Chemistry & Biology
Article

Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases

Hugo G. Menzella,[1,*] John R. Carney,[1] and Daniel V. Santi[1,2,*]
[1] Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA
[2] Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94143, USA
*Correspondence: menzella@kosan.com (H.G.M.), daniel.v.santi@gmail.com (D.V.S.)
DOI 10.1016/j.chembiol.2006.12.002

SUMMARY

Type I polyketide synthases (PKSs) consist of modules that add two-carbon units in polyketide backbones. Rearranging modules from different sources can yield novel enzymes that produce unnatural products, but the rules that govern module-module communication are still not well known. The construction and assay of hybrid bimodular units with synthetic PKS genes were recently reported. Here, we describe the rational design of trimodular PKSs by combining bimodular units. A cloning-expression system was developed to assemble and test 54 unnatural trimodular PKSs flanked by the loading module and the thioesterase from the erythromycin synthase. Remarkably, 96% of them produced the expected polyketide. The obtained results represent an important milestone toward the ultimate goal of making new bioactive polyketides by rational design. Additionally, these results show a path for the production of customized tetraketides by fermentation, which can be an important source of advanced intermediates to facilitate the synthesis of complex products.

INTRODUCTION

Modular polyketide synthase (PKS) genes determine the biosynthesis of important natural products like erythromycin, tacrolimus, and many others [1]. They encode giant enzymes consisting of sets (modules) of active sites (domains), forming an "assembly line" that builds the carbon chain of the final product in a stepwise fashion. The canonical modular PKS gene set encodes multiple extender modules preceded by a loading module (LM) and terminated with a thioesterase (TE) [2, 3]. The LM selects and transfers the starter acyl group to the first extender module. All extender modules contain a ketosynthase (KS), an acyl transferase (AT), and an acyl carrier protein (ACP). Most also contain up to three additional reductase/dehydratase domains that modify the β-carbonyl group of the growing ketide chain. The KS receives the acyl unit from the preceding module, while the AT transfers an appropriate acyl-extender unit to the prosthetic group of the ACP. The KS then catalyzes a condensation between the acyl KS and the extender unit to give a β-keto-acyl-ACP. The polyketide chain is successively elongated by downstream extender modules, and, at the end of the assembly line, the TE domain releases and often cyclizes the products [2–4]. Thus, the specificity of the AT domain, the complement of reductive domains, and carbon branch stereochemistry dictate the structure of the two-carbon unit that each module adds; the order of the modules determines the sequence of units in the polyketide product, and the number of modules determines carbon chain length.

Growth of the polyketide chain requires its transfer from the ACP of one module to the KS of the next. Adjacent modules can be present on the same or different proteins, necessitating either intra- and interpolypeptide chain transfers, promoted by appropriate linkers to facilitate proximity [5, 6]. Intrapeptide linkers (LI) are spacers of ~20 amino acids that separate the ACP of one module from the KS of the next. Interpeptide linkers (docking domains) consist of an ~80–130 residue domain at the C terminus of one module (DC) that interacts with a cognate ~30–50 residue domain at the N terminus of the downstream module (DN). Importantly, cognate linker sets enable productive module-module interactions between foreign module pairs [5, 7, 8].

The modular nature of polyketide biosynthesis has facilitated genetic engineering of PKS genes to modify polyketide structure. Because PKS modules comprise natural, integrated catalytic units, rearranging intact modules is an attractive approach for combinatorial biosynthesis [6, 9, 10] that, in theory, could generate virtually any polyketide. Since current knowledge is insufficient to rationally design functional module-module interfaces a priori, we initially approached the problem empirically by using combinatorial biosynthesis. The strategy was to make and test large numbers of module-module combinations in order to identify enough productive interfaces to be used in the assembly of diverse, complex polyketides. We previously reported [11] a two-plasmid assay to rapidly construct and scan the ability of two heterologous PKS modules to interact and produce a polyketide product. By creating libraries of synthetic modules in each of these vectors,

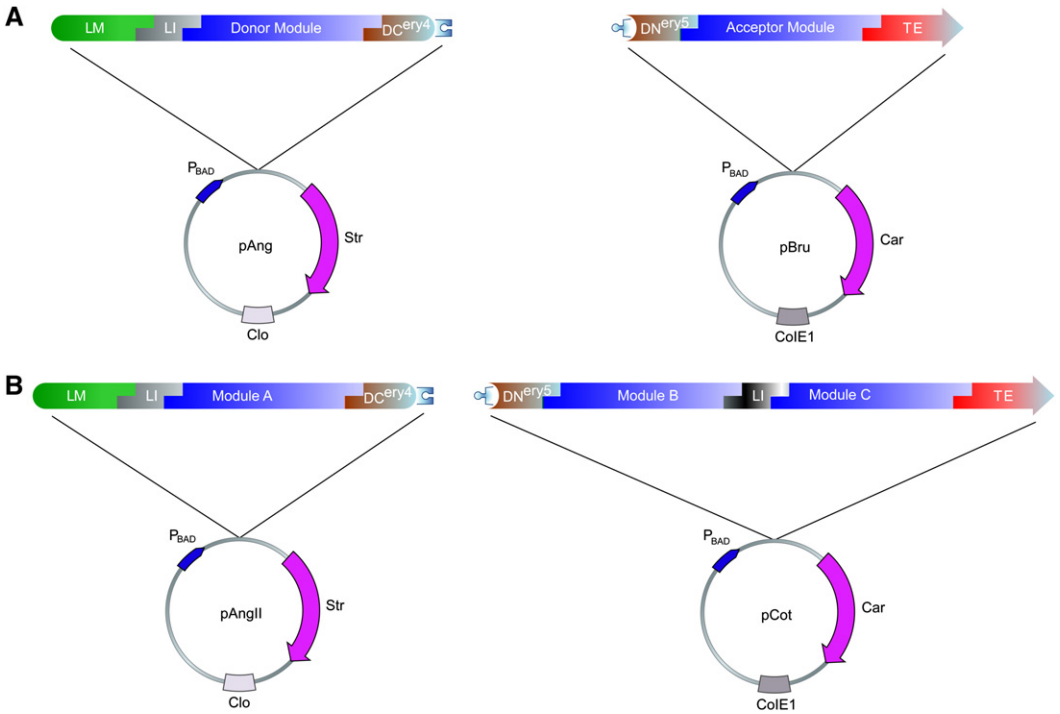

Figure 1. Generic Plasmids Used to Express Synthetic PKSs in *E. coli*

(A) The two classes of expression plasmids 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.

we tested over 150 module-module combinations for activity and found that ~45% formed the expected product.

In the present work, we describe an experimental system that allows combinatorial exploration of PKSs containing three modules and two intermodular interfaces. We show that, by applying knowledge gained from the bimodular system and certain design constraints (notably by overlapping active bimodules), the rate of successfully producing functional PKSs can be greatly enhanced.

RESULTS AND DISCUSSION

Abbreviations and Nomenclature

The abbreviations used in this article are as follows: PKS, polyketide synthase; KS, keto synthase, AT, acyl transferase; ACP, acyl carrier protein; Mod, extender module; LM, loading module; TE, thioesterase; DN, N-terminal docking domain, DC, C-terminal docking domain; LI, intrapeptide linker; ery, erythromycin; sor, soraphen; gdm, geldanamycin; rap, rapamycin; SNAC, *N*-acetylcysteamine thioester; ORF, open reading frame.

The PKS source of a component is indicated by the first three letters of the PKS name. Modules are named by their PKS source, followed by the number of the module in the source (e.g., eryM3); LI, DC, and DN are superscripted with the PKS source and module number (e.g., DN^{eryM3}).

Cloning and Expression Systems

We previously described a two-vector, two-open reading frame (ORF) cloning and expression system that allows combinatorial biosynthesis of polyketides encoded by a LM and two extender modules [11]. In this system, two ORFs encoding PKS genes to be coexpressed were assembled in the expression vectors pAng and pBru (Figure 1A). The pAng "donor" contains the ORF encoding the N-terminal PKS fusion protein composed of a LM, an LI, an extender module, and a C-terminal docking domain. The pBru "acceptor" encodes a fusion protein containing an N-terminal docking domain cognate to the one on pAng, followed by an extender module and a C-terminal TE domain. The two PKS polypeptides expressed from pAng and pBru in the engineered *Escherichia coli* K207-3 strain [12] interact via their docking domains to form the holo-PKS, which, if catalytically active, produces a triketide lactone polyketide. In this system, combinatorial biosynthesis is achieved by independent coexpression of a series of PKS ORFs in pAng with a series in pBru—an approach dubbed "combinatorial transformation" [11, 13].

Implementation of the system is facilitated by the use of chemically synthesized gene components that contain sets of standard, unique restriction sites flanking domains, linkers, and modules to allow easy assembly and interchange (Figure 1) [11]. The composite design has been

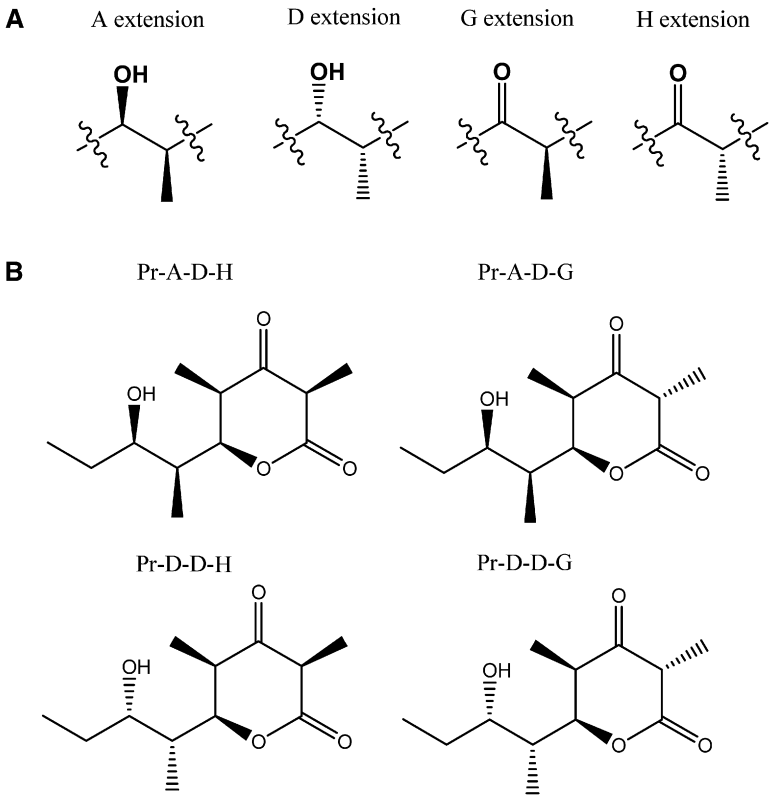

Figure 2. Predicted Products of PKSs Created in This Work

(A) Structures of two-carbon units encoded by PKS modules. A, D, G, and H are one-letter codes describing the extensions added by the modules used in this study.

(B) Structures of predicted tetraketides obtained by the combination of modules with A-type activity (eryM1), D-type activity (eryM2, eryM5, eryM6, sorM6, and gelM3), G-type activity (eryM3, rapM3, and rapM6), or H-type activity (pikM6).

likened to a set of "lego" blocks, and the process has been called "lego-ization" of polyketide combinatorial biosynthesis [14, 15].

To expand this system to three PKS modules, we considered constructing three compatible vectors that would each express a separate module; however, this would have required the development of promoter combinations to provide similar multiple-protein expression [16] and introduction of a second pair of docking domains. To avoid some of these issues, we chose instead to adapt the current system to one in which the acceptor vector contained two extender modules encoded in a single ORF (pCot in Figure 1B). The pCot series of vectors for the expression of ORFs with the structure DN-Mod-LI-Mod-TE were created by sequential cloning, by using the unique restriction sites at the borders of each building block. Trimodular PKSs were created by combinatorial transformation and expression of single-module "donor" pAng vectors with these bimodular pCot vectors.

Modules

In this work, we used ten different PKS modules that naturally encode ketide structures of four kinds, dubbed A, D, H, and G (Figure 2A). Nine of these were previously reported and validated as catalytically active in many different contexts [11]. The tenth, rapM6—encoding ketide G—was synthesized in the present work and validated as active by using the bimodular system previously described [11]. To do so, rapM6 cloned into the pBru vector was coexpressed in *E. coli* K207-3 with pAng vectors containing donor modules of interest, and the anticipated triketide lactones were quantified. For rapM6 in this system, the donor modules in pAng and the yields of expected polyketide products (given in parentheses) were: eryM1 (0.33 mg/l), eryM2 (1.4 mg/l), eryM5 (0.36 mg/l), eryM6 (0.53 mg/l), sorM6 (0.37 mg/l), and gdmM3 (0.19 mg/l).

Generic Building Block Components

Our approach for combinatorial biosynthesis requires that PKS components other than modules—LM, linkers, and TE—be generically useful in multiple contexts, so that diversity can be achieved by varying the type of modules that add two-carbon units (ketides) within a uniform background of other PKS components.

Given that the erythromycin LM and TE are functional in many heterologous PKS constructs [11], we used these domains in all of the PKSs examined here. The docking domains derived from eryM2 (DC^{eryM2}) and eryM3 (DN^{eryM3}) were successful at forming functional interactions between most single modules tested [11]. However, we observed that the larger PKSs encoded by bimodular constructs in pCot were better expressed when the N-terminal linker of eryM5 (DN^{eryM5}) was used instead of DN^{eryM3} (data not shown). Since the eryM2-3 and eryM4-5 linker pairs appeared to be equiefficient [5, 8], we used the latter for interpeptide connections of the trimodular constructs reported here. The series of plasmids for expressing LM-Mod-DC^{eryM4} was named pAngII (Figure 1B).

Table 1. Production of Triketide Lactones by Bimodular PKSs

		Separated Proteins		Fusion Proteins	
Entry	Module Pairs	Docking Domains	Triketide Lactone Production (mg/l)	Intrapeptide Linker	Triketide Lactone Production (mg/l)
1	eryM2, eryM3	DC^{eryM2}, DN^{eryM3}	23.5	LI^{eryM6}	14.6
				LI^{eryM1}	13.2
2	eryM2, pikM6	DC^{eryM2}, DN^{eryM3}	4.1	LI^{eryM6}	2.9
				LI^{eryM1}	2.8
3	eryM2, rapM3	DC^{eryM2}, DN^{eryM3}	2.1	LI^{eryM6}	1.7
				LI^{rapM3}	1.9
4	eryM2, rapM6	DC^{eryM2}, DN^{eryM3}	1.4	LI^{eryM6}	1.2
				LI^{rapM6}	1.4
5	eryM5, eryM3	DC^{eryM2}, DN^{eryM3}	2.4	LI^{eryM6}	1.5
6	eryM5, pikM6	DC^{eryM2}, DN^{eryM3}	0.2	LI^{eryM6}	0.2
7	eryM6, eryM3	DC^{eryM2}, DN^{eryM3}	3.2	LI^{eryM6}	2.4
				LI^{eryM1}	2.7
8	eryM6, pikM6	DC^{eryM2}, DN^{eryM3}	0.8	LI^{eryM6}	0.7
				LI^{eryM1}	0.6
9	eryM6, rapM3	DC^{eryM2}, DN^{eryM3}	0.9	LI^{eryM6}	0.7
				LI^{rapM3}	0.8
10	eryM6, rapM6	DC^{eryM2}, DN^{eryM3}	0.6	LI^{eryM6}	0.6
				LI^{rapM6}	0.5

Bimodules were expressed as separated proteins or as a fusion protein with different intrapeptide linkers. When expressed as separated proteins from pAng and pBru plasmids (Figure 1), the first module of the PKS was preceded by an LM. When expressed as fusion proteins from pCot plasmids (Figure 1), the cultures were supplemented with 1 mM propionyl-SNAC to prime the synthesis of triketide lactones.

Choosing a connector for two extender modules contained within a protein (LIs)—as in the current pCot constructs—is more problematic. Such sequences are generally not conserved, and the junctions that connect them to flanking modules are not obvious [6]; indeed, work aimed at creating a truly generic LI has not been successful to date.

Using the unique restriction sites in our generic design, we examined several LIs to join the two modules in pCot. First, we tried the linker naturally associated with the N terminus of the second module in this plasmid. This design provided one rather than two unnatural sequence junctions that resulted from inserting an entirely foreign linker between two modules. For comparison, we inserted foreign LIs (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 by feeding propionyl-*N*-acetylcysteamine thioester (SNAC) to expressing cultures and assaying for production of the anticipated triketide lactone by LC/MS/MS.

We initially compared triketide lactone production levels from bimodule-TEs on single polypeptides with LIs to the same bimodules in two polypeptides connected by docking domains. As shown in Table 1 and Figure 3, the product yields in the two systems are very well correlated ($R^2 = 0.99$), and the LIs provided ~65% of the product produced by the corresponding interpeptide system. Next, we compared LIs naturally present at the N terminus of the second modules in pCot (rapM3, rapM6) to a completely foreign linker (eryM6) and found that yields were essentially identical (entries 3, 4, 9, and 10 in Table 1). Finally, using identical bimodules, we observed no difference in triketide lactone yields with foreign linkers derived from the N terminus of eryM1 and eryM6 (entries 1, 2, 7, and 8 in Table 1). Thus, in the present context, all LIs tested appeared to be equally efficient at facilitating the transfer of a polyketide chain from one module to the next. Where possible, we used the LI naturally found on the N terminus of the second module in pCot; when no such linker existed (eryM3, pikM6), we used the LI associated with eryM6.

Trimodular PKS Design and Assessment

The objective of the present work was to determine whether information obtained in studies of heterologous bimodular PKSs could be used to improve the success rate in achieving active trimodular PKSs over arbitrary

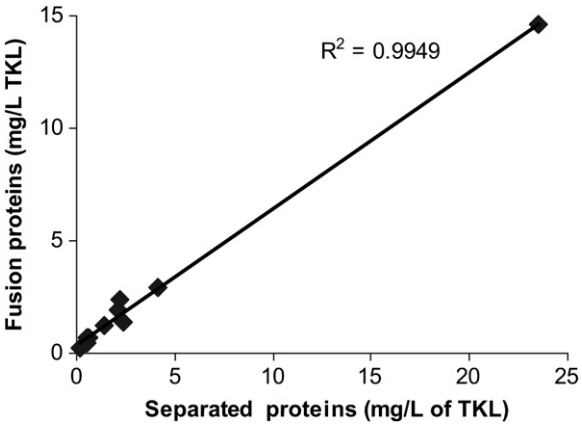

Figure 3. Product Yields of Bimodular PKSs Expressed as Separated or as Fusion Proteins

When expressed as separated proteins, modules are connected with the DC^{eryM2} and DN^{eryM3} docking domains; when expressed as fusion proteins, modules are connected with the LI^{eryM6} intrapeptide linker. TKL, triketide lactone.

Table 2. All Possible Trimodular PKSs that Can Be Assembled by Overlapping Functional Bimodules in Our Current Library

Entry	Trimodules	Number of Possible PKSs
1	D-A-D	3
2	D-A-G	2
3	D-A-H	1
4	A-D-D	11
5	A-D-G	5
6	A-D-H	3
7	D-D-D	66
8	D-D-G	38
9	D-D-H	18
	Total	147

Only bimodules with A, D, G, and H activities are considered, since bimodular PKSs containing E-type modules showed poor activity [11].

constructs. If we assume that the activity of a multimodule construct depends only on the activities of adjacent module pairs, the expected fraction of active multimodules (F_N) is $F_N = (F_{BM})^{(N-1)}$, where F_{BM} is the fraction of active bimodular interfaces obtained upon random combination of N modules, and $(N-1)$ is the number of module interfaces in the PKS. Since 45% of the random combinations of two modules with foreign interfaces gave active PKSs, random combinations of three heterologous modules may be estimated to achieve a 20% (0.45^{3-1}) success rate in producing active trimodules. We recognize that this is likely to be an overestimate, since factors other than active adjacent modules (e.g., substrate specificity) are likely to have negative impacts on activity.

The three-module PKSs were designed with certain constraints to increase the probability of success. First, initial modules had all been shown to accept the propionyl starter unit (Pr-) from the erythromycin LM to provide the appropriate Pr-ketide acyl-enzyme intermediate [11]. Second, all adjacent module pairs (bimodules) used in the trimodular constructions were known to be active in the context of both inter- and intrapeptide modules. Thus, no single module interface in the trimodule per se was expected to be a barrier to activity of the trimodule. Moreover, when tested as bimodules, the adjacent modules yielded sufficient triketide lactone ($\geq$0.2 mg/l) such that they were likely to continue to yield detectable product after an expected loss in yield upon introduction of the second modular interface. Finally, all trimodules were constructed from overlapping active bimodules. Thus, in any trimodular construct the central module is a known acceptor of the Pr-ketide from the first module—assuring reaction through the first two modules—and a known donor of its encoded ketide to the last module. Thus, provided the third module can tolerate the Pr-ketide extension on the substrate resulting from the LM and the first module, each trimodule should be active.

Applying these constraints to our available data on bimodular PKS [11], we selected target trimodular candidates with sequences of activities of A/D-D-G/H that were expected to produce tetraketides Pr-A/D-D-G/H as six-member lactones. Only module types A, D, and E have been examined as donors in the bimodular system [11]; thus, these modules were contenders for the first module in the three-module system. Type E modules were removed from consideration as components of the trimodular PKSs in the current study because of the poor activity of this class of modules in our previous experiments [11], which might be attributed to their specificity for malonyl-CoA and the low levels of this substrate in our *E. coli* host (J. Kealey, personal communication).

The D-type module was chosen as a common second (connecting) module since it provides the possibility of rationally designing a wider variety of trimodular PKSs with the available data. It serves as either a starter or terminator in 63 of 72 (88%) of the active bimodules in our current library [11], and it could be used as the connector to design 141 trimodular PKSs (entries 4–9 in Table 2) within the aforementioned constraints, with the potential to produce 6 different tetraketides. Using A-type modules as connectors, only six PKSs could be designed, and these would, in theory, produce only three different products (entries 1–3 in Table 2). The G or H modules were not considered as connectors since they were not tested as donors in the bimodular system, and the anticipated products are unusual eight- rather than six-membered lactones. For the last module in the trimodules, D, G, and H were candidates by our criteria. For convenience of analysis, we chose to study trimodules with H or G as terminal modules; these provide single equilibrated forms of the R and S α-methyl

Table 3. Production of Tetraketides by Trimodular PKSs

		DN-eryM2-eryM3-TE (D-G)	DN-eryM2-pikM6-TE (D-H)	DN-eryM2-rapM3-TE (D-G)	DN-eryM2-rapM6-TE (D-H)	DN-eryM5-eryM3-TE (D-G)	DN-eryM5-pikM6-TE (D-G)	DN-eryM6-eryM3-TE (D-G)	DN-eryM6-pikM6-TE (D-G)	DN-eryM6-rapM3-TE (D-G)	DN-eryM6-rapM6-TE (D-G)
1	LM-eryM1-DC (A)	+++	++	++	++	+	+	+	+	+	+
2	LM-eryM2-DC (D)	+	+	+	+	+	+	+	+	+	+
3	LM-eryM5-DC (D)	+	+	+	+	nd	+	++	++	++	++
4	LM-eryM6-DC (D)	+	+	+	+	+	+	+	+	+	+
5	LM-sorM6-DC (D)	nt[a]	nt[a]	nt[a]	nt[a]	nd	+	+	+	+	+
6	LM-gdmM3-DC (D)	+	+	+	+	nt[a]	nt[a]	+	+	+	+

The one-letter code describes the activity of each extension module. Symbols indicate the production level ranges of tetraketide by the trimodular PKSs: +++, more than 10 mg/l; ++, 1–10 mg/l; +, less than 1 mg/l; nd, not detected.
[a] The first two modules did not produce a triketide lactone when tested as a bimodular PKS. In these cases, connectivity of overlapping bimodules was not tested.

ketone that can be analyzed as a single product (see below). Thus, we decided to construct 54 trimodular PKSs with A or D modules in the first position, a D module in the second position, and G or H modules in the third position. The structures of the four expected tetraketide products are shown in Figure 2B.

Analysis of the four tetraketide products—Pr-A/D-D-G/H—was simplified by keto-enol equilibration of the terminal G or H ketide. The resulting products, Pr-A-D-G/H or Pr-D-D-G/H, elute as single peaks by HPLC with a mass spectrum consistent with the product due to either a single stable isomer or rapidly equilibrating forms. An authentic sample was used to identify Pr-A-D-G/H. Presumed Pr-D-D-G/H products showed the same parent ion and fragmentation as Pr-A-D-G/H, with different relative intensities of the parent/daughter ions (Figure S1; see the Supplemental Data available with this article online). These characteristics have previously been shown to be typical of different stereoisomers of methyl and hydroxyl substituents on triketide lactones [11]. In addition, feeding of (2*R*,3*S*)-2-methyl-3-hydroxyhexanoic acid SNAC to DN^{eryM5}-eryM2-eryM3-TE, which fixes the stereochemistry of the first D ketide unit, provided a product with an identical mass, retention time, and relative intensities of the parent/daughter ions to those presumed to be Pr-D-D-G/H. Taken together, these results provide strong evidence that the tetraketide product has the anticipated Pr-D-D-G/H structure.

Table 3 shows the matrix of six single-module donor pAng (rows) and ten bimodule acceptor pCot (columns) constructs that were coexpressed in *E. coli* K207-3, along with their yields of tetraketide products. Of the 60 possible trimodular constructs, 6 were not tested because the combinations of the first and second module of these were known to be nonfunctional [11]. Of the 54 tested combinations, 52 (96%) formed a product with LC/MS/MS profiles consistent with the expected tetraketide. Twelve of the trimodules contain contiguous extender modules in the same order in which they are found in the native PKS, but there are no natural contiguous modules in the remaining 42; nevertheless, 40 (95%) of these constructs also produced the expected polyketide.

We previously observed that, although separated by a transplanted pair of docking domains, modules that are naturally contiguous in a PKS consistently gave higher yields of polyketide products [11]. Likewise, the catalytic activity of trimodules tested seems to be related to the number of contiguous natural modules. For example, of the 54 trimodules tested, ery1-ery2-ery3—containing three sequential natural modules—provides the highest yield; next are PKSs with a sequence of two natural modules—three with eryM1-eryM2-ModX, and four with eryM5-eryM6-ModX (Table 3, rows 1 and 3).

Thus, by preselecting modular interfaces known to be productive in bimodules, a striking 96% success rate in production of tetraketides was observed in three-module systems. However, since yields of polyketide products were seen to decrease as a function of unnatural junctions introduced, we expect the efficiency of product formation to continue to decrease when additional modules and heterologous interfaces are added. This limitation might be circumvented by using only the highest-yield trimodules as components, or by incorporating strings of naturally contiguous modules when designing long chimeric PKSs to reduce the number of unnatural interfaces. Additionally, knowledge gained from recent advances in the elucidation

of the three-dimensional structure of PKS modules will continue to provide details of how domains and modules interact [17–19], which may help in the near future to improve the efficiency of our approach by enabling the rational engineering of bimodular interfaces.

SIGNIFICANCE

The ability to manipulate intact PKS modules to generate novel assembly lines, while of great value for the biosynthesis of unnatural compounds, is limited by an incomplete understanding of module-module communication. To partially overcome this problem, we developed a system using synthetic genes to rapidly scan bimodular interactions that enabled the identification of many functional hybrid, bimodular PKSs [11]. Here, we made a conceptual advance in developing the concept of "connectivity" and in showing that functional bimodular units can be overlapped to build functional trimodular PKSs. Connectivity suggests that if, in a given bimodule PKS, module B extends the ketide offered by module A, and ModC extends the ketide offered by module B in another bimodule, then a functional trimodular PKS can be made by "connecting" ModA to ModC via ModB. The library of bimodules that we previously reported was constructed by a random approach, and it gave a success rate of ~45%. Constructing trimodules by simply adding another module randomly should have, at most, provided ~20% success; here, we show that using data from the bimodular library and the "connectivity" principle provides an additional functional heterologous interface with a striking 95% success rate.

The rational assembly of trimodular PKSs represents an important step toward the goal of making bioactive molecules by genetically engineering both polyketides that cannot be isolated in significant quantities [20, 21] and novel, "designer" molecules. The principle/rationale should, for example, be applicable to build unnatural PKSs from scratch by combining isolated modules or by 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. Although beyond the scope of the present work, we project the achievement of such milestones in the near future.

EXPERIMENTAL PROCEDURES

Host and Vectors

The *E. coli* polyketide production strain K207-3 [*BL21ΔprpBCD::T7prom prpE, T7prom accA1-pccB, T7prom sfp*], as well as the pAng plasmids containing LM^{ery}-Mod-DC^{eryM2} with eryM1 (pKOS422-108-1), eryM2 (pKOS422-99-2), eryM5 (pKOS422-126-2), eryM6 (pKOS422-99-3), sorM6 (pKOS422-99-4), or gdmM3 (pKOS554-52-4), and the pBru plasmids containing DN^{eryM3}-Mod-TE^{ery} with eryM2 (pKOS422-100-1), eryM5 (pKOS422-100-2), eryM6 (pKOS422-100-3), sorM6 (pKOS422-114-3), gdmM3 (pKOS422-114-4), pikM6 (pKOS422-171-6), eryM3 (pKOS422-171-5), or rapM3 (pKOS422-171-7) have been described [11, 12]. DNA-manipulation procedures were performed as previously described [22].

The SpeI-XbaI fragment of MGP064 containing the synthetic version of rapM6 was synthesized as described [23] and was cloned into the XbaI-SpeI sites of the pAng plasmid to create pKOS554-48-7 for the expression of DN-rapM6-TE.

The construct the pAngII series of vectors, the DC^{eryM2} of the pAng plasmids was removed with SpeI-EcoRI and was replaced by the DC^{eryM4} from the MGP011 digested with the same enzymes to create the expression vectors containing LM^{ery}-Mod-DC^{eryM4} with eryM1 (pKOS554-160-1), eryM2 (pKOS554-160-2), eryM5 (pKOS554-160-3), eryM6 (pKOS554-160-4), sorM6 (pKOS554-160-5), or gdmM3 (pKOS554-160-6).

The pCot series of plasmids for the expression of DN^{eryM5}-ModB-ModC-TE was created as follows: (1) the generic pCot plasmid was made by removing the DN^{eryM3} of a pBru plasmid with NdeI-MfeI and by cloning the NdeI-MfeI fragment of MGP009, containing DN^{eryM5}; (2) the SpeI-EcoRI fragments of plasmids MGP007, MGP034, MGP041, and MGP064, containing the synthetic eryM3, pikM6, rapM3, and rapM6, respectively, were cloned into the XbaI-EcoRI sites of MGP001, MGP009, or MGP010 to create the following intermediate plasmids containing bimodular fusions: pKOS554-110-1 (eryM2-eryM3), pKOS554-110-2 (eryM2-pikM6), pKOS554-110-3 (eryM2-rapM3), pKOS554-110-4 (eryM2-rapM6), pKOS554-110-5 (eryM5-eryM3), pKOS554-110-6 (eryM5-rapM6), pKOS554-110-7 (eryM6-eryM3), pKOS554-110-8 (eryM6-pikM6), pKOS554-110-9 (eryM6-rapM3), pKOS554-110-10 (eryM6-rapM6); (3) finally all intermediate plasmids described in (2) were digested with SpeI-XbaI, and the fragments containing bimodular fusions were cloned into XbaI-SpeI sites of the pCot plasmid to create the expression vectors pKOS554-121-1 (DN^{eryM5}-eryM2-eryM3-TE), pKOS554-121-2 (DN^{eryM5}-eryM2-pikM6-TE), pKOS554-121-3 (DN^{eryM5}-eryM2-rapM3-TE), pKOS554-121-4 (DN^{eryM5}-eryM2-rapM6-TE), pKOS554-121-5 (DN^{eryM5}-eryM5-eryM3-TE), pKOS554-121-6 (DN^{eryM5}-eryM5-pikM6-TE), pKOS554-121-7 (DN^{eryM5}-eryM6-eryM3-TE), pKOS554-121-8 (DN^{eryM5}-eryM6-pikM6-TE), pKOS554-121-9 (DN^{eryM5}-eryM6-rapM3-TE), and pKOS554-121-10 (DN^{eryM5}-eryM6-rapM6-TE).

SNAC Feeding to Bimodules

K207-3 bacteria harboring pCot expression plasmids were grown in 2.5 ml LB with carbenicillin (50 μg/ml) at 37°C to an OD_{600} of 0.5. Cultures were induced with IPTG (0.5 mM) and arabinose (2 mg/ml), and 0.5 ml of a mixture of sodium glutamate (50 mM), sodium succinate (50 mM), sodium propionate (5 mM), and propionic acid *N*-acetylcysteamine thioester (1 mM) (propionyl-SNAC) [24] was added. After incubation at 22°C for 72 hr with agitation, bacteria were removed by centrifugation, and supernatants were acidified with phosphoric acid to pH 2.5 and analyzed after at least 30 min for polyketide production by LC/MS/MS as previously described [11, 25].

Protein Expression Analysis

Samples (1 ml) of each culture were centrifuged at 14,000 × g for 3 min, resuspended in 1 ml 20 mM Tris, 150 mM NaCl (pH 7.5), and lysed by sonication. After 10 min of centrifugation at 14,000 × g, soluble fractions equivalent to a 10 μl cell suspension were separated on NuPAGE Novex 3%–8% Tris-acetate gels (Invitrogen), stained with Sypro-Red Staining (Molecular Probes), and quantified with a Typhoon scanner with BSA standards.

Activity of Bimodular Combinations

K207-3 bacteria harboring pAng donor plasmids and pBru acceptor plasmids were grown in 3 ml LB containing carbenicillin (50 μg/ml) and streptomycin (20 μg/ml) at 37°C to an OD_{600} of 0.5. Cultures were induced with IPTG (0.5 mM) and arabinose (2 mg/ml), and 0.5 ml of a mixture of sodium glutamate (50 mM), sodium succinate (50 mM), and sodium propionate (5 mM) was added. After 72 hr at 22°C, culture supernatants were analyzed for triketide lactone production by LC/MS/MS as previously described [11, 25].

Activity of Trimodular Combinations

K207-3 bacteria harboring pAngII donor plasmids and pCot acceptor plasmids were grown in 3 ml LB containing carbenicillin (50 μg/ml) and streptomycin (20 μg/ml) at 37°C to an OD_{600} of 0.5. Cultures were induced with IPTG (0.5 mM) and arabinose (2 mg/ml), and 0.5 ml of a mixture of sodium glutamate (50 mM), sodium succinate (50 mM), and sodium propionate (5 mM) was added. After 72 hr at 22°C, culture supernatants were acidified with phosphoric acid to pH 2.5 and analyzed after at least 30 min for tetyraketide production by LC/MS/MS.

Tetraketide Detection

Samples were analyzed by using a system consisting of a Leap Technologies HTC PAL sample handler, an Agilent 1100 HPLC pump, and an Applied Biosystems API-3000 triple quadrupole mass spectrometer equipped with a Turbo-ionspray source. For identification/characterization of tetraketides, samples (10 μl) were chromatographed on an Agilent Zorbax Eclipse XDB-C8 column (3.5 μm, 2.1 × 150 mm) at 250 μl/min by holding a mobile phase of 10% MeCN (0.1% HOAc) in H_2O (0.1% HOAc) for 3 min, followed by a linear gradient to MeCN (0.1% HOAc) over 9 min. Positive-ion product ion mass spectra of *m/z* 229 ($[M+H]^+$ of the tetraketides) were acquired with unit mass resolution in the first and third quadrupole by scanning a mass range of *m/z* 50–250 at a rate of 1 scan/second. Additional conditions were as follows: source temperature, 375°C; declustering and focusing potentials, 51 and 180 V, respectively; spray tip potential, 5000 V; and collision energy, 15 eV.

The same LC/MS system was used for detection of the tetraketides with different HPLC and mass spectrometry conditions. Samples (10 μl) were injected onto the same column, and its temperature was maintained at 45°C; the mobile phase used was a linear gradient from 10% MeCN (0.1% HOAc) in H_2O (0.1% HOAc) to MeCN (0.1% HOAc) over 10 min. Multiple reaction monitoring (MRM) in positive-ion mode was used for detection. The parent/daughter pairs of *m/z* 229/211, 229/193, 229/165, and 229/127 were each acquired with a dwell time of 200 ms and at unit resolution in the first and third quadrupoles. Additional conditions were: source temperature, 375°C; declustering and focusing potentials, 26 and 200 V, respectively; spray tip potential, 4600 V; and collision energy, 19 eV.

The tetraketides were identified by comparing their characteristic mass spectra to that of an authentic synthetic standard corresponding to the Pr-A-D-G-derived tetraketide. Diastereomers were readily distinguishable by the ratios of the relative intensities of the parent/daughter ions. Concentrations were estimated from the MRM data by comparing the area response of samples to that of the standard at a known concentration.

Supplemental Data

Supplemental Data include mass spectrometry results and are available at http://www.chembiol.com/cgi/content/full/14/2/143/DC1/.

ACKNOWLEDGMENTS

We thank David Hopwood, Chris Reeves, Gary Ashley, and Sumati Murli for the critical review of this manuscript and Chaitan Khosla for providing the synthetic standard for tetraketide analysis. This work was supported in part by the National Institute of Technology Advanced Program grant 70NANB2H3014.

Received: September 25, 2006
Revised: November 22, 2006
Accepted: December 4, 2006
Published: February 23, 2007

REFERENCES

1. Walsh, C. (2003). Antibiotics: Actions, Origins, Resistance (Washington, D.C.: ASM Press).
2. Reeves, C.D. (2003). The enzymology of combinatorial biosynthesis. Crit. Rev. Biotechnol. *23*, 95–147.
3. Khosla, C., Gokhale, R.S., Jacobsen, J.R., and Cane, D.E. (1999). Tolerance and specificity of polyketide synthases. Annu. Rev. Biochem. *68*, 219–253.
4. Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.
5. Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001). Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. Biochemistry *40*, 2326–2331.
6. 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.
7. Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F., and Weissman, K.J. (2003). The structure of docking domains in modular polyketide synthases. Chem. Biol. *10*, 723–731.
8. Weissman, K.J. (2006). The structural basis for docking in modular polyketide biosynthesis. ChemBioChem *7*, 485–494.
9. 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 ketosynthase domains. Chem. Biol. *13*, 469–474.
10. 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.
11. 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.
12. 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.
13. Xue, Q., Ashley, G., Hutchinson, C.R., and Santi, D.V. (1999). A multiplasmid approach to preparing large libraries of polyketides. Proc. Natl. Acad. Sci. USA *96*, 11740–11745.
14. Sherman, D.H. (2005). The Lego-ization of polyketide biosynthesis. Nat. Biotechnol. *23*, 1083–1084.
15. Kittendorf, J.D., and Sherman, D.H. (2006). Developing tools for engineering hybrid polyketide synthetic pathways. Curr. Opin. Biotechnol. *17*, 597–605.
16. 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.
17. Keatinge-Clay, A.T., Maltby, D.A., Medzihradszky, K.F., Khosla, C., and Stroud, R.M. (2004). An antibiotic factory caught in action. Nat. Struct. Mol. Biol. *11*, 888–893.
18. Keatinge-Clay, A.T., and Stroud, R.M. (2006). The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. Structure *14*, 737–748.
19. Tang, Y., Kim, C.Y., Mathews, I.I., Cane, D.E., and Khosla, C. (2006). The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. USA *103*, 11124–11129.
20. Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F., and Hutchinson, C.R. (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl. Environ. Microbiol. *71*, 4840–4849.

21. Salomon, C.E., Magarvey, N.A., and Sherman, D.H. (2004). Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. Nat. Prod. Rep. *21*, 105–121.

22. Sambrook, J., Maniatis, T., and Fritsch, E.F. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

23. 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.

24. Murli, S., MacMillan, K.S., Hu, Z., Ashley, G.W., Dong, S.D., Kealey, J.T., Reeves, C.D., and Kennedy, J. (2005). Chemobiosynthesis of novel 6-deoxyerythronolide B analogues by mutation of the loading module of 6-deoxyerythronolide B synthase 1. Appl. Environ. Microbiol. *71*, 4503–4509.

25. 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.