Thioesterase Portability and Peptidyl Carrier Protein Swapping in Yersiniabactin Synthetase from *Yersinia pestis*†

Zucai Suo*

Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210

Received November 23, 2004; Revised Manuscript Received January 23, 2005

ABSTRACT: Multimodular enzymes, including polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), and mixed PKS/NRPS systems, contain functional domains with similar functions. Domain swapping and module fusion are potential powerful strategies for creating hybrid enzymes to synthesize modified natural products. To explore these strategies, yersiniabactin (Ybt) synthetase containing two subunits, HMWP2 [two NRPS modules (N-terminus-ArCP-Cy1-A-PCP1 and Cy2-PCP2-C-terminus)] and HMWP1 [one PKS (N-terminus-KS-AT-MT1-KR-ACP) one NRPS module (Cy3-MT2-PCP3-TE-C-terminus)], was used as a model system to study peptidyl carrier protein (PCP) domain swapping, thioesterase (TE) portability, and module−module fusion. The PCP1 domain of the N-terminal NRPS module of HMWP2 was swapped with either PCP2 or PCP3. The fusion proteins were 3−8-fold less active than the wild-type protein. The swapping of PCP2 of HMWP2 abolished the heterocyclization activity of the Cy2 domain while retaining its condensation function. When the two PCPs of HMWP2 were swapped by PCP3TE, it created two active fusion proteins: one or two NRPS modules fused to the TE domain. The internal TE domain of the two fusion proteins catalyzed the hydrolysis of enzyme-bound intermediates HPT-*S*-PCP3 to form HPT-COOH and HPTT-*S*-PCP3 to form HPTT-COOH. The TE activity was eliminated by the S2980A point mutation at its active site. Therefore, the three PCPs of the Ybt synthetase were swappable, and its lone TE domain was portable. The reasons for the observed low activities of the fusion proteins and lessons for protein engineering in generating novel modular enzymes were discussed.

Yersinia pestis, the causative agent of human plague, produces and secretes an iron-chelating siderophore, yersiniabactin (Ybt) (Chart 1), under iron starvation (*1*−*3*). Ybt binds ferric iron tightly with an equilibrium constant of 10^{35} M^{-1} (*4*). The biosynthetic genes of Ybt in *Y. pestis* have been identified as *ybtE*, *irp1*, *irp2*, *ybtU*, *ybtT*, *ybtS*, and *ybtD* (*5*, *6*). The *irp1* and *irp2* genes encode two high-molecular weight proteins, HMWP1[1] (349 kDa) and HMWP2 (229 kDa) (Figure 1), respectively (*7*). The other five genes encode five different single-domain enzymes shown in Figure 1. YbtD (∼28 kDa) is a phosphopantetheinylate transferase which catalyzes the transfer of phosphopantetheinyl moieties from coenzyme A molecules to the carrier domains of HMWP1 and HMWP2, leading to the conversion of these carrier domains from their apo forms to their holo forms (*6*). YbtS likely participates in salicylate biosynthesis (*3*, *7*). YbtE (57 kDa) activates salicylate and ligates a salicyl moiety to the holo aryl carrier domain (ArCP) of HMWP2 to initiate Ybt biosynthesis (Figure 1) (*8*). HMWP1 and HMWP2 form an assembly line of 16 predicted domains to convert salicylate, three cysteines, and malonyl-CoA into the mixed nonribosomal peptide/polyketide Ybt (Chart 1) (*7*, *9*−*11*). The four N-terminal domains of HMWP2 [ArCP, cyclization domain 1, adenylation, and peptidyl carrier protein 1 (ArCP-Cy1-A-PCP1)] (Figure 1) form the first nonribosomal peptide synthetase (NRPS) module which activates and incorporates the first cysteine and synthesizes the HPT-*S*-PCP1 intermediate (Figure 1) (*8*, *11*). The two C-terminal domains of HMWP2 (Cy2-PCP2) (Figure 1) form the second NRPS module, incorporating the second cysteine and elongating HPT-*S*-PCP1 to form the HPTT-*S*-PCP2 intermediate (Figure 1) (*9*−*12*). HMWP1 is a hybrid of a polyketide synthase (PKS) with five domains [ketoacyl synthase, acyltransferase, MT1, ketoacyl reductase, and acyl carrier protein (KS-AT-MT1-KR-ACP)] fused to a four-domain NRPS module [Cy3, MT2, PCP3, and thioesterase (Cy3-MT2-PCP3-TE)] (Figure 1) (*10*). The PKS module synthesizes the *tert*-butyl linker, while the NRPS module incorporates the third cysteine and assembles the methylated thiazoline ring of Ybt (Figure 1) (*10*). The TE domain catalyzes the hydrolytic release of Ybt from the Ybt-*S*-PCP3 intermediate (Figure 1) (*13*). YbtU, a stand-alone reductase, reduces the second thiazoline ring of

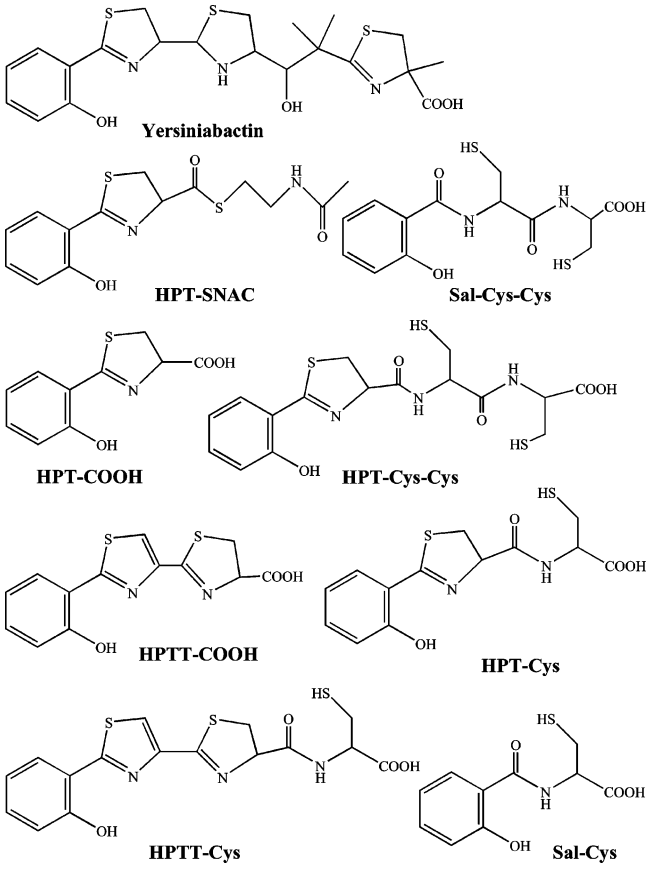
† This work was supported in part by the Jane Coffin Childs Memorial Fund for Medical Research and the startup fund provided by The Ohio State University.

* To whom correspondence should be addressed. Telephone: (614) 688-3706. Fax: (614) 292-6773. E-mail: suo.3@osu.edu.

[1] Abbreviations: A, adenylation; ACP, acyl carrier protein; ArCP, aryl carrier protein; AT, acyltransferase; HPLC, high-performance liquid chromatography; Cy, condensation/cyclization; HPT or HPT-COOH, 4,5-dihydro-2-(2-hydroxyphenyl)-4-thiazolecarboxylic acid; HPT-Cys, *N*-{[2-(2-hydroxyphenyl)]-4,5-dihydrothiazole-4-carbonyl}cysteine; HPTT or HPTT-COOH, 2′-(2-hydroxyphenyl)thiazolyl-2,4-thiazolinyl-4-caboxylic acid; HMWP, high-molecular weight protein; IPTG, isopropyl *β*-D-thiogalactopyranoside; KR, *β*-ketoacyl reductase; KS, *β*-ketoacyl synthase; MT, methyltransferase; NRPS, nonribosomal peptide synthetase; PCP, peptidyl carrier protein; PCR, polymerase chain reaction; PKS, polyketide synthase; PPTase, phosphopantetheinyl transferase; Sal, salicylate; SDS−PAGE, SDS−polyacrylamide gel electrophoresis; SNAC, synthetic *N*-acetylcysteamine thioester; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; TE, thioesterase.

10.1021/bi047538s CCC: $30.25
Published on Web 03/04/2005

Chart 1



the growing enzyme-tethered intermediate to thiazolidine before, during, or after the action of HMWP1 (*10*). YbtT represents an external TE domain thought to be responsible for *in vivo* biosynthetic editing (*10*, *14*). Recently, Ybt has been successfully reconstituted *in vitro* from four recombinant proteins [YbtE, HMWP2, HMWP1, and YbtU (*10*)]. The Ybt synthetase genes have been introduced into heterologous host *Escherichia coli* to produce Ybt (*15*).

Previously, the TE domain of HMWP1 has been found to selectively hydrolyze synthetic *N*-acetylcysteamine thioesters (SNAC), mimicking moieties of Ybt (*13*). The ACP and PCP3 domains of HMWP1 are shown to exhibit different substrate specificity in phosphopantetheinylation by different exogenous phosphopantetheinyl transferases, acylation with malonyl CoA by the AT domain, and aminoacylation with cysteine by the adenylation domain of HMWP2 (*16*). The Cy domains of HMWP2 recognize three PCPs of the Ybt synthetase *in trans* (*17*). However, these *in trans* complementation systems are not efficient enzymes (*17*). In this paper, I examined *in cis* complementation of the three PCPs in the NRPS modules of both HMWP1 and HMWP2. The PCPs were able to functionally replace each other, and the hybrid enzymes are significantly more active than similar modules formed from *in trans* complementation. I also tested the portability of the C-terminal TE domain of HMWP1. The TE domain was active after being fused to the two NRPS modules of HMWP2. The *in trans* fragment·fragment combinations had catalytic activities quite different from those of their corresponding *in cis* fusion proteins.

EXPERIMENTAL PROCEDURES

Materials. ATP, coenzyme A, imidazole, and magnesium chloride were purchased from Sigma Chemical Co. Tris was purchased from J. T. Baker. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Molecular Probes (Eugene, OR). All DNA oligomers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). The competent *E. coli* BL21(DE3) and pET22b plasmid were purchased from Novagen. All restriction enzymes were purchased from New England Biolabs.

Cloning of the Fusion Proteins Shown in Figure 2A

Cloning of 1−1382-PCP1′. This was a two-step cloning process. First, a pET22b-1-1382 plasmid was constructed from ligation of a 192 bp insert and a vector from digestion of pET22b-irp2/1−1491 (*8*) by *Not*I and *Xho*I. The 192 bp

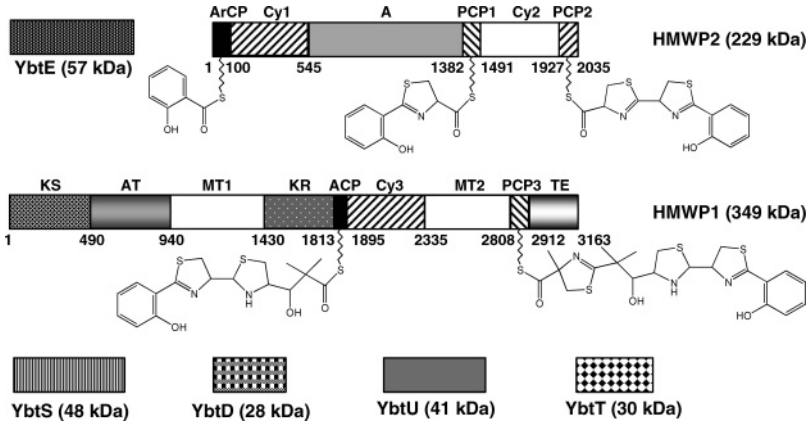


FIGURE 1: Seven subunits of Ybt synthetase and the enzyme-bound intermediates. All domains of both HMWP1 and HMWP2 with amino acid residue numbers and domain names are shown as distinct rectangles. HMWP2 contains the following domains: ArCP, aryl carrier protein (residues 1−100); Cy1, condensation/cyclization 1 (residues 101−545); A, adenylation (residues 546−1382); PCP1, peptidyl carrier protein 1 (residues 1383−1491); Cy2, condensation/cyclization 2 (residues 1492−1927); PCP2, peptidyl carrier protein 2 (residues 1928−2035). HMWP1 contains the following domains: KS, β-ketoacyl synthase (residues 1−490); AT, acyltransferase (residues 491−940); MT1, methyltransferase 1 (residues 941−1430); KR, β-ketoacyl reductase (residues 1431−1813); ACP, acyl carrier protein (residues 1814−1895); Cy3, condensation/cyclization 3 (residues 1896−2335); MT2, methyltransferase 2 (residues 2336−2808); PCP3, peptidyl carrier protein 3 (residues 2809−2912); TE, thioesterase (residues 2913−3163).

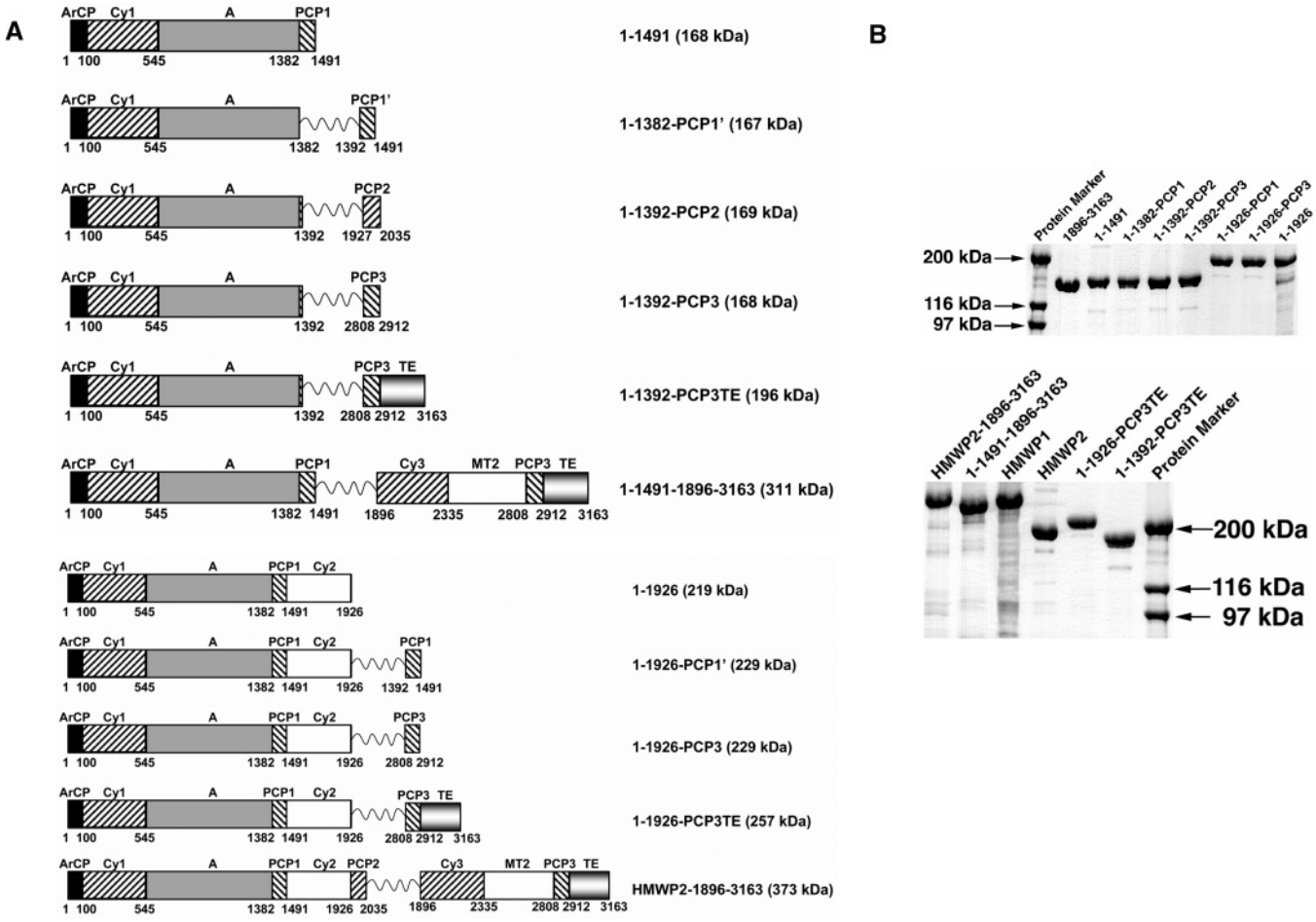


FIGURE 2: (A) Domain structures of fusion proteins and fragments. Each fusion site is represented by a wavy line. The name and molecular mass of each construct are given to the right of its domain structure. All domains with amino acid residue numbers and domain names are shown as distinct rectangles. (B) SDS−PAGE gels of purified fusion proteins. All constructs except 1−1926 were overexpressed in *E. coli* and purified as C-terminal His_6-tagged fusion proteins using a nickel affinity column. 1−1926 was purified as an N-terminal His_6-tagged fusion protein.

insert was the PCR product using pET22b-1−1491 as the template and primers *Not*I (5′-GGGCGGCATTATTTAGC-**GCGGCCGC**CTTC-3′) and 1382*Xho*I (5′-CGC GCG G**CTCGAG**TT TTC CCG TTA GCC GTC AGC G-3′). Second, the PCP1′ gene fragment corresponding to HMWP2 residues 1392−1491 was amplified from pET22b-irp2/1−1491 by PCR using primers PCP1*Xho*I.for (5′-GAAGCGT-**CTCGAG** AC CCC TGA AGC GGA AAA CCC GG-3′) and PCP1*Xho*I.rev (5′-TGA CCG **CTC GAG** TTC TTC AGG GGA GTG GAC GAA TG-3′). The PCR product was ligated to the *Xho*I site of pET22b-1−1382 to make plasmid pET22b-1−1382-PCP1′. The DNA sequences of the ligation sites and the insert (single copy, correct orientation) were confirmed by sequencing. All DNA sequencing in this study was performed by the Dana Farber Molecular Biology Core Facility (Boston, MA). The plasmid pET22b-1−1382-PCP1′ directs production of 1−1382-PCP1′ with a C-terminal LEHHHHHH tag. My fusion strategy added two extra amino acid residues (LE) at each fusion site of every fusion protein construct described in this paper.

Cloning of 1−1392-PCP2, 1−1392-PCP3, and 1−1392-PCP3TE. First, a plasmid pET22b-1−1392 was constructed from ligation of a 222 bp insert and a vector from digestion of pET22b-irp2/1−1491 by *Not*I and *Xho*I. The 222 bp insert was the PCR product using pET22b-1−1491 as the template and primers *Not*I and 1392Rev (5′-CCG CTT CCT CGA GAT GAC GAC GCT TCA GCG C-3′). Second, the gene fragments, PCP2 corresponding to HMWP2 residues 1927−2035, PCP3 corresponding to HMWP1 residues 2808−2912, and PCP3TE corresponding to HMWP1 residues 2808−3163, were amplified from pIRP2 (*4*) for PCP2 or pSDR498.4 for PCP3 and PCP3TE using primer pairs: PCP2for-1 (5′-GGA GAT ATA **CTCGAG** GAG CGC TCA CCG CG-3′) and PCP2rev-1 (5′-GGT GGT GGT G**CTC-GAG**TA TCC GCC G-3′) for the PCP2 insert, PCP3TEfor-1 (5′-GGA GAT ATA **CTCGAG** GCC CCG TCT GAT GCG-3′) and irp1PCP3.rev (5′-TGA CCG **CTCGAG** ACC GTC GCC CTG GCA AAG GGG T-3′) for the PCP3 insert, and PCP3TEfor-1 and 1351-Rev (5′-CCA CCG **CTCGAG** TAA CGT GTT CTC CGG TTG CGT TG-3′) for the PCP3TE insert. The PCR products were ligated to the *Xho*I site of pET22b-1−1392 to make plasmids pET22b-1−1392-PCP2, pET22b-1−1392-PCP3, and pET22b-1−1392-PCP3TE. The DNA sequences of the ligation sites and the insert (single copy, correct orientation) were confirmed by sequencing. Plasmids pET22b-1−1392-PCP2, pET22b-1−1392-PCP3, and pET22b-1−1392-PCP3TE direct production of 1−1392-PCP2, 1−1392-PCP3, and 1−1392-PCP3TE, respectively, each with a C-terminal LEHHHHHH tag.

Construction of the TE S2980A Mutant of 1−1392-PCP3TE. The mutation was performed by using the QuikChange site-directed mutagenesis kit (Stratagene). The TE mutant plasmid pET22b-1−1392-PCP3TE-Temut′ was synthesized by PCR using Pfu polymerase (Stratagene), the pET22b-1−1392-PCP3TE template, and primers TEmut.top (5′-GTG CTG GCG GGT TGG **G**CG TAT GGC GCG TTT CTT G-3′) and TEmut.bottom (5′-CAA GAA ACG CGC CAT ACG **C**CCAACCCGCCAGCAC-3′). The mutation was confirmed by DNA sequencing. The *Nhe*I−*Sty*I DNA fragment of pET22b-1−1392-PCP3TE-Temut′ was ligated to the *Nhe*I and *Sty*I sites of pET22b-1−1392-PCP3TE to make the final TE mutant plasmid pET22b-1−1392-PCP3TE-TEmut. The DNA sequences of the ligation sites and the insert were confirmed by sequencing.

Cloning of HMWP2-1−1926. pET28b-1−1926 was constructed from ligation of a 1.8 kb insert and a vector from digestion of pET28b-1−1382 (*9*) by *Not*I and *Xho*I. The 1.8 kb insert was the PCR product using pIRP2 as the template and primers *Not*I and irp1926.rev (5′-CGCGCGG**CTCGAG**TTACTAGCATTGCCCTGCGTGTTT-3′). pET22b-1−1926 was constructed from ligation of a 1.8 kb insert and a vector from digestion of pET22b-1−1392 by *Not*I and *Xho*I. The 1.8 kb insert was the PCR product using pIRP2 as the template and primers *Not*I and 1926*Xho*I.rev (5′-CGC GCG GCT CGA GGC ATT GCC CTG CGT GTT TAA C-3′). The DNA sequences of the ligation sites and the insert were confirmed by sequencing. The pET28b-1−1926 plasmid directs production of the HMWP2-1−1926 fragment as a histidine tag fusion with the MGSSHHHHHHSSGLVPRGSH amino acid sequence appended to the N-terminus.

Cloning of 1−1926-PCP1′, 1−1926-PCP3, and 1−1926-PCP3TE (both the Wild-Type and TE S2980A Mutant). Inserts PCP1′, PCP3, and PCP3TE (both the wild-type and TE mutant) were prepared from *Xho*I digestion of plasmids pET22b-1−1392-PCP1′, pET22b-1−1392-PCP3, pET22b-1−1392-PCP3TE, and pET22b-1−1392-PCP3TE-Temut. These inserts were ligated to the *Xho*I site of pET22b-1−1926 to generate plasmids pET22b-1−1926-PCP1′, pET22b-1−1926-PCP3, pET22b-1−1926-PCP3TE, and pET22b-1−1926-PCP3TE-Temut. Each plasmid directs production of a fusion protein with a C-terminal LEHHHHHH tag. The DNA sequences of the ligation sites were confirmed by sequencing.

Cloning of 1−1491-1896−3163 and HMWP2-1896−3163. The insert corresponding to HMWP1 residues 1896−3163 was synthesized by PCR using the pET22b-1812−3163 template (*16*) and primers 1896*Xho*I.for (5′-GTT CTA AGC **CTCGAG** GCA CGC CTG CCG CTT CG-3′) and 1351-Rev. The insert was then ligated to the *Xho*I site of pET22b-1−1491 and pET22b-HMWP2 to make plasmids pET22b-1−1491-1896−3163 and pET22b-HMWP2-1896−3163, respectively. The DNA sequences of the ligation sites and the insert were confirmed by sequencing. Each of the two plasmids directs production of an extremely large protein with a C-terminal LEHHHHHH tag.

Overexpression and Purification of Fusion Proteins (Figure 2A)

Competent cells of *E. coli* strain BL21(DE3) were transformed separately with pET22b-1−1382-PCP1′, pET22b-1−1392-PCP2, pET22b-1−1392-PCP3, pET22b-1−1392-PCP3TE, pET22b-1−1392-PCP3TE (TE mutant), pET22b-1−1491-1896−3163, pET28b-1−1926, pET22b-1−1926-PCP1′, pET22b-1−1926-PCP3, pET22b-1−1926-PCP3TE, pET22b-1−1926-PCP3TE (TE mutant), pET22b-HMWP2-1896−3163, and pET37b-HMWP2. These fusion proteins with sizes up to 373 kDa posed a serious challenge in their expression and purification. Both the 349 kDa HMWP1 (*13*, *16*) and 229 kDa HMWP2 (*11*) (Figure 1) have been previously expressed under slow induction conditions (18 °C and 75 μM IPTG) and purified with a modified protocol as soluble and active proteins despite their enormous sizes (*13*, *16*). By following the same protocols, all fusion proteins in Figure 2A were successfully expressed and purified as C-terminal hexahistidine-tagged proteins (Figure 2B). After dialysis and concentration, the purified proteins were stored in buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM Mg^{2+}, 0.1 mM EDTA, 10% glycerol, and 1 mM TCEP at −80 °C. Concentrations of the purified proteins were measured spectrophotometrically at 280 nm using calculated extinction coefficients of 0.26646, 0.26518, 0.26101, 0.26542, 0.31818, 0.48434, 0.36405, 0.36253, 0.3695, 0.36846, 0.42122, and 0.58193 μM^{-1} cm^{-1} for 1−1491, 1−1382-PCP1′, 1−1392-PCP2, 1−1392-PCP3, 1−1392-PCP3TE (both the wild-type and its TE mutant), 1−1491-1896−3163, HMWP2, 1−1926, 1−1926-PCP1′, 1−1926-PCP3, 1−1926-PCP3TE (both the wild-type and its TE mutant), and HMWP2-1896−3163, respectively.

The preparations of 1−1382 (*9*), 1383−2035 (*9*), PCP3TE (*12*), 1896−3163 (*16*), HMWP1 (*13*), PCP1, PCP2, and PCP3 (*7*, *8*) have been described elsewhere. *Bacillus subtilis* Sfp was prepared previously (*18*).

HPT-SNAC Hydrolysis by Fusion Proteins at 23 °C

A reaction mixture (100 μL) of 1 μM enzyme, 0.5 mM HPT-SNAC, 5 mM TCEP, and 2.5% dimethyl sulfoxide (DMSO) in reaction buffer [225 mM KH_2PO_4 (pH 7.0)] was incubated at 23 °C for 1 h prior to being quenched by addition of 20 μL of 8.5% phosphoric acid (final pH of ≈1.5) and then frozen with liquid nitrogen. Each thawed sample (120 μL) was injected directly and analyzed by HPLC (Waters) connected to a C18 reverse phase column (VYDAC, 250 mm × 5 mm) with a detector wavelength of 254 nm and a mobile phase with the following gradient: 4% B and 1 mL/min flow rate at 0.0 min, 80% B and rate of 1 mL/min at 23.0 min, 4% B and rate of 1.5 mL/min at 23.1 min, and 4% B and rate of 1 mL/min at 31.0 min. Mobile phase A was 25 mM KH_2PO_4 (pH 5.4). Mobile phase B was 100% acetonitrile. The product peaks were identified by standard co-injection.

Isolation and Identification of Products Synthesized by the Fusion Proteins

A solution (500 μL) of 5 μM fusion protein, 0.3 μM Sfp, 0.5 μM YbtE, 75 mM Tris-HCl (pH 8.8), 10 mM $MgCl_2$, 5 mM TCEP, 1 mM CoASH, 5 mM L-cysteine, 1 mM salicylate, and 10 mM ATP was incubated at 30 °C for 13 or 20 h. The reaction was quenched with 100 μL of 8.5% phosphoric acid. The acidified mixture was extracted three times with 1 mL of ethyl acetate. The organic layers were combined in a glass vial and dried under reduced pressure. The dried residue was submitted directly for LC−MS

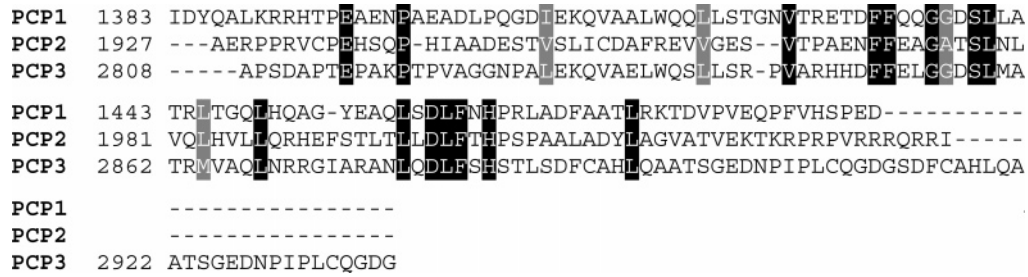


FIGURE 3: Protein sequence alignment. The alignment of the sequences of PCP1 and PCP2 of HMWP2 and PCP3 of HMWP1 reveals very little conservation except in the PCP signature sequence region.

positive ion trap mass spectrometry analysis (Chemistry Department, Harvard University, Cambridge, MA).

Time Courses of Product Formation Catalyzed by the Fusion Proteins at 30 °C

Each fusion protein was incubated with 75 mM Tris-HCl (pH 8.8), 10 mM $MgCl_2$, 5 mM TCEP, 0.3 μM Sfp, 1 mM CoASH, 0.5 μM YbtE, 5 mM L-cysteine, and 1 mM salicylate at 30 °C for 30 min to allow phosphopantetheinylation of each fusion protein. Then 10 mM ATP was added to start the reaction (500 μL mixture) at 30 °C (without ATP in the control reaction mixture). At various times, each 54 μL sample was acidified by addition of 10 μL of 8.5% phosphoric acid. The acidified mixture (pH ≈1.5) was extracted with 1 mL of ethyl acetate. A portion of the organic layer (950 μL) was dried in a SpeedVac. The dried residue was dissolved in 200 μL of a 30% acetonitrile/water mixture, and 180 μL of the solution was analyzed by HPLC connected to a C18 reverse phase column (VYDAC, 250 mm × 4.6 mm) with a detector wavelength of 254 nm and a mobile phase with the following gradient: mobile phase A, a mixture of formic acid (0.1 mL) and triethylamine (0.2 mL) in water (1 L) (pH 3.3); and mobile phase B, a 4:1 mixture of acetonitrile and mobile phase A; flow rate of 1 mL/min at 0.0 (10% B), 23.0 (100% B), and 23.1 min (10% B) and flow rate of 1.5 mL/min at 31.0 min (10% B), and at 1 mL/min.

Measurement of the Affinity of Two Fragments at 30 °C

For the determination of each K_D, varied concentrations of one fragment were used while the other was held at 2 μM. The reactions were performed in a manner similar to that described above for time courses except that the volume of each reaction was 100 μL, and the reaction time was constant at 150 min. The reaction mixtures were analyzed by HPLC as described above.

Product Quantitation

Each product peak from the HPLC traces was integrated by computer to obtain its peak area (*S*). The integrated area (*S*) was converted into quantity using the corresponding standard calculation curve: $(4.8400 \times 10^{-6})S$ nmol for Sal-Cys, $(1.5716 \times 10^{-6})S$ nmol for HPT-Cys, $(1.8426 \times 10^{-6})S$ nmol for HPT-COOH or HPT, and $(4.2751 \times 10^{-6})S$ nmol for HPTT-COOH or HPTT. Due to a lack of synthetic standards, the peaks of Sal-Cys-Cys, HPT-Cys-Cys, and HPTT-Cys were quantitated using the Sal-Cys, HPT-Cys, and HPTT-COOH standard curves, respectively.

Data Analysis

For the HPLC time courses, the data were fit to the linear equation [product] = constant + $k_{obs}E_0t$, where E_0 represents the enzyme concentration and k_{obs} the observed steady-state rate constant. For the measurement of K_D of 1−1491·1896−3163, the k_{obs} of HPT-Cys formation was fit to the equation $k_{obs} = 2m_1 - 0.5m_1\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$, where m_2 represents K_D, m_0 the concentration of 1896−3163, and m_1 the steady-state rate constant of HPT-Cys formation catalyzed by free 1−1491 (see the Appendix). For the measurement of K_D of the HMWP2·PCP3TE complex, the k_{obs} of HPT-Cys formation was fit to the equation $k_{obs} = 2m_1 + 0.5m_3\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$, where m_2 represents K_D, m_0 the concentration of PCP3TE, and m_3 the steady-state rate constant of HPT-Cys formation catalyzed by the HMWP2·PCP3TE complex (see the Appendix). For the measurement of the K_D values of the HMWP2·1896−3163 and HMWP2·HMWP1 complexes, the k_{obs} of HPTT formation was fit to the equation $k_{obs} = 2m_1 + 0.5(m_3 - m_1)\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$, where m_2 represents K_D, m_0 represents the concentration of 1896−3163 or HMWP1, and m_1 and m_3 represent the steady-state rate constants of HPTT formation catalyzed by free HMWP2 and the HMWP2·1896−3163 or HMWP2·HMWP1 protein complex, respectively (see the Appendix).

RESULTS

Product Formation at 30 °C Catalyzed by the Fusion Proteins from PCP Swapping at the PCP1 Site of HMWP2. Due to lack of structural information, domain boundaries in NRPS and PKS modules are unclear and are usually assigned through sequence alignment analysis. Uncertainty in domain boundaries thereby becomes a big hurdle for the formation of modified natural products through domain swapping and module fusion. It is plausible that neighboring domains in PKS and NRPS modules are connected through short peptide loops or linkers (∼15 residues) with random amino acid sequences (*19*). Because these linker regions are thought to be nonfunctional, they are the optimal fusion sites for the construction of artificial synthetases (*20*). The 1−1491 fragment of HMWP2 (Figure 2A) was shown previously to be a fully active NRPS module (*8*). To determine if there is a linker between the A domain and the PCP1 domain, I constructed a 1−1382-PCP1′ fusion protein with a nine-amino acid residue (1383−1391) deletion (Figure 2A). The sequence alignment of the three PCPs shown in Figure 3 reveals that other than the Ppant-attachment motif, G(G/A)-

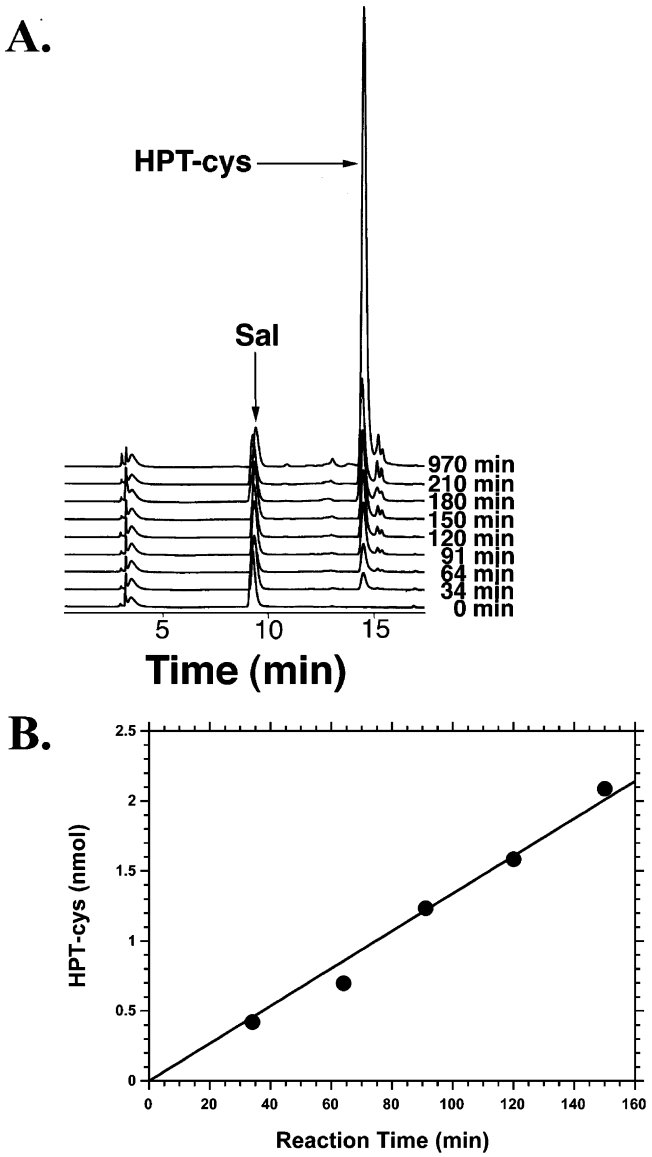


FIGURE 4: Time course of HPT-Cys formation catalyzed by 5 μM 1−1382-PCP1′ at 30 °C in the presence of 75 mM Tris-HCl (pH 8.8), 5 mM TCEP, 10 mM ATP, 0.5 μM YbtE, 5 mM L-cysteine, 1 mM salicylate, and 10 mM $MgCl_2$. 1−1382-PCP3TE was phosphopantetheinylated for 30 min by 0.3 μM *B. subtilis* Sfp and 1 mM CoA prior to the addition of ATP. (A) HPLC traces. The major peak was identified as HPT-Cys by both co-injection with a synthetic standard and LC−MS. (B) Plot of the amount of HPT-Cys vs reaction time. A k_{obs} of 0.054 min^{-1} was calculated from a linear fit of the data (●).

(D/T)SL, there is limited sequence homology among them. To evaluate whether PCP1 is replaceable, the 1−1392 fragment of HMWP2 was fused to either the PCP2 domain (amino acids 2808−2912) of HMWP2 or the PCP3 domain (amino acids 2808−2912) of HMWP1 to construct 1−1392-PCP2 or 1−1392-PCP3 (Figure 2A), respectively.

If 1−1382-PCP1′, 1−1392-PCP2, and 1−1392-PCP3 (Figure 2A) are active, each of them is expected to behave as an NRPS module like 1−1491 (*7*) to synthesize HPT-Cys and a small amount of HPT-COOH (Chart 1) from salicylate, ATP, and cysteine. Cosubstrate cysteine and water molecules cleave HPT-*S*-PCP1 (Figure 1) as nucleophiles and release HPT-Cys and HPT-COOH, respectively (*8*, *9*). HPT-Cys was indeed produced by 1−1382-PCP1′ (Figure 4A) with an observed rate constant of 0.054 min^{-1} (Figure 4B and Table 1) at 30 °C. Thus, the deletion of the 1383−1391 peptide produced an active protein 1−1382-PCP1′. However, this deletion mutant retained only 60% of the activity of the intact 1−1491 (Table 1), suggesting that the 1383−1391 peptide formed an essential boundary between the A domain and PCP1, rather than a nonfunctional linker (*20*). This important peptide was thus retained in three fusion protein constructs (1−1392-PCP2, 1−1392-PCP3, and 1−1392-PCP3TE) described below.

Table 1: Observed Rate Constants of Product Formation at 30 °C by Fusion Proteins from PCP Swapping at the PCP1 Site of HMWP2

enzyme	Sal-Cys (min^{-1})	Sal-Cys-Cys (min^{-1})	HPT-Cys (min^{-1})
1−1491[a]			0.087
1−1382-PCP1′			0.054
1−1392-PCP2	0.013	0.009	0.026
1−1392-PCP3			0.010

[a] Data are from ref *9*.

Both 1−1392-PCP2 and 1−1392-PCP3 (Figure 2A) were evaluated in the same manner as 1−1382-PCP1′ (data not shown), and the observed rate constants are listed in Table 1. The two fusion proteins were also active and slowly produced HPT-Cys, indicating PCP1 can be swapped with PCP2 and PCP3. However, the fusions were 3−8-fold less active than wild-type 1−1491 (Table 1). The artificial junctions and different PCP domains could contribute to the low activity of the fusion proteins. In addition, Sal-Cys and Sal-Cys-Cys were also synthesized slowly by 1−1392-PCP2 (Table 1), suggesting there was leakage at both ArCP and PCP2, and the product flux catalyzed by the two fusion proteins was not as efficient as in 1−1491.

Activity of 1−1392-PCP3TE at 30 °C. Previous experiments have shown that when fused to the C-termini of appropriate PKS modules, the TE domain can release enzymatic intermediates bound to the PKS modules (*21*−*25*). This could be applied to the TE domains of NRPS systems or mixed PKS/NRPS systems. To test this hypothesis, I intended to fuse the TE domain of HMWP1 to the NRPS modules of HMWP2. However, the domain boundary between PCP3 and TE was not clearly defined. The fused TE domain may be inactive if it is fused directly to the PCPs of HMWP2 with a wrong boundary. Because all PCPs are similar in function and because their boundaries may be more flexible than those of enzymatic domains, 1−1392-PCP3TE and 1−1926-PCP3TE (Figure 2A) were thereby constructed and believed to be more likely to have active TE domains than 1−1491-TE and 1−1926-TE.

If TE is active, 1−1392-PCP3TE should mostly synthesize HPT-COOH (or HPT), and the product flux may be accelerated. The results shown in Figure 5A confirmed that HPT-COOH was indeed the dominant product at long reaction time points. Interestingly, HPT-Cys was produced more than HPT-COOH in the early time points, and there was a puzzling ~60 min delay for HPT-COOH formation (Figure 5B). The total turnover rate constants (0.018 min^{-1}) of HPT-COOH (0.013 min^{-1}) and HPT-Cys (0.005 min^{-1}) (Table 2) were slightly higher than that of 1−1392-PCP3 (0.010 min^{-1}) (Table 1), suggesting the TE domain indeed

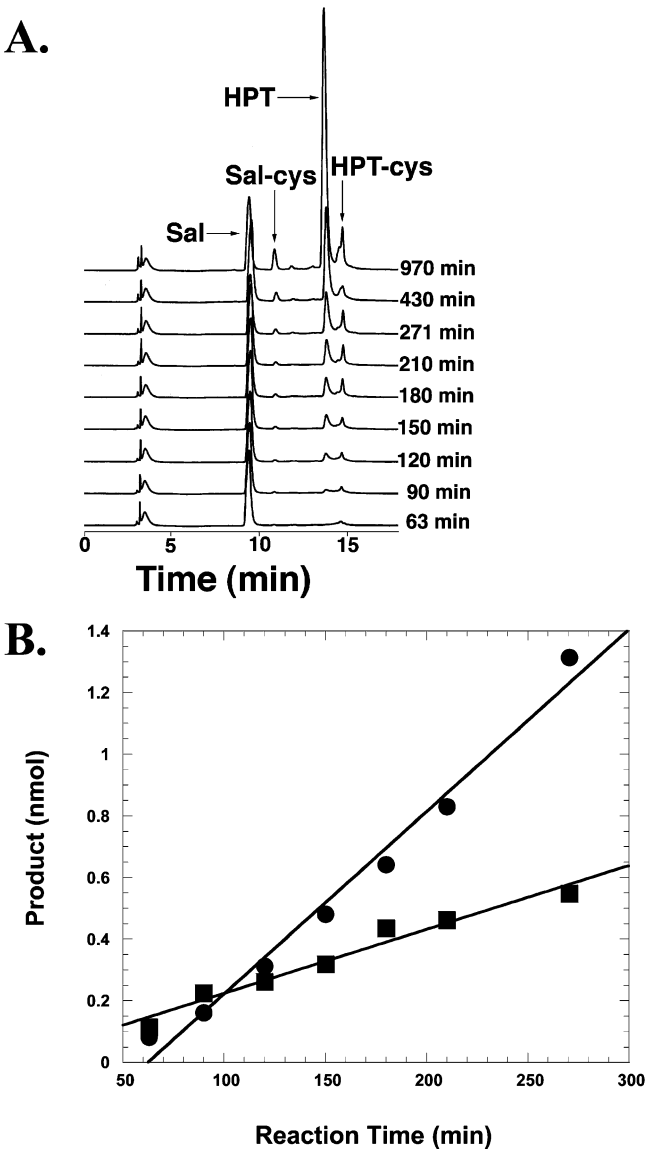


FIGURE 5: Time course of product formation catalyzed by 5 μM 1−1392-PCP3TE at 30 °C. The reaction conditions are the same as those described in the legend of Figure 4. (A) HPLC traces. The peaks were identified as Sal-Cys, HPT, and HPT-Cys by both co-injection with synthetic standards and LC−MS. (B) Plot of quantities of HPT (●) and HPT-Cys (■) vs reaction time. Linear fits of data yield values of k_{obs} of 0.013 and 0.005 min^{-1} for formation of HPT and HPT-Cys, respectively.

Table 2: Observed Rate Constants of Product Formation at 30 °C by Fusion Proteins from PCP3TE Swapping at the PCP1 and PCP2 Sites of HMWP2

enzyme	HPT (min^{-1})	HPT-Cys (min^{-1})	HPT-Cys-Cys (min^{-1})	HPTT (min^{-1})
1−1392-PCP3TE	0.013	0.005		
1−1392-PCP3TE (TE mutant)		0.009		
1−1926-PCP3TE	0.0022	0.024	0.0031	0.0054
1−1926-PCP3TE (TE mutant)		0.010	0.010	

accelerated the product flux through the hybrid NRPS module.

To further confirm that the production of HPT-COOH in Figure 5A was due to the presence of the TE domain in

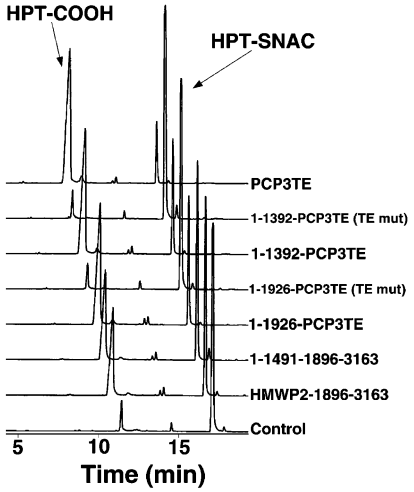


FIGURE 6: Hydrolysis of 0.5 mM HPT-SNAC by 1 μM fusion protein at 23 °C for 60 min in the presence of 5 mM TCEP, 2.5% DMSO, and 225 mM KH_2PO_4 (pH 7.0). There was no enzyme in the control reaction mixture. The reaction mixtures were injected directly and analyzed by HPLC. The HPLC traces were shifted to the left side by 0.5 min relative to the control reaction. The peaks were identified by co-injection with synthetic standards.

1−1392-PCP3TE and not due to the hydrolytic cleavage of the HPT-*S*-PCP3 intermediate, I constructed the active site S2980A mutant of the TE domain of 1−1392-PCP3TE. The inactivation of the TE domain by the S2980A mutation was confirmed by its inability to hydrolyze HPT-SNAC (Figure 6) (*13*). Only HPT-Cys was synthesized by the point mutant (Figure 7A). This suggested that the production of HPT-COOH by 1−1392-PCP3TE was indeed due to the internal TE domain (Figure 5A). The observed rate constant for HPT-Cys production by the mutant was 0.009 min^{-1} (Figure 5B and Table 2) which was close to the rate constant observed with 1−1392-PCP3 (Table 1). This was reasonable since 1−1392-PCP3 and 1−1392-PCP3TE (TE mutant) were expected to have similar enzymatic activities.

Tandem Heterocyclization Activity of 1−1926 and PCP2 in Trans at 30 °C. Fragments 1−1382 and 1383−2035 from dissection of HMWP2 have been previously shown to catalyze the synthesis of HPTT-Cys and HPTT-COOH *in trans* (*9*). To explore whether 1−1926·PCP2 (Figure 2A) could also reconstitute the activity of full-length HMWP2, the two fragments were prepared and examined in a manner similar to that for 1−1382·1383−2035. The results shown in Figure 8 demonstrated that 1−1926·PCP2 (trace b), like 1−1382·1383−2035 (trace a), synthesized HPT-Cys, HPT-Cys-Cys, and HPTT-Cys·HPTT-COOH (Chart 1). These products indicated product flux in the ArCP → PCP1 → PCP2 direction. More HPT-Cys-Cys, HPTT-Cys, and HPTT-COOH and less HPT-Cys in trace b than in trace a indicated more efficient product flux in 1−1926·PCP2 than in 1−1382·1383−2035. Interestingly, 1−1926 alone (trace c) synthesized only Sal-Cys, although it contained the first NRPS module of HMWP2. This was probably due to the protection of the PCP1 site by the Cy2 domain in 1−1926 which blocked the access of nucleophiles such as cosubstrate cysteine or water molecules to PCP1 which in turn inhibited the product flux from ArCP to PCP1. The presence of PCP2 *in trans* of 1−1926 facilitated the ArCP → PCP1 → PCP2 product flux and the formation of the enzymatic intermediates HPT-Cys-PCP2 and HPTT-*S*-PCP2.

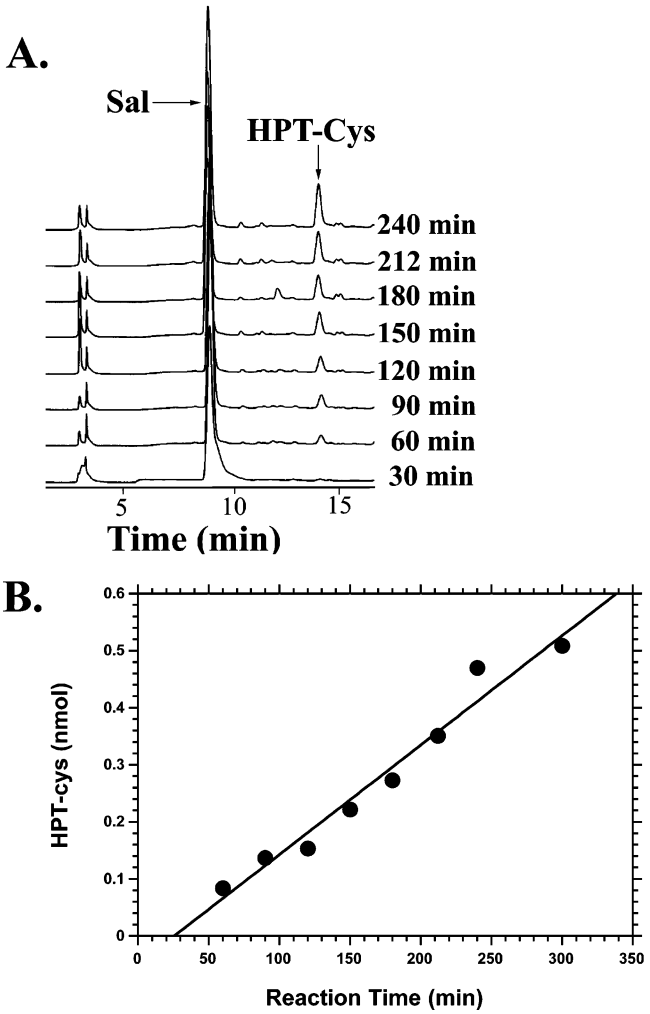


FIGURE 7: Time course of product formation catalyzed by 5 μM 1−1392-PCP3TE (TE mutant) at 30 °C. The reaction conditions are the same as those described in the legend of Figure 4. (A) HPLC traces. The major peak was identified as HPT-Cys by both co-injection with a synthetic standard and LC−MS. (B) Plot of the quantity of HPT-Cys vs reaction time. A k_{obs} of 0.009 min^{-1} is calculated from a linear fit of the data (●).

Activity of 1−1926-PCP3TE at 30 °C. To examine whether the TE domain was also portable at the second NRPS module of HMWP2, I constructed 1−1926-PCP3TE (see above). HPLC trace c in Figure 9 showed that the TE domain in 1−1926-PCP3TE was indeed active and catalyzed the formation of HPTT-COOH from cleavage of the HPTT-*S*-PCP3 intermediate while HPTT-Cys from nucleophilic cleavage of the intermediate by cysteine was undetectable. The reaction rate constants (Table 2) obtained from time courses of 1−1926-PCP3TE (Figure 10) were small.

To further confirm that the TE domain and not water molecules catalyzed the cleavage of HPTT-*S*-PCP3TE to form HPTT-COOH, I constructed a point mutant (S2980A) of 1−1926-PCP3TE. The results of the HPT-SNAC hydrolysis shown in Figure 6 confirmed that the TE domain was inactivated by the mutation. HPT-Cys and HPT-Cys-Cys, rather than HPTT-Cys and HPTT-COOH, were synthesized by the point mutant (Figure 11A), indicating an ArCP → PCP1 → PCP3 product flux and an inactive cyclization function of Cy2. Moreover, the observed rate constants for both products (Table 2) obtained from the time courses shown in Figure 11B were smaller than those

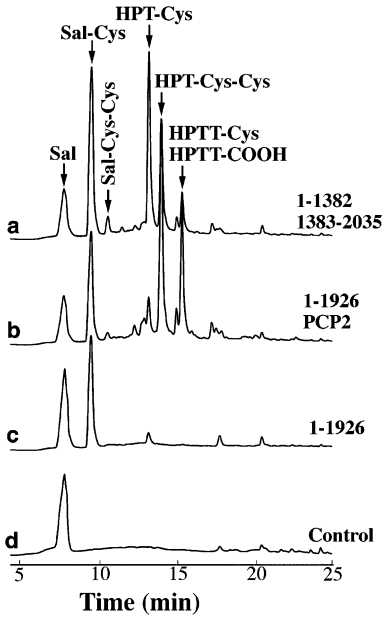


FIGURE 8: HPLC analysis of reaction products by different protein fragments at 30 °C for 13 h in the presence of 75 mM Tris-HCl (pH 8.8), 10 mM $MgCl_2$, 1 mM CoA, 5 mM TCEP, 0.77 μM YbtE, 5 mM L-cysteine, 1 mM salicylate, and 10 mM ATP. The protein fragments were phosphopantetheinylated for 30 min by 0.5 μM *B. subtilis* Sfp and 1 mM CoA prior to the addition of ATP to initiate the reactions: (a) 1−1382 (5 μM) with 1383−2035 (5 μM), (b) 1−1926 (4 μM) with PCP2 (4.9 μM), (c) 1−1926 (5 μM), and (d) control. The peaks were identified by standard co-injection and LC−MS.

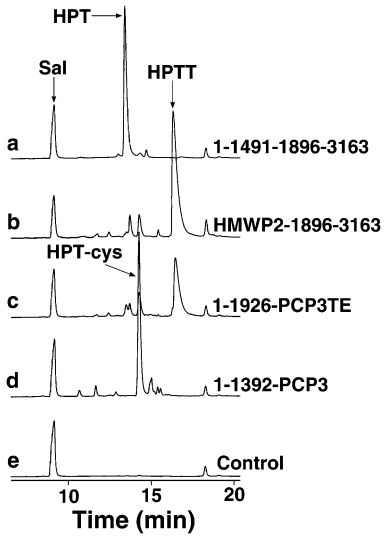


FIGURE 9: HPLC analysis of reaction products by different fusion proteins at 30 °C for 20 h. The reaction conditions were the same as those described in the legend of Figure 8: (a) 1−1491-1896−3163 (4 μM), (b) HMWP2-1896−3163 (5 μM), (c) 1−1926-PCP3TE (4 μM), (d) 1−1392-PCP3 (10 μM), and (e) control. The peaks were identified by standard co-injection and LC−MS.

corresponding rate constants of wild-type 1−1926-PCP3TE. This suggested that the active TE domain in the fusion slightly increased the product flux.

Product Formation at 30 °C by the Fusion Proteins from PCP Swapping at the PCP2 Site of HMWP2. With the success of PCP swapping at the PCP1 site and the active fusion of 1−1926-PCP3TE, I further tested whether the PCP2 domain of HMWP2 could be swapped with either the PCP1 or PCP3 domain. The two large fusion proteins, 1−1926-

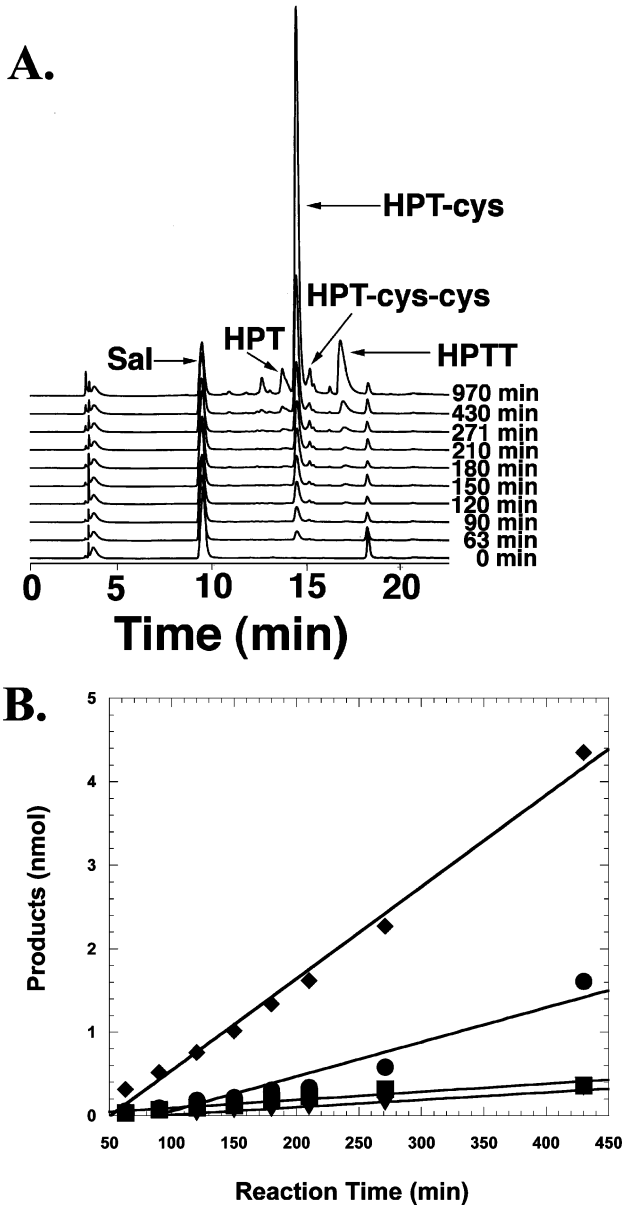


FIGURE 10: Time course of product formation catalyzed by 5 μM 1−1926-PCP3TE at 30 °C. The reaction conditions are the same as those described in the legend of Figure 4. (A) HPLC traces. The peaks were identified as HPT, HPT-Cys, HPT-Cys-Cys, and HPTT by both co-injection with synthetic standards and LC−MS. (B) Plot of quantities of HPT (▼), HPT-Cys (◆), HPT-Cys-Cys (■), and HPTT (●) vs reaction time. Linear fits of data yield values of k_{obs} of 0.0022, 0.024, 0.0031, and 0.0054 min^{-1} for the formation of HPT, HPT-Cys, HPT-Cys-Cys, and HPTT, respectively.

PCP1′ and 1−1926-PCP3 (Figure 2A), like 1−1926-PCP3TE (TE mutant), were found to synthesize only HPT-Cys and HPT-Cys-Cys, but HPTT-COOH and HPTT-Cys were not detected (data not shown). Additionally, the observed rate constants of product formation listed in Table 3 were similar to those of the TE mutant of 1−1926-PCP3TE, suggesting inefficient product flux catalyzed by the two fusion proteins.

Product Formation Catalyzed by the Hybrids from Whole-Module Fusions at 30 °C. The NRPS module of HMWP1 has been shown to be enzymatically active (*10*). It is possible that the *in cis* fusion of these modules from HMWP2 and HMWP1 could form modular enzymes with novel activities. Two large fusion proteins, 1−1491-1896−3163 and HMWP2-1896−3163 (Figure 2A), were thus constructed, and their

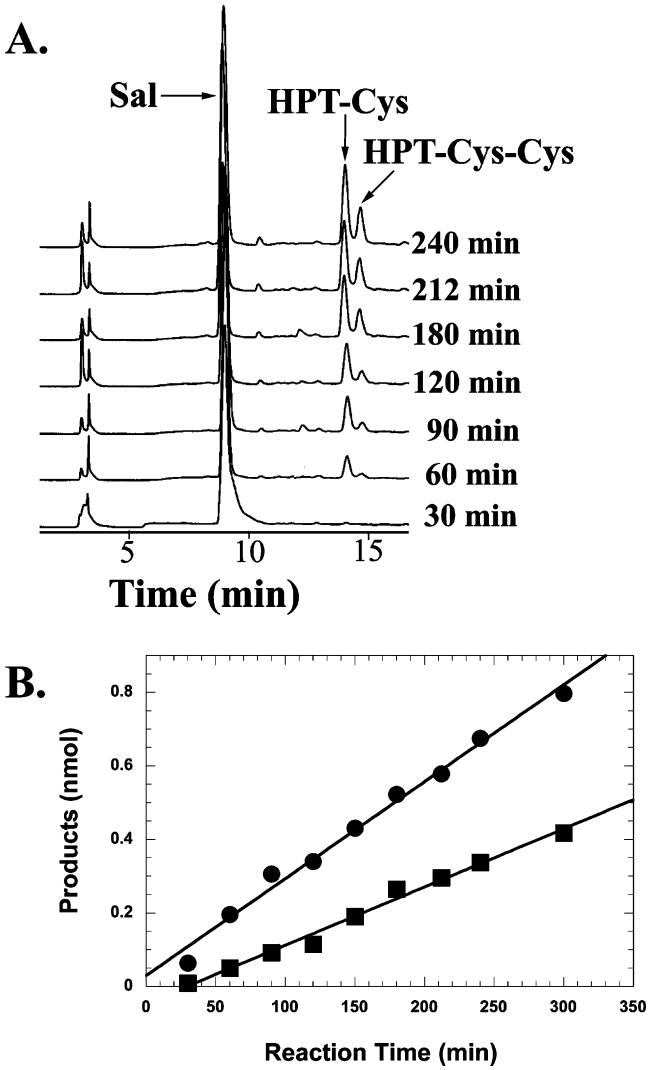


FIGURE 11: Time course of product formation catalyzed by 5 μM 1−1926-PCP3TE (TE mutant) at 30 °C. The reaction conditions are the same as those described in the legend of Figure 4. (A) HPLC traces. The peaks were identified as HPT-Cys and HPT-Cys-Cys by both co-injection with synthetic standards and LC−MS. (B) Plot of quantities of HPT-Cys (●) and HPT-Cys-Cys (■) vs reaction time. Linear fits of data yield values of k_{obs} of 0.010 and 0.006 min^{-1} for the formation of HPT-Cys and HPT-Cys-Cys, respectively.

Table 3: Observed Rate Constants of Product Formation at 30 °C by Fusion Proteins from PCP Swapping at the PCP2 Site of HMWP2

enzyme	HPT-Cys (min^{-1})	HPT-Cys-Cys (min^{-1})	HPTT/HPTT-Cys (min^{-1})
1−1382·1383−2035[a]	0.057	0.035	0.033
1−1926-PCP1′	0.015	0.024	
1−1926-PCP3	0.011	0.014	

[a] Data are from ref *9*.

final products were expected to be HPTT-COOH and HPTTT-COOH, respectively. However, the major products from 20 h reactions catalyzed by the two proteins were HPT-COOH and HPTT-COOH, respectively (Figure 9). This suggests that the NRPS module of HMWP1 in the fusion proteins was inactive. However, its TE domain was active in the two large hybrids due to the formation of HPT-COOH and HPTT-COOH (Figure 9). Neither HPT-Cys nor HPTT-Cys was detected (Figure 9). The active TE domain was

Table 4: Observed Rate Constants of Product Formation at 30 °C by Two Whole-Module Fusion Proteins

enzyme	HPT (min^{-1})	HPT-Cys (min^{-1})	HPT-Cys-Cys (min^{-1})	HPTT (min^{-1})
1−1491-1896−3163	0.016	0.0081		
HMWP2-1896−3163		0.040	0.0042	0.029

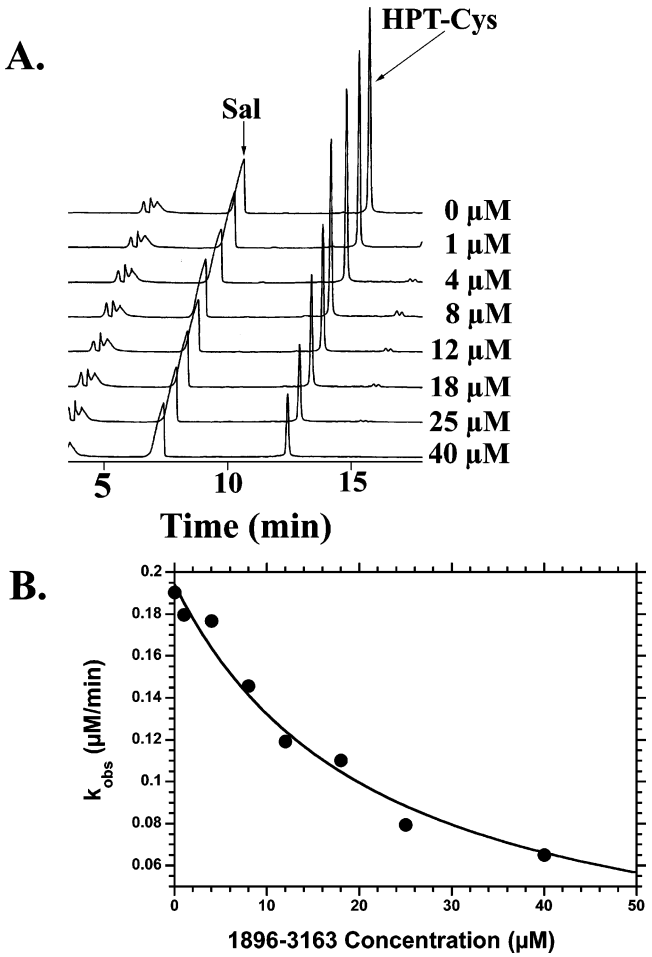


FIGURE 12: Determination of the affinity of 1−1491 and 1896−3163 at 30 °C in the presence of 75 mM Tris-HCl (pH 8.8), 5 mM TCEP, 0.5 μM YbtE, 5 mM L-cysteine, 1 mM salicylate, 10 mM ATP, and 10 mM $MgCl_2$. The protein fragments were phosphopantetheinylated for 30 min by 0.3 μM *B. subtilis* Sfp and 1 mM CoA prior to the addition of ATP to initiate the reactions. The concentration of 1896−3163 was varied from 0 to 40 μM as indicated, while the concentration of 1−1491 was held constant at 2 μM. The reactions were quenched with 8.5% H_3PO_4 after 150 min. (A) HPLC traces. The only product peak was identified as HPT-Cys by both co-injection with a synthetic standard and LC−MS. (B) Plot of k_{obs} vs the 1896−3163 concentration. A K_d of 20 ± 2 μM was calculated from a fit of the data (●) to an equation described in the Appendix.

further confirmed by the hydrolysis of HPT-SNAC by the two fusions (Figure 6). The observed rate constants of product formation (Table 4) catalyzed by 1−1491-1896−3163 and HMWP2-1896−3163 were similar to those of 1−1392-PCP3TE and 1−1926-PCP3TE (Table 2), respectively. This suggested that the 1896−2912 region of the NRPS module of HMWP1 served as a large linker in each of the two large hybrids.

Product Formation Catalyzed by Protein Fragments in Trans at 30 °C. So far, all fusion proteins (Figure 2A) were

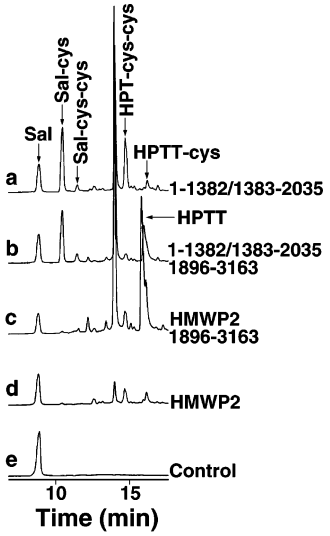


FIGURE 13: HPLC analysis of reaction products by different protein combinations at 30 °C for 13 h. The reaction conditions are the same as those described in the legend of Figure 8: (a) 1−1382 (5 μM) with 1383−2035 (5 μM), (b) 1−1382 (5 μM) with 1383−2035 (5 μM) and 1896−3163 (20 μM), (c) HMWP2 (5 μM) with 1896−3163 (20 μM), (d) HMWP2 (5 μM), and (e) control. The peaks were identified by co-injection with synthetic standards and LC−MS.

shown to be active (see above). It would be interesting to compare the activities of those fusions and their corresponding fragments *in trans*. The 1−1392·PCP2, 1−1392·PCP3, 1−1382·PCP3TE, 1−1926·PCP1′, 1−1926·PCP3, and 1−1926·PCP3TE combinations were examined, and there was no detectable product resulting from product flux between the fragments (data not shown). This was strikingly different from what was observed with the fusion proteins described above. This was also in contrast to the 1−1382·PCP1 (*9*), 1−1382·1383−2035 (*9*), and 1−1926·PCP2 (Figure 8) active combinations which reconstituted the activities of the corresponding intact proteins.

When 1−1491 (*8*) was combined with 1896−3163, the production of HPT-Cys by 1−1491 (*8*) was surprisingly inhibited. There was product flux between the two fragments *in trans*, and the TE domain of 1896−3163 was not actively involved in catalysis (Figure 12A). The inhibition was more severe at higher concentrations of 1896−3163 (Figure 12A). The affinity of the two fragments was estimated to be 20 μM from the dependence of the observed rate constants of HPT-Cys formation on the concentrations of 1896−3163 (Figure 12B).

The HMWP2·1896−3163 combination was more active in the synthesis of HPT-Cys and HPTT-COOH than HMWP2 alone (Figure 13), but less active than HMWP2-1896−3163 in the synthesis of HPTT-COOH (Figure 9). The 1−1382·1383−2035·1896−3163 triple combination had activity similar to that of HMWP2·1896−3163 (Figure 13). HMWP2 was also more active in catalyzing the synthesis of HPT-Cys and HPTT-COOH if incubated with either PCP3TE or HMWP1 (data not shown). Interestingly, the TE domain in these combinations, unlike 1−1491·1896−3163 (Figure 12A), was active as a hydrolase *in trans* to produce HPTT-COOH (Figure 13). Moreover, there was no product flux between the fragments from HMWP1 and HMWP2. The affinity of each combination was determined in a manner similar to that used for 1−1491·1896−3163 (data not shown)

Table 5: Affinity of Two Interacting Fragments at 30 °C

enzymes	K_D (μM)
1−1491·1896−3163	20 ± 2
HMWP2·1896−3163	8 ± 7
HMWP2·PCP3TE	63 ± 19
HMWP2·HMWP1	10 ± 2

and listed in Table 5. The K_D values suggest that the larger the fragments, the higher their affinity for each other.

DISCUSSION

A vast number of polyketides and short peptides that are pharmaceutically and biotechnologically important are synthesized by large modular enzymes PKSs and NRPSs, respectively. Interestingly, nature has created another large class of natural products synthesized by the mixed PKS/NRPS systems to provide evolutionary advantage to certain microorganisms. To learn from nature, myself and others have attempted to develop the following approaches to further increase the structural diversity of natural products and hopefully find novel, pharmaceutically active compounds. (i) Genetically create hybrid enzymes by swapping domains or modules from homologous and heterologous biosynthetic clusters. For example, replacing the second module of bacitracin synthetase A with modules of MbtB and MtaD from the biosynthesis systems of mycobactin and myxothiazol, respectively, creates hybrid synthetases to produce novel heterocyclic dipeptides *in vitro* (*26*). (ii) Mutate the active sites of enzymatic domains of modular enzymes to alter substrate specificity. For example, a single-point H322E mutation in the adenylation domain of the SrfA-B2 module of the surfactin synthetase A alters the substrate specificity of the whole module from L-Asp to L-Asn *in vitro* and yields an expected surfactin derivative *in vivo* (*27*). (iii) Feed enzymes with alternative synthetic substrates. For example, exogenous addition of diketide mimics to the cultures of *Streptomyces coelicolor* CH999 containing a plasmid encoding the C729A mutant of deoxyerythronolide B synthase produces multiple milligrams of unnatural polyketides (*25*). The success of these approaches has been relatively rare so far due to the limited knowledge of NRPSs, PKSs, and mixed NRPS/PKS systems at the molecular level. Specifically, the domain boundaries are not clearly defined, and compatibility of different or similar function domains is unknown. Thus, random domain swapping or module fusion often leads to inactive proteins. More research needs to be done to understand the basic mechanism of these modular enzymes. In this paper, I used the yersiniabactin synthetase, a typical mixed PKS/NRPS system, as a model to address several basic questions such as the swappability of PCPs and the portability of the TE domain in NRPSs.

The similarity between the three-dimensional NMR structures of two ACPs, *E. coli* FAS ACP (*28*) and *S. coelicolor* actinorhodin PKS ACP (*29*), and one PCP from *Bacillus brevis*, tyrocidine synthetase III (*30*), suggests that all carrier proteins, including the three PCPs of the Ybt synthetase, possess similar overall architecture. The three PCPs of the Ybt synthetase have been found to function similarly as cysteine carriers and thiol way stations during Ybt synthesis (*7*). On the basis of their similar functions and structures, these PCPs may be exchangeable despite having low levels of sequence homology (Figure 3). This hypothesis was confirmed by the results obtained with 1−1392-PCP2 and 1−1392-PCP3 generated from PCP swapping at the PCP1 site of HMWP2 (Tables 1 and 3). However, each fusion protein was less active than wild-type 1−1491 by 3−8-fold (Table 1). The lack of a clearly defined boundary between the adenylation and PCP1 domains and non-native PCP occupying the PCP1 site could slow the condensation/heterocyclization activities of Cy1. The boundary effect was clearly demonstrated by the activity of 1−1382-PCP1′ with a nine-residue deletion at the boundary between the adenylation and PCP1 domains, which was 60% of the activity of 1−1491 (Table 1). The swapping at the PCP2 site of HMWP2 was more complicated than at the PCP1 site. Although 1−1926-PCP1′ and 1−1926-PCP3 synthesized HPT-Cys-Cys, there was no detectable formation of HPTT-COOH and HPTT-Cys which HMWP2 is capable of synthesizing (*11*). This suggests that non-native PCPs at the PCP2 site of HMWP2 abolished the heterocyclization activity of Cy2 without affecting its condensation activity. This surprising separation of the two activities of Cy2 of the two fusion proteins has not been achieved by the previous mutagenesis of the conserved DX_4DX_2S motif of Cy1 (*11*). The heterocyclization activity of Cy2 could be more sensitive to the artificial junctions and non-native PCPs at the PCP2 site than the same activity of Cy1. Nevertheless, these results demonstrated that the PCPs can be swapped with other PCPs, especially those from the same modular synthetase, although the fusion proteins have less (and sometimes different) activity than corresponding wild-type proteins. Since carrier protein domains are not enzymatic and are generally considered to be more tolerant than enzymatic function domains in non-native environments, it is expected to be more troublesome if the enzymatic function domains are swapped. Therefore, caution should be taken when using a domain swapping strategy to manipulate existing biosynthetic clusters to create novel hybrid enzymes.

The intermediate products are normally generated in stoichiometric equivalence by multimodular enzymes and are not wastefully discharged into solution. However, some pharmacologically important intermediates need to be produced with high turnovers and large quantities. Novel ways of releasing enzyme-bound intermediates must be found. Cosubstrates or small molecules could be used as nucleophiles to attack the enzyme-bound intermediates and release them into solution in a multiturnover fashion. For example, cysteine, water, and particularly DTT molecules were found previously (*8*, *9*, *11*) and in this paper to cleave HMWP2-bound intermediates. The disadvantages of this approach are slow turnovers and intermediate−nucleophile adducts. In PKSs, NRPSs, and mixed systems, the release of final products is generally catalyzed by a C-terminal TE domain, rather than an external TE which presumably functions as an editing enzyme to remove mischarged intermediate products (*31*−*33*). Thus, the C-terminal TE domains could be moved around and used as tools to release important intermediates. This has been successfully achieved in PKSs previously (*21*−*25*). My results demonstrate that the TE domains are also portable in NRPSs. The TE domain located at the C-terminus of the NRPS module of HMWP1 has been shown previously (*13*) to hydrolyze acyl-SNACs as a hydrolase and has been postulated to cleave the final enzyme-

bound intermediate, Ybt-*S*-PCP3, to produce Ybt (*7*, *13*). This TE was also active in the hydrolytic release of HPT-COOH and HPTT-COOH from enzyme-bound intermediates after being fused to the two NRPS modules of HMWP2 (Figures 5A, 9, and 10A). Although the product flux was slightly facilitated by the fused TE domain (Table 2), the overall turnover rate constants were small. The slow turnover has also been observed with the module 2−module 9−TE hybrid derived from tyrocidine synthetase of *Bacillus brevis*, where the TE is originally attached to module 10 (*19*). The artificial junctions and substrate specificity of the internal TE domains in these fusion proteins may lead to slow product flux. But the slow catalysis by 1−1491-1896−3163 and HMWP2-1896−3163 (Table 4) argues against the importance of the first factor since the TE domain was fused to native proteins 1−1491 and HMWP2 through a long and inactive protein linker, 1896−2912 (Figure 2A). The second factor is supported by the fact that the HMWP1 TE domain has been found previously to be a substrate specific hydrolase (*13*). The native substrate for the TE domain is Ybt-*S*-PCP3, not HPT-*S*-PCP3 or HPTT-*S*-PCP3. The suboptimal substrates, HPT-*S*-PCP3 and HPTT-*S*-PCP3, could thereby lead to slow rate constants observed with the four hybrids [1−1392-PCP3TE, 1−1926-PCP3TE, 1−1491-1896−3163, and HMWP2-1896−3163 (Tables 2 and 4)]. Therefore, other than the domain boundary issue, the substrate specificity of the TE domains or other catalytic domains should also be considered before these domains are fused or swapped.

The domain swapping and fusion approach should have a kinetic advantage over the fragment combination approach in creating more products since the prior depends on *in cis* catalysis while the latter on *in trans* catalysis. One fragment in each combination has to be used in molar excess to saturate the other fragment to constitute the full activity of the corresponding intact protein. This has been demonstrated by two examples: 1−1491 and 1−1382·PCP1 (*9*), HMWP2 (*11*), and 1−1382·1383−2035 (*9*) or 1−1926·PCP2 (Figure 8). If the protein−protein recognition is very weak, saturation of the fragments is not achievable and no product flux between fragments can be observed due to the severe kinetic penalty. This often happens to the fragments derived either from heterologous systems or from different modules of the same systems. There were no detectable product fluxes between the fragments in the 1−1392·PCP2, 1−1392·PCP3, 1−1392·PCP3TE, 1−1926·PCP1′, 1−1926·PCP3, and 1−1926·PCP3TE combinations, while the corresponding fusions were catalytically active (Tables 1 and 2).

The *in trans* interaction of the large protein fragments (Table 5) was quite different from that of their corresponding fusion proteins *in cis* due to the interaction of unintended domains (Table 5). The first example was the whole-module fusion 1−1491-1896−3163 which catalyzed the formation of HPT-COOH (Figure 9) as if the 1−1491 fragment fused to the TE domain directly. However, in the 1−1491·1896−3163 fragment combination, the 1896−3163 fragment inhibited the catalysis of 1−1491 (Figure 11). The interaction between the PCP3 domain of 1896−3163 and the adenylation domain of 1−1491 fragment could lead to the blockage of the PCP1 site of the 1−1491 fragment by the 1896−3163 fragment, which in turn slowed the cleavage of HPT-*S*-PCP1 by cysteine and the production rate constant of HPT-Cys. The second example was another whole-module fusion, HMWP2-1896−3163, which catalyzed the synthesis of HPTT-COOH as if HMWP2 fused to TE directly (Figure 12). HMWP2 alone (Figure 12) was less active in catalysis than in the HMWP2·PCP3TE, HMWP2·1896−3163, and HMWP2·HMWP1 combinations. The higher activity of HMWP2 in these combinations is probably due to protein−protein interactions which extended the structure of HMWP2, and thus, the PCP2 site was more exposed and accessible to nucleophiles for hydrolysis. These two examples not only demonstrate the different activities of the protein fragments *in trans* and *in cis* but also provide a lesson for the combinatory approach in the hope of synthesizing novel compounds.

In summary, I used Ybt synthetase as a model system to demonstrate that PCPs were swappable and TE domains were portable in NRPSs or mixed PKS/NRPS systems, although the fusion proteins were less active than wild-type proteins. The protein−protein interactions *in trans* (fragments) and *in cis* (fusion proteins) were sometimes similar and sometimes very different. My work provides several lessons for the protein engineering approach to rational design of "non-natural" natural products.

ACKNOWLEDGMENT

I thank Dr. Christopher T. Walsh at Harvard Medical School (Boston, MA) for his mentoring and comments on the manuscript.

APPENDIX

The observed rate constants of product formation catalyzed by protein fragment 1 (m_1) and the complex of two protein fragments (m_3), the K_d of the complex (m_2), and the initial concentrations of protein fragment 1 (2 μM) and protein fragment 2 (m_0) have following relationship:

	m_1	m_2	m_3
	Product ← Protein1 + Protein2 = Protein1•Protein2 → Product		
Initial concentration:	2 μM	m_0	0
At time t:	2 - X	m_0 - X	X

where X is the concentration of the protein complex at time t. Thus, $(2 - X)(m_0 - X) = Xm_2$. Then, $X = 0.5\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$.

Since the observed rate constant of product formation (k_{obs}) is the sum of product formation rate constants from both protein fragment 1 and the complex

$$\begin{aligned} k_{obs} &= m_1[\text{protein 1}]_{free} + m_3[\text{protein 1·protein 2}] \\ &= m_1(2 - X) + m_3X \\ &= 2m_1 + (m_3 - m_1)X \\ &= 2m_1 + 0.5(m_3 - m_1)\{2 + m_0 + m_2 - \\ &\qquad [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\} \end{aligned}$$

If $m_1 \gg m_3$, $k_{obs} = 2m_1 - 0.5m_1\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$. If $m_3 \gg m_1$, $k_{obs} = 2m_1 + 0.5m_3\{2 + m_0 + m_2 - [(2 + m_0 + m_2)^2 - 8m_0]^{0.5}\}$.

REFERENCES

1. Fetherston, J. D., Lillard, J. W., Jr., and Perry, R. D. (1995) Analysis of the pesticin receptor from *Yersinia pestis*: Role in

iron-deficient growth and possible regulation by its siderophore, *J. Bacteriol. 177*, 1824−1833.
2. Perry, R. D., and Fetherston, J. D. (1997) *Yersinia pestis*: Etiologic agent of plague, *Clin. Microbiol. Rev. 10*, 35−66.
3. Bearden, S. W., Fetherston, J. D., and Perry, R. D. (1997) Genetic organization of the yersiniabactin biosynthetic region and construction of avirulent mutants in *Yersinia pestis*, *Infect. Immun. 65*, 1659−1668.
4. Perry, R. D., Balbo, P. B., Jones, H. A., Fetherston, J. D., and DeMoll, E. (1999) Yersiniabactin from *Yersinia pestis*: Biochemical characterization of the siderophore and its role in iron transport and regulation, *Microbiology 145* (Part 5), 1181−1190.
5. Crosa, J. H., and Walsh, C. T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria, *Microbiol. Mol. Biol. Rev. 66*, 223−249.
6. Bobrov, A. G., Geoffroy, V. A., and Perry, R. D. (2002) Yersiniabactin production requires the thioesterase domain of HMWP2 and YbtD, a putative phosphopantetheinylate transferase, *Infect. Immun. 70*, 4204−4214.
7. Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T., and Perry, R. D. (1998) Iron acquisition in plague: Modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*, *Chem. Biol. 5*, 573−586.
8. Gehring, A. M., Mori, I., Perry, R. D., and Walsh, C. T. (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*, *Biochemistry 37*, 11637−11650.
9. Suo, Z., Walsh, C. T., and Miller, D. A. (1999) Tandem heterocyclization activity of the multidomain 230 kDa HMWP2 subunit of *Yersinia pestis* yersiniabactin synthetase: Interaction of the 1−1382 and 1383−2035 fragments, *Biochemistry 38*, 17000.
10. Miller, D. A., Luo, L., Hillson, N., Keating, T. A., and Walsh, C. T. (2002) Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*, *Chem. Biol. 9*, 333−344.
11. Keating, T. A., Miller, D. A., and Walsh, C. T. (2000) Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase, *Biochemistry 39*, 4729−4739.
12. Keating, T. A., Suo, Z., Ehmann, D. E., and Walsh, C. T. (2000) Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain, *Biochemistry 39*, 2297−2306.
13. Suo, Z., Chen, H., and Walsh, C. T. (2000) Acyl-CoA hydrolysis by the high molecular weight protein 1 subunit of yersiniabactin synthetase: Mutational evidence for a cascade of four acyl-enzyme intermediates during hydrolytic editing, *Proc. Natl. Acad. Sci. U.S.A. 97*, 14188−14193.
14. Geoffroy, V. A., Fetherston, J. D., and Perry, R. D. (2000) *Yersinia pestis* YbtU and YbtT are involved in synthesis of the siderophore yersiniabactin but have different effects on regulation, *Infect. Immun. 68*, 4452−4461.
15. Pfeifer, B. A., Wang, C. C., Walsh, C. T., and Khosla, C. (2003) Biosynthesis of Yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host, *Appl. Environ. Microbiol. 69*, 6698−6702.
16. Suo, Z., Tseng, C. C., and Walsh, C. T. (2001) Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase, *Proc. Natl. Acad. Sci. U.S.A. 98*, 99−104.
17. Miller, D. A., and Walsh, C. T. (2001) Yersiniabactin synthetase: Probing the recognition of carrier protein domains by the catalytic heterocyclization domains, Cy1 and Cy2, in the chain-initiating HWMP2 subunit, *Biochemistry 40*, 5313−5321.
18. 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) A new enzyme superfamily: The phosphopantetheinyl transferases, *Chem. Biol. 3*, 923−936.
19. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) Construction of hybrid peptide synthetases by module and domain fusions, *Proc. Natl. Acad. Sci. U.S.A. 97*, 5848−5853.
20. Finking, R., and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides1, *Annu. Rev. Microbiol. 58*, 453−488.
21. Cortes, J., Wiesmann, K. E., Roberts, G. A., Brown, M. J., Staunton, J., and Leadlay, P. F. (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage, *Science 268*, 1487−1489.
22. Kao, C. M., Luo, G., Katz, L., Cane, D. E., and Khosla, C. (1995) Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase, *J. Am. Chem. Soc. 117*, 9105−9106.
23. Kao, C. M., Luo, G., Katz, L., Cane, D. E., and Khosla, C. (1996) Engineered Biosynthesis of Structurally Diverse Tetraketides by a Trimodular Polyketide Synthase, *J. Am. Chem. Soc. 118*, 9184−9185.
24. Kao, C. M., McPherson, M., McDaniel, R. N., Fu, H., Cane, D. E., and Khosla, C. (1997) Gain of Function Mutagenesis of the Erythromycin Polyketide Synthase. 2. Engineered Biosynthesis of an Eight-Membered Ring Tetraketide Lactone, *J. Am. Chem. Soc. 119*, 11339−11340.
25. Jacobsen, J. R., Hutchinson, C. R., Cane, D. E., and Khosla, C. (1997) Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase, *Science 277*, 367−369.
26. Duerfahrt, T., Eppelmann, K., Muller, R., and Marahiel, M. A. (2004) Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis, *Chem. Biol. 11*, 261−271.
27. Eppelmann, K., Stachelhaus, T., and Marahiel, M. A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics, *Biochemistry 41*, 9718−9726.
28. Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., and Clore, G. M. (1988) Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry, *Eur. J. Biochem. 175*, 9−15.
29. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A., and Simpson, T. J. (1997) Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2), *Biochemistry 36*, 6000−6008.
30. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases, *Struct. Folding Des. 8*, 407−418.
31. Butler, A. R., Bate, N., and Cundliffe, E. (1999) Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*, *Chem. Biol. 6*, 287−292.
32. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, *Chem. Rev. 97*, 2651−2674.
33. Schneider, A., and Marahiel, M. A. (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*, *Arch. Microbiol. 169*, 404−410.

BI047538S