The NDP sugars cannot be fed to cells exogenously, but recent gene-knockout and gene-transfer experiments have shown that



Scheme 2. *Acyl-CoAs used for the loading domains of polyketide synthases and incorporation of the acyl units into the final products (marked with the shaded circles).*



Scheme 3. *Nonproteinogenic amino acids at positions 2, 4, 5, 6, and 7 of vancomycin.*

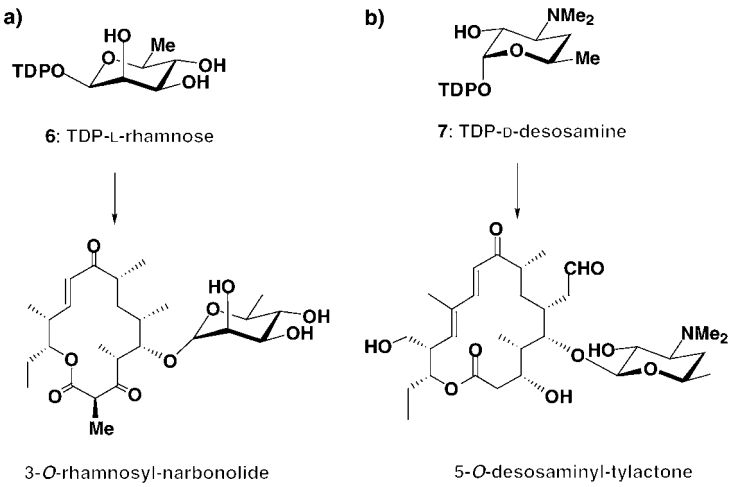
S. venezuelae can be tricked into producing dTDP-L-rhamnose **6** or dTDP-3-dimethylamino-4,6-dideoxyglucose **7** and that these sugars will be transferred by the desosaminyl transferase DesVII that glycosylates the aglycon of pikrimycin to produce hybrid glycosylated macrolides[30, 31] (Scheme 4). Analogously the genes encoding conversion of dTDP-glucose into dTDP-desosamine have been moved as a cassette and shown to be utilized with aglycons distinct from the cognate erythronolide scaffold.[32] In vitro studies do have a place in determining glycosyl transferase specificity before the pathway for sugar biosynthesis can be reengineered, and Thorson and colleagues have recently reported[33] a preparative-scale enzymatic synthesis of many TDP sugars in a one-step reaction.

4. Reengineering the Assembly Line Domains and Modules

4.1. Chain initiation

The loading modules for polyketide synthase (PKS) assembly lines typically have an inactive form of keto synthase (KS), an acyl transferase (AT) domain, and an ACP domain (see Scheme 1 a). The AT catalyzes transthiolation, for example, of propionyl-CoA to propionyl-S-ACP. The first extender module would typically have a KS-AT-ACP core set of domains for transferring in malonyl/methylmalonyl to the holoACP domain and catalyzing condensation to the β-keto-acyl-S-ACP on the extender module. The fate of the β-keto group is determined by the presence and activity of additional domains in this module, including keto reductase (KR: β-keto $\rightarrow\beta$-OH), dehydratase (DH: β-OH $\rightarrow\alpha,\beta$-ene), and enoyl reductase (ER: α,β-ene $\rightarrow\beta$-CH_2) domains. There are natural mutations that inactivate KR, DH, or ER domains in various PKS assembly lines, preserving the β-keto, β-hydroxy, or α,β-olefinic moiety for transfer to the next downstream module. (For reviews, see refs. [4, 34, 35].)

Efforts at reengineering the DEBS assembly line (see Scheme 1 a) have built on such mutational precedents to inactivate



Scheme 4. *Utilization of alternate TDP-deoxyhexoses by tailoring glycosyl transferases in macrolide maturation to create novel hybrids: a) TDP-L-rhamnose →3-O-rhamnosyl-narbonolide; b) TDP-D-desosamine →5-O-desosaminyl-tylactone.*

specific domains in particular modules to create libraries of erythronolide skeletons altered at specific β-carbon sites.[14, 15, 20, 36] When the KS domain in the first extender module is inactivated by mutation it is possible to bypass the block with an exogenous diketide thioester (an acyl-*S*-*N*-acetylcysteamine (SNAc)), for example, the natural β-hydroxy-α-methylbutyryl side chain, or with alternative diketide acyl chains to smuggle in novel starter unit monomers that are taken through the whole assembly line.[23, 36]

Manipulation of chain initiation in the NRPS assembly lines (see Scheme 1 b) has not progressed so far; most studies are still at the in vitro level to evaluate rules for the selection of amino acid monomers by the adenylation (A) domains. Two approaches with some merit for bypassing or altering selectivity have been reported. The first uses synthetic aminoacyl-CoAs[37] as substrates for the apo carrier protein forms of paired A-PCP domains and the phosphopantetheinylating enzyme sfp,[38] which is tricked into transferring not the normal HS-pantetheine-P onto the serine side chain of the apo PCP, but rather the aminoacyl-S-pantetheine-P to install any aminoacyl moiety and bypass the editing function of an A domain. The second approach has been the decoding of the specificity rules of A domains for amino acid selection.[39, 40] The X-ray crystal structure analysis of a Phe adenylation domain[41] and a bioinformatics analysis of more than 150 A-domain sequences aided the decoding of the side chains in the A domains' active sites; these are crucial for recognition of amino acid substrates. The code has been validated experimentally by mutation and alteration of the specificity of the amino acid activated.[39] In principle, one could inspect any A-domain sequence in any NRPS module, mutate bases encoding two to four residues, and change specificity at a particular site in the assembly line. Whether it will be possible to engineer an "all-purpose" A domain of relaxed specificity to create libraries of peptides with several amino acids, especially nonproteinogenic ones, fractionally incorporated in each elongation module, remains to be tested.

4.2. Chain elongation

The ability to mutationally inactivate KR, DH, and ER domains in any PKS module allows the control of oxidation state and sterochemistry (*R*- or *S*-OH, *E*- or *Z*-ene) of functionality at any β-carbon site during PKS-mediated chain elongation. This has been validated with many examples in the DEBS system.[15, 20, 23, 36] Also, analysis of the AT domain sequences in extender modules has revealed two subforms such that domain swaps will interconvert specificity for malonyl into specificity for methylmalonyl or vice versa.[4, 36, 42] Thus methyl groups can be added or deleted at any β-carbon site, and this has been validated in the erythronolide skeleton. In general the modules downstream of a point mutation seem to tolerate the altered acyl chain and carry it through to the end of the assembly line. This chemical competency has to be accompanied by kinetic competency; to date it is difficult to predict which alterations will have large or small effects on kinetic fluxes of chain elongation.

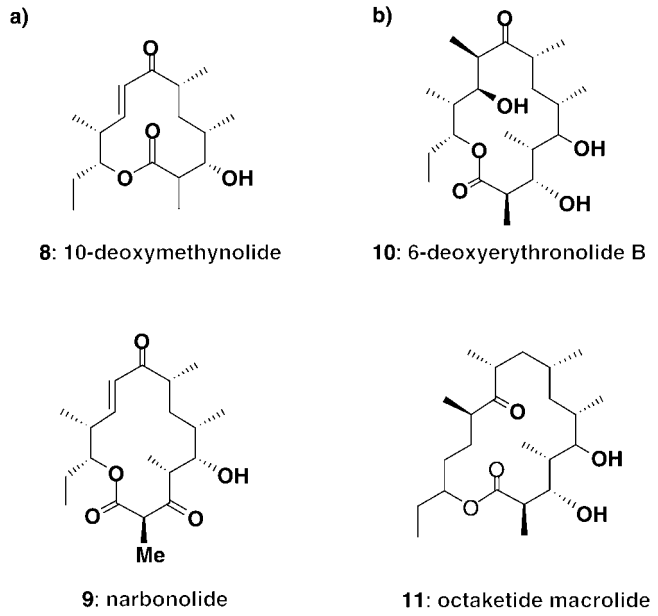
With regard to the NRPS assembly lines, work has been reported on swapping of A domains to alter the amino acid selected by a given module. The prototype has been surfactin synthetase, which makes a lipoheptapeptide that is cyclized to a macrolactone between the β-OH of the fatty acyl moiety on the NH-terminal Glu 1 and the Leu 7 carbonyl group. Replacement of Leu 7 by Cys, Phe, Orn, or Val by swapping adenylation domains from the gramicidin or ACV synthetase genes produced surfactin analogues that were detected after secretion into the culture medium.[43] Subsequent to that seminal experiment proving that NRPS domain swaps could regiospecifically reprogram an NRPS assembly line, attention has been paid to evaluating domain swaps versus swaps of a whole module, specifically, switching an A domain versus switching a C-A-PCP module, based on issues of protein – protein recognition between domains within a module and across modules, to maximize efficiency of peptide chain growth in hybrid assembly lines.[44] Studies with tyrocidine synthetase modules have demonstrated that elongation modules can be swapped and fused and the C-terminal thioesterase (TE) domain moved to create two- and three-module systems that release new di- and tripeptides catalytically. These results are presumably harbingers to reprogramming of modules in full-length NRPS assembly lines.

Several natural products of therapeutic interest are hybrid products[45, 46] of PKS and NRPS assembly lines, including the immunosuppressants rapamycin and FK506, and the antitumor antibiotics bleomycin and epothilones. The reprogramming strategies being worked out for both the PKS and NRPS elongation modules should be transferable to these mixed assembly lines.

4.3 Chain termination

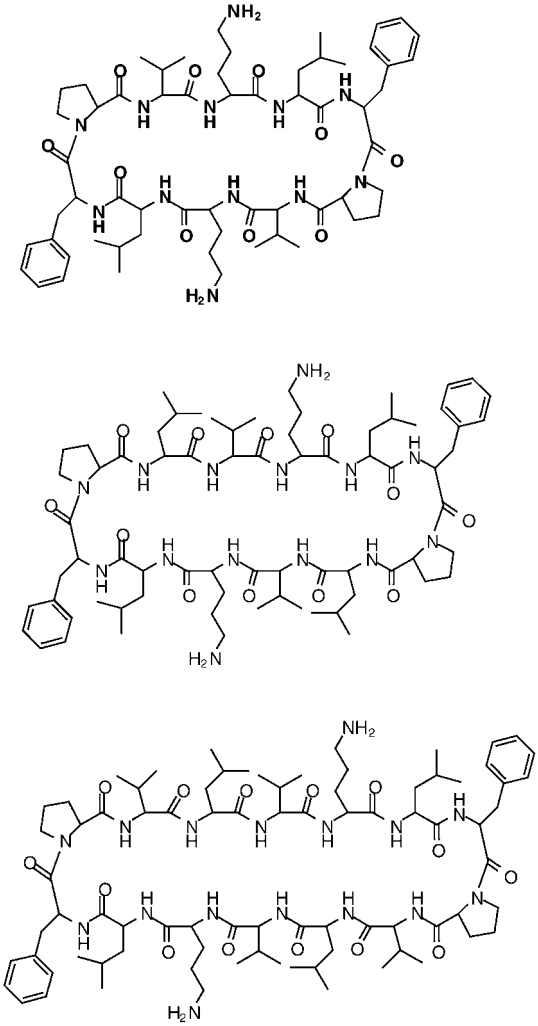
The catalytic domain for chain release and termination of the assembly line process is the TE domain closest to the C terminus of both PKS and NRPS assembly lines. Some TEs are hydrolytic, releasing the free acid, while others are cyclizing, releasing macrolactones (erythronolide) or macrolactams (for example,

bacitracin or tyrocidine[47]). Two recent studies have examined the ability of pikromycin synthase (Pik) and erythromycin synthase to produce macrolactones of altered ring size in the PKS assembly lines. In the Pik system, the PikIII and PikIV subunits contain the last two PKS modules (modules 5 and 6). If PikIV is truncated by initiation at an internal start codon, it produces only the 12-membered lactone **8** (Scheme 5 a), in which the acyl chain on PikIII is transferred to the TE domain. In the assembly line containing the native PikIV subunit some chains get cyclized from module 5, while others proceed to module 6, and the full-length 14-membered ring product **9** is released.[48, 49] In an analogy a mutant of the erythromycin-producing *Sacchropolyspora erythrea* yields minor amounts of the 16-ring lactone **11** as well as the normal 14-ring erythronolide **10**, which has been traced to a stuttering in module 4 of the DEBS assembly line; the chain transfer across the DEBS B subunit to module 5 in the DEBS C subunit must be slow enough that a second elongation cycle can occur on module 4.[50] Whether either of these ring-size-altering pathways can be optimized and controlled to yield macrocycles of different sizes at will remains to be shown. It is known that the PKS C-terminal TE domain is portable, and moving it upstream leads to catalytic release of shorter chains.[51, 52]



Scheme 5. *Alterations in the size of the macrocycle. a) Pik produces 12- and 14-membered rings (**8** and **9**, respectively) depending on the integrity of the PikIV subunit. b) With erythromycin synthase 14- and 16-membered rings (**10** and **11**, respectively) result from a stuttering in module four.*

In the NRPS systems with cyclizing TE domains no data has yet been reported on in vivo reprogramming. But recent in vitro studies with the purified, excised TE domains as isolated catalytic fragments reveal that peptidyl thioesters can be cyclized with good rates and high efficiency. The TE domain from the tyrocidine synthetase C subunit cyclizes the linear decapeptidyl-SNAc to authentic tyrocidine at a rate of 60 min^{-1} and low micromolar levels of K_m (K_m = the substrate concentration that produces half the maximal rate),[53] and does so with chiral recognition of a D-Phe_1 side chain (Scheme 6). Evaluation of hydrogen bonding in a preorganized substrate conformer has provided some of the determinants for cyclization, and it is clear that rings from 18 to 42 atoms are cyclized with about equal efficiency and that the TE domain will work on peptide libraries.[54] The TE domains excised from gramicidin synthetase and surfactin synthetase show additional properties of note:[55] the gramicidin synthetase domain will take pentapeptidyl-SNACs and dimerize and then cyclize them to yield the cyclic 10-mer antibiotic gramicidin S. The surfactin synthetase TE domain produces a macrolactone from a β-OH fatty acyl *N*-peptidyl-SNAc,[55] opening the prospect of engineering assembly lines for macrocyclization of variants of such lipopeptide antibiotics as ramoplanin[56] and daptomycin.[57]



Scheme 6. *Formation of 10-, 12-, and 14-membered rings by the tyrocidine synthetase TE domain.*

5. Tailoring Enzymes

Some tailoring enzymes are embedded as constituent catalytic domains within NRPS or PKS assembly lines, and their placement

determines where they act on the growing chain. These include *N*-methyl transferase domains, for example in cyclosporin synthetase[58] and in cognate *C*-methyl transferases, which are part of the bleomycin and lovastatin assembly lines.[59, 60] They also include epimerase domains for equilibration of L-aminoacyl-S-PCP to give D-aminoacyl-S-PCP intermediates:[61] for example the ACV tripeptide precursor to penicillins has a D-valine residue, and one of the ten domains in the ACV synthetase enzyme is an epimerase in the valine-processing third module.[62] Additionally there are heterocyclization domains, as in the second module of bacitracin synthetase,[63, 64] which cyclodehydrate the Ile-Cys-S-PCP_2 intermediate to furnish the corresponding thiazoline, which is a crucial determinant of its cation-dependent antibiotic activity.

A second set of tailoring enzymes are separate proteins that function on the mature chains released from the PKS and NRPS assembly lines. In the synthesis of the two antibiotics erythromycin and vancomycin (see Scheme 1) these include cytochrome P450 heme protein monooxygenases and glycosyl transferases. The timing of these downstream tailoring enzymes may be set by substrate recognition requirements, which may differ for each protein catalyst.

5.1. Cytochrome P450 enzymes

In the maturation of erythromycin (**1**; see Scheme 1a) the P450 enzyme EryF is known to utilize NADPH and O_2 to convert stereo- and regiospecifically deoxyerythronolide into the 6-β-OH-erythronolide B.[65, 66] This is then the substrate for successive glycosyl transfers, the first by EryBV transferring the mycarosyl sugar to the C3-OH and the second by EryCIII to introduce D-desosamine at the C4-OH.[4] Then the second P450 enzyme in the pathway, EryK, hydroxylates the doubly glycosylated macrolide to yield the 12-β-OH product, erythromycin C (**1**).[67] The promiscuity of both P450 enzymes for alternate macrolides has been tested, and a congeneric P450 in the pikromycin[68] maturation pathway examined for selectivity. In the maturation of the macrolide scaffold of the related antibiotic oleandromycin the OleP oxygenase acts to introduce an epoxide at the 6-methyl group.[69]

In the maturation of the vancomycin family members (see Scheme 1b), the biosynthetic gene cluster for chloroeremomycin[12] reveals four Orfs provisionally assigned as P450 enzymes, and all four prove to be heme proteins on purification.[70] Orf 20 is probably involved in conversion of Tyr_2 and Tyr_6 into β-OH-Tyr residues,[71] while the tandem Orfs 7, 8, and 9 may be involved in oxidative coupling of phenolic side chains to set the three crosslinks between the aryl side chains of residues 2 and 4, 4 and 6, and 5 and 7 (see Scheme 1b). If these several families of P450 enzymes can be expressed in different producer organisms, substantial diversity in oxygenation may be effected.

5.2. Glycosyl transferases

We noted above that the biosynthetic gene clusters for glycosylated polyketides and peptide antibiotics that have been sequenced contain the Orfs needed to make the particular deoxysugars as activated NDP deoxysugars on a "just in time" basis. The genes encoding the dedicated glycosyl transferases (Gtfs) are also generally found in the same gene clusters. Relaxed specificities are beginning to be reported for such Gtfs in urdamycin biosynthesis (D- and L-rhodinose in place of D-olivose),[72] for the OleG2 Gtf which can transfer a rhamnosyl group (from **6**) instead of mycarosyl to yield 3-*O*-L-rhamnosyl-narbonolide,[73] and the misglycosylation of the tylosin scaffold by the Tyl M2 Gtf using TDP-D-desosamine (**7**) to give 5-*O*-desosaminyl-tylactone[74] (Scheme 4a). The promiscuity of DesVII for glycosyl transfers to the pikromycin aglycon was noted above.[30, 31]

In the cluster for the chloroeremomycin biosynthesis, Orfs 11, 12, and 13 encode Gtfs A, B, and C (Scheme 7),[12] where GtfB transfers the glucosyl moiety from UDP-glucose onto the phenolic oxygen of the $PheGly_4$ residue[75, 76] and GtfC is an L-epivancosaminyl transferase to make the epivancosaminyl-1,2-glucose disaccharide,[76] which is appended to residue 4 of the heptapeptide scaffold. GtfA is an epivancosaminyl transferase with different regiospecificity, which uses the β-hydroxy oxygen atom of the β-OH-Tyr 6 residue as the nucleophile (Scheme 7a). The GtfB,C pair and the homologous GtfD,E pair[76] from the vancomycin biosynthetic cluster can be used with a variant aglycon, for example, the aglycon from teicoplanin **12**, and the UDP-glucose and UDP-epivancosamine NDP sugar substrates to make the novel glycopeptide antibiotic **13** (Scheme 7b), altered in three sites from natural teicoplanin: the peptide scaffold and both hexoses.

6. Challenges to Going Fully Combinatorial in Polyketide and Polypeptide Antibiotic Biosynthesis

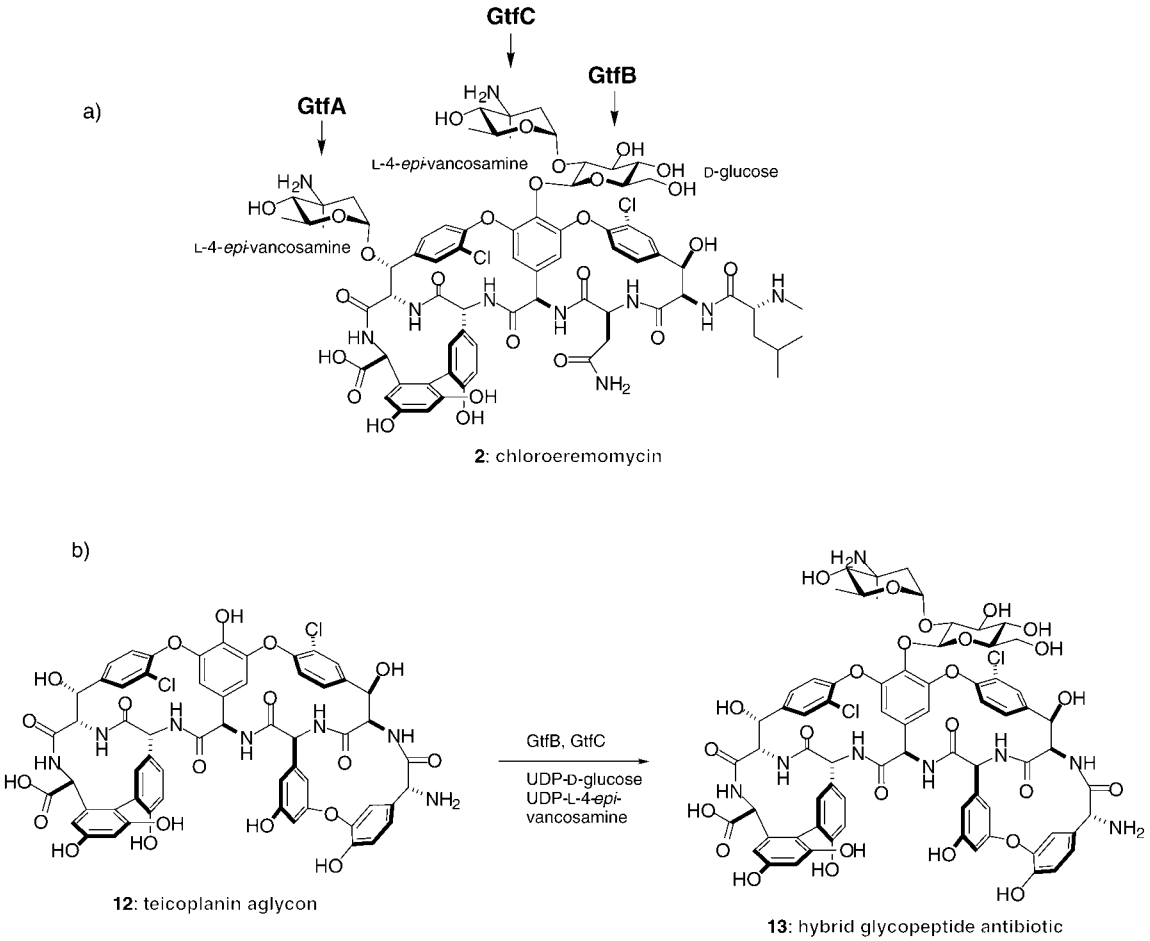
We have noted the progress that has been made in strategies for provision of new monomers in vivo to microbial cells producing polyketide, nonribosomal peptide, and hybrid PK/NRP antibiotics. For polyketides emphasis has been on providing acids as precursors of acyl-CoAs as alternate substrates for loading domains/initiation modules. For the NRPS assembly lines, non-proteinogenic amino acids may be provided to any module: initiation, elongation, or termination. It is probable that metabolic pathways for such unusual amino acids can be constructed and moved successfully into antibiotic producers. But this combinatorial approach is likely to be low throughput, since each new monomer requires its own biosynthetic pathway. For antibiotics requiring attached deoxysugar units for biological activity, while the biosynthetic pathways can be lengthy, there is now enough knowledge that specific enzymes can be replaced or deleted even in vivo to alter the outcomes predictably at carbon atoms 2–6 of the sugar backbone.

The ability to provide new sets of monomers to be selected and incorporated by the catalytic domains of the antibiotic assembly lines must be coupled to reprogramming of the specificity of the constituent enzyme modules to accept the new substrates. This can be done both by domain swaps, as in the malonyl $\Leftrightarrow$ methylmalonyl changes that have been introduced module by module,[4] or by point mutations where A-domain

selectivity and editing preferences[39] can be replaced. Further refinement in understanding the A-domain codes for the selection of amino acids to be incorporated will increase the efficiency of utilizing any desired amino acid at any point in the growing chain of the nonribosomal peptide. The corresponding reprogramming of the elongation module in PKS assembly lines is well advanced, and manipulations of the AT, KR, DH, and ER domains have been successfully conducted in almost all of the six extender modules of DEBS. The combinatorial aspect has been addressed by splitting the assembly line into pieces, for example by conveniently dispersing each of the three DEBS1 – 3 subunits (Scheme 1 a) onto three plasmids, such that a single mutation in a different domain of the three subunits can give multiplicatively a triple structural change in the final macrolide.[15] Questions to be resolved are how well tolerated are particular classes of structural mutations in the elongating chain, both in the module producing the structural alteration and by the downstream modules; this affects fluxes and product yields. Detailed understanding of the catalytic efficiency parameters of the enzyme domains in the assembly lines is yet to be extracted.

More generally one may wish to split subunits further into modules or constituent domains and replace, for example, the A domain of the fifth module of the chloroeremomycin synthetase assembly line (see Scheme 1 b) with any of ten A domains of altered specificity to give a tenfold variation at position five in the growing chain. It is unknown as yet how much protein – protein recognition and specificity exists between the C and A and the A and PCP domains in a C-A-PCP module; this is of great consequence for yields and fluxes of hybrids. Also it is not yet known which domains and modules must be in the same polypeptide to retain high efficiency in transfer of the elongating PK or NRP chains, and which domains can be on separate proteins. The more proteins of the assembly line one can set up as separate protein components, the easier combinatorial biosynthesis will be. In multisubunit assembly lines as shown in Scheme 1, the intersubunit chain transfers are separated naturally, and these are good starting points for libraries of cognate protein domains. On the other hand, there is evidence that rather than domain swaps, modules (for example, C-A-PCP or KS-AT-KR-ACP)[45, 77] may be better elements to replace, shuffle, or combinatorialize and still retain enough protein – protein recognition for hybrid assembly lines to reform and unnatural acyl chains to be elongated. In the PKS assembly lines evidence is now accruing that intermodule linkers determine recognition of partner modules and that these linkers are portable.[78]

Many of these antibiotic classes gain activity only when the aglycon product that has been released from the assembly line



Scheme 7. *a) Glycosyl transferases GtfA, B, and C in the maturation of the vancomycin family member chloroeremomycin (**2**). GtfA and C are regiospecific L-epivancosaminyl transferases; GtfB is a glucosyl transferase. b) Utilization of teicoplanin aglycon **12** and GtfB and C to create the hybrid glycopeptide antibiotic **13**.*

undergoes maturation by sets of dedicated tailoring enzymes, such as the variety of P450 hemeprotein monoxygenases, and the NDP deoxysugar glycosyl transferases (Gtfs) noted above. Combinatorial strategies that either replace the P450 and/or Gtf enzymes or alter them to relaxed specificity may be the quickest route to biologically generated sets of new hybrid antibiotics. Finally, the regiospecific acylation enzymes of such lipid-containing antibiotics as teicoplanin, surfactin, ramiplanin, and daptomycin have not been substantially investigated. They may provide another layer of combinatorial enzymatic modifications that will be useful in diversity and generation of compounds with novel activity, especially if such acyl-CoA-utilizing enzymes could be reprogrammed to use medicinally favored aryl and heteroaryl acid analogues[8] of the physiologic straight-chain fatty acids.

[1] H. S. Gold, R. C. Moellering, *N. Engl. J. Med.* **1996**, *335*, 1448 – 1453.
[2] D. Cane, C. Walsh, C. Khosla, *Science* **1998**, *282*, 63 – 68.
[3] B. Shen, *Top. Curr. Chem.* **2000**, *209*, 1 – 51.
[4] L. Katz, *Chem. Rev.* **1997**, *97*, 2557 – 2576.
[5] M. A. Marahiel, T. Stachelhaus, H. A. Mootz, *Chem. Rev.* **1997**, *97*, 2651 – 2674.
[6] H. von Döhren, U. Keller, J. Vater, R. Zocher, *Chem. Rev.* **1997**, *97*, 2675 – 2705.
[7] R. Nagarajan, A. A. Schabel, J. L. Occolowitz, F. T. Counter, J. L. Ott, A. M. Felty-Duckworth, *J. Antibiot.* **1989**, *42*, 63 – 72.
[8] R. Kerns, S. D. Dong, S. Fukuzaka, J. Carbeck, J. Kohler, L. Silver, D. Kahne, *J. Am. Chem. Soc.* **2000**, *122*, 12 608 – 12 609.
[9] S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991**, *252*, 675 – 679.
[10] J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **1990**, *348*, 176 – 178.
[11] For a review, see: K. C. Nicolaou, C. N. C. Boddy, S. F. Brase, N. Winssinger, *Angew. Chem.* **1999**, *111*, 2230 – 2287; *Angew. Chem. Int. Ed.* **1999**, *38*, 2096 – 2152.
[12] A. M. Wageningen, P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. J. M. Jones, P. J. Solenberg, *Chem. Biol.* **1998**, *5*, 155 – 162.
[13] C. M. Kao, L. Katz, C. Khosla, *Science* **1994**, *265*, 509 – 512.
[14] Q. Xue, G. Ashley, C. R. Hutchinson, D. V. Santi, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11 740 – 11 745.
[15] R. McDaniel, A. Thamchaipenet, C. Gustaffson, H. Fu, M. C. Betlach, M. Betlach, G. Ashley, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1846 – 1851.
[16] H.-w. Liu, J. Thorson, *Annu. Rev. Microbiol.* **1994**, *48*, 223 – 256.
[17] A. Trefzer, J. Salas, A. Bechtold, *Nat. Prod. Rep.* **1999**, *16*, 283 – 299.
[18] T. A. Cropp, D. J. Wilson, K. A. Reynolds, *Nature Biotechnol.* **2000**, *18*, 980 – 983.
[19] M. G. Anderson, D. Monypenny, R. W. Rickards, J. M. Rothchild, *J. Chem. Soc. Chem. Commun.* **1989**, 311 – 313.
[20] A. F. Marsden, B. Wilkinson, J. Cortes, N. J. Dunster, J. Staunton, P. F. Leadlay, *Science* **1998**, *279*, 199 – 202.
[21] S. J. Admiraal, C. Walsh, C. Khosla, *Biochemistry* **2001**, *40*, 6116 – 6123.
[22] H. A. McArthur in *Developments in Industrial Microbiology—BMP'97* (Eds.: C. R. Hutchinson, J. McAlpine), Society for Industrial Microbiology, Fairfax, VA, **1998**, pp. 43 – 48.
[23] J. R. Jacobson, C. R. Hutchinson, D. E. Cane, C. Khosla, *Science* **1997**, *277*, 367 – 369.
[24] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790 – 1792.
[25] G. M. Sheldrick, P. G. Jones, O. Kennard, D. H. Williams, G. A. Smith, *Nature* **1978**, *271*, 223 – 225.
[26] R. Ciabatti, J. K. Kettering, G. Winters, G. Tuan, L. Zerilli, B. Cavalleri, *J. Antibiot.* **1989**, *42*, 236 – 241.
[27] B. K. Hubbard, M. G. Thomas, C. T. Walsh, *Chem. Biol.* **2000**, *7*, 931 – 942.
[28] O. W. Choroba, D. H. Williams, J. B. Spencer, *J. Am. Chem. Soc.* **2000**, *122*, 5389 – 5390.
[29] L. E. Khaw, G. A. Bohm, S. Metcalfe, J. Staunton, P. F. Leadlay, *J. Bacteriol.* **1998**, *180*, 809 – 814.
[30] H. Yamase, L. Zhao, H.-w. Liu, *J. Am. Chem. Soc.* **2000**, *122*, 12 397 – 12 398.
[31] L. Zhao, J. Ahlert, Y. Xue, J. Thorson, D. H. Sherman, H.-w. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 9881 – 9882.
[32] L. R. Tang, R. McDaniel, *Chem. Biol.* **2001**, *8*, 547 – 555.
[33] J. Jiang, J. B. Biggins, J. Thorson, *J. Am. Chem. Soc.* **2000**, *122*, 6803 – 6804.
[34] C. Khosla, *Chem. Rev.* **1997**, *97*, 2577 – 2590.
[35] J. Staunton, B. Wilkinson, *Chem. Rev.* **1997**, *97*, 2611 – 2630.
[36] L. Katz, R. McDaniel, *Med. Res. Rev.* **1999**, *19*, 543 – 558.
[37] P. Belshaw, C. Walsh, T. Stachelhaus, *Science* **1999**, *284*, 486 – 489.
[38] R. Lambalot, A. Gehring, R. Flugel, P. Zuber, M. Lacelle, M. A. Marahiel, R. Reed, C. Khosla, C. Walsh, *Chem. Biol.* **1996**, *3*, 923 – 936.
[39] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493 – 505.
[40] G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211 – 224.
[41] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174 – 4183.
[42] X. Ruan, D. L. Stassi, A. Pereda, D. Zeidner, A. Shvakumar, S. Kakavas, M. Jackson, R. G. Summers, L. Katz, *J. Bacteriol.* **1997**, *179*, 6416 – 6425.
[43] T. Stachelhaus, A. Schneider, M. A. Marahiel, *Science* **1995**, *269*, 69 – 72.
[44] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5848 – 5853.
[45] D. E. Cane, C. Walsh, *Chem. Biol.* **1999**, *6*, R319 – 325.
[46] L. Du, C. Sanchez, B. Shen, *Metab. Eng.* **2001**, *3*, 78 – 95.
[47] T. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger, C. Walsh, *ChemBioChem* **2001**, *2*, 101 – 109.
[48] Y. Xue, D. Wilson, D. H. Sherman, *Gene* **2000**, *245*, 203 – 211.
[49] Y. Xue, D. H. Sherman, *Nature* **2000**, *403*, 571 – 574.
[50] B. Wilkinson, G. Foster, B. A. M. Rudd, N. L. Taylor, A. P. Blackaby, P. J. Sidebottom, D. J. Cooper, M. J. Dawson, A. D. Buss, S. Gaisser, I. U. Bohm, C. J. Rowe, J. Cortes, P. F. Leadlay, J. Staunton, *Chem. Biol.* **2000**, *7*, 111 – 117.
[51] J. Cortes, K. E. H. Weismann, G. A. Roberts, M. J. B. Brown, J. Staunton, P. F. Leadlay, *Science* **1995**, *268*, 1487 – 1489.
[52] C. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **1997**, *119*, 11 339 – 11 340.
[53] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. Walsh, *Nature* **2000**, *407*, 215 – 217.
[54] J. W. Trauger, R. M. Kohli, C. Walsh, *Biochemistry* **2001**, *40*, 7092 – 7098.
[55] R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel, C. Walsh, *Biochemistry* **2001**, *40*, 7099 – 7108.
[56] M. Kurz, W. Gurba, *Biochemistry* **1996**, *35*, 12 570 – 12 575.
[57] H. van Dohren in *Genetics and Biochemistry of Antibiotic Production* (Eds.: L. C. Vining, C. Stuttard), Butterworth – Heinemann, Newton, MA, **1995**, pp. 129 – 171.
[58] G. Weber, K. Schorgendorfer, E. Schneider-Scherzer, E. Leitner, *Curr. Genet.* **1994**, *26*, 120 – 125.
[59] L. Du, C. Sanchez, M. Chen, D. J. Edwards, B. Shen, *Chem. Biol.* **2000**, *7*, 623 – 642.
[60] L. Hendrickson, C. R. Davis, C. Roach, D. K. Nguyen, T. Aldrich, P. C. McAda, C. D. Reeves, *Chem. Biol.* **1999**, *6*, 429 – 439.
[61] T. Stachelhaus, C. Walsh, *Biochemistry* **2000**, *39*, 5775 – 5787.
[62] D. J. Smith, J. E. Alison, G. Turner, *EMBO J.* **1990**, *9*, 2743 – 2750.
[63] D. Konz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, R39 – R48.
[64] T. Keating, C. Walsh, *Curr. Opin. Chem. Biol.* **1999**, *3*, 598 – 606.
[65] J. M. Weber, J. O. Leung, S. J. Swanson, K. B. Idler, J. B. McAlpine, *Science* **1991**, *252*, 114.
[66] J. F. Anderson, C. R. Hutchinson, *J. Bacteriol.* **1992**, *174*, 725 – 735.
[67] D. L. Stassi, S. Donadio, M. J. Staver, L. Katz, *J. Bacteriol.* **1993**, *175*, 182 – 189.
[68] M. C. Betlach, J. T. Kealy, M. Betlach, G. W. Ashley, R. McDaniel, *Biochemistry* **1998**, *37*, 14 937 – 14 942.
[69] S. Shah, Q. Xue, L. Tang, J. R. Carney, M. Betlach, R. McDaniel, *J. Antibiot.* **2000**, *53*, 502 – 508.
[70] B. Hubbard, C. Walsh, unpublished results.

[71] H.-w. Chen, C. Walsh, *Chem. Biol*. **2001**, *8*, 301 – 312.
[72] D. Hoffmeister, K. Ichinose, S. Domann, B. Faust, A. Trefzer, G. Drager, A. Kirschning, C. Fischer, E. Kunzel, D. W. Bearden, J. Rohr, A. Bechtold, *Chem. Biol*. **2000**, *7*, 821 – 831.
[73] M. Doumith, R. Legrand, C. Lang, J. Salas, M. C. Raynal, *Mol. Microbiol*. **1999**, *34*, 1039 – 1048.
[74] S. Gaisser, J. Reather, G. Wirtz, L. Kellenberger, J. Staunton, P. F. Leadlay, *Mol. Microbiol.* **2000**, *36*, 391 – 401.
[75] P. J. Solenberg, P. Matsushima, D. R. Stack, S. C. Wilkie, R. C. Thompson, R. H. Baltz, *Chem. Biol.* **1997**, *4*, 195 – 202.
[76] H. Losey, M. Peczuh, Z. Chen, U. S. Eggert, S. D. Dong, I. Pelczer, D. Kahne, C. Walsh, *Biochemistry* **2001**, *40*, 4745 – 4755.
[77] R. S. Gokhale, S. Y. Tsuji, D. E. Cane, C. Khosla, *Science* **1999**, *284*, 482 – 484.
[78] S. Y. Tsuji, D. E. Cane, C. Khosla, *Biochemistry* **2001**, *40*, 2326 – 2331.

Received: November 19, 2001 [A 322]