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†

Nicholas Wu,‡ David E. Cane,§ and Chaitan Khosla*,‡,||,⊥

Departments of Chemistry, Chemical Engineering, and Biochemistry, Stanford University, Stanford, California 94305, and Department of Chemistry, Brown University, Providence, Rhode Island 02912

Received November 26, 2001; Revised Manuscript Received January 31, 2002

ABSTRACT: 6-Deoxyerythronolide B synthase (DEBS) is the modular polyketide synthase (PKS) responsible for the biosynthesis of 6-dEB, the aglycon core of the antibiotic erythromycin. The biosynthesis of 6-dEB proceeds in an assembly-line fashion through the six modules of DEBS, each of which catalyzes a dedicated set of reactions, such that the structure of the final product is determined by the arrangement of modules along the assembly line. This transparent relationship between protein sequence and enzyme function is common to all modular PKSs and makes these enzymes an attractive scaffold for protein engineering through module swapping. One of the fundamental issues relating to module swapping that still needs to be addressed is the mechanism by which intermediates are channeled from one module to the next. While it has been previously shown that short linker regions at the N- and C-termini of adjacent polypeptides play an important role in mediating intermodular transfer, the contributions of other protein−protein interactions have not yet been probed. Here, we investigate the roles of the linker interactions as well as the interactions between the donor acyl carrier protein (ACP) domain and the downstream ketosynthase (KS) domain in various contexts. Linker interactions and ACP−KS interactions make relatively equal contributions at the module 2−module 3 and the module 4−module 5 interfaces in DEBS. In contrast, modules 2 and 6 are more tolerant toward substrates presented by nonnatural ACP domains. This tolerance was exploited for engineering hybrid PKS−PKS and PKS−NRPS (nonribosomal peptide synthetase) junctions and suggests fundamental ground rules for engineering novel chimeric PKSs in the future.

Modular polyketide synthases (PKSs)[1] are multienzyme assemblies responsible for the biosynthesis of numerous pharmacologically relevant natural products including the antibiotic erythromycin and the immunosuppressant FK506. As shown in the schematic diagram of the 6-deoxyerythronolide B synthase (DEBS) in Figure 1, the active sites of these enzymes are organized into distinct modules, each of which is responsible for elongating the polyketide chain by one ketide unit through the coordinated action of the three core active sites—a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). In addition to these three core active sites, there are a variable number of postcondensational active sites within each module—including a ketoreductase (KR), a dehydratase (DH), and an enoylreductase (ER)—that generate structural diversity in the final product. The growing polyketide chain is processively elongated as it passes through each of the modules in an assembly-line fashion such that the number of extensions is dictated by the number of modules in the enzyme system. The choices of building blocks made by each module and the number and types of domains within each module catalyzing postcondensation reactions dictate the chemical functionality at each carbon atom in the final product.

The unique organization of modular PKSs and the transparency of the functional code offer tremendous potential for the use of these enzyme systems as a scaffold for the generation of novel small molecules through combinatorial biosynthesis. Of all possible strategies for generating new natural product-like molecules, the fusion of intact modules from different sources (also referred to as "module swapping") presents one of the most appealing methods of generating new compounds. According to this strategy, since each module controls the functionality and stereochemistry of two adjacent carbon atoms, novel compounds can be generated by simply rearranging the order of modules along the assembly line. While there are a few examples of successful use of this strategy (*1*−*3*), it is still not clear what factors are important in mediating intermodular transfer, and how much of a role each factor plays.

It has been previously shown that while individual modules of DEBS have inherent specificities for small molecule substrates, these modules are still tolerant of a wide variety

† This research was supported by grants from the National Institutes of Health (CA 66736 to C.K. and GM 22172 to D.E.C.). N.W. is a recipient of a predoctoral fellowship from the Stanford-NIH Biotechnology Training Grant.

* Corresponding author. Phone/Fax: 650-723-6538. E-mail: ck@chemeng.stanford.edu.

‡ Department of Chemistry, Stanford University.

§ Department of Chemistry, Brown University.

|| Department of Chemical Engineering, Stanford University.

⊥ Department of Biochemistry, Stanford University.

[1] Abbreviations: 6-dEB, 6-deoxyerythronolide B; ACP, acyl carrier protein; AT, acyltransferase; DEBS, deoxyerythronolide B synthase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; KS, ketosynthase; NAC, *N*-acetylcysteamine; NRPS, nonribosomal peptide synthetase; PCP, peptidyl carrier protein; PKS, polyketide synthase.

10.1021/bi012086u CCC: $22.00
Published on Web 03/19/2002

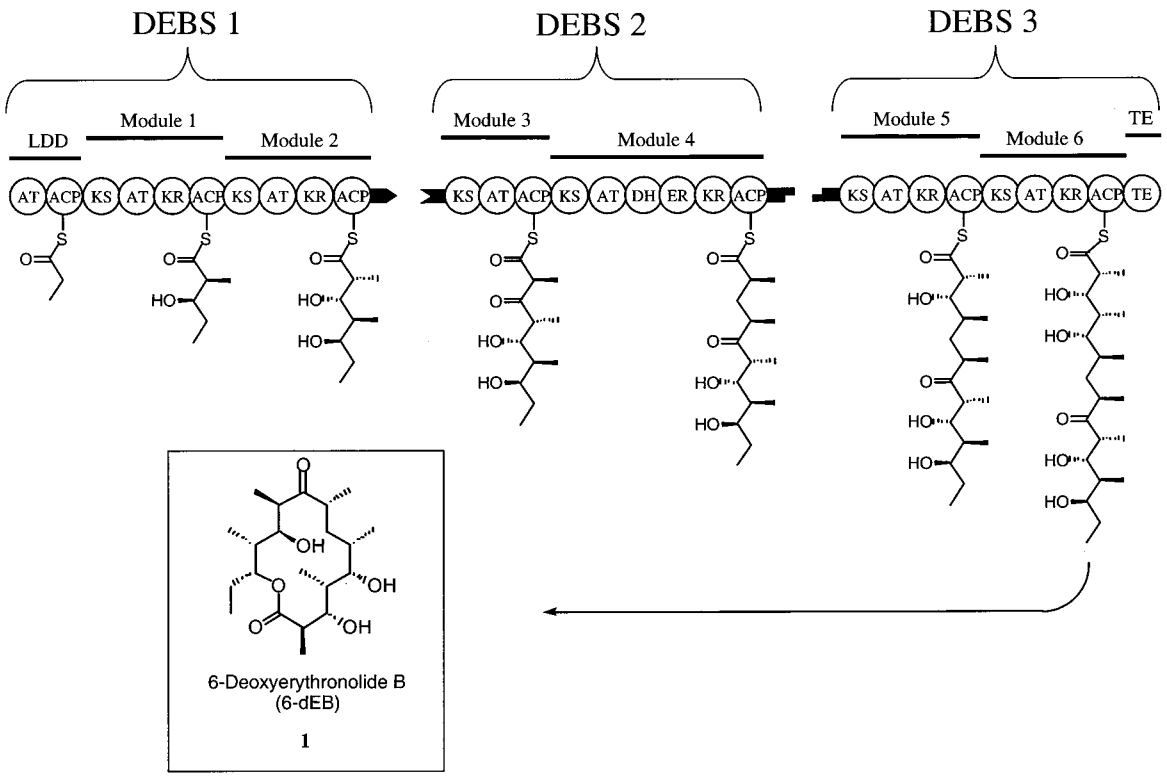

FIGURE 1: Deoxyerythronolide B synthase (DEBS) catalyzes the biosynthesis of 6-dEB (**1**), the aglycon precursor of the antibiotic erythromycin. DEBS is composed of three polypeptides—DEBS1, DEBS2, and DEBS3—each of which comprises two modules for a total of six modules (modules 1−6). Individual catalytic domains are represented by circles, and linker regions are represented by solid tabs between DEBS1 and DEBS2 and between DEBS2 and DEBS3. Each module contains three core catalytic domains—ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP)—as well as a variable number of optional domains—ketoreductase (KR), dehydratase (DH), and enoylreductase (ER). Polyketide biosynthesis is initiated by the action of the loading didomain (LDD) at the N-terminus of DEBS1, which primes the synthase with C_3-subunit derived from propionyl-CoA. Biosynthesis of **1** then proceeds in an assembly-line fashion such that the incoming polyketide chain is loaded onto the KS of an extending module from the ACP of the previous module. This is followed by a decarboxylative condensation reaction between the growing chain and a methylmalonyl-derived C3 extender unit that has been loaded onto the ACP by the AT. This C−C bond-forming reaction places the growing chain on the ACP, where it can then undergo unique functionalization catalyzed by KR, DH, and ER before being passed to the KS of the downstream module. This processive cycle of elongation and functionalization occurs until the penultimate intermediate reaches the thioesterase (TE), which catalyzes macrocyclization and product release to yield **1**.

of stereochemical variation in the substrate, and their inherent small molecule specificities do not appear to be limiting in intermodular transfer (*3*). By using a novel assay system in which small molecule diketide substrates were covalently attached to donor ACP proteins as shown in Figure 2A, comparisons of kinetic parameters of donor protein-assisted substrate loading (Figure 2B) versus diffusive substrate loading (Figure 2C) were made. These experiments indicated that a channeling mechanism not only increases the specificity of a module for a particular substrate by approximately 3−4 orders of magnitude but also can facilitate incorporation of otherwise poor substrates (*3*, *4*).

Linker regions at the N- and C-termini of each polypeptide interface (shown as matching tabs in Figure 1) have been previously identified as important factors for mediating specific channeling between polypeptides. Consisting of approximately 30−90 hypervariable residues, these linker regions have been suggested to form coiled-coils and have been shown to interact pairwise and specifically with each other (i.e., the C-terminal linker of module 2 interacts specifically with the N-terminal linker of module 3, and the C-terminal linker of module 4 interacts specifically with the N-terminal linker of module 5) (*1*, *5*). While the importance of these linker regions in mediating intermodular specificity has been demonstrated, other interpolypeptide interactions have not been ruled out. The most likely candidate for relevant intermodular interactions is the interface between the ACP domain of the upstream module and the KS domain of the downstream module, since these two active sites are involved in forming the tetrahedral intermediate of the transthioesterification reaction as the polyketide intermediate is channeled from one module to the next.

To evaluate the relative contributions of the linker interactions and the donor ACP−acceptor KS interactions, we used the assay system illustrated in Figure 2B. All donor proteins were loaded with the same (2*S*,3*R*)-2-methyl-3-hydroxypentanoyl thioester (hereafter referred to as "diketide"), which is derived from **2** and which has been shown to be a good substrate for DEBS modules 2, 3, 5, and 6 (*4*). Kinetic parameters relating to the substrate transfer, elongation, and release were measured in the presence of different combinations of donor ACPs, acceptor modules, and linkers. From these data, a distinct pattern emerged, providing the framework for basic ground rules for engineering novel PKSs by module swapping.

MATERIALS AND METHODS

Nomenclature. The nomenclature used in this report for proteins containing linker regions is identical to that used

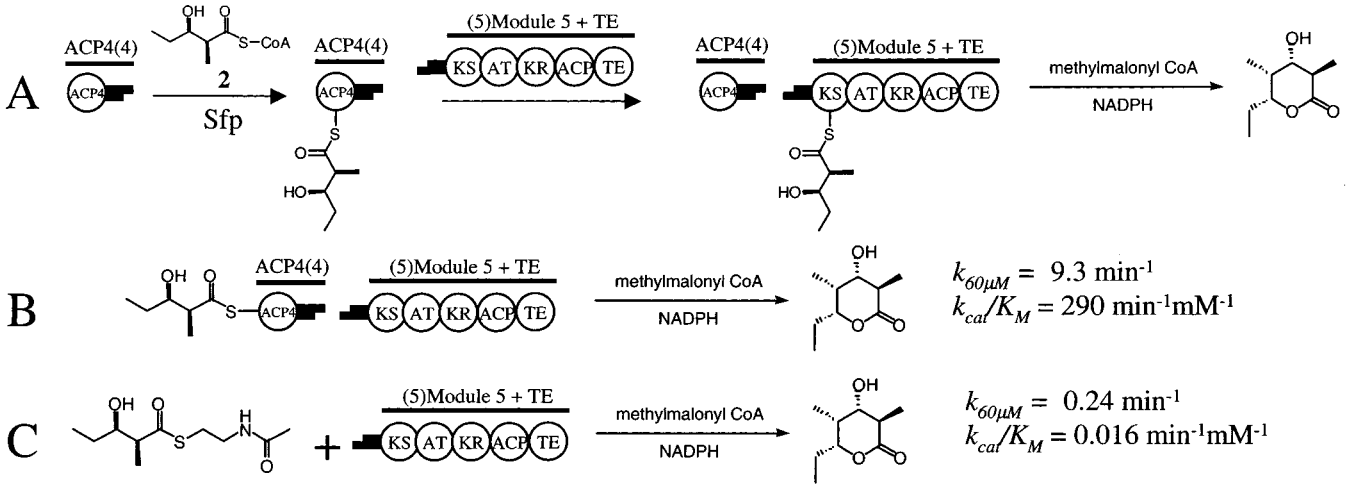

FIGURE 2: (A) Setup and mechanism for the intermodular transfer and elongation assay. Diketide-S-CoA is covalently attached to apo-ACP with the phosphopantetheinyl transferase *Sfp* to yield diketide-ACP. When this substrate is added to a module with complementary protein−protein interactions, the diketide is transferred to the KS of the acceptor module, where in the presence of methylmalonyl-CoA extender units it will be elongated one time and cyclized to release the six-membered triketide lactone. (B) Reaction of diketide-ACP4(4) + (5)M5+TE as previously reported (*3*). In this system, the diketide is channeled from the donor ACP4(4) protein to the acceptor (5)M5+TE protein. (C) Reaction of diketide-S-*N*-acetylcysteamine + (5)M5+TE as previously reported (*4*). In this reaction, the diketide is diffusively loaded onto the acceptor (5)M5+TE protein.

previously (*3*, *5*). Specifically, the module of origin of the linker is placed in parentheses either before or after the name of the domain or module to which it is attached, depending on whether it is an N- or a C-terminal linker, respectively. The boundaries of ACP domains, KS domains, and linkers are defined as before (*1*, *5*). For a protein whose linker region has been deleted, a null set symbol (∅) is placed in the parentheses. Accordingly, module 6 that has been engineered with the N-terminal linker from module 5 is represented as (5)M6; likewise, ACP2 with no linker regions is represented as ACP2(∅). If a thioesterase domain is fused to the C-terminal end of a module, it is indicated as such [e.g., (5)M5+TE].

Reagents and Chemicals. DL-[2-*methyl*-^{14}C]Methylmalonyl-CoA (56 mCi/mmol) was purchased from ARC, Inc. All other chemicals were purchased from Sigma-Aldrich. Buffer A: 100 mM NaH_2PO_4, 2.5 mM DTT, 1 mM EDTA, 20% glycerol, pH 7.1. Buffer B: 100 mM NaH_2PO_4, 10 mM imidazole, 1 M NaCl, 20% glycerol, pH 8.0. Buffer C: 400 mM NaH_2PO_4, 1 mM EDTA, 2.5 mM DTT, 20% glycerol, pH 7.1.

Construction of Plasmids. The construction of genes encoding (5)M2+TE, (3)M3+TE, (5)M5+TE, and (5)-M6+TE (pRSG64, pRSG34, pRSG46, and pRSG54, respectively) (*1*); (5)M3+TE (pST132) (*5*); ACP4(4) (pNW8) (*3*); eryLDD (pJL636) (*6*); and NovH(∅) (*7*) has been previously described. (3)M5+TE encodes a derivative of DEBS module 5 in which its natural N-terminal linker has been replaced with the N-terminal linker from module 3. The N-terminal linker of module 3 was excised from pRSG34 (*1*) [which encodes (3)M3+TE] as an *Nde*I−*Bsa*BI fragment. The resulting fragment was used to replace the corresponding *Nde*I−*Bsa*BI fragment in pRSG45, which encodes (5)-M5+TE (*1*), to yield pST133. ACP2(2) encodes the ACP domain of DEBS module 2 through its natural stop codon. This sequence was extracted from the gene cluster as an *Nde*I−*Eco*RI fragment by PCR using the primers 5′-C CAT ATG **CTG CGC GAC CGG CTG**-3′ and 5′-GAA TTC **TCA ATC GCC GTC GAG CTC C**-3′. ACP2(4) encodes the ACP domain of DEBS module 2 with its natural C-terminal linker replaced with the corresponding linker from module 4 using an engineered *Spe*I site at the junction. The ACP domain was obtained as an *Nde*I−*Spe*I fragment by PCR using the primers 5′-C CAT ATG **GTG GTC GAC CGG CTC G**-3′ and 5′-ACT AGT **GAG GAA ACC GGC GAC CG**-3′ (sequences complementary to DEBS shown in boldface type). Generation of the C-terminal linker region as an *Spe*I−*Eco*RI fragment by PCR has been previously described (*5*). These two fragments were cloned into pET28a to give pNW19. ACP2(∅) and ACP4(∅) encode the ACP domain of DEBS module 2 and module 4, respectively, with stop codons engineered at the end of the regions of homology. The coding regions were obtained as *Nde*I−*Eco*RI fragments by PCR using the primers 5′-C CAT ATG **CTG CGC GAC CGG CTG**-3′ and 5′-GAA TTC TTA **GCC GAG CTC GGC GTC**-3′ for ACP2(∅) and the primers 5′-C CAT ATG **GTG GTC GAC CGG CTC G**-3′ and 5′-GAA TTC TTA **GAA CAG CCT GTC CCG CAG**-3′ for ACP4-(∅). The PCR products were cloned into pET28a to afford pNW6 [ACP2(2)], pNW7 [ACP2(∅)], and pNW9 [ACP4-(∅)]. NovH(4) encodes the adenylation (A) and peptidyl carrier protein (PCP) domains of the NovH open reading frame (ORF) from the novobiocin pathway (*7*). It was fused to the C-terminal linker of module 4 of DEBS as follows. DNA encoding NovH was derived from pHC10 (*7*) as an *Nde*I−*Xho*I fragment. The linker region was obtained as an *Xho*I−*Bpu*1102I fragment using the primers 5′-CTG CTC GAG **AGG CTG TTC GCG GCC TCA**-3′ and 5′-C CCG CTG AGC **CTA CAG GTC CTC TCC CC**-3′. These two fragments were cloned into pET28a to yield pNW35.

Expression and Purification of Individual Modules. All previously characterized single modules were expressed and purified as previously described (*4*, *5*). (3)M5+TE (pST133) was expressed using a slightly modified version of the protocol used for previously characterized individual modules (*4*, *5*). This protein was expressed in *E. coli* BAP1 (*8*) in which the *sfp* phosphopantetheinyl transferase gene from *Bacillus subtilis* (*9*) has been inserted into the chromosome.

BAP1/pST133 cells were grown at 37 °C in LB media with 100 mg/L carbenicillin to an $OD_{600} = 0.5$, at which point they were cooled to 22 °C in a water bath and then induced with 0.7 mM IPTG for 12 h. The cells were harvested by centrifugation, washed with 50 mM Tris/1 mM EDTA (pH 8), and then resuspended in disruption buffer [100 mM NaH_2PO_4 (pH 7.2), 100 mM NaCl, 1.2 mM DTT, 1.2 mM EDTA, 0.7 mM benzamidine, 1 mg/L pepstatin, 1 mg/mL leupeptin, and 15% glycerol] before lysis by French Press (2×). After the cell debris was removed by centrifugation, the supernatant was treated with a 0.1% PEI precipitation followed by a 60% $(NH_4)_2SO_4$ precipitation for 2 h. The resulting $(NH_4)_2SO_4$ pellet was resuspended in buffer A (see Reagents and Chemicals section above for composition), flash-frozen in liquid nitrogen, and stored at −80 °C until ready for further purification. The crude protein was purified by FPLC on a hydrophobic butyl Sepharose column followed by a Resource Q anion exchange column as previously described (*4, 5*) to yield a 10 mg/L culture of purified (3)M5+TE.

Expression and Purification of ACP and PCP Proteins. Apo-ACP4(4) and apo-NovH(∅) were expressed in the *E. coli* strain BL21(DE3) and purified as previously described (*4, 7*). Apo-ACP2(2), apo-ACP2(∅), apo-ACP4(∅), apo-ACP2(4), and apo-NovH(4) were obtained by overexpression of pNW6, pNW7, pNW9, and pNW19, respectively, in the *E. coli* strain BL21(DE3). After growth in LB (50 mg/L kanamycin) at 37 °C to $OD_{600} = 0.5-0.7$, the cells were cooled in a water bath to 22 °C and then induced with 1 mM IPTG for 12 h at 22 °C. The cells were then harvested by centrifugation, washed with 50 mM Tris (pH 8), and then resuspended in buffer B before lysis by French Press (2×). The cell debris was cleared by centrifugation and the supernatant batch-loaded onto Ni NTA-agarose (Qiagen) resin (4 mL/L of culture) for 1 h. The resin was loaded into a Flex-column (Kontes) and washed with 10 column volumes of 35 mM imidazole in buffer B (see Reagents and Chemicals section above for composition), and then the desired N-terminal His_6-tagged proteins were eluted with 100 mM imizadole in buffer B. The appropriate fractions were concentrated, and the buffers were exchanged with buffer A (see Reagents and Chemicals section above for composition) + 1.5 M $(NH_4)_2SO_4$ in Centriprep spin columns (Amicon). Using an Akta FLPC system (Amersham Pharmacia Biotech AB), the concentrated protein was loaded at 1 mL/min onto a XK 16/20 column packed with 30 mL of Phenyl Sepharose High Performance resin and equilibrated with the same buffer. A gradient from 1 M $(NH_4)_2SO_4$ to 0 M $(NH_4)_2SO_4$ in buffer A was applied, resulting in the elution of the desired proteins between 150 and 0 mM $(NH_4)_2SO_4$. The appropriate fractions were concentrated and buffer-exchanged with buffer A in Centriprep spin columns to yield approximately 6 mg/L ACP2(2), a 15 mg/L culture of ACP2-(4), a 5 mg/L culture of purified ACP2(∅), and a 3 mg/L culture of ACP4(∅). These purified proteins were then flash-frozen in liquid nitrogen and stored at −80 °C. Expression and purification of apo-NovH(4) were performed under the same condition as described for the ACP proteins, except expression was induced with 0.1 mM IPTG at 15 °C. These conditions yielded a 25 mg/L culture of purified NovH(4) The masses of these proteins were confirmed by ESI-MS or MALDI-MS. The parent masses of the proteins were found in all cases. Mass peaks 178 daltons less than the parent masses were found in some cases, corresponding to loss of N-terminal *N*-formylmethionines: apo-ACP2(∅), observed mass = 12 073 Da (parent mass) and 11 895 Da (mass − formylmethionine), calculated mass = 12 027 Da; apo-ACP4(∅), observed mass = 11 917 Da (parent mass), calculated mass = 11 901 Da; apo-ACP2(2), observed mass = 20 532 Da (parent mass) and 20 354 Da (mass − formylmethionine), calculated mass = 20 495 Da; apo-ACP2(4), observed mass = 20 635 Da (parent mass) and 20 457 Da (mass − formylmethionine), calculated mass = 20 661 Da; apo-NovH(4), observed mass = 74 502 Da (parent mass) and 74 323 Da (mass − formylmethionine), calculated mass = 74 626 Da.

Chemoenzymatic Synthesis of Diketide-ACP and Diketide-PCP Substrates. The apo-PCP and apo-ACP proteins were converted to their respective diketide-ACP forms as previously described and as shown in Figure 2A (*3*). Briefly, phosphopantetheinylation of each active site serine residue was catalyzed by *sfp* in the presence of **2**, which was synthesized as previously described (*3*). The diketide-ACP/PCP substrates were either immediately used in the module substrate incorporation assays or used after purification by ion exchange chromatography. Purified protein concentrations were determined by Lowry assay. The masses as well as complete phosphopantetheinylation were confirmed by ESI-MS or MALDI-MS. A SDS−PAGE gel of the purified proteins is shown in Figure 3. Representative mass spectral data are shown in Figure 4C, illustrating the purity and the complete conversion from the apo-ACPs. The parent masses of the proteins were found in all cases. Mass peaks 178 daltons less than the parent masses were found in some cases, corresponding to loss of N-terminal *N*-formylmethionines. All observed parent masses are within the 1% error range that is expected from the spectrometers: diketide-ACP2(∅), observed mass = 12 528 Da (parent mass) and 12 350 (mass − formylmethionine), calculated mass = 12 498 Da; diketide-ACP4(∅), observed mass = 12 372 Da (parent mass), calculated mass = 12 353 Da; diketide-ACP2(2), observed mass = 20 809 Da (parent mass) and 20 988 Da (parent mass − formylmethionine), calculated mass = 20 947 Da; diketide-ACP2(4), observed mass = 21 089 Da (parent mass) and 20 910 Da (parent mass − formylmethionine), calculated mass = 21 113 Da; diketide-NovH(4), observed mass = 74 783 Da (parent mass), calculated mass = 75 078 Da.

Substrate Transfer and Elongation Assays. Qualitative assays were performed with a diketide-ACP or diketide-PCP substrate either taken directly from the *sfp* phosphopantetheinlyation reaction or taken after further purification of the substrate. These assays were performed with 20 μM diketide-ACP/PCP substrate for 2 h in the following reaction conditions: 1 μM acceptor module, 0.5 mM ^{14}C-methylmalonyl-CoA, 4 mM NADPH in buffer C, 30 °C. After quenching by addition of 250 μL of EtOAc and vortexing, the products were extracted with 2 × 250 μL of EtOAc, resolved on a silica gel TLC plate, and visualized on a Packard InstantImager. A representative TLC plate image is shown in Figure 8B.

Kinetic parameters were measured using purified diketide-ACP/PCP substrates in identical reactions conditions as described above for the qualitative assays. $k_{60\mu M}$ values were

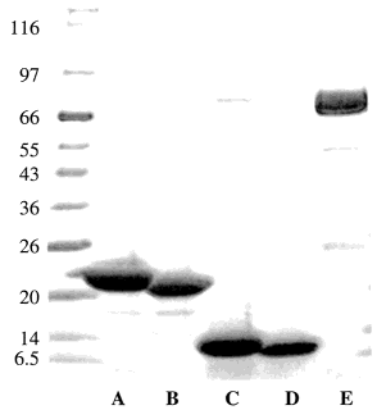

FIGURE 3: SDS−PAGE image of the purified protein substrates. Only proteins which have not been previously reported are shown. A protein ladder is shown in the left-most lane. Lane A: Diketide-ACP2(2). Lane B: Diketide-ACP2(4). Lane C: Diketide-ACP2-(∅). Lane D: Diketide-ACP4(∅). Lane E: Diketide-NovH(4).

measured using 60 μM diketide-ACP substrate. A representative time course is shown in Figure 6E. These reactions were quenched by the addition of 80 μL of 12.5% SDS to a 20 μL reaction mixture and immediate vortexing. The products were then extracted from the aqueous phase with 2 × 250 μL of EtOAc. After removing the organic solvents in vacuo, the residual products were then spotted onto a TLC plate (Baker-flex 250 μM silica gel), resolved in 60% EtOAc/40% hexanes, and the radioactive spots were visualized and quantified on a Packard InstantImager. As previously described (*3*), k_{cat}/K_M values were determined by competitive assay of the substrate of interest against a substrate with known k_{cat}/K_M parameters. Representative liquid scintillation counting data are shown in Figure 6F.

RESULTS

Protein Preparations. ACP4(4) (i.e., DEBS ACP4 with its natural C-terminal linker) (*3*) and eryLDD(∅) (i.e., the DEBS loading didomain with no C-terminal linker) (*6*) were constructed and expressed as previously described. ACP2-(2) includes the DEBS ACP2 domain and its natural C-terminal linker. The linker is defined as the sequence from the end of the ACP consensus sequence to the natural stop codon (*5*). ACP2(4) was constructed as a fusion protein between ACP2 and the C-terminal linker of ACP4. ACP2(∅) and ACP4(∅) are isolated ACP domains without linker regions. All proteins were expressed as N-terminally His_6-tagged apo proteins that could subsequently be purified by Ni-affinity chromatography to yield a 6 mg/L culture of ACP2(2), a 15 mg/L culture of ACP2(4), a 5 mg/L culture of purified ACP2(∅), a 3 mg/L culture of ACP4(∅), and a 25 mg/L culture of NovH(4). These proteins were converted to diketide-ACPs and diketide-PCP substrates by phosphopantetheinylation with *sfp* in the presence of **2**, as previously described (*3*). An SDS−PAGE gel of the purified protein substrates is shown in Figure 3. In addition, representative mass spectral data of diketide-ACP2(2) and diketide-ACP2-(4) are shown in Figure 4C to demonstrate quantitative phosphopantetheinlyation by *sfp*.

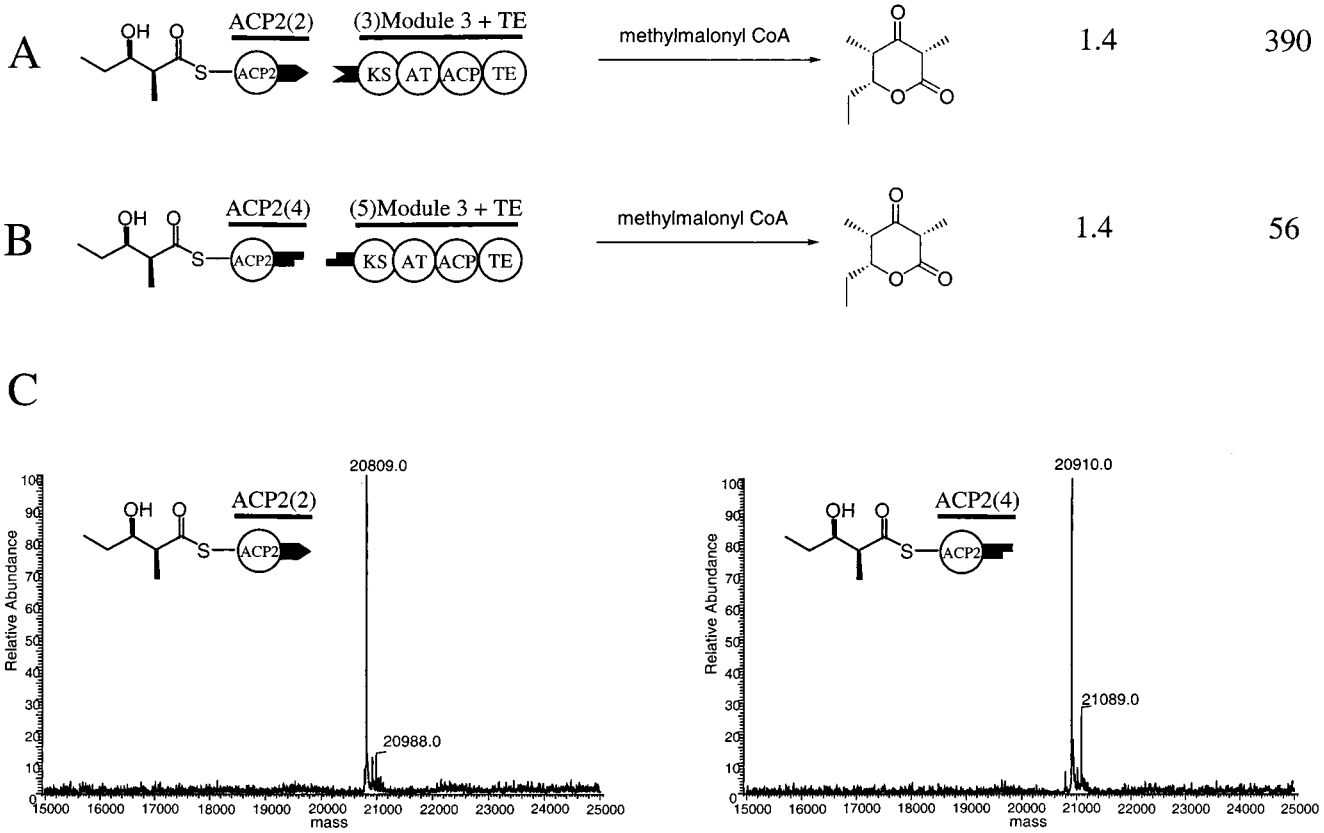

FIGURE 4: Modularity of the linker regions in the ACP2−module 3 interface. (A) Reaction of diketide-ACP2(2) + (3)M3+TE with the natural module 2−module 3 linker pair. (B) Reaction of diketide-ACP2(4) + (5)M3+TE with the alternate module 4−module 5 linker pair. (C) Representative mass spectra of the purified diketide−ACP substrates showing purity as well as complete conversion from the apo-ACPs. Data for diketide-ACP2(2) and diketide-ACP2(4) are shown here. In both cases, the major peak corresponds to the parent mass minus 177, indicating loss of *N*-formylmethionine.

(5)M2+TE, (3)M3+TE, (5)M3+TE, (5)M5+TE, and (5)M6+TE were constructed and expressed as previously described (*4*, *5*). pST133 encodes (3)M5+TE, which is a fusion protein of module 5 covalently attached to the thioesterase domain to facilitate turnover. In addition, the natural N-terminal linker of module 5 is replaced with the N-terminal linker of module 3. Expression and purification of this protein were carried out according to the previously reported protocol (*4*).

Analysis of the Modularity of Linker Regions. The linker regions have previously been suggested to be modular, or functionally independent (*1*, *5*). The kinetics of substrate transfer at the module 2−module 3 interface followed by elongation and product release were examined as a function of the $k_{60\mu M}$ and k_{cat}/K_M values of the overall reaction. The $k_{60\mu M}$ values reported here represent the apparent overall rate of product formation at an initial substrate concentration of 60 μM. In many cases, the $k_{60\mu M}$ values approximate the maximal overall turnover rates, as determined by back-calculating the K_M value for the reactions. True saturation kinetics were not practical because of the technical limitations (e.g., solubility) and limited supply associated with high molecular weight substrates such as diketide-ACP and diketide-PCP. k_{cat}/K_M values were determined by competitive assay of the substrate of interest against a substrate with a known k_{cat}/K_M value, as previously described (*3*). This method for determining k_{cat}/K_M values was chosen because it allowed us to conserve our limited supply of protein-based substrates compared with a direct measurement of the initial slope of a full v vs [S] plot. A representative time course and liquid scintillation counting data used to determine $k_{60\mu M}$ values and k_{cat}/K_M values are shown in Figure 6E,F, respectively.

In the first reaction, shown in Figure 4A, diketide-ACP2 and module 3 with their natural linker regions manifest $k_{60\mu M}$ and k_{cat}/K_M values of 1.4 min^{-1} and 390 min^{-1} mM^{-1}, respectively. When the module 4−module 5 linker pairs are transplanted into the module 2−module 3 interface as shown in Figure 4B, the $k_{60\mu M}$ value remains approximately the same, but the k_{cat}/K_M value decreases to 56 min^{-1} mM^{-1}. This comparison suggests that swapping out natural linker pairs for alternative linker pairs affects the K_M value of the transfer and elongation reaction, but not the maximum rate.

Analysis of the Relative Contributions of the Donor ACP, Acceptor KS, and Linkers to Chain Elongation. Various donor ACP−acceptor module pairs were examined for their ability to transfer substrates from the donor ACPs to the acceptor modules, which could then elongate and release triketide lactone product. Two sets of reactions were carried out—one in which the acceptor module was DEBS module 3 and the other in which the acceptor module was DEBS module 5. For each set of reactions, reactions were performed representing one of the following conditions: (A) matched linkers and matched donor ACP−acceptor KS pairs; (B) mismatched linkers and matched ACP−KS pairs; (C) matched linkers and mismatched ACP−KS pairs; or (D) mismatched linkers and mismatched ACP−KS pairs. As indicated by the formation of the expected triketide lactone product, transfer of diketide from the donor ACP to the acceptor module occurred at 20 μM substrate concentration in the reactions shown in Figure 5A−C and Figure 6A−C. These successful reactions represent conditions A−C (as defined above), and their kinetic parameters were further investigated. In contrast, no product was detected at the same substrate concentrations from the reactions in Figures 5D and 6D (representing condition D), indicating that transfer did not occur in the presence of both mismatched linkers and ACP−KS pairs. These qualitative data indicate the diketide substrate can be transferred to module 3 or 5 as long as either the linkers are matched or the ACP−KS pairs are matched.

To quantify the relative contributions of the linker pairs versus the ACP−KS pairs to the efficient channeling of substrates, $k_{60\mu M}$ and k_{cat}/K_M values were measured for the reactions shown in Figures 5A−C and 6A−C. The reactions of diketide-ACP2(2) + (3)M3+TE (Figure 5A) and diketide-ACP4(4) + (5)M5+TE (Figure 6A) manifest $k_{60\mu M}$ values of 1.4 and 9.3 min^{-1} and k_{cat}/K_M values of 390 and 290 min^{-1} mM^{-1}, respectively. In contrast to these reactions comprising matched linkers and matched ACP−KS pairs, the reactions in which either the linkers are mismatched or the ACP−KS pairs are mismatched (but not both) manifest significant and similar decreases in catalytic efficiencies and specificities. While the $k_{60\mu M}$ and k_{cat}/K_M values for the mismatched reactions shown in Figure 5B,C fell approximately 3−5-fold and 80−200-fold, respectively, the corresponding values for the mismatched reactions shown in Figure 6B,C fell approximately 20-fold and 150-fold, respectively. These data suggest that for both module 3 and module 5, the linker interactions and the donor ACP−acceptor KS interactions play significant and approximately equal roles in the channeling of substrates between modules.

Analysis of Chain Elongation by Various Acceptor Modules in the Presence of a Linkerless ACP4. Linker interactions were eliminated entirely from the transfer and elongation assays in the reaction of linkerless diketide-ACP4(∅) with (5)M2+TE, (5)M5+TE, (3)M5+TE, and (5)M6+TE. Formation of the expected triketide lactone was observed from the reactions of diketide-ACP4(∅) with (5)M5+TE and (3)M5+TE (Figure 7A,B), both of which contained matched ACP−KS pairs. The reaction shown in Figure 7A has $k_{60\mu M}$ and k_{cat}/K_M values of 0.49 min^{-1} and 4.1 min^{-1} mM^{-1}, respectively, and the reaction in Figure 7B has corresponding values of 0.27 min^{-1} and 2.5 min^{-1} mM^{-1}, respectively. These values are comparable to those observed when the linkers are mismatched and the ACP−KS pairs are matched (Figures 5B and 6B), indicating that, in this case, the presence of mismatched linkers and the deletion of complete linker pairs are kinetically equivalent.

ACP4(∅) was not able to efficiently transfer substrates to module 3, regardless of which N-terminal linker was covalently fused to the module (Figure 7C,D). This result was expected based on the above observation that channeling to module 3 is eliminated in the absence of both matched ACP−KS pairs and matched linker pairs. In contrast, transfer of diketide from ACP4(∅) to modules 2 and 6 was observed (Figure 7E,F, respectively), despite the elimination of linker interactions and the nonconsecutive ACP−KS pairs. By comparison to the kinetics parameters for the same reaction catalyzed by modules 2 and 6 in the presence of matched linkers (Figure 7G,H) (*3*), we note that the $k_{60\mu M}$ values drop approximately 10-fold and the k_{cat}/K_M values drop approximately 70−300-fold when the linker interactions are eliminated. These data suggest that modules 2 and 6 are

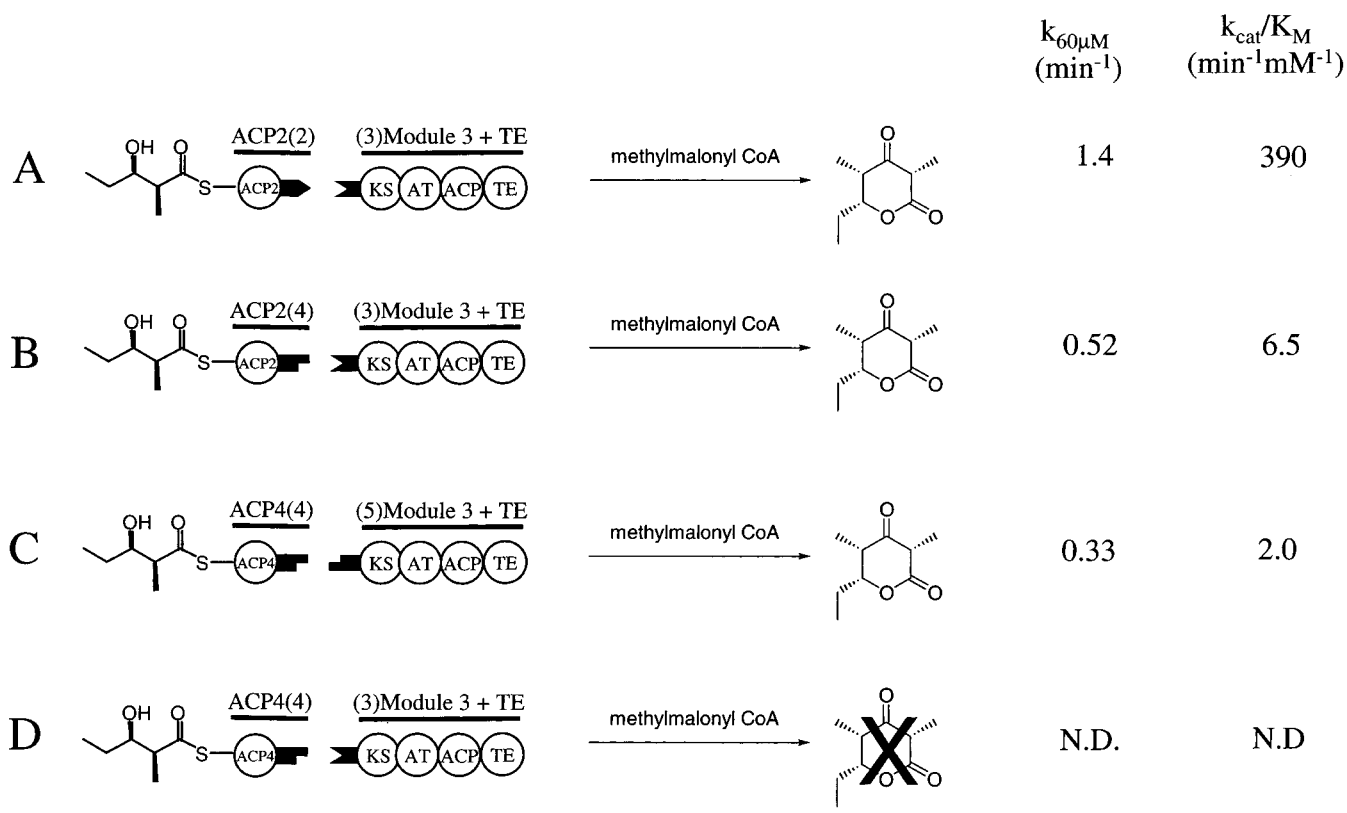

FIGURE 5: Schematic diagram and kinetic parameters of the four combinations of matched and mismatched linker regions and matched and mismatched ACP−KS pairs with module 3 as the acceptor module. (A) Reaction of diketide-ACP2(2) + (3)M3+TE with matched linkers and matched ACP−KS pairs. (B) Reaction of diketide-ACP2(4) + (3)M3+TE with mismatched linkers and matched ACP−KS pairs. (C) Reaction of diketide-ACP4(4) + (5)M3+TE with matched linkers and mismatched ACP−KS pairs. (D) Reaction of diketide-ACP4(4) + (3)M3+TE with mismatched linkers and mismatched ACP−KS pairs.

weakly, but demonstrably more tolerant to unnatural donor proteins than modules 3 and 5.

Tolerance of Modules 2 and 6 for Unnatural Donor Proteins. ACP2(∅), eryLDD(∅), NovH(∅), and NovH(4) were examined as potential donor proteins for the transfer of diketide to modules 2 and 6 (Figure 8A; a radio-TLC image is shown in Figure 8B). Reactions of these same ACPs and PCPs were also performed with (3)M3+TE, (5)M3+TE, (5)M5+TE, and (3)M5+TE. As predicted by previous experiments, substrate transfer from any of these linkerless donor proteins to module 3 or 5 was not observed (data not shown). In contrast, both ACP domains [ACP2(∅), eryLDD-(∅)] were able to channel the diketide substrate to both (5)-M2+TE and (5)M6+TE, despite the absence of matched linker pairs.

NovH(∅) is an adenylation-peptidyl carrier protein (A-PCP) didomain involved in the biosynthesis of the coumarin ring of novobiocin (*7*). This protein has no apparent C-terminal linker region as determined by sequence alignment and does not naturally interact with any known PKS domain in its role in novobiocin biosynthesis. In our assays, NovH(∅) was not able to transfer the diketide substrate to either (5)M2+TE or (5)M6+TE without the benefit of linker interactions. However, interaction between the NRPS-derived donor protein and PKS modules could be induced by engineering the C-terminal linker from DEBS module 4 onto the C-terminal end of NovH to create NovH(4). With the benefit of matched linker pairs, NovH(4) was able to channel the diketide substrate to module 2 with a $k_{60\mu M}$ value of 0.16 min^{-1} and a k_{cat}/K_M value of 3.5 min^{-1} mM^{-1} and to module 6 with a $k_{60\mu M}$ value of 0.53 min^{-1} and a k_{cat}/K_M value of 8.7 min^{-1} mM^{-1}. As the first demonstration of engineered interface involving the interaction of an NRPS domain that does not naturally interact with any PKS domains and a PKS domain that does not naturally interact with any NRPS domains, the experiment illustrates the power and utility of the linker regions for engineering artificial interpolypeptide junctions.

DISCUSSION

Understanding the factors that control the specificity of intermodular chain transfer is fundamental to the ability to rationally engineer novel polyketide synthases via module swapping. Among the factors to be considered are small molecule substrate specificity as well as protein−protein interactions between the donor and acceptor modules. It has been previously shown that while individual modules have defined specificities for small molecules, there is considerable tolerance toward less favored stereochemical configurations (*3*). In addition, 30−90 residue linker regions at the N- and C-termini of the bimodular polypeptides of DEBS have been identified and shown to contribute to the specificity of intermodular transfers between the two proteins (*1, 5*). While these linker regions are potentially powerful tools for enhancing specificity at engineered intermodular junctions, it is likely that other protein−protein interactions are involved in mediating the specificity of chain transfer. One of the most plausible candidates for relevant protein−protein interactions is the interaction between the ACP domain of the donor module and the KS domain of the acceptor module. These

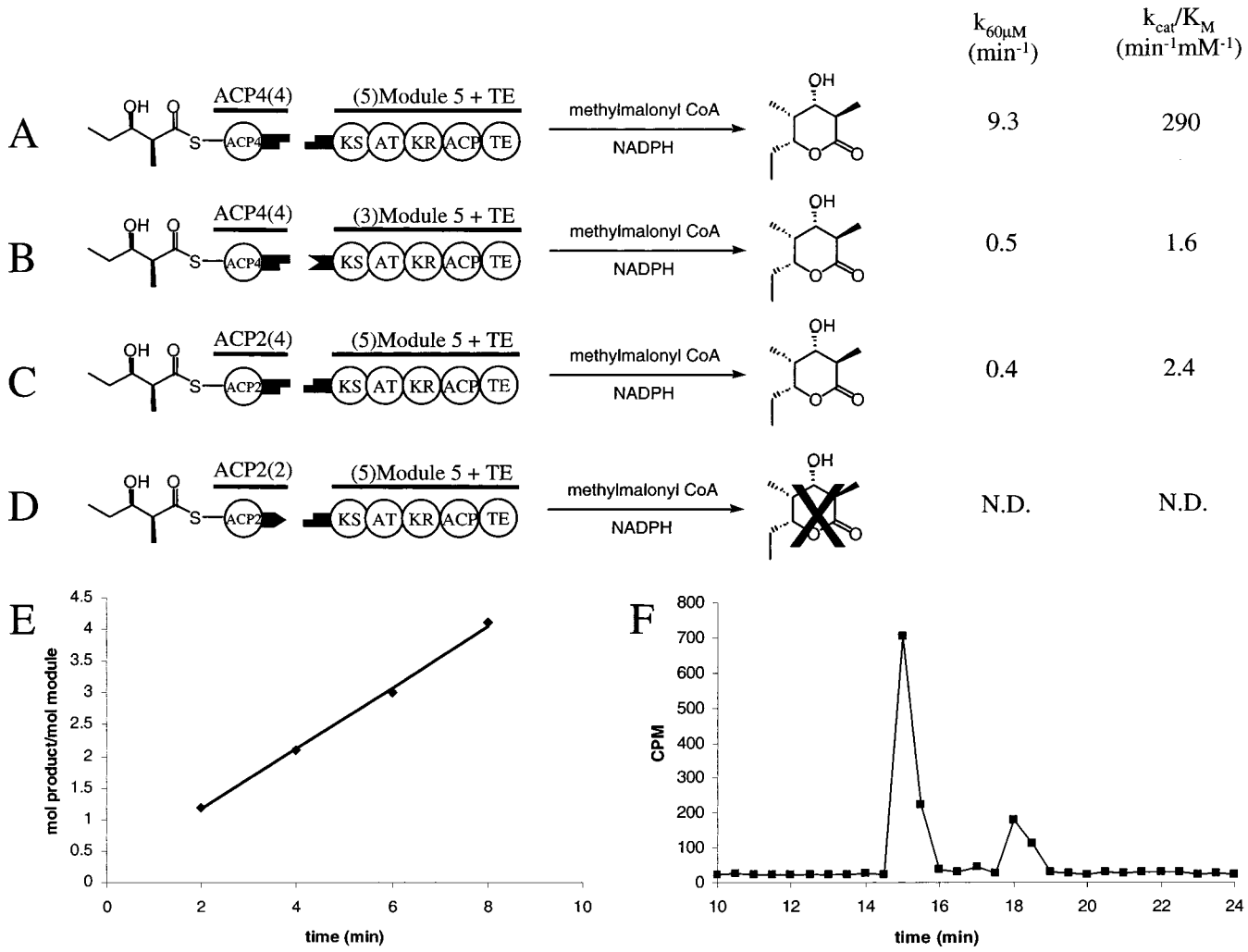

FIGURE 6: Schematic diagram and kinetic parameters of the four combinations of matched and mismatched linker regions and matched and mismatched ACP−KS pairs with module 5 as the acceptor module. (A) Reaction of diketide-ACP4(4) + (5)M5+TE with matched linkers and matched ACP−KS pairs. (B) Reaction of diketide-ACP4(4) + (3)M5+TE with mismatched linkers and matched ACP−KS pairs. (C) Reaction of diketide-ACP2(4) + (5)M5+TE with matched linkers and mismatched ACP−KS pairs. (D) Reaction of diketide-ACP2(2) + (5)M5+TE with mismatched linkers and mismatched ACP−KS pairs. (E) Representative time course used to determine $k_{60\mu M}$ values. The data here correspond to the reaction of diketide-ACP4(4) + (3)M5+TE. All reactions were performed in duplicate to confirm reproducibility. (F) Representative liquid scintillation counting data from the competitive assays used to determine k_{cat}/K_M values. The data shown here correspond to the reaction of 2 mM VDK-SNAC + 25 μM diketide-ACP4(4) + (3)M5+TE. The peak at 15 min corresponds to the product derived from VDK-SNAC [(2*S*,3*R*)-2-methyl-3-hydroxy-*S*-(*N*-acetylcysteamine)heptanethioate], and the peak at 18 min corresponds to the product derived from diketide-ACP4(4). By measuring the initial slope of a v vs [S] plot, the k_{cat}/K_M value for VDK-SNAC + (3)M5+TE was previously determined to be 0.078 min^{-1} mM^{-1} (data not shown). All reactions were performed in duplicate at different ratios of competing substrates to confirm reproducibility.

two domains presumably dock together as the substrate is channeled from the ACP to the KS domain via a tetrahedral transition state; therefore, a certain degree of spatial proximity can be inferred, suggesting the existence and relevance of additional protein−protein interactions at the ACP−KS interface.

Modularity of the linker regions is essential for their use in mediating unnatural interactions between modules from different sources. That is, engineering of the linker regions onto heterologous protein must be accompanied by a minimal kinetic penalty. To assess the modularity of the two linker pairs from DEBS (i.e., the linker pair at the module 2−module 3 interface and the linker pair at the module 4−module 5 interface), kinetic parameters describing the transfer from ACP2 to module 3 were determined for the two reactions in which each matched linker pair was inserted into the module 2−module 3 interface. Engineering of the heterologous module 4−module 5 linker pair into the module 2−module 3 junction had no effect on the maximal rate of transfer and elongation as compared to the natural module 2−module 3 linker pairs (Figure 4); this is consistent with a previous experiment examining the transfer over the same interface but using the full module 2 donor protein (*5*). However, replacing the natural linker pair with the heterologous linker pair increases the K_M for the ACP2−module 3 reaction by approximately 7-fold. The contrast between the uniformity of the $k_{60\mu M}$ term (which approximates the maximal rate) and the variability of the K_M term in the presence of different linker pairs suggests that swapping out the natural module 2−module 3 linker pair for the alternate module 4−module 5 linker pair perturbs only the initial association−dissociation of the ACP2 and module 3. When using the full module 2 protein, the increase in K_M value upon swapping in the alternate linker pair is a more modest

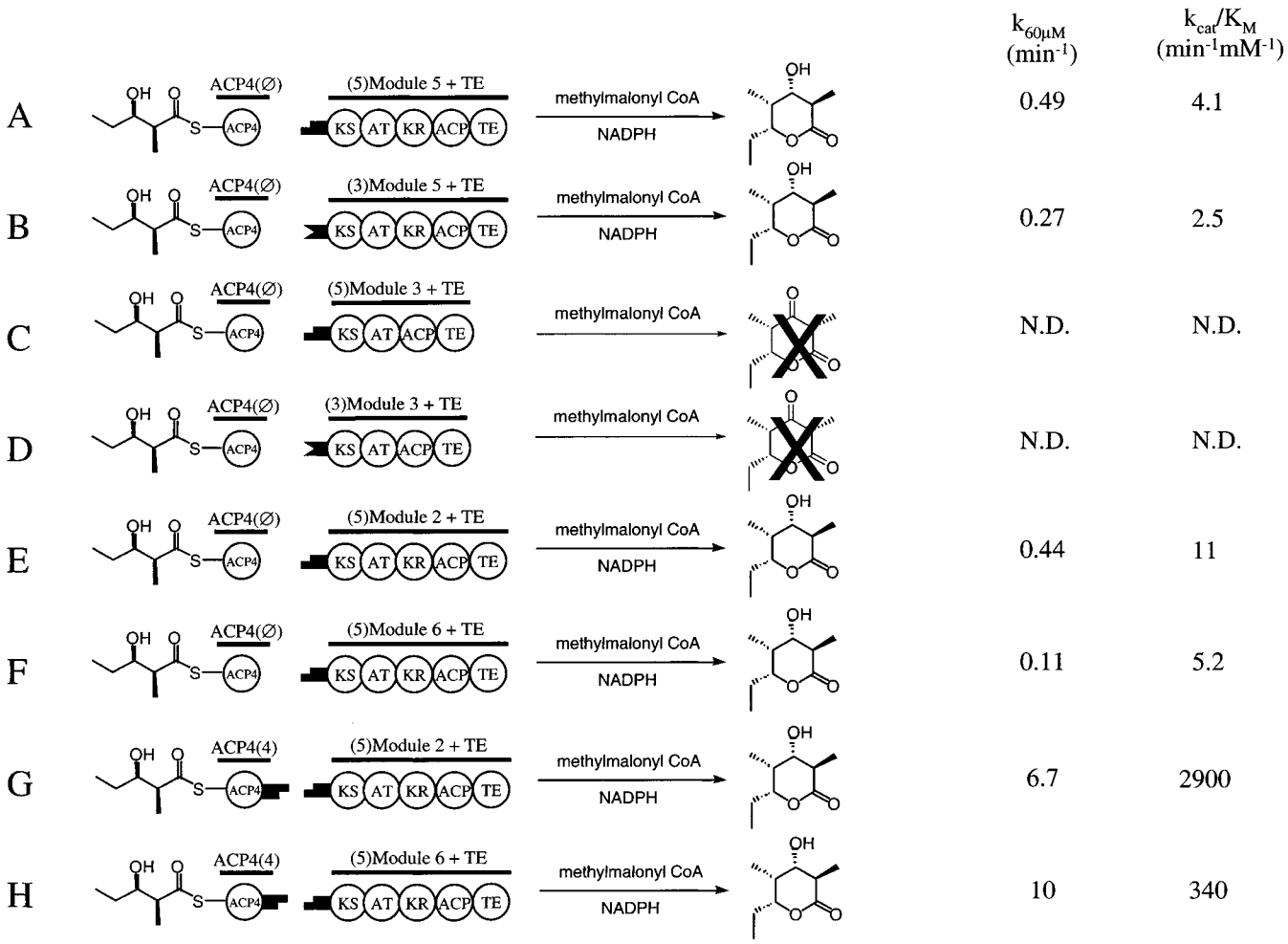

FIGURE 7: Linker-less ACP4(∅) as the donor protein. (A) Diketide-ACP4(∅) + (5)M5+TE. (B) Diketide-ACP4(∅) + (3)M5+TE. (C) Diketide-ACP4(∅) + (5)M3+TE. (D) Diketide-ACP4(∅) + (3)M3+TE. (E) Diketide-ACP4(∅) + (5)M2+TE. (F) Diketide-ACP4(∅) + (5)M6+TE. (G) Diketide-ACP4(4) + (5)M2+TE, shown for reference (*3*). (H) Diketide-ACP4(4) + (5)M6+TE, shown for reference (*3*).

2-fold, suggesting the more significant K_M effect when using isolated ACP2 may be an artifact of the truncated upstream protein.

To identify and quantify the relative contributions of various protein−protein interactions involved in mediating substrate channeling, we have replaced the linkers on two donor ACP domains (ACP2 and ACP4) as well as corresponding acceptor modules in a modified version of the minimal donor ACP system that had been previously developed (*3*). In two independent data sets using the N-terminal modules 3 and 5 as the acceptor modules, baseline kinetic parameters were first measured for reactions comprising both matched linkers and consecutive ACP−KS domains (Figures 5A and 6A). When either the linker regions or the donor ACP was swapped such that either the linker pairs or the ACP−KS domains were now mismatched, comparable attenuation of kinetic parameters was observed (Figures 5B,C and 6B,C), indicating that for modules 3 and 5, the ACP−KS interactions and linker interactions contribute comparably to the specificity of intermodular chain transfer.

The reactions of linkerless ACP4 [i.e., ACP4(∅)] with (5)M5+TE and (3)M5+TE (Figure 7A,B) demonstrated comparable kinetic parameters to the reactions between ACP4 and module 5 comprising mismatched linkers (Figure 6B). This indicates that eliminating linker interactions through mismatched linkers is kinetically comparable to eliminating linker interactions through physical deletion of the region. Furthermore, the kinetic effects observed in the mismatched linker reactions are probably a result of the elimination of protein−protein interactions rather than an artifact of protein engineering.

Whereas the KS domains of the N-terminal modules 3 and 5 are specific for their natural upstream ACP domains, the KS domains of the C-terminal modules 2 and 6 are promiscuous toward heterologous upstream ACP domains. ACP4(∅) was observed to be capable of transferring substrates to both (5)M2+TE and (5)M6+TE, despite the absence of matched linker interactions (Figure 7E,F). Kinetic analysis of these two reactions indicates that the attenuation of kinetic efficiency and specificity compared to the corresponding reactions comprising matched linkers [i.e., ACP4(4)+(5)M2+TE and ACP4(4)+(5)M6+TE] can be accounted for entirely by the elimination of linker interactions. The dichotomy between N-terminal modules (e.g., modules 3 and 5) and C-terminal modules (e.g., modules 2 and 6) of DEBS can perhaps be rationalized in the context of their natural positions in the assembly line. N-terminal modules such as DEBS modules 3 and 5 naturally accept incoming substrates from an upstream module on a different polypeptide. Therefore, built-in specificity for donor ACP domains would be highly advantageous for maintaining specific intermodular transfers. On the other hand, C-terminal

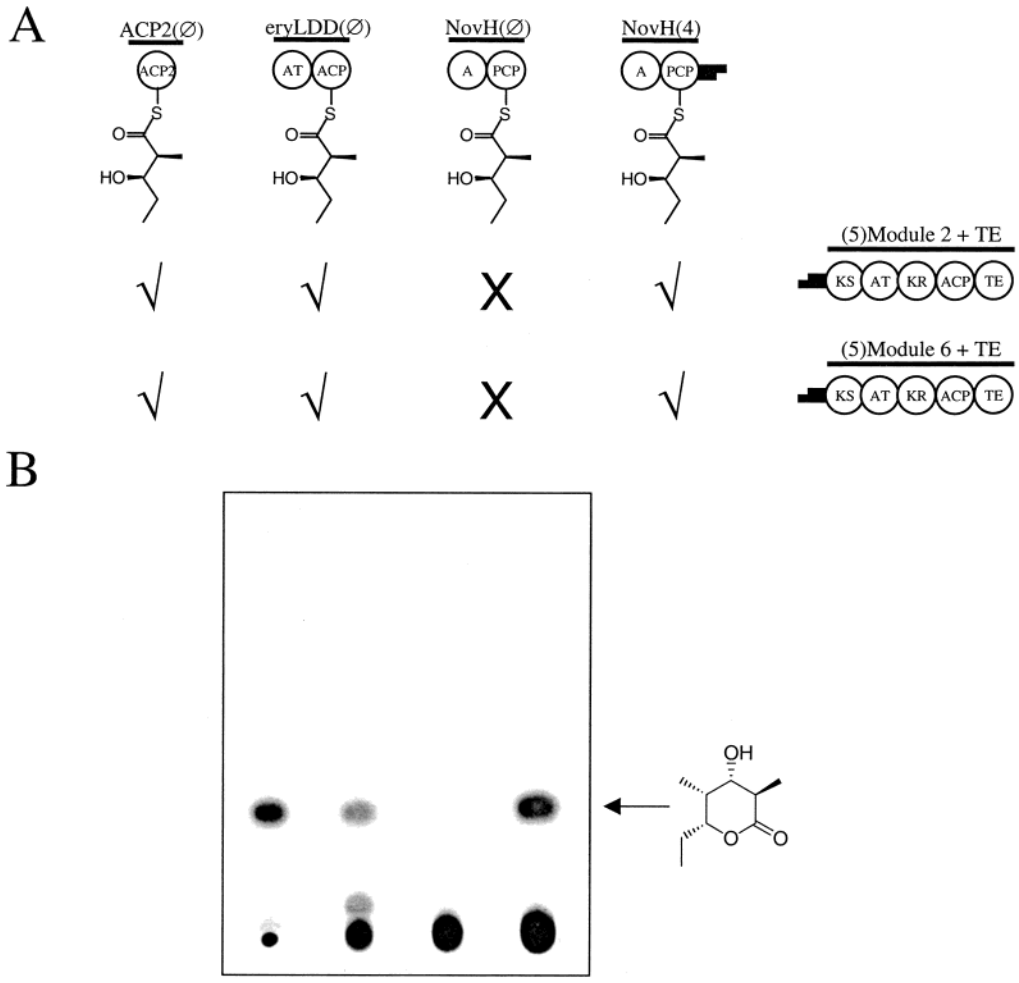

FIGURE 8: (A) Qualitative assessment of the ability of various donor proteins to transfer diketide substrates to modules 2 and 6. In the columns, going left to right: diketide-ACP2(∅), diketide-eryLDD(∅), diketide-NovH(∅), diketide-NovH(4). In the rows, going down: (5)M2+TE, (5)M6+TE. (B) Representative radio-TLC image of qualitative assays. From left to right, the lanes correspond to the reactions of diketide-ACP2(∅), diketide-eryLDD(∅), diketide-NovH(∅), and diketide-NovH(4) with (5)M2+TE. All reactions were performed under the conditions described under Materials and Methods. The heavy spots at the baseline correspond to methylmalonyl-CoA and propionyl-CoA (derived from decarboxylation of methylmalonyl-CoA) that were adventitiously extracted into the organic layers. The spot at R_f = 0.05 in the diketide-eryLDD(∅) reaction was not identified. The reactions of the same substrates with (5)M6+TE afforded similar raw data.

modules such as modules 2, 4, and 6 naturally accept incoming substrates from covalently attached upstream modules, making the specificity between the donor ACP and the acceptor module less essential.

The generality of the tolerance of modules 2 and 6 for unnatural donor ACP domains was elaborated using the linkerless, heterologous ACP domains ACP2(∅) and eryLDD-(∅). In all tested cases, channeling was observed even in the absence of matched linkers and consecutive ACP−KS pairs (Figure 8A). A natural extension of these observations is to explore the tolerance of these KS domains for peptidyl carrier protein (PCP) domains derived from nonribosomal peptide synthetases (NRPSs). While PCP and ACP domains share similar three-dimensional structural folds and are functionally analogous, the homology of PCPs to ACPs is generally relatively low (approximately 15−30%), contributing to very disparate surface polarities (*10, 11*). There are also numerous examples of hybrid NRPS−PKS gene clusters in which PCP domains transfer substrates to KS domains (*12*−*25*). NovH comprises adenylation (A) and peptidyl carrier protein (PCP) domains and is involved in the formation of the coumarin ring in the biosynthesis of novobiocin. As there are no PKS genes in the novobiocin gene cluster, it is assumed that this A-PCP didomain does not naturally interact with any PKS proteins during novobiocin biosynthesis. While NovH(∅) failed to channel substrates to (5)M2+TE or (5)M6+TE in the absence of matched linkers (Figure 8A), interaction between NovH and modules 2 and 6 could be effected by engineering the C-terminal linker of DEBS module 4 onto the end of NovH such that the resulting NovH(4) protein was capable of efficiently transferring substrates to modules 2 and 6. Although an artificial intrapolypeptide NRPS−PKS interface has previously been created by replacing the DEBS loading didomain with the rifamycin synthetase A-PCP loading didomain (*8*), the rifamycin A-PCP didomain naturally interacts with PKS domains on the same polypeptide, indicating that it may be inherently more amenable to engineering into alternate NRPS−PKS junctions. In contrast, this experiment with NovH(4) is to our knowledge the first example of engineering a functional NRPS−PKS interface involving an NRPS domain that does not naturally interact with any PKS proteins and a PKS domain that does not naturally interact with any NRPS proteins. While this

experiment biases the transfer reaction by eliminating the small molecule recognition component of a true NRPS−PKS transfer, it indicates that the heterologous linker regions are sufficient for inducing interaction between two naturally noninteracting proteins and illustrates the potential of these linker regions for future engineering of artificial interpolypeptide junctions.

The aggregate of these data begins to provide basic ground rules for the development of novel polyketide synthases via module replacement. As mentioned above, it has been previously demonstrated that linker pairs can be powerful tools for creating specificity in artificial interpolypeptide junctions (*1*, *5*). However, it is also essential to consider the origin of the modules in the engineered junction as well as the modules in any competing junctions. Whereas natural interpolypeptide junctions comprise a C-terminal module that channels substrates to an N-terminal module (represented as C → N), artificial junctions should be designed to represent one of the other three combinations (N → N, C → C, or N → C) in order to maximize specificity in the engineered assembly line. Practical module swapping experiments are underway to evaluate these proposals. In addition, complementary studies of intrapolypeptide junctions will provide additional and essential foundations for the development of modular PKSs as a scaffold for combinatorial biosynthesis.

ACKNOWLEDGMENT

We thank Professor Christopher Walsh for providing pHC10 and Dr. Stuart Tsuji for providing pST133 and (3)M5+TE.

REFERENCES

1. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) *Science 284*, 482−485.
2. 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) *Chem. Biol. 6*, 731−741.
3. Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) *J. Am. Chem. Soc. 123*, 6465−6474.
4. Wu, N., Kudo, F., Cane, D. E., and Khosla, C. (2000) *J. Am. Chem. Soc. 122*, 4847−4852.
5. Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) *Biochemistry 40*, 2317−2325.
6. Lau, J., Cane, D. E., and Khosla, C. (2000) *Biochemistry 39*, 10514−10520.
7. Chen, H., and Walsh, C. T. (2001) *Chem. Biol. 74*, 1−12.
8. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) *Science 291*, 1790−1792.
9. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) *Chem. Biol. 3*, 923−936.
10. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A., and Simpson, T. J. (1997) *Biochemistry 36*, 6000−6008.
11. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) *Struct. Fold. Des. 8*, 407−418.
12. Schwecke, T., Aparicio, J. F., Molnar, I., Konig, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortes, J., Lester, J. B., Bohm, G. A., and Staunton, J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7839−7843.
13. Beyer, S. J., Kunze, B., Silakowski, B., and Muller, R. (1999) *Biochim. Biophys. Acta 1445*, 185−195.
14. Cane, D. E., and Walsh, C. T. (1999) *Chem. Biol. 6*, R319−R325.
15. Paitan, Y., Alon, G., Orr, E., Ron, E. Z., and Rosenberg, E. (1999) *J. Mol. Biol. 286*, 465−474.
16. Bender, C. L., AlarconChaidez, F., and Gross, D. C. (1999) *Microbiol. Mol. Biol. Rev. 63*, 266.
17. Quadri, L. E. N. (2000) *Mol. Microbiol. 38*, 1−12.
18. Molnar, I., Schupp, T., Ono, M., Zirkle, R. E., Milnamow, M., NowakThompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., Gorlach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J., and Ligon, J. M. (2000) *Chem. Biol. 7*, 97−109.
19. Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000) *Gene 249*, 153−160.
20. Du, L. C., Sanchez, C., Chen, M., Edwards, D. J., and Shen, B. (2000) *Chem. Biol. 7*, 623−642.
21. Tillett, D., Dittmann, E., Erhard, M., vonDohren, H., Borner, T., and Neilan, B. A. (2000) *Chem. Biol. 7*, 753−764.
22. Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K., Ochi, K., and Shirai, M. (2000) *J. Biochem. 127*, 779−789.
23. Moffitt, M. C., and Neilan, B. A. (2001) *FEMS Microbiol. Lett. 196*, 207−214.
24. Huang, G. Z., Zhand, L. H., and Birch, R. G. (2001) *Microbiology (Reading, U.K.) 147*, 631−642.
25. Schwarzer, D., and Marahiel, M. A. (2001) *Naturwissenschaften 88*, 93−101.

BI012086U