CHAPTER FOURTEEN

Design and Synthesis of Pathway Genes for Polyketide Biosynthesis

Salvador Peirú, Hugo Gramajo, *and* Hugo G. Menzella

Contents

Abstract

In this chapter we describe novel methods for the design and assembly of synthetic pathways for the synthesis of polyketides and tailoring sugars. First, a generic design for type I polyketide synthase genes is presented that allows their facile assembly for the expression of chimeric enzymes in an engineered *Escherichia coli* host. The sequences of the synthetic genes are based on naturally occurring polyketide synthase genes but they are redesigned by custom-made software to optimize codon usage to maximize expression in *E. coli* and to provide a standard set of restriction sites to allow combinatorial assembly into unnatural enzymes. The methodology has been validated by building a large number of bimodular mini-PKSs that make easily assayed triketide products. Learning from the successful bimodules, a conceptual advance was made by assembling genes encoding functional trimodular enzymes, capable of making tetraketide products. Second, methods for the rapid assembly and exchange of sugar pathway genes into functional operons

Instituto de Biología Molecular y Celular de Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

Methods in Enzymology, Volume 459
ISSN 0076-6879, DOI: 10.1016/S0076-6879(09)04614-X
© 2009 Elsevier Inc.
All rights reserved.

are described. The approach was validated by the assembly of the 15 genes for the synthesis of mycarose and desosamine in two operons, which yielded erythromycin C when coexpressed with the corresponding PKS genes.

These methods are important enabling steps toward the goals of making designer drugs by polyketide synthase and sugar pathway engineering and, in the shorter term, producing by fermentation advanced intermediates for the synthesis of compounds that otherwise require large numbers of chemical steps.

1. Introduction

Type I polyketide synthases (PKSs) that determine the biosynthesis of valuable natural products like erythromycin, epothilone, tacrolimus and many others consist of multiple catalytic domains forming a protein "assembly line." The domains are associated in modules that determine the structure of a 2-carbon extender unit of the polyketide chain. Thus, the sequence of domains in the assembly line, directly reflecting the DNA sequence of the corresponding genes, determines the structure of the polyketide product. This simple logic led to genetic engineering of PKSs to make "unnatural natural products" (Walsh, 2003, 2004; Weissman and Leadlay, 2005) as shown in the early 1990s by research groups at Stanford and Cambridge (Cortez *et al.*, 1995; Kao *et al.*, 1994). They demonstrated that domains and modules could be reorganized to create novel enzymes capable of making a variety of unnatural compounds.

The initial approaches in the field of genetic engineering of polyketides included adding, deleting or interchanging domains or whole modules in a PKS aimed to alter the number and structure of the 2-C units built into a polyketide. Whereas these might appear simple tasks for current genetic engineering technologies, they are encumbered by difficulties, and PKS engineering is far from being established in the high-throughput format required to provide the necessary flow of unnatural compounds in order to obtain "hits" with new biological properties.

One problem is that most natural polyketide producers are not readily amenable to a complete range of genetic methodologies without considerable time-consuming effort. Most of them are slow-growing compared with commonly used laboratory strains, often requiring more than a month for a round of genetic manipulation, and many target organisms are so far uncultured. Therefore, there is a need for a "universal host" in which PKS components from different sources can be expressed and where combinatorial experiments can be performed rapidly and ideally in a high-throughput fashion (Pfeifer and Khosla, 2001).

A second problem is that natural PKS gene clusters are extraordinarily long (up to 200 kb), usually with a high G+C content and with highly repetitive coding sequences for successive modules. Rarely are there conserved restriction sites to facilitate domain or module exchange. Thus, there is a need for molecular biological tools to accelerate the assembly of chimeric PKS genes in order to find productive combinations.

Expressing PKS genes in a more tractable host can circumvent some of these difficulties. Recently, *Escherichia coli* has been successfully engineered to produce polyketides (Murli *et al.*, 2003; Pfeifer and Khosla, 2001). Since its genetic manipulation is so straightforward, it was chosen as a host for developing new methods for combinatorial biosynthesis of polyketides in the work described here.

To avoid the problems of manipulating natural PKS genes, a first step was to develop a technology for facile synthesis of error-free ~5 kb sequences of DNA—the average for complete PKS modules (Kodumal *et al.*, 2004; Reisinger *et al.*, 2006). Next, a generic design of PKS modules was created with unique restriction sites flanking their components to allow exchange as building blocks in an approach that has been dubbed "Legoization" (Menzella *et al.*, 2005; Sherman, 2005). Finally, methods were developed to rapidly assemble and express simple chimeric PKSs and to screen for their activity in order to identify productive combinations to be used in the construction of larger, designer enzymes (Menzella *et al.*, 2005; Menzella and Reeves, 2007).

Another important aspect to be considered during polyketide engineering is that, in order to be fully active, many of these molecules require the attachment of sugar moieties on their core structure. The presence of these sugars, as well as other post-PKS modifications, is often essential to impart or enhance the biological activity of the molecule. Thus, heterologous expression of PKSs also requires the incorporation of complete glycosylation pathways into the host, to achieve the synthesis of a fully decorated molecule. The possibility of altering the glycosylation pattern of polyketides through genetic engineering is another promising tool for diversification of these compounds.

Sugar biosynthesis genes are normally located within the polyketide gene cluster (Peirú *et al.*, 2005). The heterologous expression of these sugar pathways typically requires the cloning of individual genes and their assembly into operons in order to have a controlled expression system under inducible promoters. As with PKS modules, several issues need to be considered for the construction of sugar operons.

Here we describe practical experience with these genetic engineering tools. The ultimate goal is to produce complex molecules by creating multidomain PKSs combined with sugar tailoring genes to be used directly as drugs or as lead compounds for chemical optimization. Meanwhile, at a more modest level, even the combination of a few PKS modules can

produce molecules with multiple chiral centers (up to two per module) that are difficult to obtain by ordinary chemical synthesis (McDaniel *et al.*, 2005). The production of such molecules made by fermentation to be used as intermediates for the synthesis of complex structures can avoid large numbers of chemical steps and significantly reduce manufacturing costs. Therefore, engineering synthetic polyketide synthases is a promising strategy to produce novel molecules for multiple applications.

2. Redesign of PKS Genes to Accommodate Unique Restriction Sites Flanking Individual Components and for Efficient Expression in *E. coli*

The strategy involves the synthesis of redesigned genes containing a set of standard unique restriction sites at the borders of each PKS component, including domains, linkers and modules, to allow their facile interchange. To allocate these restriction sites, sequence conservation at domain edges was analyzed (Fig. 14.1) by aligning the amino acid sequences of more than 300 PKS modules. When the identified conserved sequences were reverse-translated to all possible DNA sequences, a conserved six-base-pair restriction site could be found in some cases. In modules where that sequence was not present, amino acids were changed to a sequence resembling one present in at least some PKSs or which was already known, from prior experimental evidence, not to disturb functionality.

For the ketosynthase (KS) domain, an *Mfe*I site was incorporated near the 5′ edge by using bases 2 to 7 of nine bases coding for the amino acid sequence PIA at the beginning of this domain. Of the KS domains analyzed, 70% needed no change in amino acid sequence, and 30% required only conservative changes from V, L, or M. On the 3′ edge of all KSs, the conserved GT can be encoded by the sequence for a *Kpn*I site.

At the 3′ side of the acyl transferase (AT) domain, a *Pst*I site was placed where *Pst*I and *Xho*I have been used successfully by others (McDaniel *et al.*, 1999; Oliynyk *et al.*, 1996). A highly conserved sequence occurs in 90% of the modules as R-Y/F-W. The codons specifying the two amino acids C-terminal to the W of this motif were modified to introduce the *Pst*I site coding for LQ. The dipeptide at this position is variable in the native sequence. However, in more than half the cases the first amino acid is L, and in 80% of these the second is hydrophilic or even charged. Thus, the LQ dipeptide was likely to be well tolerated.

No constant sites were found allowing independent exchange of the reductive domains ketoreductase (KR), dehydratase (DH), enoyl reductase (ER). For modules containing a KR domain, an *Age*I site could be

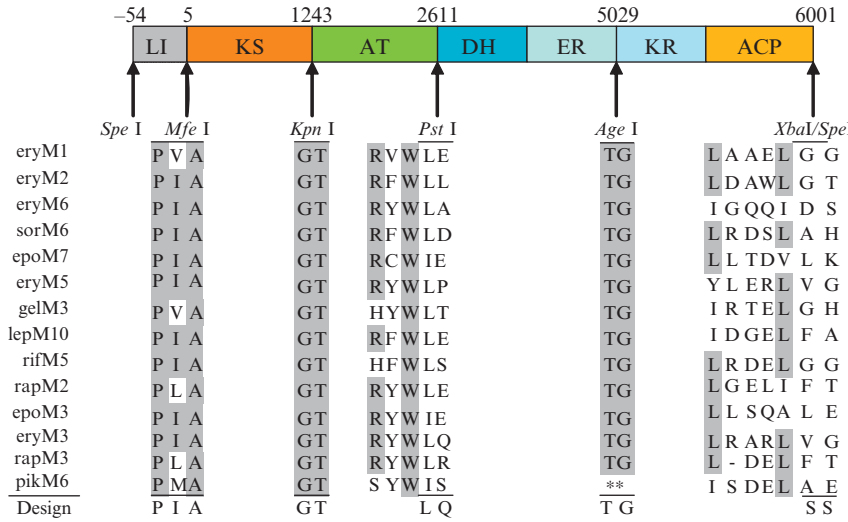


Figure 14.1 Conserved amino acid sequences at domain borders and the chosen design to generate the desired restriction sites for domain assembly. Restriction site positions are referenced to module 4 of DEBS. Numbering begins at the first residue of the motif EPIAIV on the N-terminal edge of the KS domain. Sequence homologues for 14 of the 140 modules used for the generic module design are shown. eryM3 and rapM3 contain an inactive KR domain. ery, erythromycin; sor, soraphen; epo, epothilone; gel, geldanamycin; lep, leptomycin; rif, rifamycin; rap, rapamycin; pik, pikromycin. (From Menzella *et al.*, 2005. Reproduced with permission from Nature Journals.)

incorporated at its 5′ edge to encode the TG dipeptide found in 132 of 137 native KR domains analyzed, while in the remaining five this domain was predicted to be inactive. When required, DH/ER pairs could be added, deleted, or exchanged using the *Pst*I and *Age*I sites.

Finally, since the 5′ boundary of the acyl carrier protein (ACP) domain is poorly defined, no unique restriction site could be placed there, so domain exchange required the use of an upstream unique site. Except for the minimal module containing only AT, KS, and ACP, this required comobilization of the reductive domains. At the 3′ end of the module, an *Xba*I site was incorporated at a well-defined position next to the carboxy side of the ACP. There are two conserved L residues at positions 36 and 40 downstream of the active site serine of ACPs. The codons of the two amino acids at positions 41 and 42 were changed to the *Xba*I sequence coding for SS, already known not to affect functionality (Gokhale *et al.*, 1999).

To facilitate gene expression in *E. coli*, all the PKS genes were codon-optimized. For the gene design, a software package called GeMS (Jayaraj *et al.*, 2005) was developed, comprising a set of programs to predict restriction sites, perform codon optimization for expression in any host, assign or

eliminate restriction sites, separate long sequences into smaller fragments to facilitate synthesis, carry out Tm and stem-loop determinations, and design oligonucleotides to be used as raw material. GeMS accepts DNA or protein sequences and its output is a complete design report with a list of oligonucleotides to be used in the gene synthesis process.

Using GeMS a complete collection of PKS components was designed and later synthesized, including 50 extender modules, 5 loading modules, 5 thioesterases, 30 intrapeptide linkers, and 20 pairs of docking domains, totaling ~400 kb of synthetic DNA. All the building blocks contain the standard set of restriction sites to facilitate the assembly of hybrid PKSs. For codon optimization, a codon table was used representing the *E. coli* genome. For each amino acid in the target protein, the software randomly selects a codon in accordance with its distribution in the table. The optimized genes thus contain codons in the same proportion as in *E. coli*. Although all the gene design work described here was made using GeMS, many other free web-based programs like, for example, Gene Designer (http://www.dna20.com/tools/genedesigner) and Optimizer (http://genomes.urv.es/OPTIMIZER) may serve well for this purpose.

3. Validation of Synthetic PKS Gene Design

In the first attempt to validate the generic PKS gene design, the 31-kb gene cluster encoding the 6-deoxyerythronolide B (6dEB) synthase (DEBS) was assembled and expressed in an *E. coli* strain engineered for polyketide production (Menzella *et al.*, 2006). In the redesigned cluster the total set of 15 restriction sites were accommodated flanking the individual domains (Fig. 14.1). For this, 20 amino acid substitutions were required to insert some of the *Mfe*I and *Pst*I sites and all of the *Spe*I/*Xba*I fusions. The PIA tripeptide needed for the *Mfe*I site is naturally present in five of the six DEBS modules, and module 1 (Mod1) has a conserved PVA. The LQ dipeptide required for the *Pst*I site is found naturally only in DEBS Mod3; four modules have L followed by E (Mod1), L (Mod2), P (Mod5) or A (Mod6), and Mod4 has an unusual PR sequence found in only 2% of PKS modules. The *Spe*I/*Xba*I fusion used to connect all redesigned modules results in an SS dipeptide that is unnatural to all modules. Overall, in the final redesigned DEBS, 13 of the 20 amino acid changes accommodated the *Spe*I/*Xba*I junction, one accommodated the *Mfe*I site of Mod1 and six accommodated the *Pst*I sites in Mod 1, 2, 4, 5, and 6. For the rapid assembly of building blocks to create chimeric open reading frames a suitable method is described below.

In the described gene design, NdeI and XbaI sites flank the coding region of the N-terminal PKS components (LM and docking domains); the coding regions of internal modules are flanked by SpeI and XbaI sites;

and the coding region of C-terminal components (docking domains and TE) are flanked by SpeI and EcoRI sites. All the PKS components were synthesized and cloned into pUC-based vectors containing an additional NotI site upstream of the coding region.

1. Create a multicopy vector with a polylinker containing NdeI-NotI-SpeI-EcoRI sites.
2. Initiate the assembly by inserting the C-terminal component of the ORF into the SpeI-EcoRI sites of the vector.
3. Clone the internal components sequentially as NcoI-XbaI fragments into the NcoI-SpeI sites of the vector. This SpeI/XbaI ligation destroys both sites, allowing the repetitive use of these enzymes to insert as many components as required.
4. Finally, clone the N-terminal component as an NdeI-XbaI fragment into the NdeI-SpeI sites to complete the assembly of the ORFs, which can be mobilized as NdeI-EcoRI fragments into expression vectors or used to assemble operons.

To determine whether these amino acid changes introduced to accommodate the standard set of restriction sites affected function of the enzymes, protein expression levels of the three synthetic DEBS ORFs as well as 6dEB production were compared with those of the native sequences using the following method:

1. Introduce expression vectors for synthetic PKSs by co-transformation into *E. coli* K207-3[BL21ΔprpBCD::T7prom prpE, T7prom accA1-pccB, T7prom sfp]. This strain is a derivative of BL21(DE3) with the propionate utilization operon (prpRBCD) and the methylmalonyl-CoA decarboxylase gene (ygfG) deleted and with chromosomal copies of sfp (phosphopantetheine transferase from *Bacillus subtilis*), prpE (endogenous propionyl-CoA ligase), and accA1/pccB (propionyl-CoA carboxylase from *Streptomyces coelicolor*) under the control of T7 promoters (Murli *et al.*, 2003). Feeding propionate to this strain results in the accumulation of methylmalonyl-CoA, a polyketide synthase substrate produced by wildtype *E. coli* at very low levels.
2. Grow the resulting colonies in 3 ml LB containing appropriate antibiotics for selection at 37° to an $OD_{600} = 0.5$.
3. Induce with IPTG (0.5 m*M*), and a mixture of sodium glutamate (50 m*M*), sodium succinate (50 m*M*), and sodium propionate (5 m*M*).
4. Incubate the induced cultures at 22° for 72 h with agitation.
5. Centrifuge 10 min at 10,000 g; remove the supernatant for polyketide analysis and cell pellet for protein expression analysis.
6. Extract cell-free supernatants (1 ml) twice with an equal volume of ethyl acetate. Dry the extracts *in vacuo* and dissolve in 1 ml of methanol. Analyze for 6dEB production as described in Menzella *et al.* (2006).

7. Re-suspend cell pellets in 1 ml 20-m*M* Tris, 150 m*M* NaCl, pH 7.5, containing 1 tablet of Complete^{TM} EDTA-free protease inhibitor cocktail (Roche) per 50 ml.
8. Disrupt by sonication, and centrifuge at 14,000×*g* for 3 min.
9. Analyze soluble and insoluble fractions on NuPAGE Novex 3 to 8% Tris-Acetate gels (Invitrogen). Gels can be stained using Simply Blue SafeStain (Invitrogen) and analyzed using a UVP Bioimaging System with Labworks 4.0 software (UVP, Inc.).

The synthetic components of DEBS ORFs were expressed using three compatible vectors (ColE1 for DEBS1, P15A for DEBS2, and RSF1010 for DEBS3), all under control of a T7 promoter. When individually expressed in the engineered *E. coli* strain K207-3, each codon-optimized ORF produced about fivefold more protein than its native counterpart, as did the complete set of three coexpressed synthetic ORFs. However, no polyketide product could be detected from the three coexpressed synthetic ORFs, while the wildtype PKS gene set produced 20 mg/l of 6dEB in shake-flask fermentations.

To identify the defective module(s), chimeras of wildtype and synthetic ORFs were constructed by cotransformation of combinations using the three-plasmid system. Of the three possible chimeras containing one synthetic and two natural ORFs, only the combination containing synthetic DEBS2 produced no 6dEB. Since the synthetic DEBS2 contained the nonconserved PR-to-LQ mutation to accommodate the Mod4 *Pst*I site, we converted the sequence back to PR. With this change, synthetic DEBS2 combined with DEBS1 and DEBS3 produced 6dEB, but at a level far below that obtained with the wildtype sequences. Further studies showed that expression of synthetic DEBS from the strong T7 promoters led to a significant reduction in the expression of the accessory genes *accA* and *pccB* required to produce methylmalonyl CoA in the *E. coli* K207-3 strain, and this could presumably account for the low levels of 6dEB observed. When expression of the three synthetic DEBS genes was driven by the *lac*UV5 promoter, three- to four-fold less PKS protein was observed than when expressed under control of the T7 promoter, but both the production levels of 6dEB and the expression of the *accA* and *pccB* genes were comparable to those observed with the wildtype sequences.

In another validation example, the entire synthetic epothilone gene cluster—*epoA, epoB, epoC, epoD*, *epoE*, and *epoF*—and the C-terminal thioesterase were assembled and expressed in the *E. coli* K207-3 strain (Mutka *et al.*, 2006). The expected products, epothilone C and D, were obtained, indicating that the 29 amino acid changes needed to insert the restriction sites into the redesigned genes could be tolerated to obtain a catalytically active PKS complex of approximately 2 MDa (monomeric mass). As in the case of DEBS, titers were improved when gene expression

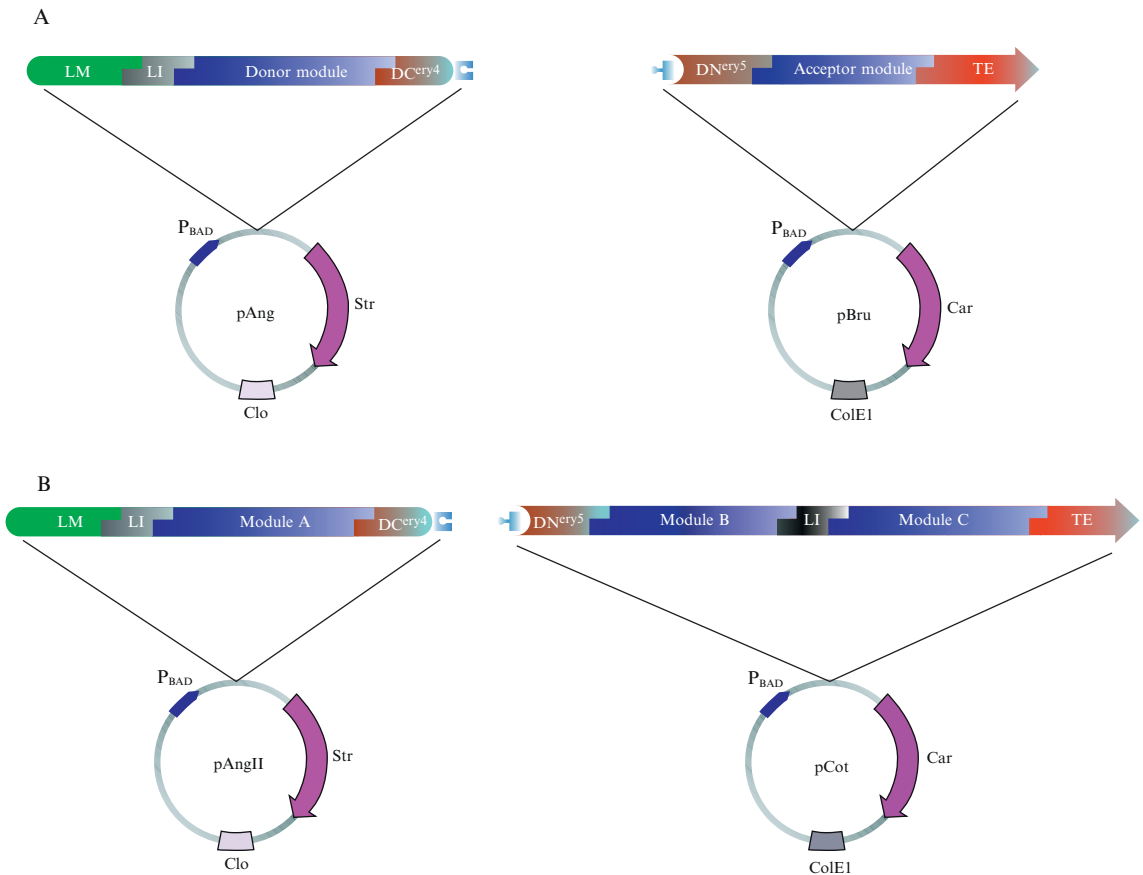
was down-regulated by replacing the strong T7 promoters by an alternative, in this case by the P_{BAD} arabinose-inducible promoter.

In summary, with the sole exception of the *Pst*I site in DEBS module 4, all the amino acid changes introduced to accommodate the restriction sites in the more than 70 kb of PKS genes for DEBS and epothilone were well tolerated. Strong evidence was also obtained to suggest that weaker promoters are needed to balance expression of the codon-optimized genes with that of the accessory proteins required for polyketide production, while stronger promoters may be necessary to achieve balance when the high-GC native genes are used.

4. A Rapid Assay to Identify Productive Combinations of PKS Modules

In spite of initial excitement, in more than 15 years of PKS engineering only a few successful attempts to create functional unnatural module–module interfaces using wildtype PKS sequences have been reported (Gokhale *et al.*, 1999; Kim *et al.*, 2002; McDaniel *et al.*, 1997; Ranganathan *et al.*, 1999; Watanabe *et al.*, 2003). Since unsuccessful attempts to create hybrid PKSs are rarely published, it is hard to anticipate how often two unrelated PKS modules could be successfully connected to yield the expected polyketide. Thus the possibilities and limitations of the combinatorial approach remain unclear. In order to address this fundamental issue, an *in vivo* assay was developed to rapidly screen interactions between unrelated PKS modules (Menzella *et al.*, 2005). Here, two expression vectors with compatible origins of replication and different selectable markers capable of expressing equivalent amounts of protein from arabinose-inducible promoters were used (Fig. 14.2A). The "donor" vector contains the loading module (LM) of DEBS and the C-terminal docking domain of DEBS1 (DC^{eryM2}), separated by unique *Xba*I and *Spe*I sites, so any synthetic module can be inserted into this plasmid in one step to create a fusion ORF of the class LM^{ery}-Module-DC^{eryM2}. The "acceptor" vector contains the DEBS2 N-terminal linker (LN^{eryM3}) and the DEBS thioesterase (TE), separated by unique *Mfe*I and *Spe*I sites. Any synthetic PKS module can be cloned into this plasmid to create a second type of fusion ORF designated LN^{eryM3}-Module-TE^{ery}. In this *in vivo* assay, the two fusion proteins are coexpressed and the rapid detection of the predicted triketide product by LC/MS/MS reveals productive module–module interactions.

To validate the assay 154 mini-PKSs of this kind were assembled, introduced into *E. coli* K207-3 (Murli *et al.*, 2003) and the resulting clones cultured and tested for protein expression and production of the expected compounds using the following method:


A
LM
LI
Donor module
DCery4
DNery5
Acceptor module
TE
PBAD
pAng
Str
Clo
PBAD
pBru
Car
ColE1
B
LM
LI
Module A
DCery4
DNery5
Module B
LI
Module C
TE
PBAD
pAngII
Str
Clo
PBAD
pCot
Car
ColE1

1. Introduce the two expression vectors by co-transformation into *E. coli* K207-3.
2. Grow the resulting colonies in 3 ml LB containing antibiotics for selection at 37° to an OD_{600}=0.5.
3. Induce with IPTG (0.5 m*M*), arabinose (2 mg/ml) and a mixture of sodium glutamate (50 m*M*), sodium succinate (50 m*M*), and sodium propionate (5 m*M*).
4. Incubate the induced cultures at 22° for 72 h with agitation.
5. Centrifuge 10 min at 10,000*g*; remove the supernatant for polyketide analysis and cell pellet for protein expression analysis.
6. Acidify culture supernatants with phosphoric acid to pH 2.5.
7. Analyze for triketide production y LC/MS/MS as described in Ashley and Carney (2004).
8. Protein expression can be assessed following the same procedure described above for DEBS.

Remarkably, almost half of the 154 bimodules produced the expected polyketide and each tested module functioned as a donor or acceptor in one or more contexts, indicating that all redesigned modules were catalytically competent, and further validating the generic PKS gene design. Interestingly, all the modules processed unnatural substrates, confirming their relaxed specificity (Gokhale *et al.*, 1999; McDaniel *et al.*, 1999; Ranganathan *et al.*, 1999).

Production of a triketide by the hybrid mini-PKSs must have involved overcoming several barriers. Potential obstacles to transfer of the propionyl moiety from the LM to the KS of each attached module include the foreign intrapeptide linkers, potential substrate intolerance by noncognate KS domains, or a requirement for possible ACP-KS interactions. Some of the unnatural fusions used the intrapeptide linker naturally associated with the module, and some others had the linker that normally separates LM from eryM1. Except for eryM1, none of the modules naturally accept and process a propionyl group, but all module fusions to LM produced the expected triketide in some bimodular combinations. The obtained results show that there is considerable flexibility in the intrapeptide linkers used to separate the LM and the first module, and the KS in all tested modules fused to the LM can elongate the propionyl moiety. Therefore, all donor modules form

Figure 14.2 (A) The two classes of expression plasmid used to test bimodular interactions in *E. coli*: pAng vectors contain a CloDF13 replication origin, a streptomycin resistance selection marker and a P_{BAD} promoter to drive expression of LM-Module-DC^{eryM2} ORFs. pBru vectors contain a ColE1 replication origin, a carbenicillin resistance selection marker and a P_{BAD} promoter to drive expression of DN^{eryM3}-Module-TE ORFs. (B) The two types of vectors used to express trimodular PKSs: pAngII vectors are similar to pAng, but DC^{eryM2} was replaced by DC^{eryM4}. pCot plasmids are similar to pBru, but they express ORFs of the class DN^{eryM5}-Mod-LI-Mod-TE. (From Menzella *et al.*, 2007. Reproduced with permission from Elsevier.)

an acyl-ACP at the C terminus that can donate the diketide to a subsequent module. Transfer of the diketide through the hybrid interface to the KS of the acceptor module requires physical proximity of the ACP and KS, and possible protein–protein interactions between them (Wu *et al.*, 2002); and elongation requires substrate tolerance by and catalytic activity of the KS (Watanabe *et al.*, 2003). Since about half of the bimodules produced the predicted polyketide, and the modules were always connected by the eryM2-eryM3 linkers, this pair serves well as generic linkers to connect unrelated modules.

The results of this study provided a clear indication of the possibilities of the PKS combinatorial biosynthesis approach since, to the best of our knowledge, this is the first analysis of the productivity of a large number of random module–module combinations. Encouragingly, it seems that unrelated PKS modules can be successfully connected to create functional pairs in many cases. On the other hand, the yields of polyketides obtained from the hybrid PKSs are typically at least one order of magnitude lower than those obtained from the natural bimodular pairs (Menzella *et al.*, 2005). Importantly, the chosen assay can be easily adapted to a high-throughput format to rapidly screen for the most productive unrelated combinations of modules.

In the above tests, all the acceptor modules processed a particular diketide from some donor modules but not from others, so recognition and elongation of the incoming diketide by the acceptor KS was unlikely to explain the failures. We attributed them instead to the absence of an appropriate ACP-KS protein–protein interaction (Wu *et al.*, 2002). To address this issue, KS domains were combinatorially changed to activate inactive module pairs and optimize the yield of a desired polyketide product (Chandran *et al.*, 2006). The unique *Mfe*I-*Kpn*I restriction sites flanking the KS domain in all the synthetic modules made this a simple task (Fig. 14.2).

The feasibility of the approach was validated by resuscitating nine previously inactive bimodular PKSs. Such activation of hybrid modular interfaces by KS replacements provides a powerful tool to engineer synthetic PKSs. Of particular interest is the use of this procedure to incorporate into assembly lines unusual modules with rare activities that cannot be replaced easily just by seeking a more adaptable natural module with the same activity.

5. Assembly of Larger Polyketide Synthases Using Information Gained with the Bimodular Assay

Polyketides with biological activity are much more complex than the triketides produced by the bimodular PKSs described above, and biosynthesis of such molecules requires the assembly of synthetic enzymes

containing multiple modules with several hybrid interfaces. Thus, it is imperative to employ the information gained with the simple test system to assemble larger PKSs. For this, a conceptual advance was made by developing the concept of ''connectivity'' to allow the overlapping of functional bimodules to build active trimodular PKSs. Connectivity describes the situation in which module B in a given bimodule extends the ketide offered by module A, and in another bimodule, module C extends the ketide offered by module B; then a trimodular PKS made by connecting module A to module C via module B should also be active. The library of bimodules previously described was constructed at random, and gave a success rate of ~45%. Constructing trimodules by simply adding another module randomly should have had, at most, a ~20% success rate. A recent publication (Menzella *et al.*, 2007) showed that, using data from the bimodular library and the connectivity principle, an additional functional heterologous interface was constructed with a remarkably higher success rate.

To prove the concept of connectivity, an expression system was developed to express trimodular PKSs (Fig. 14.2B). In this case a two-plasmid system was again used for the expression. The first plasmid encoded a fusion of the class LM-Mod-DC and the second plasmid a chimeric protein comprising DN-Mod-LI-Mod-TE. Given that the erythromycin LM and TE are functional in many heterologous bimodular PKS constructs, these domains were used for all the trimodular PKSs. The docking domains derived from eryM2 (DC^{eryM2}) and eryM3 (DN^{eryM3}) were successful at forming functional interactions between most single modules tested. However, we observed that the larger protein encoded by bimodular constructs in the pCot vector were better expressed when the N-terminal linker of eryM5 (DN^{eryM5}) was used instead of DN^{eryM3}. Using the unique restriction sites in our generic design, we examined several intrapeptide linkers to join the two modules in pCot, including the linker naturally associated with the first module and the one naturally present upstream of the N-terminus of the second module in this plasmid as well as foreign intrapeptide linkers (LI^{eryM6}, LI^{eryM1}) between the two modules, creating two unnatural junctions. The ability of bimodular constructs in pCot to interact and produce a polyketide was assessed and shown to be similar in all cases. Thus, it seems that all the tested linkers are equally efficient in connecting the modules.

Several pairs of functional bimodules were overlapped to create 54 trimodular PKSs flanked by the DEBS LM and TE. Remarkably, 52 (96%) of the synthetic enzymes formed the expected tetraketide product. Twelve of the trimodules contain contiguous extender modules in the order in which they occur in their native PKS, but in 42 there are no naturally contiguous modules; nevertheless, 40 (95%) of these also produced the expected polyketide. Interestingly, in terms of substrate specificity these results suggest that the portion of the upcoming polyketide recognized by a

given module is in most cases the extension added by the previous module regardless of the length and complexity of the polyketide chain.

The rational assembly of trimodular PKSs represents an important step toward the goal of making bioactive molecules by genetic engineering, both complete polyketides that cannot be isolated in significant quantities, and novel, ''designer'' molecules. The rationale should, for example, be applicable to building unnatural PKSs from scratch by combining individual artificial modules, as described here, as well as to adding or replacing ketide units in preexisting chains. It also provides a rational approach for ''stitching'' two or more strings of naturally contiguous modules together to make novel or inaccessible large polyketides. However, since yields of polyketide products were seen to decrease as a function of the number of unnatural junctions introduced, the efficiency of product formation will probably continue to decrease when further modules and heterologous interfaces are added. This limitation might be circumvented by using only the highest yielding trimodules as components, or by incorporating strings of as many naturally contiguous modules as possible when designing long chimeric PKSs so as to reduce the number of unnatural interfaces

6. Design and Construction of Synthetic Operons for the Expression of Sugar Pathway Genes

The increasing knowledge and technologies related to PKS engineering represent a significant advance toward the ultimate goal of synthesizing rationally designed polyketides in *E. coli*. However, as in many naturally occurring polyketides, such novel compounds might also require several post-PKS biosynthetic steps in order to become bioactive. These polyketide ''decoration'' steps, which generally involve complex glycosylation processes, are absent in *E. coli* and thus need to be incorporated in this heterologous host in combination with the PKSs to produce fully decorated polyketides.

Glycosylation of polyketides has been validated in *E. coli*, through the production of the antibiotic erythromycin C (EryC) (Peirú *et al.* 2005). This process involves the attachment of the deoxysugars TDP-L-mycarose and TDP-D-desosamine to the aglycon 6dEB (firstly hydroxylated to EB), to yield erythromycin D (EryD), which is further hydroxylated at C12 to obtain EryC. Two operons were constructed for this purpose, in order to have separate control of the two sugar biosynthetic pathways. The complete conversion of 6dEB to EryC requires functional expression of 17 genes with a variety of catalytic activities, comprising TDP-sugar biosynthesis, glycosyltransferases and hydroxylases. In this seminal work, all the

genes were individually amplified by PCR from *Micromonospora megalomicea* chromosomal DNA and a simple method was designed to allow sequential cloning of genes to build synthetic operons by adding suitable flanking restriction sites and translation signaling sequences (Fig. 14.3). *Nde*I and *Spe*I sites were introduced immediately 5′ and 3′ of each gene, respectively, to enable cloning into a special expression plasmid named pD2, containing *Xba*I-RBS-His_5 coding sequence-*Nde*I-*Spe*I. The N-terminal His_5 fusion, which can be omitted, introduces a short codon-optimized sequence that we have observed improves gene expression in most cases. It also provides a tag for further protein expression analysis and/or protein purification by means of anti-His_5 antibodies. The flanking *Xba*I and *Spe*I sites that result from cloning the PCR-amplified ORFs into this plasmid permit the sequential cloning of genes as *Xba*I/*Spe*I fragments into the *Spe*I site of the acceptor vector, resulting in a growing string of genes separated by a minimal sequence. The obtained operons can be further mobilized as *Xba*I/*Spe*I fragments to vectors with different promoters, selection markers, or replication origins if required. A detailed method is described below.

1. Amplify by PCR the individual genes to be included in the synthetic operon. For the design of the upper primer, engineer an NdeI site overlapping the translational initiation codon (changing GTG start codons to ATG when necessary). For the design of the lower primer, engineer a SpeI site immediately downstream of the stop codon.
2. Create a special vector for the sequential cloning of the PCR products. For this purpose the pET24a expression vector (Novagen) can be modified by replacing the BamHI site in the polylinker by a SpeI site, and by deleting the 21-bp region between the XbaI site and the RBS. Alternatively, a His_5 coding sequence can be added upstream of the NdeI site.
3. Clone each PCR-amplified sugar biosynthetic gene as an NdeI/SpeI fragment into the same sites of the modified pET24a vector.
4. Initiate the assembly by sequentially cloning each gene as XbaI/SpeI fragments into the SpeI site of the acceptor vector, making sure that the inserted gene is properly orientated. The intergenic XbaI/SpeI ligation destroys both sites, allowing the repetitive use of the remaining 3′ SpeI site to insert as many components as required.
5. Finally, subclone the resulting operon as an XbaI/SpeI fragment into the expression vector with the appropriate promoter, selection marker and replication origins.

Using this method, two operons were built: the ''mycarose'' operon, for the conversion of 6dEB up to 3-α-mycarosyl-EB (MEB), and the ''desosamine'' operon, for the conversion of MEB to EryC. The activity of both operons was found to be optimal in *E. coli* cultures grown at 22°, when induced in the presence of a plasmid for the overexpression of the *E. coli* chaperones GroES/GroEL, which improve solubility of the heterologous

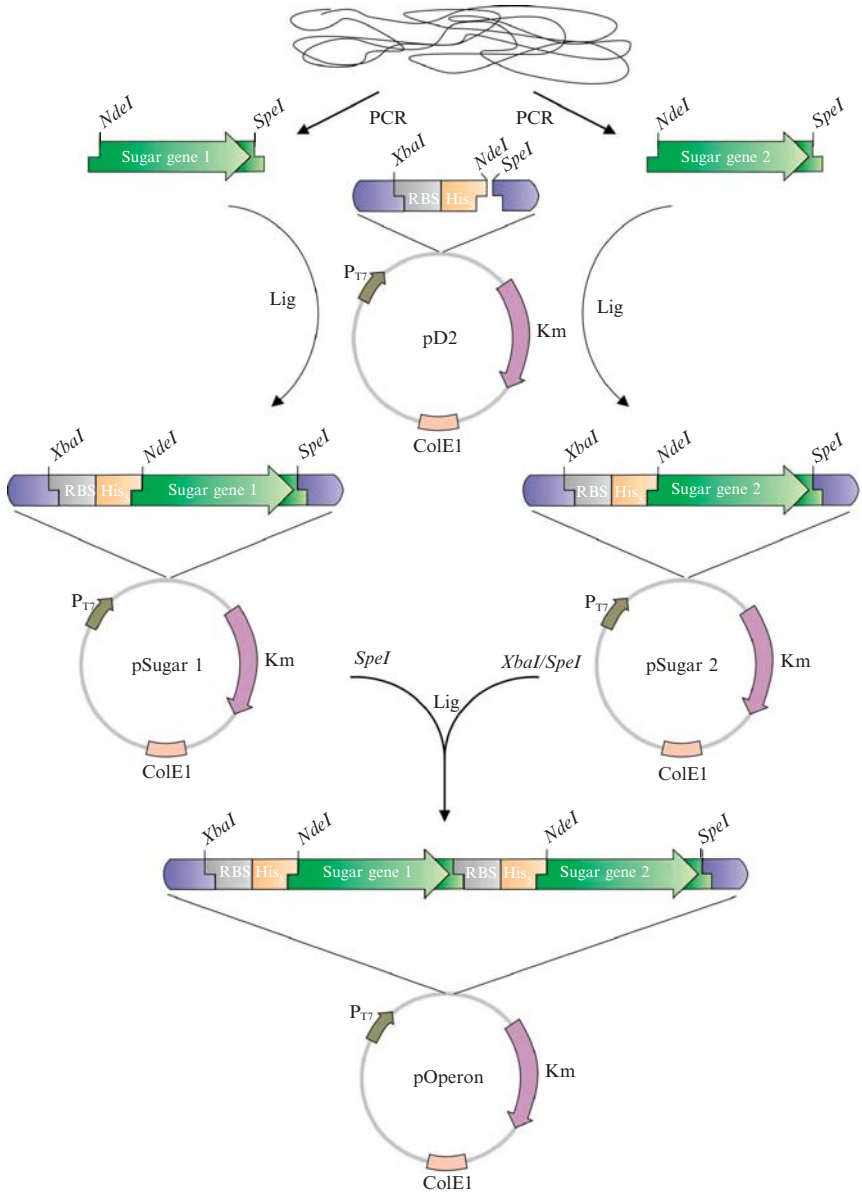


Figure 14.3 Cloning strategy for the construction of sugar biosynthesis operons. Sugar genes are PCR-amplified, with flanking *Nde*I and *Spe*I sites, and cloned into pD2. This expression vector contains a ColE1 replication origin, a kanamycin resistance selection marker, and adds an RBS and a His_5 N-terminal fusion to the cloned gene, whose expression is driven by a P_{T7} promoter. The acceptor vector (pSugar 1) is digested with *Spe*I, and the second gene in the nascent operon is introduced as an *Xba*I/*Spe*I fragment. The process can be repeated with the resulting pOperon plasmid in a recursive fashion, until the entire operon is constructed.

proteins. Biosynthesis of EryC from propionate in K207-3 required cloning of the mycarose and desosamine operons into compatible vectors (RSF1030 and CloDF13, respectively), to allow coexpression with the three plasmids for the synthesis of DEBS and chaperones containing ColE1, RSF1010, and P15A origins of replication (Peirú *et al.*, 2005). Codon-optimized versions of the genes involved in these pathways were not tested since the expression levels were satisfactory in all cases.

Using a wildtype background the glycosylation yields achieved in the initial experiments were low (<1 mg/l of EryC). Later, a second-generation strain was created where several metabolic constraints related to the availability of TDP-sugar precursors in the *E. coli* working host were addressed. The *E. coli* K207-3 first-generation strain has been recently optimized for polyketide glycosylation by metabolic engineering, through the blocking of endogenous pathways that utilize the common TDP-sugar biosynthesis precursor, TDP-4-keto-6-deoxyglucose (Peirú *et al.*, 2008). In addition, the macrolide efflux pump AcrAB-TolC was also disrupted in this strain, to avoid a premature export of the partially glycosylated precursors, which was also shown to be detrimental to the overall efficiency of the glycosylation process. The resulting *E. coli* strain, named LB19b, showed a significant increase in the production of TDP-L-mycarose and TDP-D-desosamine, which was reflected in a remarkable 100-fold improvement of EryD production from 6dEB in bioconversion experiments, compared with the parental K207-3 strain.

The higher TDP-sugar production achievable with the LB19b strain permits an easier analysis of sugar pathways by LC/MS/MS. This, in combination with the described universal cloning strategy, allows a step-by-step monitoring of the growing operon, when adding genes following their predicted order in the biosynthetic pathway, by looking for the expected TDP-sugar intermediate (Peirú *et al.*, 2007; Rodriguez *et al.*, 2006). In this way, incorrect combinations or nonfunctional genes can be easily detected, and novel enzymatic activities or biosynthetic routes determined, providing new tools for gene function analysis, sugar pathway engineering, and production of novel glycosylated polyketides.

REFERENCES

Ashley, G. W., and Carney, J. R. (2004). API-mass spectrometry of polyketides. I. A study on the fragmentation of triketide lactones. *J. Antibiot. (Tokyo)* **57,** 224–234.

Chandran, S. S., Menzella, H. G., Carney, J. R., and Santi, D. V. (2006). Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of keto-synthase domains. *Chem. Biol.* **13,** 469–474.

Cortes, J., Wiesmann, K. E., Roberts, G. A., Brown, M. J., Staunton, J., and Leadlay, P. F. (1995). Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268,** 1487–1489.

Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284,** 482–485.

Jayaraj, S., Reid, R., and Santi, D. V. (2005). GeMS: An advanced software package for designing synthetic genes. *Nucleic Acids Res.* **33,** 3011–3016.

Kao, C. M., Katz, L., and Khosla, C. (1994). Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265,** 509–512.

Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004). Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. USA* **101,** 15573–15578.

Kim, B. S., Cropp, T. A., Florova, G., Lindsay, Y., Sherman, D. H., and Reynolds, K. A. (2002). An unexpected interaction between the modular polyketide synthases, erythromycin DEBS1 and pikromycin PikAIV, leads to efficient triketide lactone synthesis. *Biochemistry* **41,** 10827–10833.

McDaniel, R., Kao, C. M., Hwang, S. J., and Khosla, C. (1997). Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4,** 667–674.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA* **96,** 1846–1851.

McDaniel, R., Welch, M., and Hutchinson, C. R. (2005). Genetic approaches to polyketide antibiotics. 1. *Chem. Rev.* **105,** 543–558.

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23,** 1171–1176.

Menzella, H. G., Carney, J. R., and Santi, D. V. (2007). Rational design and assembly of synthetic trimodular polyketide synthases. *Chem. Biol.* **14,** 143–151.

Menzella, H. G., and Reeves, C. D. (2007). Combinatorial biosynthesis for drug development. *Curr. Opin. Microbiol.* **10,** 238–245.

Menzella, H. G., Reisinger, S. J., Welch, M., Kealey, J. T., Kennedy, J., Reid, R., Tran, C. Q., and Santi, D. V. (2006). Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. *J. Ind. Microbiol. Biotechnol.* **33,** 22–28.

Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., and Kealey, J. T. (2003). Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. *J. Ind. Microbiol. Biotechnol.* **30,** 500–509.

Mutka, S. C., Carney, J. R., Liu, Y., and Kennedy, J. (2006). Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry* **45,** 1321–1330.

Oliynyk, M., Brown, M. J., Cortes, J., Staunton, J., and Leadlay, P. F. (1996). A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3,** 833–839.

Peirú, S., Menzella, H., Rodríguez, E., Carney, J., and Gramajo, H (2005). •Production of the potent antibacterial polyketide erythromycin C in *Escherichia coli*. *Appl. Environ. Microbiol.* **71,** 2539–2547.

Peirú, S., Rodriguez, E., Tran, C. Q., Carney, J., and Gramajo, H (2007). Characterization of the heterodimeric MegBIIa:MegBIIb aldo-keto reductase involved in the biosynthesis of L-mycarose from *Micromonospora megalomicea*. *Biochemistry* **46,** 8100–8109.

Peirú, S., Rodriguez, E., Menzella, H., Carney, J., and Gramajo, H (2008). Metabolically engineered *Escherichia coli* for efficient production of glycosylated natural products. *Microb. Biotechnol.* **1**(6), 476–486.

Pfeifer B. A., and Khosla, C. (2001). Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **65**(1), 106–118.

Ranganathan, A., Timoney, M., Bycroft, M., Cortes, J., Thomas, I. P., Wilkinson, B., Kellenberger, L., Hanefeld, U., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1999).

Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: A route to simple statin analogues. *Chem. Biol.* **6,** 731–741.
Reisinger, S. J., Patel, K. G., and Santi, D. V. (2006). Total synthesis of multi-kilobase DNA sequences from oligonucleotides. *Nat. Protoc.* **1,** 2596–2603.
Rodríguez, E., Peirú, S., Carney, J., and Gramajo, H (2006). *In vivo* Characterization of the dTDP-D-desosamine Biosynthetic Pathway of the Megalomicin Cluster from *Micromonospora megalomicea*. *Microbiology* **152**(Pt3), 667–673.
Sherman, D. H. (2005). The Lego-ization of polyketide biosynthesis. *Nat. Biotechnol.* **23,** 1083–1084.
Walsh, C (2003). ''Antibiotics: Actions, origins, resistance.'' ASM Press, Washington, DC.
Walsh, C. (2004). Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. *Science* **303,** 1805–1810.
Watanabe, K., Wang, C. C., Boddy, C. N., Cane, D. E., and Khosla, C. (2003). Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. *J. Biol. Chem.* **278,** 42020–42026.
Weissman, K. J., and Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3,** 925–936.
Wu, N., Cane, D. E., and Khosla, C. (2002). Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41,** 5056–5066.