Expression, Purification, and Characterization of HMWP2, a 229 kDa, Six Domain Protein Subunit of Yersiniabactin Synthetase †

ARTICLE *in* BIOCHEMISTRY · MAY 2000

Impact Factor: 3.02 · DOI: 10.1021/bi992923g · Source: PubMed

3 AUTHORS, INCLUDING:

Thomas A Keating
ImmunoGen, Inc.
42 PUBLICATIONS **2,924** CITATIONS

SEE PROFILE

Expression, Purification, and Characterization of HMWP2, a 229 kDa, Six Domain Protein Subunit of Yersiniabactin Synthetase†

Thomas A. Keating, Deborah Ann Miller, and Christopher T. Walsh*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received December 21, 1999

ABSTRACT: The six domain, 229 kDa HMWP2 subunit of the *Yersinia pestis* yersiniabactin (Ybt) synthetase has been expressed in soluble, full-length form in *E. coli* as a C-terminal His8 construct at low growth temperatures and with attenuated induction. All six domains of this nonribosomal peptide synthetase subunit, three phosphopantetheinylatable carrier protein domains (ArCP, PCP1, PCP2), one adenylation (A) domain, and two cyclization domains (Cy1, Cy2), have been assayed and are functional. Mutants that convert the phosphopantetheinylatable serine residue to alanine in each of the carrier protein domains accumulate acyl-*S*-enzyme intermediates upstream of the blocked apo carrier protein site. The ArCP mutant cannot be salicylated by the adenylation protein YbtE; the PCP1 mutant releases salicyl-cysteine from thiolysis of the Sal-*S*-ArCP intermediate; and the PCP2 mutant releases hydroxyphenyl-thiazolinyl-cysteine from the HPT-*S*-PCP1 acyl enzyme intermediate, all of which demonstrates processivity and directionality of chain growth. Restoration of the ArCP mutant's function was accomplished with the native ArCP fragment added in trans. The wild-type HMWP2 subunit accumulates hydroxyphenyl-4,2-bithiazolinyl-*S*-enzyme at its most downstream PCP2 carrier site, presumably for transfer to the next subunit, HMWP1. The A domain was found to activate and transfer to PCP1 and PCP2 not only the natural L-Cys but also *S*-2-aminobutyrate, L-*β*-chloroalanine, and L-Ser, enabling testing of the substrate specificity of the Cy domain. Probes of Cy domain function include mutagenesis of the Cy1 domain's conserved signature motif DX_4-DX_2S to show that both D residues but not the S are crucial for both amide bond formation and heterocyclization. Also the Cy1 domain would accept an alternate upstream electrophilic donor substrate (2,3-dihydroxybenzoyl-*S*-ArCP) but would not process any of the three alternate downstream nucleophilic acceptors in place of Cys-*S*-PCP1, even for the amide bond-forming step in chain elongation.

Yersiniabactin (Ybt,[1] **1**) is a small molecule siderophore (*1*) (iron chelator) from the causative agent of the plague, *Yersinia pestis*, and has been shown to be a virulence factor (*2*), contributing to pathogen growth under the iron-limited conditions in hosts. Although stereochemical information and iron-coordination sites have not yet been revealed, the crystal structure of the closely related micacocidin A (*3*) indicates that Ybt derives its high affinity for Fe(III) [$K_{assoc} \sim 4 \times 10^{36}$ (*4*)] from its aryl and aliphatic hydroxyls, its three heterocyclic nitrogens, and the terminal carboxylate.

Ybt is biosynthesized by a mixed nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) system. Likened to an assembly line, an NRPS or PKS involves covalent attachment of monomers through a 4′-phosphopantetheine (P-pant) tether to conserved carrier domains, followed by bond formation by condensation or ketosynthase domains, and ending with release of the mature product by a terminating thioesterase (TE) domain (*5*). In an analysis of the 3 central proteins of yersiniabactin synthetase (*6*), we proposed assignments for 16 domains in the assembly line: 1 in YbtE, 6 in HMWP2 (high molecular weight protein 2), and 9 in HMWP1 (Figure 1). Each of these domains was prospectively assigned a function based on homology and expected requirements for Ybt biosynthesis. Structural and biochemical features of interest include the mixed NRPS/PKS organization, the single adenylation (A) domain, which loads three peptidyl carrier protein (PCP) domains with L-cysteine, and the timing and mechanism of both heterocycle formation and multiple C-methylations seen in **1**. To this point, we have demonstrated the ability of the 1−1491 residue fragment of HMWP2 to form 2-hydroxyphenyl-thiazolinyl-cysteine (HPT-Cys, **6**) from salicylate and L-Cys (*7*), the interaction of the 1−1382 and 1383−2035 fragments of HMWP2 to yield the bisheterocyclic 2-hydroxyphenyl-thiazolyl-thiazolinyl-carboxylate and -cysteine (HPTT-

† This work has been supported by the National Institutes of Health (Grant GM20011 to C.T.W.). T.A.K. is a Fellow of the Cancer Research Fund of the Damon Runyon−Walter Winchell Foundation (DRG-1483).

* To whom correspondence should be addressed. Phone: 617-432-1715. Fax: 617-432-0438. E-mail: walsh@walsh.med.harvard.edu.

[1] Abbreviations: NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; HMWP, high molecular weight protein; CoA, coenzyme A; A, adenylation domain; Cy, cyclization domain; PPTase, phosphopantetheinyl transferase; P-pant, 4′-phosphopantetheine; PP_i, inorganic pyrophosphate; ArCP, aryl carrier protein domain; PCP, peptidyl carrier protein domain; TE, thioesterase domain; HPT-, 2-(2-hydroxyphenyl)thiazolinyl-; HPTT-, 2-(2-hydroxyphenyl)thiazolyl-2,4-thiazolinyl-; DHPT-, 2-(2,3-dihydroxyphenyl)thiazolinyl-; SOE, splicing by overlap extension; IPTG, isopropyl *β*-D-thiogalactopyranoside; PNP, calf spleen purine nucleoside phosphorylase; PP_i-ase, bakers yeast inorganic pyrophosphatase; MESG, 2-amino-6-mercapto-7-methylpurine ribonucleoside; DTT, dithiothreitol; TCEP, tris(carboxyethyl)-phosphine; TCA, trichloroacetic acid; DHB, 2,3-dihydroxybenzoate.

10.1021/bi992923g CCC: $19.00
Published on Web 03/30/2000

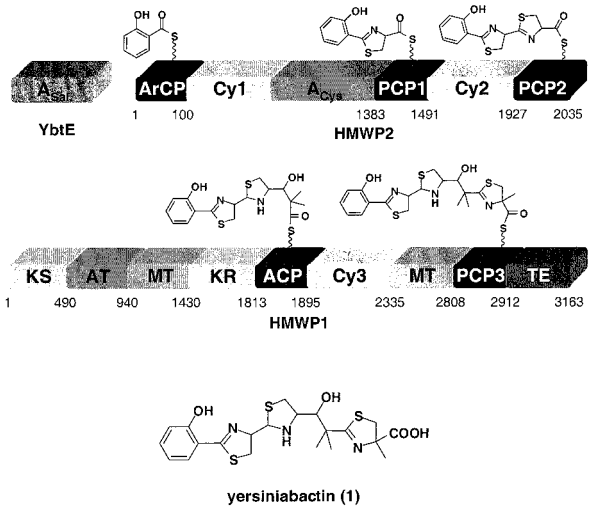

FIGURE 1: Yersiniabactin (Ybt, **1**) and the three proteins of yersiniabactin synthetase. YbtE is a single-domain, salicyl-AMP ligase. HMWP2 and HMWP1 encompass the remaining 15 domains. Numbers below the proteins indicate the residue marking each domain boundary. Predicted intermediates in Ybt assembly are drawn covalently tethered to the appropriate carrier protein domains through a 4′-phosphopantetheine (wavy line). A = adenylation; ArCP = aryl carrier protein; Cy = cyclization; PCP = peptidyl carrier protein; KS = ketosynthase; AT = acyltransferase; MT = methyltransferase; KR = ketoreductase; ACP = acyl carrier protein; TE = thioesterase. The absolute and relative stereochemistry of Ybt is as yet unknown.

COOH, **3**, and HPTT-Cys, **4**, respectively) (*8*), and also the competence of the HMWP2 adenylation domain to recognize and covalently cysteinylate the PCP3-TE fragment of HMWP1 in trans at rates significantly elevated over background (*9*).

Herein, we focus on the intact HMWP2, a 229 kDa, 2035 residue protein whose heterologous expression in soluble form and subsequent functional characterization posed several challenges. First, successful heterologous expression and purification of a protein this size is nonroutine. But second, and more importantly, we wish to prove the properties, capabilities, and specificity of this subunit absent its downstream partner HMWP1. This entails a domain-by-domain analysis employing both assays and mutagenesis tailored for the specific functions being examined. HMWP2 consists of the following: (1) three carrier protein domains, all of which require posttranslational attachment of a coenzyme A-derived P-pant prosthetic arm by a phosphopantetheinyl transferase (PPTase); the first, aryl carrier protein (ArCP), domain carries salicylate, while the downstream two, PCP1 and PCP2, carry L-Cys; (2) a single A domain, which is responsible for activating and transferring L-Cys to both PCP1 and PCP2 (and PCP3, on HMWP1); this raises topological and timing issues of cysteinyl adenylate capture; and (3) two cyclization domains (Cy1 and Cy2), which are proposed both to catalyze amide bond formation between the monomers and to heterocyclize the cysteines to form the thiazoline−thiazolidine rings. To analyze these domains and their interactions, we have inactivated each of the three carrier protein domains by mutagenesis, examined the substrate specificity of the A domain in amino acid activation and transfer, and then used both unnatural monomers and site-directed mutagenesis to probe Cy domain specificity and to define essential residues for condensation/heterocyclization.

EXPERIMENTAL PROCEDURES

Materials and Methods. Luria−Bertani (LB) medium was prepared and used for culturing *E. coli* strains as previously described (*10*). Competent cells of *E. coli* strains DH5α and DM1 were purchased from GibcoBRL; cells of BL21(DE3) were purchased from Novagen. Expression plasmid pET37b was obtained from Novagen. Restriction endonucleases, T4 DNA ligase, and pUC18 were purchased from New England Biolabs. CoA, TCEP, *S*-2-aminobutyric acid, and *β*-Cl-L-alanine were purchased from Sigma/Aldrich. [^{3}H]-CoA (57 μM, 56 Ci/mol), [^{14}C]-salicylate (1.8 mM, 55.5 Ci/mol), [^{35}S]-L-cysteine (10 μM, 1075 Ci/mmol), [^{3}H]-L-serine (50 μM, 25.8 Ci/mmol), and sodium [^{32}P]-pyrophosphate (534 μM, 19 Ci/mmol) were purchased from New England Nuclear. [1-^{14}C]-L/D-2-Aminobutyric acid (15.6 mM, 7.6 Ci/mol) was purchased from ICN. 2-Amino-6-mercapto-7-methylpurine ribonucleoside (MESG) was synthesized from 2-amino-6-chloropurine ribonucleoside according to published procedures (*11*). Phosphate-free, calf spleen purine nucleoside phosphorylase (PNP) and bakers yeast inorganic pyrophosphatase (PP_i-ase) were purchased lyophilized from Sigma and stored as 0.01 unit/μL stocks in 75 mM Tris, pH 7.5, 10 mM $MgCl_2$. Sfp, a phosphopantetheinyl transferase from *Bacillus subtilis*, has been described previously (*12*) and was a gift from Luis E. N. Quadri. EntD, a phosphopantetheinyl transferase from *E. coli*, has also been described previously (*13*). The ArCP fragment (residues 1−100) of HMWP2 was overproduced and purified according to published protocols (*7*). Standard recombinant DNA techniques were performed as described (*10*). Preparation of plasmid DNA, gel purification of DNA fragments, and purification of PCR products were performed using QIAprep plasmid miniprep kits, QIAquick gel extraction kits, and QIAquick DNA purification kits, respectively (Qiagen). PCRs were carried out with *Pfu* polymerase according to its supplier's instructions (Stratagene), except for the addition of 5% DMSO. Oligonucleotide primers were purchased from Integrated DNA Technologies, and DNA sequencing was performed on double-stranded DNA by the Molecular Biology Core Facilities of the Dana-Farber Cancer Institute (Boston, MA). Protein concentrations were determined from the predicted molar extinction coefficient (ϵ) and absorbance at 280 nm. HPLC analyses were performed on a Beckman System Gold equipped with a VYDAC C18 reverse phase analytical column. Detection was at 254 nm, with mobile phase A: 100 μL of formic acid/200 μL of triethylamine in 1 L of water; mobile phase B: 4:1 acetonitrile:A. At a flow rate of 1 mL/min, a linear gradient was maintained from 10% B to 100% B over 23 min. ^{1}H NMR spectra were acquired in CD_3OD with a Varian Unity 500 spectrometer and referenced to residual CHD_2OD. Coupling constants are listed in hertz. Positive-ion detection electrospray mass spectrometry (ESMS) was performed by the Mass Spectrometry Facility of the Department of Chemistry & Chemical Biology, Harvard University.

Cloning of HMWP2. The C-terminal ~2100 bp of *irp2*, encompassing the PCP1-Cy2-PCP2 fragment of HMWP2, was amplified from pIRP2 (*2*) using the primer pair (5′-TTAACCAATGGGCAGCCGGAG-3′, 1F; 5′-TGACCGCTC-

GAGTATCCGCCGCTGACGACGG-3′, 1R; restriction sites underlined, mutations in boldface). The amplified fragment was digested with *Not*I and *Xho*I and ligated to pET22b-irp2/1−1491 (*7*) digested with the same enzymes to generate pHMWP2.CH6, which expresses the translational fusion HMWP2−His6, in which the sequence LEHHHHHH is appended to the C-terminus of HMWP2. The entire *irp2* gene was then excised from pHMWP2.CH6 by digestion with *Nde*I and *Xho*I, and ligated to the corresponding sites in pET37b. The resulting plasmid, pHMWP2.CH8, expresses the same translational fusion as pHMWP2.CH6, with a His8 tag appended. The plasmids were introduced into *E. coli* DH5α, and then transferred to BL21(DE3) for expression.

Construction of Mutants: D246A, D251A, S254A, and D246H/L247H/A252G in HMWP2 1−1491; S52A, S1439A, and S1977A in HMWP2. The splicing-by-overlap-extension (SOE) method (*14*) was used to construct Cy1 domain mutants in the HMWP2 1−1491 protein fragment. To obtain useful cloning sites, the *Nde*I to *Xma*I fragment of pET22b-irp2/1−1491 was excised by restriction digestion and ligated to the corresponding sites in pUC18, yielding pTAK1. Using pET22b-irp2/1−1491 as template, 5′ and 3′ fragments of mutant inserts were amplified in the first round of PCR using the primer pairs (D246A: 5′-AAGTACCCCCTTCCCATTGACGCCAGTA-3′, 2F; 5′-TCCATAATCAGCAGG**G**CAATATTGACATGGA-3′, 3R; 5′-TCCATGTCAATATTG**C**CCTGCTGATTATGGA-3′, 3F; 5′-TTCCGGCTCGCCCAACGGCGATT-3′, 2R); (D251A: 2F; 5′-GTAAAGCTGGAGGCA**G**CCATAATCAGCAGGT-3′, 4R; 5′-ACCTGCTGATTATGG**C**TGCCTCCAGCTTTAC-3′, 4F; 2R); (S254A: 2F; 5′-AAGAAAAGCGTAAAG**GC**GGAGGCATCCATAAT-3′, 5R; 5′-ATTATGGATGCCTCC**GC**CTTTACGCTTTTCTT-3′, 5F, 2R); (D246H/L247H/A252G: 2F; 5′-GAG**C**CATCCATAATCAG**GT**GGT**G**AATATTGACAT-3′, 6R; 5′-ATT**C**ACC**AC**CTGATTATGGATG**G**CTCCAGCTTTAC-3′, 6F; 2R). The fragments were then used as templates for the second round of PCR, in each case using 2F and 2R as primers. The mutant inserts were then digested with *Sph*I and *Bst*XI and ligated to the corresponding sites in pTAK1, yielding the four plasmids pTAK2−5. Each of these plasmids was digested with *Nde*I and *Age*I and ligated to the corresponding sites in pET22b-irp2/1−1491, yielding pTAK6 (D246A), pTAK7 (D251A), pTAK8 (S264A), and pTAK9 (D246H/L247H/A252G). Each of these plasmids was propagated as described above.

The S52A and S1439A mutants in HMWP2 were also constructed using SOE. In the first round of PCR, the sequences upstream and downstream of the mutation were amplified separately using pIRP2 as a template and the primer pairs (S52A: 5′-GAATTCCATATGATTTCTGGCGCA-3′, 7F; 5′-GTGTAACCATCTCATCAATCTTAT**GGC**ATCCAGG-3′, 7R; 5′-CTGGAT**GCC**ATAAGATTGATGAGATGGTTACA-3′, 8F; 5′-CTGGCGCAGTGCGTCCAGATAGGCCTGGCGGCTTTC-3′, 8R); (S1439A: 5′-CTCGGCTGCTACTGGCCAGACGGCAC-3′, 9F; 5′-AGACGGGTCGCCAGCAGG**GC**ATCGCCG-3′, 9R; 5′-CGGCGAT**GC**CCTGCTGGCGACCCGTCT-3′, 10F; 5′-TGACCGCTCGAGTATCCGCCGCTGACGACGG-3′, 10R). Using these fragments as templates, the second round of PCR was performed with the primer pairs (S52A: 7F and 8R) and (S1439A: 9F and 10R). The mutant insert for S1439A was digested with *Not*I and *Xho*I and subcloned into pHMWP2.CH8 digested with the same enzymes. The mutant insert for S52A was digested with *Nde*I and *Stu*I and subcloned into pET22b-1−1061 (*7*) using DM1 cells (*dam*$^-$/*dcm*$^-$), forming pET22b-1−1061-S52A. This plasmid was digested with *Nde*I and *Msc*I and ligated to the corresponding sites in pHMWP2.CH8.

S1977A was constructed by amplifying the 1231−1959 residue region from pHMWP2.CH8 using the primer pair (S1977A: 5′-CTGGCGCCCATATGTCCCTCTGGAATGCAGACAC-3′, 11F; 5′-TTTCAGCCGATGCCAGCGCGTGCTGTCC-3′, 11R), digesting with *Nde*I and *Bsu*36I, and ligating to the corresponding sites in pET22b-(1383−2035)-S1977A (*8*) to create pET22b-(1231−2035)-S1977A. This plasmid was then digested with *Not*I and *Xho*I and subcloned into pHMWP2.CH8.

Expression and Purification of HMWP2 and Mutants, and 1−1491 and Mutants. Overproduction and purification of HMWP2 1−1491 has been described previously (*7*). Cy1 domain mutants were expressed and purified by the same procedure. For overproduction of HMWP2 and its carrier protein domain mutants, BL21(DE3) strains carrying the pHMWP2.CH8 plasmids were cultivated (6 × 1 L) with shaking (250 rpm) at 30 °C in 2 × YT broth containing 2 mM $MgCl_2$ and 40 μg/mL kanamycin. Upon reaching an OD_{600} of 0.4, the temperature was lowered to 18 °C. Cultures were induced with 50 μM IPTG at $OD_{600} = 0.8$, and shaking proceeded at this temperature for 4 h. Cells were harvested by centrifugation (10 min at 2000*g*) and resuspended in 60 mL of 20 mM Tris, pH 8.0, 500 mM NaCl, 5 mM imidazole (buffer A). Following resuspension, cells were disrupted by two passages through a French pressure cell, and the lysate was clarified by centrifugation (30 min at 95000*g*). HMWP2 and mutants were purified by nickel chelate chromatography using Ni−NTA Superflow resin (Qiagen) (all steps at 4 °C). After overnight incubation with the clarified lysate, 5 mL of resin was packed into a column and washed extensively with buffer A, followed by buffer A + 20 mM imidazole. Elution of the proteins from the column was accomplished in buffer A with a linear imidazole gradient of 20−300 mM over 100 mL. Fractions of the eluant were analyzed by SDS−PAGE, and those containing HMWP2 were pooled and dialyzed against 25 mM Tris, pH 8.0, 2 mM DTT. This material was then applied to a 6 mL Resource Q anion exchange column (Pharmacia Biotech) equilibrated with 20 mM Tris, pH 8.0 (buffer B), washed extensively with buffer B, and then eluted in buffer B with a linear NaCl gradient of 0−500 mM over 110 mL. Fractions were again analyzed by SDS−PAGE, and those containing HMWP2 were pooled and dialyzed against 25 mM Tris, pH 8.0, 2 mM DTT, 10% glycerol. The dialyzed material was concentrated using a Centriprep 50 (Amicon), aliquotted, flash-frozen in liquid nitrogen, and stored at −80 °C.

ATP−PP_i Exchange Assays for A Domain Specificity. Reactions were carried out at 30 °C in 100 μL total volume that contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 3 mM ATP, 36 nM HMWP2, 1 mM [^{32}P]-pyrophosphate, and varying concentrations of amino acid substrate. The reactions were initiated by addition of [^{32}P]-pyrophosphate, allowed to proceed for 2−5 min, and then quenched [900 μL of 1.6% (w/v) activated charcoal, 4.5% (w/v) tetrasodium pyrophosphate, 3.5% perchloric acid in water]. The charcoal was then pelleted by centrifugation, washed

twice with the quench mixture without added charcoal, and then resuspended in 0.5 mL of water and submitted for liquid scintillation counting. Reactions were typically performed in duplicate. The amount of bound radioactivity was converted into reaction velocity using the specific radioactivity of the [^{32}P]-pyrophosphate. The velocity was plotted against substrate concentration and fit to the Michaelis−Menten equation. Saturation kinetics for L-Ser could not be determined (maximum concentration tested was 100 mM), and thus the Michaelis−Menten equation fit was extrapolated to the estimated kinetic constants.

Amino Acid-Dependent ATP Usage by Apo-HMWP2. A coupled, continuous, spectrophotometric assay for inorganic pyrophosphate was employed (*9, 15*). Reactions were carried out at 30 °C in 150 μL total volume in a quartz cuvette that contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 200 μM MESG, 0.15 unit of PNP, 0.15 unit of PP_i-ase, 2 mM ATP, either 5 mM L-Cys, 10 mM β-Cl-L-Ala, 10 mM *S*-2-aminobutyrate, or 20 mM L-Ser, and varying concentrations of HMWP2. Reactions were initiated by addition of the amino acid substrate after a 10 min incubation to allow the PP_i-ase/PNP/MESG couple to remove any contaminating PP_i or P_i. A spectrophotometer monitored absorbance at 360 nm and computed rates of A_{360} increase. Substrate-specific pyrophosphate release was measured by subtracting the rate of A_{360} increase before amino acid addition from the ΔA_{360} rate after addition. This net rate was converted to reaction velocity through the molar extinction coefficient (ϵ) of MESG. This value was obtained from a standard curve constructed from known PP_i concentrations added to the assay mixture described above plotted against ΔA_{360}. This plot was linear between 0 and 24 μM pyrophosphate, and yielded $\epsilon = 17\,600\ (M\cdot cm)^{-1}$. At least three velocities measured at different HMWP2 concentrations were averaged to obtain a final velocity.

Amino Acid-Dependent ATP Usage by Holo-HMWP2. Reactions were performed as above except reactions contained in addition 67 μM CoA and 86 nM Sfp and were incubated for 20 min at 25 °C to allow for phosphopantetheinylation before reaction initiation.

Stoichiometry of Phosphopantetheinylation. To measure the time course and stoichiometry of the apo to holo conversion of HMWP2 and mutants by PPTases Sfp and EntD, reactions (100 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP, 3 mM ATP, 31 μM [^{3}H]-CoA or 41 μM [^{3}H]-acetyl-CoA, 390 nM Sfp or 333 nM EntD, and 2 μM HMWP2. Reactions (in duplicate) were initiated by PPTase addition, and 10 μL aliquots were removed at various time points and quenched with 0.2 mL of 10% trichloroacetic acid (TCA). The precipitated protein was pelleted and washed twice with fresh 10% TCA. The pellet was redissolved in 150 μL of 88% formic acid and submitted for liquid scintillation counting. Percent modification of HMWP2 and mutants was calculated from the specific activity of the CoA and the concentrations of the protein substrates.

Stoichiometry of Covalent Aminoacylation. To measure the time course and stoichiometry of salicylate loading on HMWP2 by YbtE, reactions (200 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP, 100 μM CoA, 266 nM Sfp, 4 μM HMWP2, 1 μM YbtE, 54 μM [^{14}C]-salicylate, and 5 mM ATP. Reactions were incubated for 20 min at 25 °C to allow for phosphopantetheinylation before initiation by ATP addition. Aliquots (20 μL) were removed at various time points and worked up as above.

To measure the time course and stoichiometry of amino acid loading self-catalyzed by HMWP2, reactions (200 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP, 100 μM CoA, 266 nM Sfp, 4 μM HMWP2, 5 mM ATP, and either 312 μM [^{14}C]-L/D-2-aminobutyrate, 505 μM [^{3}H]-L-Ser, or 250 μM [^{35}S]-L-Cys. Reactions were initiated and aliquots were worked up as above.

HPLC and Mass Spectrometric Analysis of Ybt Intermediates Produced by HMWP2, the 1−1491 Fragment, and Mutants. To examine the substrate specificity of the Cy1 domain in the HMWP2 1−1491 fragment and mutants by product analysis, reactions (100 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP, 10 mM DTT, 100 μM CoA, 266 nM Sfp, 4 μM HMWP2 1−1491, 5 mM ATP, 555 nM YbtE, either 1 mM salicylate or 5 mM 2,3-dihydroxybenzoate (DHB), and either 5 mM L-Cys, 10 mM *S*-2-aminobutyrate, 5 mM L-β-chloroalanine, or 20 mM L-Ser. Reactions were incubated at 30 °C for 3 h before quenching with 10 μL of 8.5% H_3PO_4. The mixture was extracted with 1 mL of ethyl acetate, which was subsequently evaporated under vacuum. The residue was redissolved in 30% acetonitrile/water and analyzed by HPLC. To obtain a time course of HPT-Cys and DHPT-Cys, 400 μL reactions as above were performed, with aliquots taken and analyzed at 30 min intervals. Integrated peak areas were converted to product amounts by calibrating with authentic standards. To obtain DHPT-Cys sufficient for NMR spectroscopy, a 2 mL reaction was performed as above, and purified by HPLC. Reactions involving HMWP2 and mutants were performed as previously described (*8*) and analyzed as above. For reactions of S52A and the ArCP fragment in trans, the concentration of S52A was held constant (4 μM) while that of ArCP was varied from 0 to 30 μM to determine an apparent K_m.

RESULTS

Construction of an Overexpression Plasmid for HMWP2, and Heterologous Expression and Purification of Wild-Type and S52A, S1439A, and S1977A Point Mutants. Construction of pHMWP2.CH8 began with pET22b-irp2/1−1491 (*7*), to which was ligated the remaining portion of the *irp2* gene, amplified from pIRP2 (*2*). The entire coding region was then transferred to the pET37b vector, which appends the sequence LEHHHHHHHH to the C-terminus of HMWP2. This octahistidine tag was chosen because of poor binding of the hexahistidine tag of HMWP2 1−1491 to a nickel affinity column (*7*). N-terminal His-tagged constructs were also expressed (data not shown), but greater levels of impurities were seen after affinity chromatography, presumably due to translation of prematurely terminated transcripts and binding of these shortened homologues. *E. coli* BL21-(DE3) cells grown at temperatures above 30 °C after induction yielded only insoluble protein. To obtain a soluble fraction, cells were cooled to 15−20 °C before induction with a low concentration of IPTG. These two factors reduced overall protein production, but the soluble fraction was increased dramatically. The two-step purification procedure (Ni-affinity chromatography followed by anion exchange) yielded fairly pure protein (Figure 2) at 0.7 mg/L of culture.

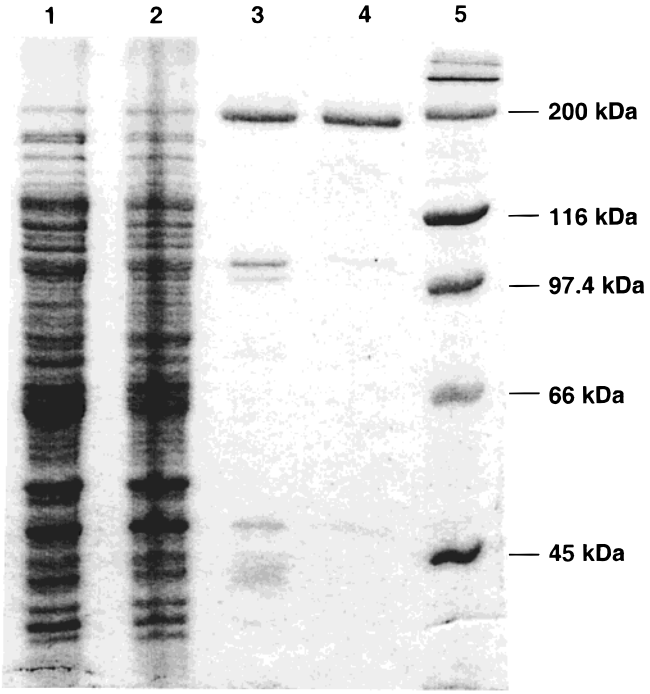

FIGURE 2: GELCODE Blue-stained 10% SDS−polyacrylamide gel depicting results of expression and purification of HMWP2 bearing a C-terminal octahistidine tag. Lane 1, crude cellular lysate; lane 2, soluble fraction of lysate; lane 3, eluant of nickel affinity column; lane 4, eluant of Resource Q anion exchange column; lane 5, molecular weight markers.

Table 1: ATP−PP_i Exchange Catalyzed by HMWP2

substrate	k_{cat} (min^{-1})	K_m (mM)	relative k_{cat}/K_m
L-cysteine	560 ± 24	0.31 ± 0.05	1
L-β-chloroalanine	257 ± 3	2.1 ± 0.1	0.068
S-2-aminobutyrate	358 ± 11	16.5 ± 1.2	0.012
L-serine	∼450	∼300	∼0.0008

After construction of the three carrier protein domain mutants S52A, S1439A, and S1977A by the SOE method, expression and purification were performed identically.

ATP−PP_i Exchange Assay for A Domain Specificity. The first step of A domain function, aminoacyl adenylate formation, was assayed by the amino acid-dependent ATP−PP_i exchange reaction (Table 1). Data for the first three substrates, the natural L-Cys, *S*-2-aminobutyrate, and L-β-chloroalanine, have been measured previously for the HMWP2 1−1491 fragment (*7*, *9*). L-Cys displayed a catalytic efficiency (k_{cat}/K_m) in line with previously examined synthetases, and some 3-fold greater than the corresponding value for the HMWP2 1−1491 fragment. *S*-2-Aminobutyrate and L-β-chloroalanine were also activated slightly more efficiently by the full-length HMWP2 (1.4× and 1.8× greater than by 1−1491), but still at only 1% and 7% of the rate of L-Cys. This supports the conclusion that the A domain in the full-length HMWP2 is folded properly and is equivalent to the 1−1491 fragment in amino acid activation. L-Serine had been previously tested as a substrate for 1−1491 and found to be inactive (*7*), but a subsequent reexamination (*9*) showed a very weak signal in the ATP−PP_i assay. Tested here with HMWP2, we were unable to measure saturation kinetics, and could only estimate k_{cat} and K_m values because of the greatly elevated K_m. Catalytic efficiency is down by 10^3 relative to L-Cys.

Table 2: Steady-State ATP Hydrolysis Rates (min^{-1}) by HMWP2 in the Absence of YbtE and Salicylate, Measured by Continuous Spectrophotometric Pyrophosphate Assay

substrate	apo-HMWP2	holo-HMWP2
L-cysteine	0.35	0.69
S-2-aminobutyrate	0.71	1.14
L-β-chloroalanine	0.70	0.92
L-serine	4.9	7.9

ATP Hydrolysis by HMWP2 Measured by PP_i Release. A coupled, spectrophotometric assay (*9*, *16*) was used to measure consumption of ATP by the A domain under steady-state conditions in both apo- and holo-HMWP2. This assay for PP_i measures the rate of loss of the aminoacyl adenylate, either by hydrolysis or by escape into solution. Apo-HMWP2, which cannot be covalently aminoacylated, exhibited low (<1 min^{-1}) ATP consumption rates for Cys, 2-aminobutyrate, and β-Cl-Ala (Table 2) that were approximately 2-fold higher than those measured previously for HMWP2 1−1491 (*9*), but this still represented loss of less that 0.5% of the flux of amino acid activation, as determined in Table 1. On the other hand, 20 mM L-Ser stimulated ATP hydrolysis to almost 5 min^{-1}, or over 10-fold greater than for Cys. With a measured velocity of 27 min^{-1} in the ATP−PP_i exchange reaction at this Ser concentration, a leak rate of 4.9 min^{-1} represents loss of 19% of the catalytic flux. Measuring PP_i release with holo-HMWP2 revealed slightly elevated rates, in each case somewhat less than doubled (Table 2). Because transfer of the aminoacyl adenylate to holo-PCP1 and -PCP2 is rapid and stoichiometric (see below), this steady-state PP_i release rate does not reflect rates of transfer, but rather some combination of the apo-HMWP2 leak rate and hydrolysis of the PCP-tethered thioester. While Cys, 2-aminobutyrate, and β-Cl-Ala still experience loss of less than 0.5% of the flux when full-length HMWP2 is phosphopantetheinylated, Ser now stimulates loss of 29%.

Covalent [^{3}H]-Phosphopantetheinylation of HMWP2 and Carrier Protein Domain Mutants. The three carrier protein domains in HMWP2—N-terminal ArCP, internal PCP1, and C-terminal PCP2—were assayed for covalent attachment of P-pant by two PPTases, Sfp from surfactin synthetase of *B. subtilis* (*12*) and EntD from enterobactin synthetase of *E. coli* (*13*). The S52A, S1439A, and S1977A single mutants change the conserved Ser, whose side chain hydroxyl is the point of P-pant attachment, to Ala, so that the apo to holo conversion cannot take place. Data are shown in Figure 3A. Under the conditions of the assay, Sfp rapidly (<1 min) modified WT and mutant HMWP2 with [^{3}H]-CoA as substrate. The stoichiometry of P-pant incorporation for the three mutants was the expected 2 equiv, while WT HMWP2 incorporated nearly 5 equiv. We obtained nearly identical results upon switching to EntD as PPTase and/or [^{3}H]-acetyl-CoA as substrate (data not shown). We cannot account for the higher than anticipated stoichiometry of WT P-pant incorporation, except through protein concentration or CoA specific activity errors.

Auto-aminoacylation of HMWP2 and Carrier Protein Domain Mutants with Radiolabeled Amino Acids. The ability of holo-HMWP2 to aminoacylate itself on PCP1 and PCP2 was assayed with [^{35}S]-L-Cys, [^{14}C]-L/D-2-aminobutyrate, and [^{3}H]-L-Ser as substrates (radiolabeled β-Cl-Ala is not com-

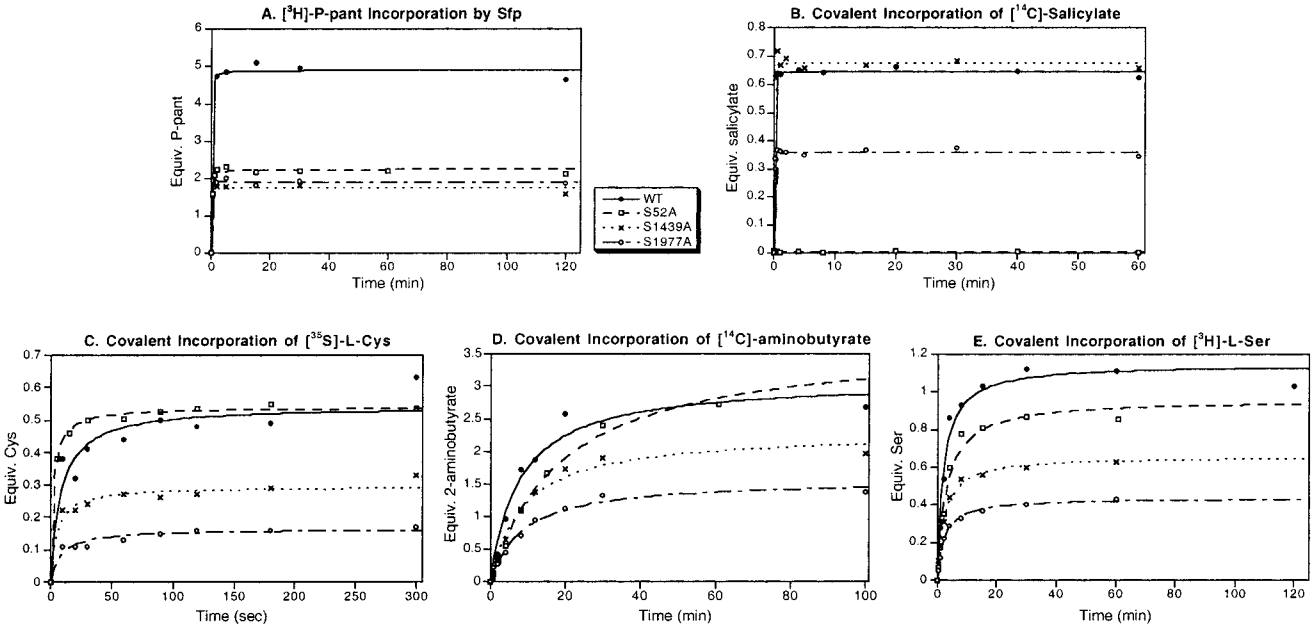

FIGURE 3: Mutational analysis of carrier protein domains and stoichiometry of covalent modification of HMWP2 and mutants. The S52A, S1439A, and S1977A mutations inactivate ArCP, PCP1, and PCP2, respectively. Radiolabeled substrate incorporation was measured by acid precipitation of the proteins and liquid scintillation counting. See Experimental Procedures for details. (A) Time course of [^{3}H]-CoA-derived phosphopantetheine incorporation into HMWP2 and carrier domain mutants by the *B. subtilis* PPTase Sfp. Wild type is expected to incorporate 3 equiv of P-pant, and each mutant, 2 equiv. (B) Time course of [^{14}C]-salicylate incorporation catalyzed by the salicyl-AMP ligase YbtE. WT and the two PCP mutants are expected to incorporate 1 equiv, and ArCP mutant, none. (C) Time course of [^{35}S]-L-Cys incorporation self-catalyzed by the A domain. WT and ArCP mutant are expected to incorporate 2 equiv, and the PCP mutants, 1 equiv each. Note time axis is in seconds. (D) Time course of [^{14}C]-2-aminobutyrate incorporation self-catalyzed by the A domain. WT and ArCP mutant are expected to incorporate 2 equiv, and the PCP mutants, 1 equiv each. (E) Time course of [^{3}H]-L-Ser incorporation self-catalyzed by the A domain. WT and ArCP mutant are expected to incorporate 2 equiv, and the PCP mutants, 1 equiv each.

mercially available). Results of covalent aminoacylation over time are shown in Figure 3C−E. For all three substrates, the concentration of amino acid was well below the measured K_m of Table 1, limited by the concentration and specific radioactivity supplied by the manufacturer. Thus, each time course of aminoacylation cannot be directly compared to the others in terms of rate. Despite this limitation, cysteinylation (Figure 3C) was very rapid, complete in under 10 s. While absolute stoichiometries did not achieve the full 2 equiv expected for WT and S52A (which possess unmodified PCP1 and PCP2), and the 1 equiv for S1439A (unmodified PCP2) and S1977A (unmodified PCP1), relative results were in line with expectations, given that PCP1 has been observed previously to be a poorer substrate for aminoacylation relative to PCP2 (*8*), a conclusion that is supported by data from the other two substrates (below).

Covalent incorporation of [^{14}C]-L/D-2-aminobutyrate and [^{3}H]-L-Ser is depicted in Figure 3D,E. Again, the absolute final stoichiometries do not match each other or Cys, but do support the notion of two aminoacylation events for WT and S52A, and one for S1439A and S1977A, with PCP1 again inferior to PCP2 as a substrate. The times required to achieve final levels of modification were longer, corresponding to poorer K_m's for these alternate substrates. Of primary importance, however, is the ability of the A domain not only to activate alternate amino acids but also to transfer them stably to holo-PCPs as well, an observation that supports earlier work on covalent aminoacylation in trans (*9*). The mutants also demonstrate that aminoacylation does not follow an obligate stepwise path: either PCP1 or PCP2 can be loaded prior to the other.

Salicylation of HMWP2 and Carrier Protein Domain Mutants by YbtE. YbtE is the 525 residue adenylation protein specific for salicylate, although it can also activate 2,3-dihydroxybenzoate (DHB), albeit at lower efficiency (*7*). It regioselectively transfers salicyl adenylate to the holo-ArCP of HMWP2, but not the PCPs, as shown in Figure 3B. The S52A mutant, lacking a holo-ArCP, does not incorporate any [^{14}C]-Sal, while WT, S1977A, and S1439A are all competent for rapid transfer.

Product Formation by HMWP2. HMWP2 was assayed by HPLC analysis of an ethyl acetate extract of an enzymatic reaction for product formation, a detection method which has been employed previously in the context of the 1−1382 and 1383−2035 fragments (*8*). A labeled trace of identified products is shown in Figure 4A. Products (Figure 4B) were identified by electrospray mass spectrometry and by HPLC comparison to synthetic standards, and were identified as HPT-Cys **6** (ESMS $[M + H]^+ = 327$), resulting from presumed thiolysis (and subsequent *S*-to-*N*-acyl shift) of the HPT-thioester on PCP1; HPTT-Cys **4** (ESMS $[M + H]^+ = 410$), resulting from corresponding thiolysis of HPTT-thioester from PCP2; and HPT-Cys-Cys (ESMS $[M + H]^+ = 430$), presumably arising from an unheterocyclized intermediate on PCP2. Small amounts of Sal-Cys **2**, from thiolysis of ArCP-bound salicylate, were also observed. Product release rates for the various species are shown in Table 3. The very slow rates can be ascribed to the absence of the second half of the Ybt synthetic machinery, HMWP1 (with its TE release domain), and the resulting slow hydrolysis/thiolysis of the products on the stalled assembly line. However, the detection of even stoichiometric amounts of

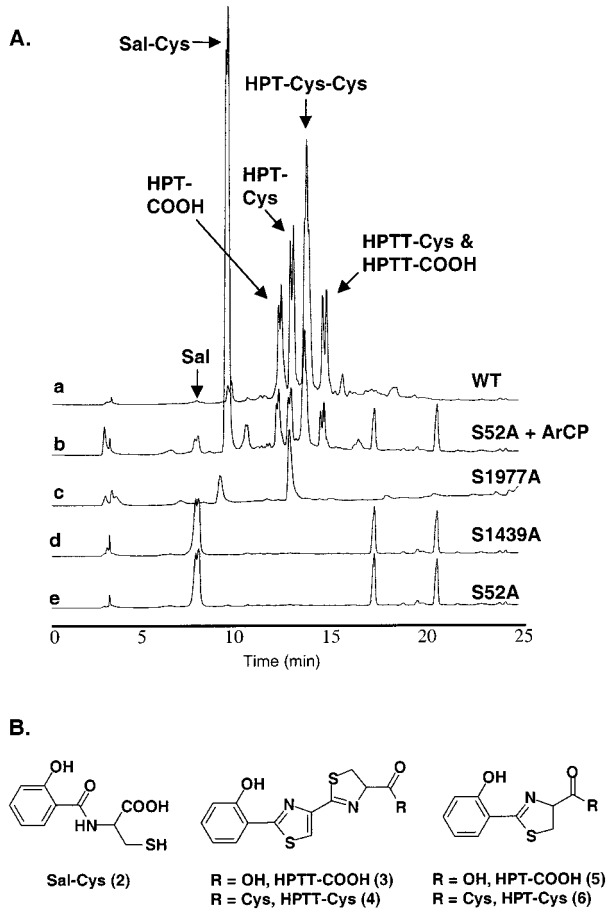

FIGURE 4: (A) Labeled HPLC traces of reactions products from (a) wild-type HMWP2 (4 μM); (b) S52A (4 μM) + ArCP (5 μM); (c) S1977A (2 μM); (d) S1439A (4 μM); and (e) S52A (4 μM). Reactions were incubated for 12 h at 30 °C prior to analysis. See Experimental Procedures for details. Split peaks observed for salicylate, Sal-Cys, and HPT-R are artifactual. (B) Structure of the enzymatic reaction products observed in this study. Sal-Cys = *N*-(2-hydroxyphenyl)cysteine; HPTT = 2-(2-hydroxyphenyl)thiazolyl-2,4-thiazolinyl-; HPT- = 2-(2-hydroxyphenyl)thiazolinyl-. Compounds **3** and **4** have spontaneously oxidized to the aromatic thiazole from the thiazoline (see text).

Table 3: Product Release Rates (min^{-1}) by HMWP2 and Carrier Domain Mutants, Measured by HPLC Analysis

HMWP2 species	HPT-R[a]	HPT-Cys-Cys	HPTT-R[b]
WT	0.038 ± 0.002	0.066 ± 0.004	0.080 ± 0.003
S52A	0	0	0
S52A + ArCP[c]	0.032 ± 0.003	0.006 ± 0.001	0.032 ± 0.003
S1439A	0	0	0
S1977A	0.0077 ± 0.0009	0	0

[a] Combined production of HPT-COOH and HPT-Cys. [b] Combined production of HPTT-COOH and HPTT-Cys [c] Rates (k_{cat}) calculated from saturation kinetics of added, in trans ArCP. Apparent K_m for ArCP is 4 ± 1 μM.

products, particularly the bisheterocyclic HPTT species, does indicate that all six domains of full-length HMWP2 are functional.

Product Formation by Carrier Protein Domain Mutants of HMWP2. The carrier protein domain mutants were used to probe the stepwise operation of the HMWP2 assembly line, as shown in Figures 4 and 5 and Table 3. S1977A, which lacks a functional PCP2, necessarily stalls with HPT-*S*-P-pant located on PCP1, which is hydrolyzed to HPT-COOH or thiolyzed by Cys or DTT from solution to HPT-Cys. S1439A, without PCP1, cannot transfer salicylate from ArCP, and thus only Sal-Cys from thiolysis of the ArCP-bound Sal is observed. Finally, S52A, without a functional ArCP, cannot be loaded by YbtE, and thus the chain cannot initiate. No products are seen.

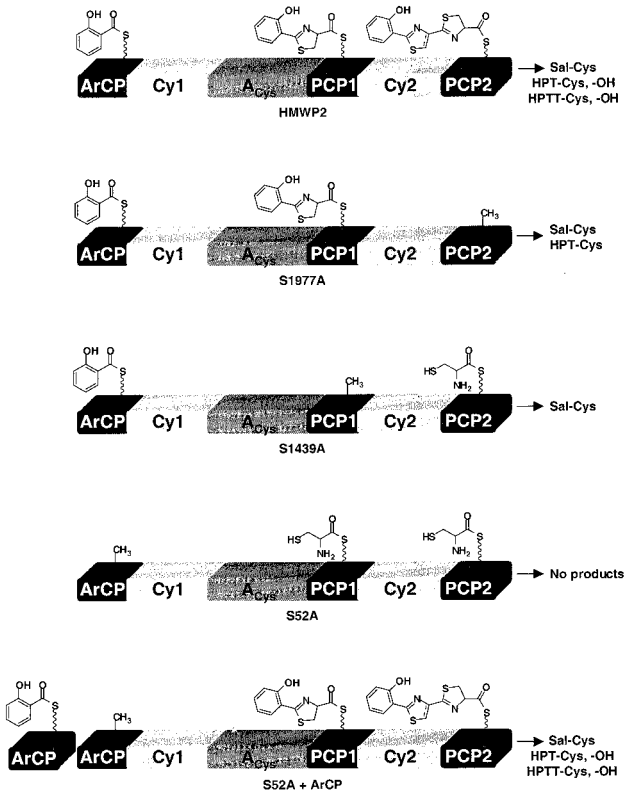

FIGURE 5: Schematic of product formation catalyzed by HMWP2 and carrier protein domain mutants. Mutants are indicated by a methyl group resulting from a Ser-to-Ala mutation that prevents attachment of a phosphopantetheine tether. These apo-domains cannot accept a monomer and thus do not permit formation of the second thiazoline (S1977A), salicylate translocation and first thiazoline formation (S1439A), or chain initiation (S52A). However, a salicylate-loaded ArCP fragment presented in trans to the S52A mutant bypasses the inactive ArCP in cis and restores the full complement of products (bottom). Covalent intermediates are drawn attached to the appropriate carrier protein domain. It is unknown whether HPTT-*S*-PCP2 oxidation to the aromatic thiazole occurs before or after thiolysis and release, and thus a dotted bond is drawn in the middle ring. Products detected by HPLC and mass spectrometry are listed to the right.

Complementation of the S52A Mutant with Holo-ArCP Added in Trans. When the 1−100 N-terminal fragment of HMWP2 (ArCP domain) was employed in trans with S52A, production of **3**−**6** and HPT-Cys-Cys was restored (Figure 4A). Analysis of the rates of product formation dependence on added ArCP concentration revealed the k_{cat} values in Table 3 and an apparent K_m for ArCP of 4 μM. This indicates that Cy1 can productively interact with ArCP as a separate protein, even in the presence of an inactivated ArCP in cis, and points to protein−protein interactions, and not covalent linkages, as the determining elements for chain transfer and assembly line function.

Mutational Analysis of the Cy1 Domain of HMWP2 1−1491. To begin to identify residues critical for the bifunctional catalytic activity of cyclization domains, condensation and heterocyclization, a series of mutants in the proposed catalytic core of Cy1 domain were constructed. This

Table 4: Mutational Analysis of Cy1 Domain of HMWP2 1−1491

HMWP2 1−1491	HPT-Cys release (min^{-1})	core sequence (245−255)[a]
WT	0.0062 ± 0.0013	I **D** L L I M **D** A S **S** F
D246A	0	I **A** L L I M **D** A S **S** F
D251A	0	I **D** L L I M **A** A S **S** F
S254A	0.0044 ± 0.0006	I **D** L L I M **D** A S **A** F
D246H/L247H/A252G	0	I **H** **H** L I M **D** **G** S **S** F[b]

[a] Boldface letters indicate highly conserved positions. Underlined letters indicate mutations. [b] Conserved residues are for condensation domains.

highly conserved region, D-x-x-x-x-D-x-x-S, has been previously identified and found to correspond in location to the highly conserved condensation domain catalytic core, H-H-x-x-x-D-G-x-S (*17*). A contemporary search of published sequences revealed 15 known and putative Cy domains, within which the first Asp was conserved in 13 (the others being Asn and Arg), the second Asp was strictly conserved, and the Ser was present in 13 (the others being Thr and Ala). The Cy1 domain mutants were constructed in the 1−1491 fragment of HMWP2, both to examine the Cy1 domain alone and to simplify the assay to a search for HPT-Cys production. Three single mutants were constructed, D246A, D251A, and S254A: each of these separately inactivates a potential catalytic residue. In addition, a triple mutant, D246H/L247H/A252G, was constructed. This triple mutant changes the core conserved sequence from the signature of a Cy domain to that of a condensation domain, to determine whether the condensation activity of the Cy domain could be selectively maintained, while deleting the heterocyclization activity. Carrier protein domain and A domain functions were confirmed in these mutants by successful radiolabeled phosphopantetheinylation, cysteinylation, and salicylation (data not shown). Upon HPLC analysis of an ethyl acetate extract of a reaction containing these mutants, it was found that the D246A and D251A mutations eliminated any HPT-Cys production, while the S254A mutation had almost no effect on activity (Table 4). The triple mutant was also inactive, as no HPT-Cys nor any alternate products were observed.

Analysis of Product Release by HMWP2 1−1491 and Alternate Substrates. Having established the stable thioesterification of three of the four amino acid substrates activated by the A domain, the specificity of the Cy1 domain toward these PCP-tethered substrates was probed with the 1−1491 fragment. As above, this fragment possesses only Cy1 for mechanistic simplicity and single product formation. Since 2-aminobutyrate and β-Cl-Ala have no side chain thiol to form the heterocyclic thiazoline, the expected products were Sal-(2-aminobutyrate) and Sal-(β-Cl-Ala), if the amide bond could be formed in the absence of the side chain thiol. Serine, on the other hand, possesses a side chain hydroxyl which can heterocyclize in other NRPS Cy domain contexts [e.g., mycobactin (*18*) and vibriobactin (*19*)] to the oxazoline. However, HPLC analysis of reactions containing each of these substrates failed to show any new products (Figure 7), even in the presence of high concentrations of DTT, which has been shown to accelerate cleavage of any acyl-*S*-PCP enzyme intermediates in this NRPS fragment (*8*).

A similar experiment was performed using DHB in place of salicylate as the chain-initiating acyl donor (Figure 7).

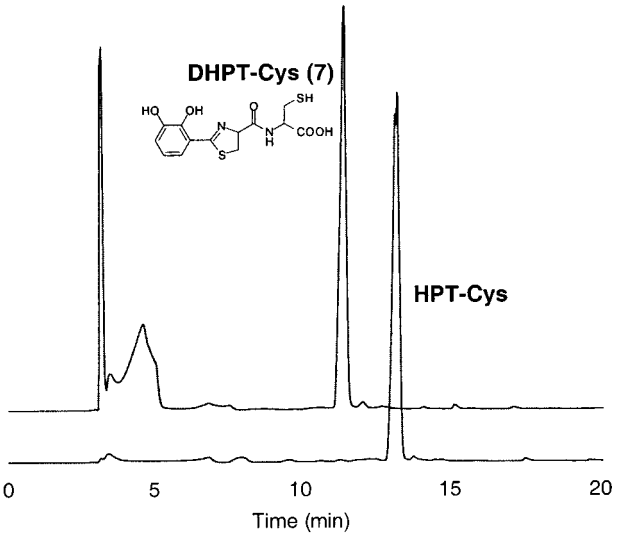

FIGURE 6: HPLC traces of the organic extractable fraction of holo-HMWP2 1−1491 reactions containing YbtE, L-Cys, and 2,3-dihydroxybenzoate (top) or salicyate (bottom) as the chain-initiating monomer. DHPT-Cys = *N*-[2-(2,3-dihydroxyphenyl)thiazoline-4-carbonyl]cysteine. DHPT-Cys and HPT-Cys product peaks are indicated, and were identified by ^{1}H NMR and mass spectrometry (DHPT-Cys) or synthetic standard co-injection and mass spectrometry (HPT-Cys). Early peaks (3−5 min) in the top trace represent the solvent front and unreacted DHB.

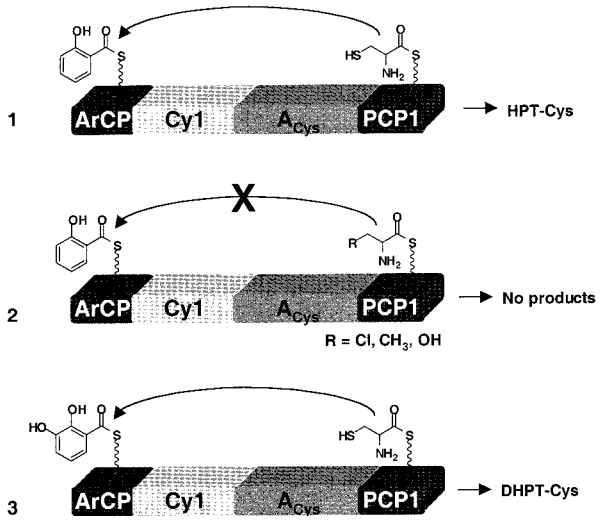

FIGURE 7: Schematic of holo-HMWP2 1−1491 reactions with covalently loaded alternate substrates. The substrates are used to probe the specificity of the Cy1 domain. 1. The natural substrates salicylate and cysteine produce HPT-Cys, as both amide bond formation and heterocyclization are catalyzed by Cy1. 2. All alternate downstream, nucleophilic substrates (β-Cl-Ala, 2-aminobutyrate, serine) fail to attack the upstream electrophilic salicylate. These substrates share the α-amine of Cys, but lack the β-thiol necessary for heterocyclization. As detailed in the text, their failure to produce condensed products could be due to a "gatekeeping" function of Cy1, a requirement for the Cys thiol to attack prior to amide bond formation, failure of Cy1 to release the condensed product, or an orienting role for the thiol in the Cy1 active site for nucleophilic amine positioning. 3. Salicylate analogue 2,3-DHB is tolerated by the Cy1 domain with downstream partner Cys to yield DHPT-Cys. Evidently Cy1 has greater tolerance for the upstream, electrophilic donor than for the downstream, nucleophilic acceptor.

YbtE has been previously shown to activate DHB to DHB-AMP by ATP−PP_i exchange (*7*), although transfer to ArCP has not been established. It has been previously shown in the actinomycin synthetase system that *p*-toluic acid can

substitute for the natural 4-methyl-3-hydroxyanthranilic acid, in activation by ACMS I, transfer to the standalone ArCP AcmACP, and in condensation with threonine by ACMS II (*20*, *21*). Employing DHB and L-Cys as substrates for YbtE and holo-HMWP2 1−1491, a single new product was formed (Figure 6), eluting some 2 min earlier than HPT-Cys. This product was collected, and electrospray mass spectrometry (ESMS $[M^+] = 342.2$) yielded a mass consistent with *N*-[2-(2,3-dihydroxyphenyl)thiazoline-4-carbonyl]cysteine (DHPT-Cys, **7**). A large-scale enzymatic reaction was performed to obtain ∼0.4 mg of this compound, and the structure assignment was confirmed by NMR spectroscopy. [^{1}H NMR: δ 6.99 (dd, 1H, $J = 8.0, 1.0$), 6.94 (dd, 1H, $J = 8.2, 1.5$), 6.76 (t, 1H, $J = 7.7$), 5.38 (t, 1H, $J = 9.0$), 4.45 (dt, 1H, $J = 11.0, 5.0$), 3.70 (m, 1H), 3.67 (m, 1H), 3.03 (dd, 1H, $J = 4.5, 1.0$), 2.98 (d, 1H, $J = 4.5$).] A time course of DHPT-Cys production by YbtE and 1−1491 revealed a linear rate of 0.5 min^{-1} through a 3.5 h reaction time, compared with a linear rate of 0.4 min^{-1} observed under the same conditions for HPT-Cys release.

DISCUSSION

The nonribosomal peptide synthetases and the polyketide synthases present tall hurdles to a complete elucidation of their structure and function. There is first their size, typically in the hundreds of kilodaltons and thousands of residues. This not only makes structural studies a formidable task, but also hinders heterologous expression and purification of the synthetase gene products. Adding to the challenge of these systems, each complete synthetase complex usually consists of several of these large proteins; higher order organization remains unclear, and each synthetase subunit may not exhibit its full complement of functions in the absence of its partners. Further complicating matters is the multidomain nature of NRPSs and PKSs: each synthetase protein has many individual enzymatic functions, each of which ideally should be assayed autonomously to build a complete kinetic and functional picture.

Several of these challenges are represented in yersiniabactin synthetase, which is comprised of 3 central proteins, 2 of which (HMWP2 and HMWP1) are 229 and 349 kDa and contain 15 of the 16 assigned functional domains. The smaller of the two, HMWP2, has six domains, and is the first half of the assembly line that biosynthesizes Ybt. We have undertaken here the heterologous expression and purification of the full-length protein, and have examined the functions of each of the six domains individually and in combination with others, to characterize the kinetics, specificity, and various acyl-enzyme intermediates that can be released as products from HMWP2.

The component domains of HMWP2 have been studied in fragment form previously, including the 1−1491 four domain fragment (*7*), the individual carrier protein domains, and the 1383−2035 three domain fragment (*8*). Heterologous expression of large synthetase proteins has proven difficult, ameliorated in individual cases by fusions to maltose binding protein (MBP) (*18*), coexpression of chaperones (*22*), and induction and growth at low temperatures (*23*). Expression of the full-length HMWP2 was accomplished in *E. coli* by growth after low-level induction at lowered temperatures, below 25 °C. Expression at temperatures over 30 °C resulted in good production, but almost entirely insoluble protein. Lowered temperatures gave less total protein, but an increased soluble fraction after clarification of the lysate (Figure 2), presumably because the slower rate of protein synthesis and lower intracellular protein concentration favored correct protein folding (a process largely unaffected at 15−20 °C) over aggregation of misfolded or partially folded protein. A two-step purification of Ni-affinity chromatography followed by anion exchange (the C-terminal, octahistidine-tagged HMWP2 does not bind well to the Ni-chelate column) yielded nearly pure protein in a final yield of 0.7 mg/L of culture (Figure 2).

The first step in NRP assembly is the conversion of the inactive, apo form of the synthetase to the active, holo form by posttranslational introduction of a CoA-derived P-pant arm onto the side chain hydroxyl of a conserved serine in each carrier protein domain, a function performed by an PPTase. We examined each of the three carrier protein domains in HMWP2 individually by constructing Ser-to-Ala mutants of each, which prevents P-pant attachment. In an assay of [^{3}H]-P-pant incorporation, we found each mutant was modified with 2 equiv of P-pant, while WT was modified by slightly under 5 equiv (Figure 3A). These results were unchanged over two different PPTases (Sfp and EntD) and two different CoA species (CoA and acetyl-CoA). While we do not know why WT HMWP2 is apparently overphosphopantetheinylated, we conclude that each of the carrier protein domains is a substrate for two different broadly specific PPTases and can be converted from apo to holo form in high stoichiometry in vitro. This high conversion efficiency is necessary for NRPS and PKS activity, since, under the thiotemplated model of NRP synthesis (*24*), a single missing P-pant tether will halt the entire assembly line, as shown in Figure 5.

The second step of NRP assembly is monomer selection, activation, and transfer by the A domains. Besides YbtE (*7*), Ybt synthetase has only one other A domain, imbedded in HMWP2, which is L-Cys-specific and loads PCP1 and PCP2 in cis, and PCP3 of HMWP1 in trans. We have examined the A domain-mediated activation and loading in the 1−1491 context previously (*9*), and here have uncovered a fourth substrate recognized by this domain, L-serine (Table 1). An examination of the relative catalytic efficiencies accorded these substrates (each of which presents a different moiety in place of cysteine's −SH: 2-aminobutyrate, $-CH_3$; β-Cl-Ala, −Cl; serine, −OH) indicates a high preference for Cys, particularly over its only proteinogenic competitor, Ser (selectivity $>10^3$). This still represents a higher misincorporation ratio than for the cysteinyl-tRNA synthetase, responsible for analogous functions in ribosomal peptide synthesis, which activates L-Ser at only 3×10^{-6} of the k_{cat}/K_m for Cys (*25*).

The question of editing/proofreading of incorrect substrates by the HMWP2 A domain was approached in two ways in order to assess the two half-reactions catalyzed by these domains. First, amino acid-dependent PP_i release rates were measured, providing an indication of ATP consumption by the A domain in a stalled mode (Table 2). A comparison of these substrate-specific "leak rates" for both apo- and holo-HMWP2 reveals tightly sequestered aminoacyl adenylates (slow escape or loss to hydrolysis) in the case of Cys, 2-aminobutyrate, and β-Cl-Ala, but a 10-fold higher rate for

Ser. A difference of this sort has been observed before for EntE, the DHB activating protein of enterobactin synthetase, with regard to its alternate substrate salicylate (*26*). At the concentration of Ser examined, loss of seryl-AMP represents almost 20% of the catalytic flux, rising to almost 30% when holo-protein is employed. The second half-reaction catalyzed by the A domain, aminoacyl-AMP transfer to the P-pant thiol, was then analyzed. The data in Figure 3C−E confirm the transfer of aminoacyl adenylates to an acid-stable thioester on PCP1 and PCP2 and further indicate that each PCP can be loaded prior to, and independently of, the other. While measured equivalents of substrate incorporation give variable end points, possibly because of specific radioactivity errors, the saturation kinetics exhibited by these three substrates support the notion that any activated amino acid is transferred to the appropriate carrier protein domain, and that while seryl-AMP is less tightly sequestered than the other substrates, its loss does not indicate active editing/proofreading since serine is loaded onto the PCPs in correct relative stoichiometry. This approach of determining and taking advantage of natural variability in A domain specificity can complement other strategies for testing unnatural substrates, such as mutational altering of A domain specificity (*27*) and incorporation of pre-acylated CoA analogues into apo-synthetases by PPTases (*28*).

After the validation of the functions of the carrier protein domains and the A domain, we turned to the two cyclization domains, Cy1 and Cy2, which were assayed by hydrolysis/thiolysis of acyl-*S*-enzyme intermediates from the various carrier domains and analysis of the released products. As examined previously in the 1−1491 fragment (*7*) and the 1−1382 and 1383−2035 fragments (*8*), the Cy domains are responsible for the dual function of monomer condensation and heterocyclization to form thiazolines. Through HPLC analysis of products extracted from an enzymatic reaction (Figure 4A), and comparison to synthetic standards and mass spectrometry, the compounds HPT-Cys and HPTT-Cys were identified. The function of Cy1 is verified by the former compound, and the processivity of synthesis confirmed by the S1977A mutant, which necessarily stalls at PCP1 and releases only the singly heterocyclized product. The identification of HPTT demonstrates the competence of Cy2. As noted before (*8, 29*), HPTT results from spontaneous oxidation of the first thiazoline to the heteroaromatic thiazole, driven in part by the gain of conjugation to the second thiazoline ring. (Both enzymatic and synthetic HPTT undergo this oxidation. Presumably, enzymatic reduction of the second ring to the thiazolidine present in Ybt—a function not found in HMWP2—reduces susceptibility toward oxidation.)

Along with the S1977A mutant, the S52A and S1439A proteins (ArCP and PCP1 mutants, respectively) behave as anticipated and validate the prediction that growing chains cannot jump an apo-carrier protein domain (Figures 4A and 5). The former mutant cannot be loaded with salicylate from YbtE, and thus cannot initiate synthesis (although this mutation can be complemented with the ArCP fragment in trans), while the latter is blocked from Cys presentation to Cy1 in the first condensation step, and yields only Sal-Cys from thiolysis of the ArCP-bound salicylate.

To begin to probe the function of the Cy domain, for which a detailed mechanism has yet to be proposed, a series of mutants were constructed in the conserved core **D**-L-L-I-M-**D**-A-S-**S** sequence of the Cy1 domain of the 1−1491 fragment (boldface letters indicate the conserved residues within Cy domains). By product assay, it was found that Asp-to-Ala mutations at residues 246 or 251 abolished HPT-Cys production, while the Ser-to-Ala mutation at residue 254 only slightly attenuated the activity (Table 4). In an attempt to selectively delete the heterocyclization function of Cy1 while retaining the amide bond-forming condensation function, the **D**-L-L-I-M-**D**-A-S-**S** core was mutated to **H**-**H**-L-I-M-**D**-**G**-S-**S**, the core sequence of a condensation domain (*30*), which lies in the same relative position within the domain (*17*). The second His (*31*) and the Asp (*30*) of this sequence have been shown to be critical for activity. However, the Cy1 triple mutant was inactive for product release, so a separation/reconstitution of C domain activity has not yet been realized. The requirement for the two Asp residues is likely to be general and making Asp-to-Ala mutations a route to knockout of Cy domain function. While this does not yet lead to a proposed mechanism, in the absence of structural data mutational analysis is a useful although slow method to identify critical residues.

Another probe for Cy domain function is the use of alternate substrates, both at the upstream (electrophilic) and at the downstream (nucleophilic) carrier protein domains. Attempts to isolate products from reactions that included HMWP2 1−1491, YbtE, and β-Cl-Ala, 2-aminobutyrate, or Ser in place of Cys all failed to identify the expected Sal-(β-Cl-Ala), Sal-(2-aminobutyrate), or hydroxyphenyl-oxazoline (Figure 7). However, when 2,3-DHB was substituted for Sal as a YbtE substrate, robust production of DHPT-Cys was observed (Figure 6). Taken with the aminoacylation data in Figure 3 that show aminoacyl-*S*-PCP1 formation, several conclusions are possible. (1) The Cy1 domain has some tolerance for variation in the upstream acyl monomer, but none for its downstream counterpart. This implies a Cy active site that can only accommodate Cys as the acceptor, but is less specific for the donor. This model postulates the Cy domains as higher-level "gatekeepers" against incorrect products beyond that offered by the A domains, and resembles the reactivity of C domains, which also display higher selectivity for the downstream acceptor (*20, 28*). (2) The mechanism of Cy domains requires cysteinyl thiol attack first, in a transthioesterification step, followed by α-nitrogen attack from the heterocycle (*7*). Lacking this nucleophilic β-SH group, 2-aminobutyrate and β-Cl-Ala cannot condense under this mechanism. Difficulty arises in explaining the failure of Ser, which possesses a nucleophilic β-OH, especially since a Ser-specific Cy domain is known from mycobactin synthetase (*18*). (3) Amide bond formation occurs, but the Sal-(aminoacyl-*S*-PCP) is not released from HMWP2 1−1491. This is unlikely since thioester cleavage from PCP1 is accelerated by DTT (*8*), which was present in the reaction. (4) The thiolate side chain of Cys in Cys-*S*-PCP1 (and perhaps also in PCP2 and PCP3) makes a key orienting interaction with the binding site in the Cy domain needed for deprotonation of the α-amino group to yield the $:NH_2$ moiety required for amide bond formation. This postulated orienting interaction could also be mediated by the PCP domain, which, if L-Cys specific, could be incompetent for correct presentation of other amino acids. Replace-

ment of PCP1 with an L-Ser-specific PCP could be used to probe this issue.

In summary, the heterologously expressed, 2035 residue *Y. pestis* HMWP2 is fully competent for biosynthesis of the first half of Ybt. Mutational analyses of the three carrier protein domains have shown their roles as sites of monomer attachment. The A domain is specific for amino acid activation, but not subsequent aminoacyl transfer to the PCPs. Finally, the Cy1 domain performs its condensation and cyclization functions on the natural monomers and a close cognate of salicylate, but refuses in cis alternate downstream aminoacyl-*S*-PCPs. The expression and characterization of HMWP1 should allow further progress toward reconstitution of the full catalytic activity of the 16 domain Ybt synthetase.

ACKNOWLEDGMENT

We thank Luis E. N. Quadri (Weill Medical College of Cornell University) for assistance with the Cy domain mutants and for many helpful discussions.

REFERENCES

1. Drechsel, H., Stephan, H., Lotz, R., Haag, H., Zähner, H., Hantke, K., and Jung, G. (1995) *Liebigs Ann.*, 1727−1733.
2. Bearden, S. W., Fetherston, J. D., and Perry, R. D. (1997) *Infect. Immun. 65*, 1659−1668.
3. Nakai, H., Kobayashi, S., Ozaki, M., Hayase, Y., and Takeda, R. (1999) *Acta Crystallogr. C55*, 54−56.
4. Perry, R. D., Balbo, P. B., Jones, H. A., Fetherston, J. D., and DeMoll, E. (1999) *Microbiology 145*, 1181−1190.
5. Keating, T. A., and Walsh, C. T. (1999) *Curr. Opin. Chem. Biol. 3*, 598−606.
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. Gehring, A. M., Mori, I., Perry, R. D., and Walsh, C. T. (1998) *Biochemistry 37*, 11637−11650.
8. Suo, Z., Walsh, C. T., and Miller, D. A. (1999) *Biochemistry 38*, 14023−14035.
9. Keating, T. A., Suo, Z., Ehmann, D. E., and Walsh, C. T. (2000) *Biochemistry 39*, 2297−2306.
10. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
11. Webb, M. R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4884−4887.
12. Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) *Biochemistry 37*, 1585−1595.
13. 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.
14. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene 77*, 51−59.
15. Upson, R. H., Haugland, R. P., Malekzadeh, M. N., and Haugland, R. P. (1996) *Anal. Biochem. 243*, 41−45.
16. Pavela-Vrancic, M., Dieckmann, R., von Döhren, H., and Kleinkauf, H. (1999) *Biochem. J. 342*, 715−719.
17. Konz, D., Klens, A., Schörgendorfer, K., and Marahiel, M. A. (1997) *Chem. Biol. 4*, 927−937.
18. Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H., and Walsh, C. T. (1998) *Chem. Biol. 5*, 631−645.
19. Griffiths, G. L., Sigel, S. P., Payne, S. M., and Neilands, J. B. (1984) *J. Biol. Chem. 259*, 383−385.
20. Pfennig, F., Schauwecker, F., and Keller, U. (1999) *J. Biol. Chem. 274*, 12508−12516.
21. Stindl, A., and Keller, U. (1993) *J. Biol. Chem. 268*, 10612−10620.
22. Symmank, H., Saenger, W., and Bernhard, F. (1999) *J. Biol. Chem. 274*, 21581−21588.
23. Makrides, S. C. (1996) *Microbiol. Rev. 60*, 512−538.
24. Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H. R. (1996) *J. Biol. Chem. 271*, 15428−15435.
25. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 1245−1249.
26. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C., and Walsh, C. T. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2509−2514.
27. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) *Chem. Biol. 6*, 493−505.
28. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) *Science 284*, 486-489.
29. Quadri, L. E. N., Keating, T. A., Patel, H. M., and Walsh, C. T. (1999) *Biochemistry 38*, 14941−14954.
30. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) *Chem. Rev. 97*, 2651−2673.
31. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) *J. Biol. Chem. 273*, 22773−22781.

BI992923G