The Loading Module of Rifamycin Synthetase Is an Adenylation−Thiolation Didomain with Substrate Tolerance for Substituted Benzoates†

Suzanne J. Admiraal,§ Christopher T. Walsh,|| and Chaitan Khosla*,§,⊥

Departments of Chemical Engineering, Chemistry, and Biochemistry, Stanford University, Stanford, California 94305-5025 and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received January 11, 2001; Revised Manuscript Received March 14, 2001

ABSTRACT: The rifamycin synthetase is primed with a 3-amino-5-hydroxybenzoate starter unit by a loading module that contains domains homologous to the adenylation and thiolation domains of nonribosomal peptide synthetases. Adenylation and thiolation activities of the loading module were reconstituted in vitro and shown to be independent of coenzyme A, countering literature proposals that the loading module is a coenzyme A ligase. Kinetic parameters for covalent arylation of the loading module were measured directly for the unnatural substrates benzoate and 3-hydroxybenzoate. This analysis was extended through competition experiments to determine the relative rates of incorporation of a series of substituted benzoates. Our results show that the loading module can accept a variety of substituted benzoates, although it exhibits a preference for the 3-, 5-, and 3,5-disubstituted benzoates that most closely resemble its biological substrate. The considerable substrate tolerance of the loading module of rifamycin synthetase suggests that the module has potential as a tool for generating substituted derivatives of natural products.

The rifamycin synthetase[1] of *Amycolatopsis mediterranei* is responsible for the biosynthesis of prosansamycin X, a precursor to the antibiotic rifamycin B (Figure 1). The rifamycin synthetase consists of a core of five multifunctional proteins, RifA, RifB, RifC, RifD, and RifE, in addition to RifF, a protein that is believed to cyclize the linear product of the other proteins via intramolecular amide formation (*1*−*4*). The five multifunctional proteins can be further subdivided into one nonribosomal peptide synthetase (NRPS)[2]-like loading module and 10 polyketide synthase (PKS) modules, based on sequence homology to other systems.

RifA, the N-terminal protein component of rifamycin synthetase, contains an NRPS-like module, the adenylation−thiolation (A−T) loading didomain, upstream of the first condensing module (Figure 1). The first such A−T type loading module was identified in the gene cluster for the natural product rapamycin (*5*). Complete gene clusters for other synthetases that contain hybrid modular interfaces have since been reported (*6*−*12*), and these synthetases produce hybrid natural products that are composed of both ketide and peptide units. The proven track record of polyketide and peptide natural products as therapeutics suggests that the increased combinatorial diversity embodied in hybrid products will advance drug discovery. A biochemical understanding of hybrid synthetases coupled with the ability to manipulate hybrid interfaces through protein engineering may enable the potential of such hybrid molecules to be realized.

The NRPS-like A−T didomain of RifA presumably primes the synthetase with 3-amino-5-hydroxybenzoate (AHB), which has been shown to be the precursor of the mC_7N structural element of rifamycin B (Figure 1) (*13*, *14*). However, the mechanism of this priming has not been established. Two alternative models can be envisioned. In a coenzyme A (CoA) ligase model present in the literature (*1*, *2*, *13*), the activated AHB-adenylate product of the A domain is attacked by CoA to generate an AHB-CoA intermediate, and the aryl thioester enzyme intermediate results from transthiolation onto the T domain (Figure 2A). In an alternative mechanism analogous to that used to prime NRPS modules, AHB is activated as the aryl-adenylate by the A domain, and the thiol of the phosphopantetheine cofactor of the T domain attacks AHB-adenylate directly to form a covalent aryl thioester enzyme intermediate (Figure 2B) (*3*).

Although AHB is the natural substrate of the A-T didomain, previous in vivo studies have revealed that RifA can be primed by the alternative substrates 3-hydroxybenzoate (3-HB) and 3,5-dihydroxybenzoate (*15*). We were intrigued by this innate substrate tolerance and its implications for the production of unnatural natural products. Therefore, we reconstituted the activity of the A−T didomain

† This work was supported by Grants AI38947 (C.K.) and GM20011 (C.T.W.) from the NIH. S.J.A. is a Fellow of the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation (DRG-1573).

* To whom correspondence should be addressed. Phone/FAX: 650-723-6538. E-mail: ck@chemeng.stanford.edu.

§ Department of Chemical Engineering, Stanford University.

|| Harvard Medical School.

⊥ Departments of Chemistry and Biochemistry, Stanford University.

[1] We refer to the protein complex responsible for biosynthesis of prosansamycin X, a rifamycin B precursor, as rifamycin synthetase because the results described herein establish that ATP is required for covalent attachment of the aryl starter unit to the loading module of the complex.

[2] Abbreviations: A, adenylation domain; AHB, 3-amino-5-hydroxybenzoate; B, benzoate; CoA, coenzyme A; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; 3-HB, 3-hydroxybenzoate; HPLC, high performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; k_{rel}, rate constant relative to analogous reaction with benzoate; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; SDS−PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; T, thiolation domain; Tris-HCl, tris(hydroxymethyl)-aminomethane hydrochloride; XB, generic substituted benzoate.

10.1021/bi010080z CCC: $20.00
Published on Web 04/20/2001

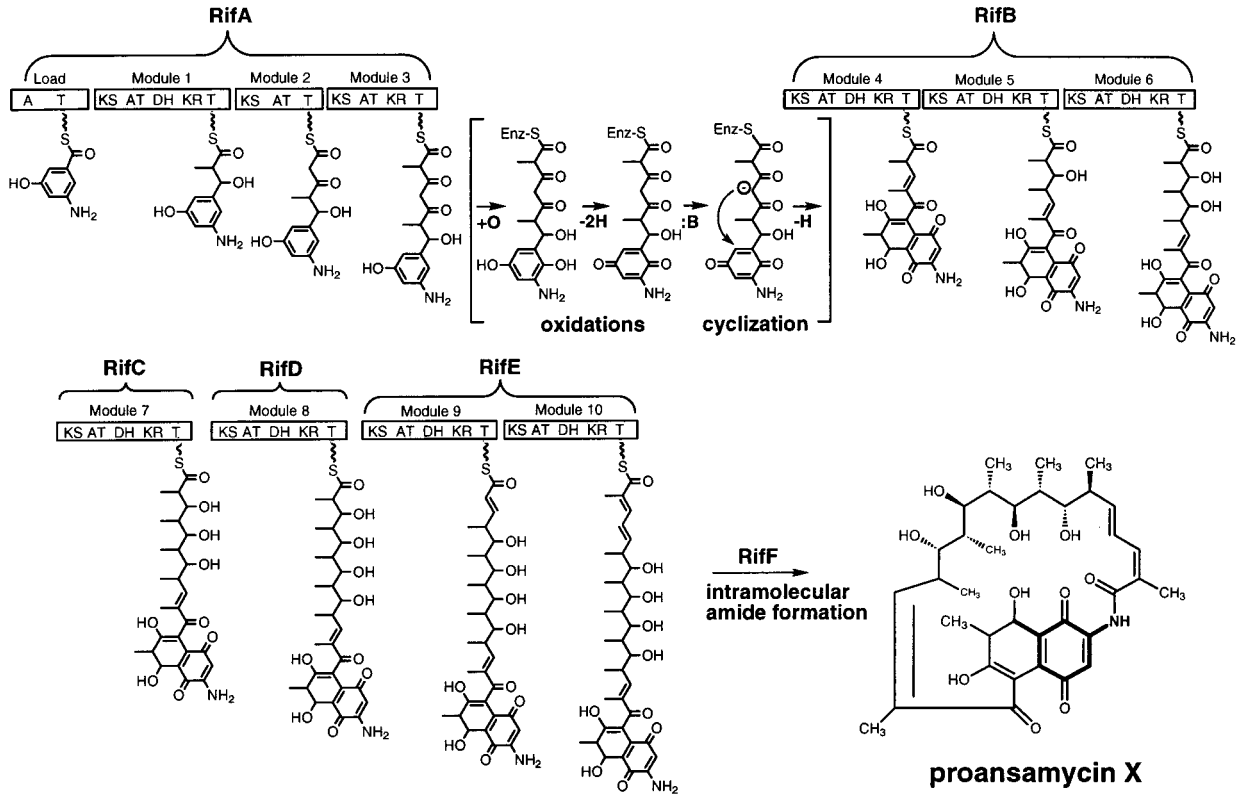


FIGURE 1: Proposed biosynthetic scheme for prosansamycin X, a precursor to rifamycin B. The rifamycin synthetase consists of a core of five large multifunctional proteins, RifA, RifB, RifC, RifD, and RifE, each containing one or more PKS modules. Each PKS module catalyzes one cycle of chain extension and associated β-ketoreduction for the biosynthesis of proansamycin X. The N-terminal A−T loading didomain of RifA primes the synthetase with AHB and is reminiscent of a minimal NRPS module. The location of the mC_7N unit derived from AHB is shown in bold in the proansamycin X structure. The active sites denote adenylation (A), thiolation (T), acyltransferase (AT), ketosynthase (KS), β-ketoreductase (KR), or dehydratase (DH) domains. As indicated, RifF is believed to catalyze cyclization via intramolecular amide formation.

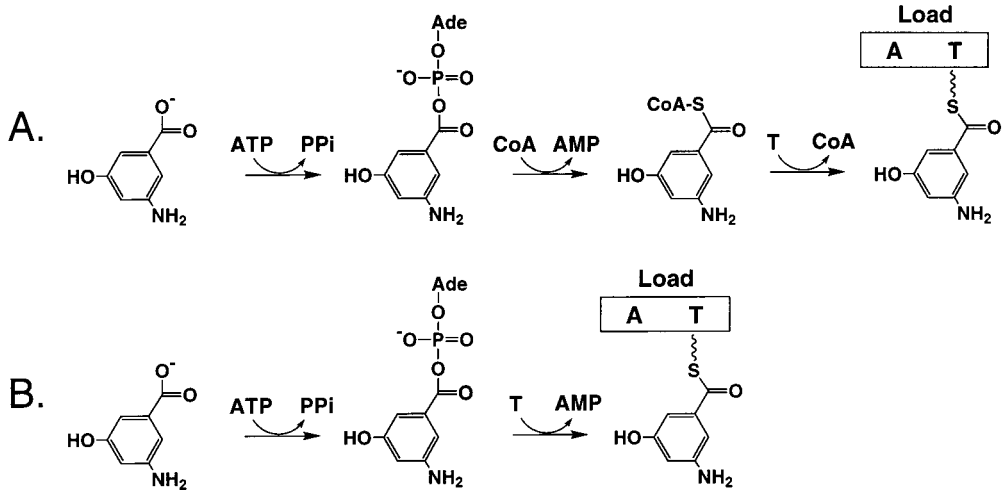


FIGURE 2: Possible mechanisms for the A-T loading didomain. (A) In the CoA ligase model, the activated AHB-adenylate product of the A domain is attacked by CoA to generate an AHB-CoA intermediate, and the aryl thioester enzyme intermediate results from transthiolation onto the T domain. (B) In the NRPS-like mechanism, AHB is activated as the AHB-adenylate by the A domain, and the thiol of the phosphopantetheine cofactor of the T domain attacks AHB-adenylate directly to form a covalent aryl thioester enzyme intermediate.

of rifamycin synthetase in vitro to establish the mechanism of this priming module and to systematically investigate its substrate tolerance.

EXPERIMENTAL PROCEDURES

Materials. [7-^{14}C]-Benzoic acid (57 mCi/mmol) and [7-^{14}C]-3-hydroxybenzoic acid (55 mCi/mmol) were obtained from American Radiolabeled Chemicals. All other substituted benzoic acids, phenylacetic acid, and 3-hydroxyphenylacetic acid were obtained from Aldrich in unlabeled form. ATP, CoA, and benzoyl-CoA were supplied by Sigma Chemical Company. AHB was synthesized according to a previously published protocol (*13*). Restriction enzymes were from New England Biolabs.

Manipulation of DNA and Strains. DNA manipulations were performed in *Escherichia coli* XL1 Blue (Stratagene)

using standard culture conditions (*16*). Polymerase chain reactions were carried out using Pfu polymerase (Stratagene) as recommended by the manufacturer.

Construction of an Expression Vector for the A-T Didomain. An *Nde*I restriction site was engineered at the start codon of the *rifA* gene using the primers 5′-GCGGC-**CATA**TGCGCACCGATCTC-3′ and 5′-AGGGCCCGCTG-GCGGGAGAAC-3′ (mutated bases are shown in bold, and the introduced *Nde*I restriction site is underlined); the amplified 2.5-kb fragment was ligated to linearized pCR-Script (Stratagene) to produce pHu29. The *rifA* gene with the engineered *Nde*I restriction site at the start codon was then reconstructed in pHu90-1, a derivative of pRM5 (*17*). Flanking restriction sites for *Pac*I and *Pst*I were used to transfer the sequence encoding the loading didomain and part of module 1 from pHu90-1 into a pUC18 derivative to produce pSA2. The loading didomain and module 1 are separated by an ~20 amino acid linker region, delineated by the C-terminal end of the consensus T domain of the loading didomain and the N-terminal end of the consensus ketosynthase domain of module 1 (GenBank accession no. AF040570). To isolate the loading didomain from module 1, a *Not*I restriction site was introduced into the linker sequence using the primers 5′-ACCGAGACCTGCGGGGC-GATCA-3′ and 5′-**GCGG**CCGCGACGGCCTGCGTG-3′ (mutated bases are shown in bold, and the introduced *Not*I restriction site is underlined); the resulting 0.94-kb fragment encodes from within the loading didomain into the linker region. This amplified fragment was ligated to linearized pCR-Blunt (Invitrogen) to produce pSA4, which was then digested with *Bam*HI and *Pst*I and ligated to pSA2 digested with the same enzymes to generate pSA6. The 1.9-kb *Nde*I−*Not*I fragment derived from pSA6 was ligated to *Nde*I−*Not*I-digested pET21c (Novagen) to produce pSA8, an expression vector for the loading didomain with hexahistidine appended to its C-terminus.

Expression and Purification of the A−T Didomain. Plasmid pSA8 was introduced via transformation into *E. coli* BL21 (Stratagene) for expression of the *apo* A−T didomain. One liter cultures of BL21/pSA8 were grown at 37 °C in 2 L flasks containing LB medium supplemented with 100 μg/mL carbenicillin. Expression of the A−T didomain was induced with 100 μM IPTG at an optical density at 600 nm of 0.7. After induction, incubation was continued for 6 h at 30 °C. The cells were then harvested by centrifugation at 2500*g* and resuspended in disruption buffer [200 mM sodium phosphate (pH 7.2), 200 mM sodium chloride, 2.5 mM DTT, 2.5 mM EDTA, 1.5 mM benzamidine, pepstatin (2 mg/L), leupeptin (2 mg/L), and 30% v/v glycerol].

All purification procedures were performed at 4 °C. The resuspended cells were disrupted by two passages through a French press at 13 000 psi, and the lysate was collected by centrifugation at 40000*g*. Nucleic acids were precipitated with polyethylenimine (0.15%) and removed via centrifugation. The supernatant was made 45% (w/v) saturated with ammonium sulfate and precipitated overnight. After centrifugation, the pellet containing protein was redissolved in 50 mM Tris-HCl (pH 8), 300 mM sodium chloride, 10 mM imidazole, and 10% v/v glycerol. This solution was loaded onto a previously equilibrated nickel-nitrilotriacetic acid (Ni-NTA) column (2 mL, Qiagen). The column was washed with 20 mM imidazole in 50 mM Tris-HCl (pH 8), 300 mM sodium chloride, and 10% v/v glycerol, and the A−T didomain was eluted with 100 mM imidazole in the same solution. Pooled fractions containing the A−T didomain were buffer exchanged into 100 mM sodium phosphate (pH 7.2), 2.5 mM DTT, 2 mM EDTA, and 20% v/v glycerol by gel filtration (PD-10, Pharmacia) and concentrated with a Centriprep-50 concentrator (Amicon). The purified protein was flash-frozen in liquid nitrogen and stored at −80 °C. Protein concentration was determined using the calculated extinction coefficient at 280 nm: 49 500 $M^{-1}cm^{-1}$ (*18*). A typical 1-L culture produced about 30 mg of purified protein.

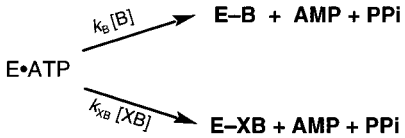
For expression of the *holo* A−T didomain, plasmid pSA8 was transformed into BL21 containing the plasmid pRSG56 (*19*), which carries a kanamycin resistance gene and the *sfp* gene. The *sfp* gene expresses Sfp, a nonspecific phosphopantetheinyl transferase from *Bacillus subtilis* that converts the *apo* protein into the *holo* protein (*20*, *21*). One liter cultures of this recombinant *E. coli* strain were grown at 37 °C in 2-L flasks containing LB medium supplemented with 100 μg/mL carbenicillin and 50 μg/mL kanamycin. The expression and purification steps for the *holo* A−T didomain were performed as described above for the *apo* A−T didomain.

Radioactive Labeling of the A−T Didomain. For qualitatively assessing the incorporation of B or 3-HB into the A−T didomain, reactions contained 5 μM *apo* or *holo* A−T didomain, 50 mM sodium phosphate (pH 7.2), 1 mM DTT, 1 mM EDTA, 15 mM $MgCl_2$, 10% glycerol, and 100 μM [7-^{14}C]-B or [7-^{14}C]-3-HB. In reactions where ATP was included, 5 mM was present. After incubation at 30 °C for 30 min, reactions were quenched with SDS−PAGE sample buffer and electrophoresed on a 4−15% gradient gel (Bio-Rad). The gel was briefly stained with Coomassie blue, destained, dried, and autoradiographed.

HPLC to Detect Possible Benzoyl-CoA Formation. Reactions contained 10 μM *apo* A−T didomain, 50 mM sodium phosphate (pH 7.2), 1 mM DTT, 1 mM EDTA, 15 mM $MgCl_2$, 5 mM ATP, 10% glycerol, 1 mM CoA, and 1 mM B. In reactions where benzoyl-CoA was included, 100 μM was present. After incubation at 30 °C for the indicated times, 20 μL samples were injected into an HPLC equipped with a C18 reverse phase column (VYDAC, 250 × 5 mm) with the detector monitoring at 254 nm. A linear gradient between buffer A (25 mM potassium phosphate, pH 5.4) and buffer B (100% acetonitrile) from 0 to 50% B was run over 14 min with a flow rate of 1 mL/min. The substrate and putative product peaks were identified by coinjection with authentic standards.

Kinetic Measurements. Typical reactions contained 1−10 μM *holo* A−T didomain, 50 mM sodium phosphate (pH 7.2), 1 mM DTT, 1 mM EDTA, 5 mM ATP, 15 mM $MgCl_2$, 10% glycerol, 0.5−5 μCi/mL [7-^{14}C]-B or [7-^{14}C]-3-HB, and varying concentrations of unlabeled B or 3-HB. Unlabeled B and 3-HB stocks were adjusted to the reaction pH prior to addition. Reactions were incubated at 30 °C, and at desired time points, 20 μL aliquots were quenched in 1 mL of ice-cold 5% trichloroacetic acid, and 200 μg of bovine serum albumin (Sigma) was added to this mixture to aid precipitation of the protein. The precipitate was pelleted by centrifugation, washed with 0.5 mL of 5% trichloroacetic acid and solubilized in 0.5 mL of a 100 mM phosphate (pH 8), 2% SDS solution. This solution was combined with 4.5 mL of

Scheme 1



$$k_{rel} = \frac{k_{XB}}{k_B} = \frac{[E\text{–}XB][B]}{[XB][E\text{–}B]}$$

(B = benzoate, XB = substituted benzoate)

liquid scintillation fluid (Formula 989, Packard), and the incorporated ^{14}C label, corresponding to E-B or E-3-HB, was quantified by liquid scintillation counting. Reaction rates were linearly dependent on enzyme concentration. Data analysis was performed using Kaleidagraph (Synergy Software), and exponential fits to the data typically gave $R \geq 0.99$.

Chase Experiment to Screen for Substrates of the A−T Didomain. Reactions were carried out in 50 mM sodium phosphate (pH 7.2), 1 mM DTT, 1 mM EDTA, 5 mM ATP, 15 mM $MgCl_2$, and 10% glycerol. Each reaction additionally contained 20 μM *holo* A−T didomain and 0.5 mM of a putative substrate, 0.5 mM unlabeled B, or no added substrate. After the sample was incubated for 30 min at 30 °C, 100 μL reaction aliquots were applied to individual G-25 microspin gel filtration columns (Pharmacia) that had been preequilibrated with the reaction buffer. The protein component of the applied sample was eluted from the microspin column in constant volume by centrifugation, according to the manufacturer's instructions. A 10 μL aliquot of each eluted protein sample was diluted with 2 μL of a [7-^{14}C]-B solution, for a final B concentration of 200 μM. These chase reactions were incubated for 15 min at 30 °C prior to analysis by SDS−PAGE autoradiography.

Determination of Relative Rate Constants for Arylation of the A−T Didomain. Reactions were performed as described above (*Kinetic Measurements*) but in the presence of 50 μM−5 mM of a series of substituted benzoates. Substituted benzoate stocks were adjusted to the reaction pH prior to addition. The relative rate constant (k_{rel}) for reaction of a given substituted benzoate with respect to B was determined from the concentrations of B and substituted benzoate in the original reaction ([B], [XB]) and the amount of product present as E−B and E-XB, according to the equation in Scheme 1 (*22*). The amount of E-XB product in each reaction at a given time point was determined by subtracting the amount of radiolabeled E-B in the presence of the competing substituted benzoate from that obtained at the same time point in an identical reaction lacking competitor. The ratio of E-B to E-XB was constant throughout a particular time course, indicating that no secondary reactions involving the reaction products were occurring. For each substituted benzoate, the same k_{rel} value, within error, was obtained for reactions performed at different substituted benzoate concentrations. The reactions were repeated for selected substituted benzoates using radiolabeled 3-HB instead of B, and the same k_{rel} values (with respect to B), within error, were obtained. Each k_{rel} value in Table 1 represents an average of at least four separate determinations.

Table 1: Relative Rate Constants for Covalent Loading of the A−T Didomain by Substituted Benzoates[a]

substrate	k_{rel}[b]
3-amino-5-hydroxybenzoate	120 ± 10
3,5-diaminobenzoate	16 ± 1
3-hydroxybenzoate	12 ± 2
3-aminobenzoate	6.6 ± 0.6
3,5-dibromobenzoate	4.1 ± 0.5
3,5-dichlorobenzoate	4.0 ± 0.5
3,5-dihydroxybenzoate	3.1 ± 0.5
3-chlorobenzoate	2.1 ± 0.2
3-bromobenzoate	1.9 ± 0.2
benzoate	(1)
2-aminobenzoate	0.62 ± 0.08
3-methoxybenzoate	0.43 ± 0.06
3-fluorobenzoate	0.42 ± 0.11
3,5-difluorobenzoate	0.13 ± 0.02
phenylacetate	<0.01
3-hydroxyphenylacetate	<0.01

[a] 30 °C, 50 mM sodium phosphate, pH 7.2, 1 mM DTT, 1 mM EDTA, 15 mM $MgCl_2$, 5 mM ATP, and 10% glycerol. [b] Rate constant for T domain arylation, relative to T domain arylation by benzoate (Scheme 1).

RESULTS

Construction and Purification of the A−T Loading Didomain. The A−T loading didomain is naturally present at the N-terminus of RifA. To investigate this didomain biochemically, it was removed from the RifA protein context. Therefore, the sequence encoding the isolated A−T didomain was subcloned into an expression vector, using an *Nde*I restriction site engineered at the transcriptional start site of RifA and a *Not*I restriction site introduced in the linker region between the C-terminal end of the consensus T domain and the N-terminal end of the consensus ketosynthase domain of module 1. Thiolation domains require covalent attachment of the 4′-phosphopantetheine moiety of CoA to a conserved serine to be active (*23*). The Sfp phosphopantetheinyl transferase from *B. subtilis*, which is capable of converting the *apo* forms of many heterologous recombinant proteins into the *holo* forms, was therefore coexpressed with the A−T didomain in the *holo* enzyme preparation (*20*, *21*). The *apo* and *holo* forms of the A−T didomain were produced in *E. coli* as C-terminal hexahistidine-tagged fusion proteins and were purified by nickel affinity chromatography to >98% homogeneity (Figure 3).

Mechanism of the A−T Didomain. As depicted in Figure 2, both models for the mechanism of the A−T didomain involve activation of AHB as the aryl-adenylate by the A domain, followed by eventual formation of a covalent aryl thioester enzyme intermediate from attack of either aryl-CoA (Figure 2A) or the aryl-adenylate (Figure 2B) by the thiol nucleophile of the phosphopantetheine cofactor of the T domain. To investigate these possible mechanisms, we sought to covalently load the A−T didomain. Although AHB is not available in radiolabeled form, in vivo feeding experiments have demonstrated that RifA can also be primed by 3-HB (*15*). Reactions containing [^{14}C]-3-HB or the putative substrate [^{14}C]-benzoate (B) and *apo* or *holo* A−T didomain were incubated in the presence or absence of Mg•ATP and subsequently analyzed by SDS−PAGE autoradiography (Figure 4). Lacking the phosphopantetheine cofactor, the *apo* A−T didomain could not be covalently loaded (lane 1). However, the *holo* A−T didomain is covalently loaded with

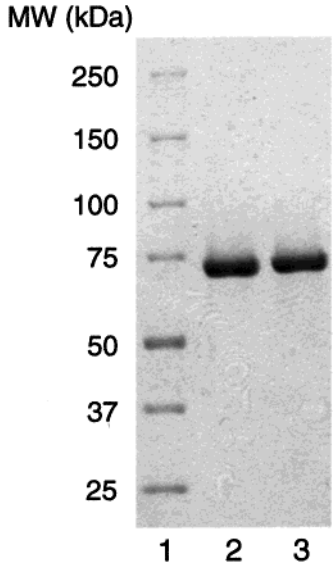


FIGURE 3: Purified recombinant *apo* and *holo* A-T didomain (encoded by plasmid pSA8) overproduced in *E. coli*. Protein samples were resolved by SDS−PAGE (4−15%, Bio-Rad) and stained with SimplyBlue Safestain (Invitrogen). Lanes 1−3 correspond to the molecular weight markers, the *apo* A−T didomain, and the *holo* A−T didomain, respectively.

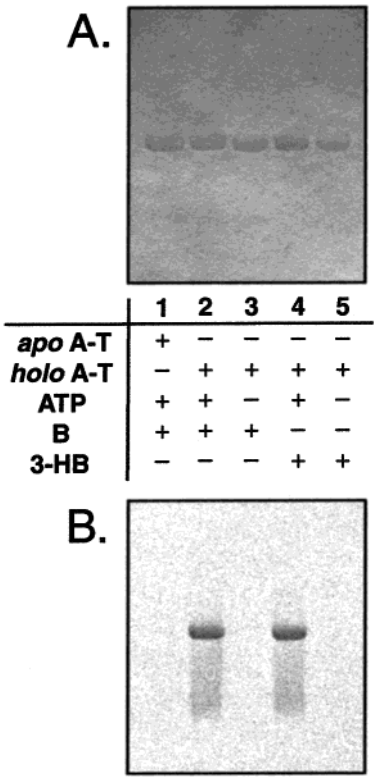


FIGURE 4: ATP-dependent covalent loading of the *holo* A−T didomain with B and 3-HB. (A) Coomassie-stained gel (4−15% gradient) of the reaction mixtures. (B) Autoradiograph of this gel. The presence or absence of *apo* or *holo* A−T didomain, ATP, and [^{14}C]-B or [^{14}C]-3-HB is indicated in the chart.

both B and 3-HB in reactions that require Mg•ATP (lanes 2−5).

CoA was not included in the labeling reactions described above, suggesting that it is not required for covalent loading of the *holo* A−T didomain. Since the loading didomain has been proposed to be a CoA ligase (Figure 2A) (*1*, *2*, *13*), we nevertheless tested the possible involvement of CoA directly. If the mechanism shown in Figure 2A is operative, the *apo* A−T didomain should be capable of producing benzoyl-CoA, as this intermediate would be expected to accumulate in the absence of the thiol nucleophile of the phosphopantetheine cofactor of the T domain. However, no benzoyl-CoA formation could be detected when the *apo* A−T didomain was incubated with ATP, B, and CoA (Figure 5, -benzoyl-CoA traces). To confirm that benzoyl-CoA, if formed, would persist in these reaction conditions, benzoyl-CoA was added to an otherwise identical reaction (Figure 5, +benzoyl-CoA traces). Benzoyl-CoA is degraded with an observed rate constant of ~0.002 min^{-1}, and this degradation is enzyme-independent since the same observed rate constant is obtained for reactions in which the *apo* A−T didomain is omitted (data not shown); this slow nonenzymatic degradation is taken into account in the k_{cat} analysis that follows.

Accumulation of 5 μM benzoyl-CoA is readily detectable using this HPLC assay (data not shown). This conservative detection limit allows an upper limit for k_{cat} for the formation of benzoyl-CoA by the *apo* A−T didomain to be calculated, as follows. Accumulation of 5 μM benzoyl-CoA would indicate that at most 10 μM benzoyl-CoA was formed during the 300-min reaction, as the half-life of benzoyl-CoA is ~300 min under these conditions ($t_{1/2} = \ln 2/k_{obs}$; $k_{obs} \approx 0.002$ min^{-1}). Therefore, the velocity of benzoyl-CoA formation is at most 0.03 μM/min (10 μM/300 min). This corresponds to $k_{cat} < 0.003$ min^{-1}, as the concentration of the *apo* A−T didomain in these reactions was 10 μM ($k_{cat} = v/[E]_t$). As described below, k_{cat} for covalent loading of the *holo* A−T didomain with B is 0.14 min^{-1}. Therefore, benzoyl-CoA is not a competent intermediate in the arylation reaction, as the rate constant for its formation is at least 50-fold less than the rate constant for formation of E−B. These results indicate that the CoA ligase model depicted in Figure 2A is not viable for the A−T loading didomain of rifamycin synthetase.

Direct Measurement of Kinetic Parameters for the Holo A−T Didomain. B and 3-HB are substrates for the *holo* A−T didomain, as shown qualitatively in Figure 4. To quantitatively assess these benzoates as substrates for aryl-adenylate formation followed by arylation of the thiol of the phosphopantetheine cofactor of the T domain, we utilized a protein precipitation assay. Aliquots from reactions containing *holo* A−T didomain, 0.5−5 μCi/mL [7-^{14}C]-B or [7-^{14}C]-3-HB, and varying concentrations of unlabeled B or 3-HB were quenched with trichloroacetic acid, and the amount of radiolabeled protein in each washed protein pellet was determined by liquid scintillation counting. Initial velocities of E-B or E-3-HB formation as a function of B or 3-HB concentrations were obtained using this method and used to generate the saturation curves shown in Figure 6. Best fits of the data to a saturation model give a k_{cat} of 1.9 min^{-1} and K_M of 180 μM for 3-HB, and a k_{cat} of 0.14 min^{-1} and K_M of 170 μM for B. The ratio of k_{cat}/K_M values for the two substrates reveals a 12-fold preference for 3-HB over B by the A−T didomain. Addition of CoA to these reactions had no effect (data not shown), consistent with the conclusion that the A−T didomain is not a CoA ligase.

Substrate Specificity of the A−T Didomain. On the basis of previous in vivo feeding experiments (*15*) and the in vitro results just described, AHB, 3-HB, B, and 3,5-dihydroxybenzoate are accepted as substrates by the A−T didomain. To screen for additional substrates that can prime the A−T

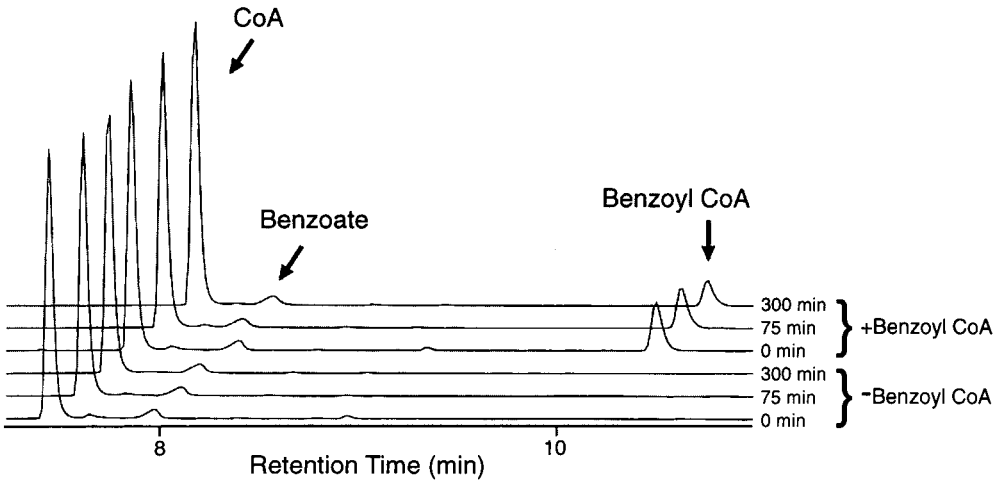


FIGURE 5: HPLC traces of time courses of reactions containing the *apo* A−T didomain. No net formation of benzoyl-CoA is observed. Labeled peaks were identified by coinjection with authentic standards of CoA, B, and benzoyl-CoA. The HPLC traces were shifted progressively by 0.15 min.

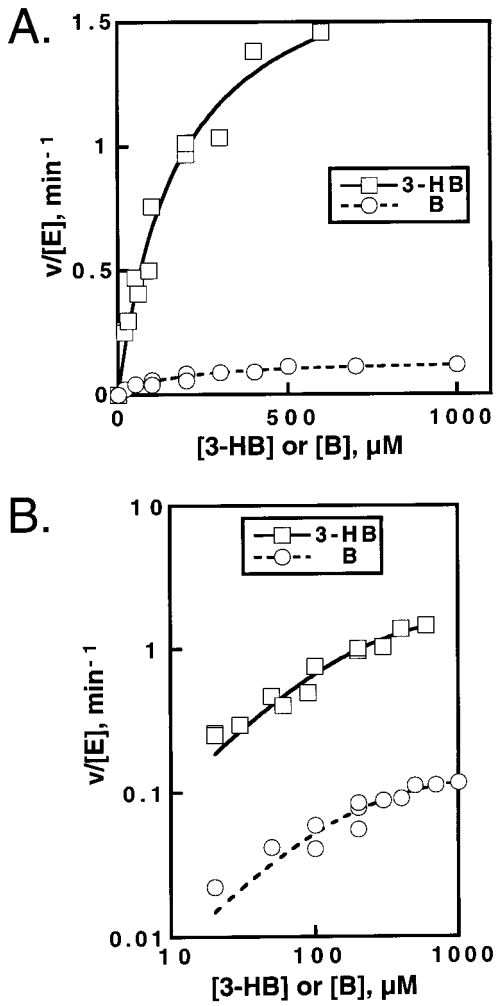


FIGURE 6: Saturation curves for covalent loading of the *holo* A−T didomain by 3-HB (□) or B (○). (A) Linear representation of the data. (B) Logarithmic representation of the data to facilitate evaluation of both data sets simultaneously. The lines are best fits of the data to a simple saturation model and give $k_{cat} = 1.9$ min^{-1} and $K_M = 180$ μM for 3-HB, and $k_{cat} = 0.14$ min^{-1} and $K_M = 170$ μM for B.

didomain, a simple chase experiment was devised. *Holo* A−T didomain was first incubated with a putative substrate under standard reaction conditions. The reaction mixture was

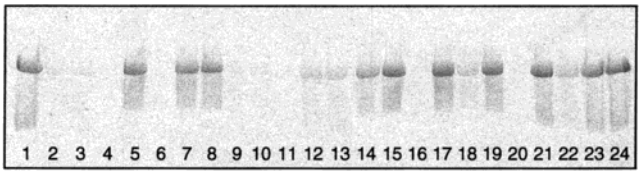


FIGURE 7: Substrate screening for the A−T didomain. Autoradiograph of a gel (4−15%, Bio-Rad) containing A−T didomain samples chased with radiolabeled B after incubation with no substrate (lane 1); unlabeled B (lane 2); 2-aminobenzoate (lane 3); 3-aminobenzoate (lane 4); 4-aminobenzoate (lane 5); AHB (lane 6); 3-amino-4-hydroxybenzoate (lane 7); 4-amino-2-hydroxybenzoate (lane 8); 3-bromobenzoate (lane 9); 3-chlorobenzoate (lane 10); 3,5-diaminobenzoate (lane 11); 3,5-dibromobenzoate (lane 12); 3,5-dichlorobenzoate (lane 13); 3,5-difluorobenzoate (lane 14); 2,3-dihydroxybenzoate (lane 15); 3,5-dihydroxybenzoate (lane 16); 3,5-dinitrobenzoate (lane 17); 3-fluorobenzoate (lane 18); 2-hydroxybenzoate (lane 19); 3-HB (lane 20); 4-hydroxybenzoate (lane 21); 3-methoxybenzoate (lane 22); 3-nitrobenzoate (lane 23); 3-sulfobenzoate (lane 24).

then passed over a microspin gel filtration column to separate the protein components from the putative unreacted substrate. Radiolabeled B was finally added to the protein fraction, and the mixture was incubated briefly prior to SDS−PAGE autoradiography. Protein samples that had originally been incubated with a substrate would contain covalently loaded E-XB, which would not react with radiolabeled B during the chase, resulting in little or no detectable E-B by SDS−PAGE autoradiography. In contrast, protein samples that had originally been incubated with a poor substrate or a non-substrate would primarily contain free E, which would readily react with radiolabeled B during the chase to form E−B, resulting in a radioactive band detectable by SDS−PAGE autoradiography.

The results of this screening experiment for a series of substituted benzoates are shown in Figure 7. The first two lanes contain control reactions in which no substrate (lane 1) or unlabeled B (lane 2) was present in the initial incubation; as expected, radiolabeled A−T didomain was formed in the no substrate control reaction but not in the unlabeled B control reaction. Radiolabeled A−T didomain is likewise absent from reactions in which the known substrates AHB (lane 6), 3,5-dihydroxybenzoate (lane 16), and 3-HB (lane 20) were present in the initial incubation. In

addition to these three substrates, 10 more likely substrates were identified for further investigation based on the absence or diminution of radiolabeled A−T didomain in Figure 7 as compared to the lane 1 control reaction. Although the simplest model for the absence of radiolabeled A−T didomain in a given reaction is that the substituted benzoate in question has been loaded onto the A−T didomain, blocking the enzyme from reaction with radiolabeled B during the chase, this experiment does not rule out the possibility that it is instead a tight binding competitive inhibitor. However, the observation described below that the competition between these substituted benzoates and the substrate B is time-independent renders the inhibition model unlikely.

Armed with this set of likely substrates, we investigated the relative specificity of the A−T didomain for aryl-adenylate formation followed by arylation of the thiol of the phosphopantetheine cofactor of the T domain. Addition of a substituted benzoate to a reaction mixture containing radiolabeled B and the *holo* A−T didomain allowed partitioning between reaction with the substituted benzoate and reaction with B to be followed. The relative rate constant (k_{rel}) for reaction of a given substituted benzoate with respect to reaction of B was determined from the concentrations of B and substituted benzoate in the original reaction ([B], [XB]) and the amount of product present as E-B and E-XB, according to the equation in Scheme 1 (*22*). The ratio of E-B to E-XB was constant throughout a particular time course, indicating that no secondary reactions involving the reaction products were occurring. The constant ratios also support the view that the substituted benzoates are true substrates and not high affinity competitive inhibitors, as E-B would continue to accumulate in the presence of a competitive inhibitor, resulting in a ratio of E-B to apparent E-XB that increases as a function of time. For each substituted benzoate, the same k_{rel} value, within error, was obtained for reactions performed at different substituted benzoate concentrations. The reactions were repeated for selected substituted benzoates using radiolabeled 3-HB instead of B, and the same k_{rel} values (with respect to B), within error, were obtained. Each k_{rel} value in Table 1 represents an average of at least four separate determinations. Competition with B for reaction with the A−T didomain by phenylacetate and 3-hydroxyphenylacetate could not be detected, so limits for k_{rel} for these compounds are reported in Table 1.

The k_{rel} values in Table 1 represent the k_{cat}/K_M ratio for a given substituted benzoate and B, and as such provide a measure of the specificity of the A−T didomain for each substrate (*22*). The validity of this approach is demonstrated by comparing the k_{rel} value of 12 obtained for 3-HB with the identical k_{cat}/K_M ratio of 12 obtained from direct measurement of k_{cat}/K_M for 3-HB and B (Figure 5). The A−T didomain exhibits a 10−1000-fold preference for AHB, its biological substrate, over all other substrates.

DISCUSSION

Mechanism of the Loading Module. The observation that CoA is not required for arylation of the T domain and that benzoyl-CoA is not a competent intermediate in this process establishes the loading module of rifamycin synthetase as an NRPS-like A−T didomain (Figure 2B) (*3*). This is in contrast to a previously proposed model of the loading module as a CoA ligase (*1*, *2*, *13*).

The conclusion that the loading module of rifamycin synthetase functions as an NRPS-like A−T didomain has implications for other sytems. Biosynthetic gene clusters for rapamycin (*24*), FK506 (*25*), ansatrienin (*26*), FK520 (*11*), microcystin (*10*), and pimaricin (*27*) all encode loading modules with homology to the A−T didomain of rifamycin synthetase. However, several of these systems have been proposed to be primed by an activated CoA substrate, presumably generated via a CoA ligase mechanism analogous to that shown in Figure 2A (*5*, *25*, *28*). A more likely mechanism for priming of these systems is the adenylation−thiolation mechanism operative for rifamycin synthetase.

Although the mechanisms shown in Figure 2 are distinct, the chemistries involved are essentially the same. In both cases, activation of AHB occurs via the aryl-adenylate, and the only difference is whether there is intermediate transfer of AHB to CoA prior to arylation of the T domain. Because the phosphopantetheine cofactor of the T domain is derived from CoA, the thiol nucleophiles of the T domain and CoA are chemically equivalent. Therefore, it is not difficult to envision how an enzyme could evolve from a CoA ligase into an A−T didomain, simply by covalent incorporation of the nucleophilic end of CoA as a phosphopantetheine cofactor. There is presumably an advantage to covalently tethering the aryl substrate moiety to the synthetase via the T domain instead of noncovalently binding it as the aryl-CoA. Nevertheless, aryl-CoA ligases are known to be involved in polyketide synthesis in the plant kingdom (see, for example, refs *29* and *30*), and benzoyl-CoA appears to be a substrate of the iterative type II PKS that produces enterocin (*31*).

Substrate Tolerance of the Loading Module. Prior to this investigation, AHB, 3-HB, and 3,5-dihydroxybenzoate were known to be substrates of the A−T didomain (*15*). Eleven additional substrates, including B, have been identified herein (Table 1). Previous work suggests that the substrate tolerance of the A−T didomain of rifamycin synthetase for alternative substituted benzoates is shared to a degree by related bacterial benzoyl-CoA ligases (*32*, *33*) and EntE (*34*), a stand-alone A domain that is a component of the enterobactin synthetase. These proteins are able to accept several alternative substituted benzoates, in addition to their biological substrates.

Although analysis of the substrate specificity results for the A−T didomain at a detailed molecular level awaits a crystal structure of this loading module, some preliminary observations can be made based on the substrate screening results in Figure 7 and the relative reactivity data in Table 1. With the exception of 2-aminobenzoate and B, only benzoates with 3-, 5-, or both 3- and 5-substituents are substrates for the A−T didomain. Binding sites that accommodate the 3-amino- and 5-hydroxy- substituents of the biological substrate AHB can apparently also accommodate alternative substituents at these positions. 3-Sulfobenzoate, 3-nitrobenzoate, and 3,5-dinitrobenzoate were likely rejected as substrates for steric reasons (Figure 7), since both sulfo- and nitro- substituents are significantly larger than the amino- and hydroxy- substituents of AHB. In this regard, it is surprising that 3-methoxybenzoate is accepted as a substrate, albeit a poor one, since the methoxy- substituent is also significantly larger than either substituent of AHB. The

3-fluoro- and 3,5-difluorobenzoates are discriminated against by factors of 5 and 30 with respect to their chlorinated and brominated counterparts (Table 1). Changes in the electronic properties of the aromatic ring upon fluorination may account for these differences. Phenylacetate and 3-hydroxyphenylacetate do not appear to be utilized as substrates by the A−T didomain, despite the reactivity of the corresponding benzoates, B and 3-HB (Table 1). This result suggests that the register of the carboxylate is a determinant of its reactivity, as the carboxylate of the phenylacetates is displaced by one methylene group relative to the benzoates. It should be noted that substituted benzoates were targeted as putative substrates in this study; the possibility that the tolerance of the A−T didomain for substituted benzoates extends to other types of aromatic substrates (e.g., heterocycles) remains to be tested.

The considerable substrate tolerance of the loading module of rifamycin synthetase for substituted benzoates has implications for the production of unnatural natural products through protein engineering. The endogenous loading module of 6-deoxyerythronolide B PKS was recently replaced by the loading module of the avermectin PKS, and the resulting hybrid synthase produced erythromycin derivatives that had incorporated branched starter units characteristic of the avermectin family (*35*). Similarly, it may be possible to exploit the priming promiscuity of the A−T didomain of rifamycin synthetase by appending it to other synthases or synthetases, with the goal of generating substituted derivatives of the original products.

Finally, this initial characterization of the loading module of rifamycin synthetase as an NRPS-like A−T didomain sets the stage for investigation of the hybrid NRPS/PKS interface in this system. Biochemical studies that combine the NRPS-like loading module and PKS module 1 of rifamycin synthetase (in cis or in trans) should allow functional and structural questions regarding NRPS/PKS biosynthetic interfaces to be addressed.

ACKNOWLEDGMENT

We thank Zhihao Hu for plasmid construction and Daniel Hunziker for synthesis of AHB.

REFERENCES

1. Schupp, T., Toupet, C., Engel, N., and Goff, S. (1998) *FEMS Microbiol. Lett. 159*, 201−207.
2. August, P. R., Tang, L., Yoon, Y. J., Ning, S., Müller, R., Yu, T.-W., Taylor, M., Hoffmann, D., Kim, C.-G., Zhang, X., Hutchinson, C. R., and Floss, H. G. (1998) *Chem. Biol. 5*, 69−79.
3. Tang, L., Yoon, Y. J., Choi, C.-Y., and Hutchinson, C. R. (1998) *Gene 216*, 255−265.
4. Floss, H. G., and Yu, T.-W. (1999) *Curr. Opin. Chem. Biol. 3*, 592−597.
5. Schwecke, T., Aparicio, J. R., Molnár, I., König, A., Khaw, L. E., Haydock, S. F., Oliykyk, M., Caffrey, P., Cortés, J., Lester, J. B., Böhm, G. A., Staunton, J., and Leadlay, P. F. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7839−7843.
6. Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T., and Perry, R. D. (1998) *Chem. Biol. 5*, 573−586.
7. Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H., and Walsh, C. T. (1998) *Chem. Biol. 5*, 631−645.
8. Silakowski, B., Schairer, H. U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blöcker, H., Höfle, G., Beyer, S., and Nüller, R. (1999) *J. Biol. Chem. 274*, 37391−37399.
9. Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000) *Gene 249*, 153−160.
10. Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B. A. (2000) *Chem. Biol. 7*, 753−764.
11. Wu, K., Chung, L., Revill, W. P., Katz, L., and Reeves, C. D. (2000) *Gene 251*, 81−90.
12. Du, L., Sánchez, C., Chen, M., Edwards, D. J., and Shen, B. (2000) *Chem. Biol. 7*, 623−640.
13. Ghisalba, O., and Nüesch, J. (1981) *J. Antibiot. 34*, 64−71.
14. Anderson, M. G., Monypenny, D., Rckards, R. W., and Rothchild, J. M. (1989) *J. Chem. Soc. Chem. Commun.* 311−313.
15. Hunziker, D., Yu, T.-W., Hutchinson, C. R., Floss, H. G., and Khosla, C. (1998) *J. Am. Chem. Soc. 120*, 1092−1093.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
17. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., and Khosla, C. (1993) *Science 262*, 1546−1550.
18. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
19. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) *Science 284*, 482−485.
20. 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.
21. Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) *Biochemistry 37*, 1585−1595.
22. Fersht, A. R. (1998) in *Structure and Mechanism in Protein Science*, pp 116−117, W. H. Freeman, New York.
23. Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. N., and Flugel, R. S. (1997) *Curr. Opin. Chem. Biol. 1*, 309−315.
24. Lowden, P. A. S., Böhm, G. A., Staunton, J., and Leadlay, P. F. (1996) *Anges. Chem. Int. Ed. Engl. 35*, 2249−2251.
25. Motamedi, H., and Shafiee, A. (1998) *Eur. J. Biochem. 256*, 528−534.
26. Chen, S., von Bamberg, D., Hale, V., Breuer, M., Hardt, B., Müller, R., Floss, H. G., Reynolds, K. A., and Leistner, E. (1999) *Eur. J. Biochem. 261*, 98−107.
27. Aparicio, J. F., Fouces, R., Mendes, M. V., Olivera, N., and Martín, J. F. (2000) *Chem. Biol. 7*, 895−905.
28. Moore, R. E., Chen, J. L., Moore, B. S., and Patterson, G. M. L. (1991) *J. Am. Chem. Soc. 113*, 5083−5084.
29. Beerhues, L. (1996) *FEBS Lett. 383*, 264−266.
30. Barillas, W., and Beerhues, L. (2000) *Biol. Chem. 381*, 155−160.
31. Hertweck, C., and Moore, B. S. (2000) *Tetrahedron 56*, 9115−9120.
32. Geissler, J. F., Harwood, C. S., and Gibson, J. (1988) *J. Bacteriol. 170*, 1709−1714.
33. Altenschmidt, U., Oswald, B., and Fuchs, G. (1991) *J. Bacteriol. 173*, 5494−5501.
34. Rusnak, R., Faraci, W. S., and Walsh, C. T. (1989) *Biochemistry 28*, 6827−6835.
35. Marsden, A. F., Wilkinson, B., Cortés, J., Dunster, N. J., Staunton, J., and Leadlay, P. F. (1998) *Science 279*, 199−202.

BI010080Z