Carrier Protein Recognition in Siderophore-Producing Nonribosomal Peptide Synthetases†

C. Gary Marshall, Michael D. Burkart, Robin K. Meray, and Christopher T. Walsh*

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

Received April 1, 2002; Revised Manuscript Received May 14, 2002

ABSTRACT: Nonribosomal peptide synthetases (NRPSs) use phosphopantetheine (pPant) bearing carrier proteins to chaperone activated aminoacyl and peptidyl intermediates to the various enzymes that effect peptide synthesis. Using components from siderophore NRPSs that synthesize vibriobactin, enterobactin, yersiniabactin, pyochelin, and anguibactin, we examined the nature of the interaction of such cofactor−carrier proteins with acyl-activating adenylation (A) domains. While VibE, EntE, and PchD were all able to utilize "carrier protein-free" pPant derivatives, the pattern of usage indicated diversity in the binding mechanism, and even the best substrates were down at least 3 log units relative to the native cofactor−carrier protein. When tested with four noncognate carrier proteins, EntE and VibE differed both in the range of substrate utilization efficiency and in the distribution of the efficiencies across this range. Correlating sequence alignments to kinetic efficiency allowed for the construction of eight point mutants of VibE's worst substrate, HMWP2 ArCP, to the corresponding residue in its native VibB. Mutants S49D and H66E combined to increase activity 6.2-fold and had similar enhancing effects on the downstream condensation domain VibH, indicating that the two NRPS enzymes share carrier protein recognition determinants. Similar mutations of HMWP2 ArCP toward EntB had little effect on EntE, suggesting that the position of recognition determinants varies across NRPS systems.

Nonribosomal peptide synthetases are responsible for the assembly of thousands of natural peptides elaborated by a wide range of organisms, many of which have important antibiotic, anticancer, immunosuppressant, and virulence-associated activities (*1*, *2*). The NRPS[1] machinery is remarkably versatile, composed of a handful of enzymatic domains of specific function that, through a combination of substrate specificity and physical arrangement, account for the tremendous array of natural products observed. NRPS domains are physically linked, typically clustered in modules, each module dedicated to the activation, condensation, and modification of a given residue in the final peptide (*3*, *4*). An additional level of physical constraint is applied in that the modules themselves are bound to each other and in an order that is collinear with the position of the residue they contribute. Thus, NRPSs utilize the specificity of a given module for its monomeric unit and the arrangement of the modules in the NRPS assembly line to achieve the efficient production of the desired peptide product.

NRPSs activate their monomeric units as thioesters via an adenylated intermediate (*3*, *5*). Every amino acid to be incorporated in the peptide product is tethered by an ATP-utilizing adenylation (A) domain to a phosphopantetheine (pPant) cofactor, itself attached to a small (ca. 10 kDa) carrier protein (*6*−*9*). This cofactor−carrier protein scaffold chaperones the activated aminoacyl group through all of the transformations it must bear during peptide assembly, which most often is condensation by a condensation (C) domain but may also include epimerization (E domain), methylation (MT domain), oxidation/reduction (ox/red domain), and esterification (TE domain) (*4*). The central role played by the cofactor−carrier protein scaffold in presenting the aminoacyl group to NRPS enzymatic domains suggests that it may serve as more than just a leaving group in peptide bond formation. Establishing the contribution of the cofactor−carrier protein component to NRPS domain substrate specificity would be an essential step in understanding to what extent substrate specificity plays a role in peptide assembly (as opposed to physical constraint which would alleviate the need for rigid specificity). This understanding would aid in the combinatorial reprogramming of NRPS systems to synthesize novel peptide products, in which the rearrangement of modules requires the use of noncognate carrier proteins by at least one NRPS domain per module.

Several catecholate siderophores are nonribosomal peptides elaborated by pathogenic bacteria to compete for iron with their hosts and include enterobactin (*Escherichia coli*), vibriobactin (*Vibrio cholerae*), yersiniabactin (*Yersinia pestis*), pyochelin (*Pseudomonas aeruginosa*), and anguibactin

† This work has been supported by the National Institutes of Health (Grant AI042738 to C.T.W.). C.G.M. is a Fellow of the Canadian Institutes of Health Research. M.D.B. is a Fellow of the National Institutes of Health.

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

[1] Abbreviations: A, adenylation; ACP, acyl carrier protein; ArCP, aryl carrier protein; C, condensation; CoA, coenzyme A; Cy, cyclization; DHB, 2,3-dihydroxybenzoate; DTT, dithiothreitol; E, epimerization; I, isochorismatase; MT, methyltransferase; NRPS, nonribosomal peptide synthetase; NSPD, norspermidine; PCP, peptidyl carrier protein; pPant, phosphopantetheine; PPTase, phosphopantetheinyl transferase; TCEP, tris(carboxyethyl)phosphine; TFA, trifluoroacetic acid.

10.1021/bi0202575 CCC: $22.00
Published on Web 06/06/2002

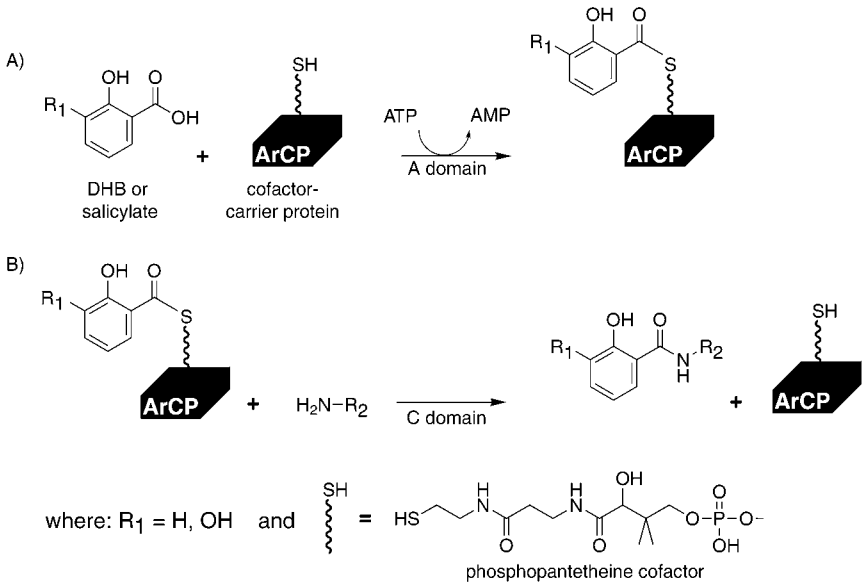

FIGURE 1: Reaction scheme for (A) A domains using DHB or salicylate as an acyl group and (B) C domains using acylated carrier protein as a donor substrate. ArCP = aryl carrier protein.

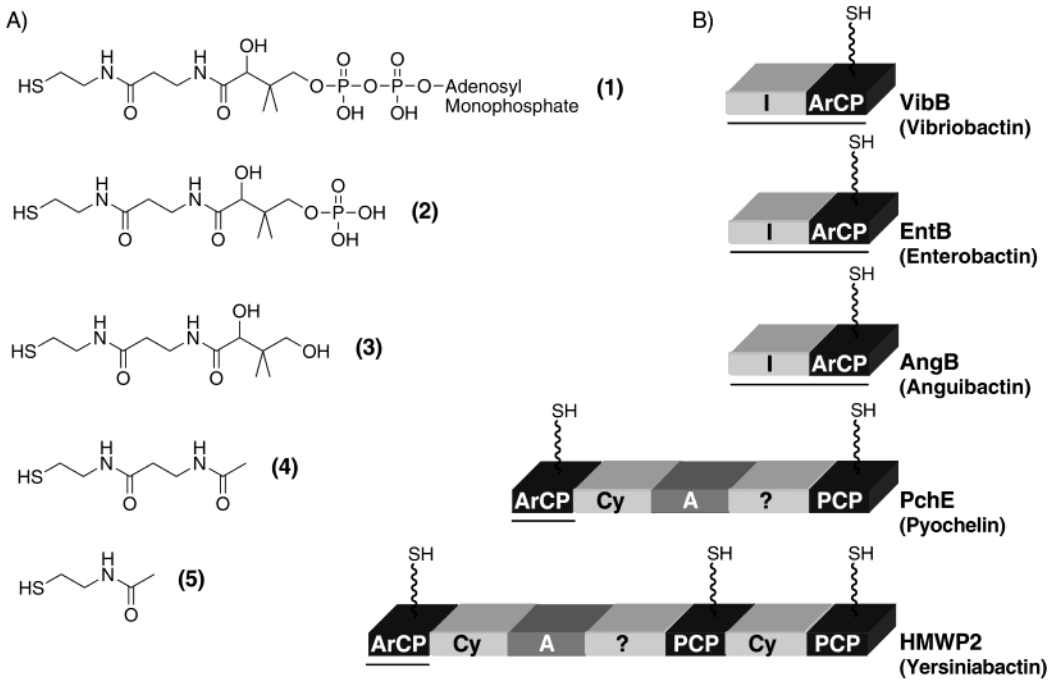

FIGURE 2: Substrates used in this study. (A) Phosphopantetheine derivatives. (B) Aryl carrier proteins from various siderophore NRPS systems shown in their natural context. The underlined region indicates the actual construct used in assays. See abbreviations for domain functional assignments.

(*Vibrio anguillarum*) (*10*−*14*). Siderophore NRPSs have fewer physical linkages between domains than the classical multimodular NRPSs. Interactions between enzyme and substrate are often intermolecular (or *in trans*) as opposed to intramolecular (or *in cis*), allowing the ready determination of kinetic parameters. All of the above siderophore peptides are N-capped by a hydroxylated phenyl ring, either 2,3-dihydroxybenzoic acid (DHB) or salicylate, and activation occurs via a free-standing adenylation domain in conjunction with a phosphopantetheine-bearing (or holo-) aryl carrier protein (ArCP) (Figure 1A). The acylated cofactor−carrier protein is then substrate for the downstream condensation processes specific to each siderophore assembling pathway (Figure 1B).

In this study, we used components of the NRPSs responsible for assembling the siderophores mentioned above to examine recognition of the cofactor−carrier protein scaffold by the first NRPS domain involved in assembly, the A domain. We tested the ability of EntE, VibE, and PchD, the aryl-acid-activating A domains of the enterobactin, vibriobactin, and pyochelin systems, respectively, to use "carrier protein-free" pPant derivatives (Figure 2A). We also tested EntE and VibE for use of noncognate carrier proteins from a variety of different systems (Figure 2B) and were further able to test VibE with carrier protein point mutants. Our findings establish the importance of the carrier protein in binding by A domains. Additionally, we identified two residues in an ArCP that are of particular importance and

show that a downstream C domain has recognition determinants similar to those of its upstream A domain.

EXPERIMENTAL PROCEDURES

Materials and General Methods. DHB, CoA, NSPD, PEP, NADH, pyruvate kinase/lactate dehydrogenase, myokinase, pantetheine, and *N*-acetylcysteamine were purchased from Sigma-Aldrich Chemical Co. [^{14}C]Salicylate was purchased from NEN. ATP was purchased from Boehringer Mannheim. TCEP was purchased from Fluka. Standard recombinant DNA techniques and microbiological procedures were performed as described (*15*). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Plasmid vectors were purchased from Novagen. Pfu polymerase and competent *E. coli* were purchased from Stratagene. Oligonucleotide primers were purchased from Integrated DNA Technologies, and DNA sequencing to verify the fidelity of amplification was performed on double-stranded DNA by the Dana-Farber/Harvard Cancer Center (Boston, MA). Ni-NTA Superflow resin was purchased from Qiagen.

Synthesis of Phosphopantetheine Derivatives. pPant was made in reactions (3 × 2 mL) containing 50 mM HEPES, pH 7.5, 2 mM TCEP, 15 mg of CoA, and 10 units of snake venom nucleotide pyrophophatase. Reactions were incubated for 18 h at 30 °C, filtered through a Centricon-10 (Amicon), and injected onto a Vydac preparative HPLC C18 column. pPant eluted as a single peak at 18.3 min with a gradient of 5%−50% acetonitrile in 0.1% TFA over 23 min at 10 mL/min, monitoring at 220 nm. The yield was 14.8 mg (70%), and UV spectral and MALDI-TOF analyses confirmed the identity of the molecule.

*Synthesis of N-Acetyl-β-aletheine (**4**)*. On the basis of previous synthetic procedures (*16*), β-alanine (50 g, 561 mmol) was suspended in cold acetic anhydride (250 mL) in an ice bath and allowed to warm to room temperature over 1.5 h with stirring. The solution was rotated in a 95 °C bath for 1 h and subsequently dried on a rotary evaporator. The yellow syrup was dissolved in 150 mL of CH_2Cl_2, and cysteamine dihydrochloride (9.4 g, 41.8 mmol) dissolved in 100 mL of water and adjusted to pH 10 with 10 N NaOH was layered over the organic phase. The biphasic mixture was vigorously stirred overnight. The organic phase was washed twice with 50 mL of water, and the aqueous phases were combined, acidified to pH 2.6 with 1 M HCl, and lyophilized. The resulting oil was stirred in 250 mL of ethyl acetate and filtered to collect precipitate. The white solid was recrystallized twice from MeOH, yielding 41.5 g of pure *N*,*N*′-diacetyl-β-aletheine in a 20% yield. MP, proton NMR, and MS were consistent with published results. Reduction of the product followed the published procedure using $NaBH_4$ (*16*), yielding *N*-acetyl-β-aletheine (**4**).

Purification of Adenylation Domains and VibH. The A domains VibE, EntE, and PchD, and the C domain VibH were purified as described previously (*17*−*19*).

Purification of Carrier Proteins. VibB, EntB, AngB, the ArCP domains of PchE and HMWP2, and all mutant carrier proteins were purified from *E. coli* BL21(DE3) containing pET expression vectors that append a $(His)_6$ tag to their C-terminus. The culture (2 × 1 L) in Luria−Bertani broth supplemented with either ampicillin or kanamycin (50 μg/mL) was grown to mid-logarithmic phase at 37 °C and then cooled to room temperature prior to induction of gene expression from the T7 promoter with 1 mM isopropyl-β-D-thiogalactopyranoside for 5 h. Harvested cells were lysed by French press into 20 mM Tris, pH 8.0, 500 mM NaCl, and 5 mM imidazole, the lysate was clarified by ultracentrifugation for 30 min at 90000*g*, and the supernatant was incubated with 0.5 mL of Qiagen Ni-NTA Superflow resin for 3 h. Carrier protein was eluted with a step gradient of 10 column volumes of lysis buffer containing 25−200 mM imidazole in 25 mM increments, and the eluant was analyzed by 15% SDS−PAGE. Fractions containing purified ArCP were pooled and dialyzed against 20 mM Tris, pH 8.0, 50 mM NaCl, 1 mM DTT, and 10% glycerol. Samples were concentrated using a Centricon-3, quantified by the Bradford protein dye assay (Bio-Rad), and aliquoted for storage at −80 °C.

Assay of Phosphopantetheine Derivatives. Transfer of DHB to pPant derivatives by A domains was coupled to NADH oxidation, and the decrease in absorbance at 340 nm was monitored. Reactions (0.1 mL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 2 mM TCEP, 200 μM 2,3-DHB (EntE and VibE) or salicylate (PchD), 5 mM ATP, various amounts of pPant derivative, and 1.0 μM A domain. To couple the reaction to NADH oxidation, reactions also contained 2.5 mM PEP, 0.25 mM NADH, 0.5 unit of myokinase, and 1.0 unit of pyruvate kinase/lactate dehydrogenase. Each mole of AMP produced results in the oxidation of 2 mol of NADH, which has an extinction coefficient of 6.22×10^3 M^{-1} cm^{-1}. In all cases except molecule **4** with EntE, the background rate was minimal and was subtracted from the experimental rate. Product formation was verified by HPLC and MALDI-TOF mass spectrometry (data not shown).

Assay of ArCP Substrates with Adenylation Domains. Acylation of holo-ArCPs with [^{14}C]salicylate was detected by TCA precipitation and scintillation counting. Reactions (25 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 2 mM TCEP, 75 μM (0.1 μCi) [^{14}C]salicylate, 5 mM ATP, various amounts of holo-ArCP, and 2−200 nM A domain. Reactions were incubated at 30 °C for a time determined to give linear turnover and quenched in 9 volumes of 10% trichloroacetic acid and 50 μg of bovine serum albumin as coprecipitant. Pellets were washed twice in 10% TCA and suspended in 88% formic acid; scintillation fluid was added and counted. Holo-ArCP was formed in reactions (25 μL) containing 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 2 mM TCEP, 200 μM CoA, up to 200 μM apo-ArCP, and 1 μM *Bacillus subtilis* Sfp. Reactions were incubated for 60 min at 30 °C and then diluted serially in similar buffer.

Bioinformatic Analysis. Amino acid alignments were performed with DNAStar MegAlign software using the Clustal X algorithm with a gap penalty of 10 and a gap length penalty of 10.

Construction of ArCP Mutant Plasmids. Plasmids containing HMWP2 ArCP point-mutant genes were created by SOE mutagenesis (*20*) of the wild-type plasmid pHMWP2ArCP using the primers listed in Table 1 and the primer pairing in Table 2. Double mutants were made using pHMWP2ArCP R78A as template, while the triple mutant was made using the H66E/R78A template. Amplification products were

Table 1: DNA Primers Used in This Study

primer name	nucleotide sequence[a]
ArCP5′f	5′-CAGCTCCGCCATCGCCGCTTC-3′
ArCP3′r	5′-GCGGTTAGCTCCTTCGGTCCTC-3′
P42VE5′r	5′-CATGTAACTGCTG**CAC**CGTTAAATTCAG-3′
P42V3′f	5′-CTGAATTTAACG**GTG**CAGCAGTTACATG-3′
H46D5′r	5′-GTTGCTCTCTTC**ATC**TAACTGCTGCGG-3′
H46D3′f	5′-CCGCAGCAGTTA**GAT**GAAGAGAGCAAC-3′
S49E5′r	5′-GCCTGGATCAGGTT**CTC**CTCTTCATGTAAC-3′
S49E3′f	5′-GTTACATGAAGAG**GAG**AACCTGATCCAGGC-3′
S49D5′r	5′-GCCTGGATCAGGTT**GTC**CTCTTCATGTAAC-3′
S49D3′f	5′-GTTACATGAAGAG**GAC**AACCTGATCCAGGC-3′
W64L5′r	5′-CGAAACCAGTGTAA**CAA**TCTCATCAATC-3′
W64L3′f	5′-GATTGATGAGA**TTG**TTACACTGGTTTCG-3′
H66E5′r	5′-CATTTTTACGAAACCA**CTC**TAACCATCTCATC-3′
H66E3′f	5′-GATGAGATGGTTA**GAG**TGGTTTCGTAAAAATG-3′
H66A5′r	5′-CATTTTTACGAAACCA**GGC**TAACCATCTCATC-3′
H66A3′f	5′-GATGAGATGGTTA**GCC**TGGTTTCGTAAAAATG-3′
W67K5′r	5′-GCCATTTTTACGAAACTTGTGTAACCATCTC-3′
W67K3′f	5′-GAGATGGTTACAC**AAG**TTTCGTAAAAATGGC-3′
R74D5′r	5′-GGCAAGGGTAAG**GTC**GTAGCCATTTTTACG-3′
R74D3′f	5′-CGTAAAAATGGCTAC**GAC**CTTACCCTTGCC-3′
R78A5′r	5′-GGCATACAGCTC**GGC**AAGGGTAAGGCG-3′
R78A3′f	5′-CGCCTTACCCTT**GCC**GAGCTGTATGCC-3′

[a] Mutation codon in bold.

Table 2: PCR Amplification Primer Pairs in the Construction of HMWP2 ArCP Mutants

mutant name	primers used
R78A	ArCP5′f and R78A5′r; R78A3′f and ArCP3′r
P42V/R78A	ArCP5′f and P42V5′r; P42V3′f and ArCP3′r
H46D/R78A	ArCP5′f and H46D5′r; H46D3′f and ArCP3′r
S49E/R78A	ArCP5′f and S49E5′r; S49E3′f and ArCP3′r
S49D/R78A	ArCP5′f and S49D5′r; S49D3′f and ArCP3′r
W64L/R78A	ArCP5′f and W64L5′r; W64L3′f and ArCP3′r
H66E/R78A	ArCP5′f and H66E5′r; H66E3′f and ArCP3′r
H66A/R78A	ArCP5′f and H66A5′r; H66A3′f and ArCP3′r
W67K/R78A	ArCP5′f and W67K5′r; W67K3′f and ArCP3′r
R74D/R78A	ArCP5′f and R74D5′r; R74D3′f and ArCP3′r
S49E/H66E/R78A	ArCP5′f and S49E5′r; S49E3′f and ArCP3′r

digested with *Xho*I and *Sph*I and ligated to the complementing product of similarly digested pHMWP2ArCP. Ligation products were transformed into *E. coli* DH5α and their identities confirmed by DNA sequencing. Successful clones were transformed into *E. coli* BL21(DE3) for heterologous expression. The host vector for all of these clones was pET22 in which a C-terminal $(His)_6$ tag is appended to the expression product.

Assay of ArCP Substrates with Condensation Domain VibH. Carrier proteins were assayed as substrates for VibH in the presence of excess VibE to maintain them in a fully acylated state. Reactions (50 μL) contained 75 mM Tris, pH 7.5, 10 mM $MgCl_2$, 2 mM TCEP, 1 mM DHB, 5 mM ATP, various amounts of holo-ArCP, 30 mM NSPD, 2 mM VibE, and 5−40 nM VibH. Reactions were incubated at 30 °C for a time determined to be in the linear turnover range, quenched with 9 volumes methanol, and centrifuged for 30 min at 12000*g* at 4 °C. The soluble fraction was dried under vacuum at 32 °C, suspended in 15% acetonitrile, and injected onto a C18 Vydac small-pore column. Peaks were eluted at 1 mL/min in a gradient from 10% to 90% acetronitrile in 0.08% TFA over 23 min. Peak integration values for the DHB product were converted to nanomoles of product on the basis of a standard curve generated with DHB−octylamine as described previously (*19*).

Table 3: Use of Phosphopantetheine Derivatives by A Domains

	V_{max}/E_TK_m ($min^{-1}\,mM^{-1}$)		
substrate	EntE[a]	VibE[a]	PchD[b]
native ArCP	$(1.0 \pm 0.1) \times 10^5$	$(5.8 \pm 0.2) \times 10^3$	$(6.5 \pm 0.2) \times 10^4$
1	0.50 ± 0.02	1.0 ± 0.1	$(18 \pm 2) \times 10^{-2}$
2	0.48 ± 0.06	1.0 ± 0.03	$(9.0 \pm 2.0) \times 10^{-2}$
3	3.3 ± 0.2	1.1 ± 0.02	$(11 \pm 1) \times 10^{-2}$
4	ND	0.17 ± 0.08	$<1.0 \times 10^{-2}$
5	1.9 ± 0.1	0.14 ± 0.001	$(1.0 \pm 0.1) \times 10^{-2}$

[a] DHB used as acylation substrate. [b] Salicylate used as acylation substrate.

RESULTS

Use of Phosphopantetheine Derivatives by EntE, VibE, and PchD. To establish the role of the pPant component of the cofactor−carrier protein substrate in A domain activity (Figure 1A), we measured the efficiency of acylation of "carrier protein-free" pPant derivatives (Figure 2A). Three of the derivatives tested were commercially available, including CoA (**1**), pantetheine (**3**), and *N*-acetylcysteamine (**5**). Phosphopantetheine (**2**) was synthesized enzymatically from CoA and a nucleotide pyrophosphatase and was obtained in good yield and purity. *N*-Acetylaletheine (**4**) was made synthetically on the basis of a published procedure (*16*) by acetylation and anhydride formation of β-alanine followed by aminolysis with cysteamine dihydrochloride to yield the disulfide *N*,*N*′-diacetyl-β-aletheine. This product was reduced with $NaBH_4$ to give **4** with an overall yield of 18%.

Table 3 lists the catalytic efficiencies measured for molecules **1**−**5** when used by EntE, VibE, and PchD A domains and compares them to the native ArCPs EntB, VibB, and PchE ArCP, respectively. Nearly all of the derivatives tested were used as substrates, with the exception of **4**, which was not a substrate for PchD and which uncoupled ATP hydrolysis from aryl acid activation in EntE. PchD was less capable of using the carrier protein-free substrates than EntE or VibE. Despite this across-the-board decrease, the trend in PchD as the pPant chain is shortened was consistent with that observed in VibE, with two distinct groups of the longer (**1**−**3**) and the shorter (**4**, **5**) derivatives separated by a gap approaching 10-fold. This is in contrast to EntE, where no such gap exists. In the case of EntE, it appears that charge and/or bulkiness is more of an obstacle to efficient binding than insufficient chain length, with CoA and pPant faring worse than pantetheine or even *N*-acetylcysteamine. The unexpected conclusion from these observations is that binding of the cofactor portion of the cofactor−carrier protein in these three A domains is functionally different, despite the invariant nature of the molecule being bound and the close relationship these enzymes maintain in amino acid sequence (average homology to each other of 45%) and in function.

From Table 3, it is immediately apparent that, for all three A domains, none of the carrier protein-free substrates were efficient substrates when compared to the native substrate. Even the best carrier protein-free substrates are 3.0×10^4, 5.3×10^3, and 3.6×10^5 worse than the native substrates for EntE, VibE, and PchD, respectively. With the exception of PchD, almost all of this difference in efficiency was a function of K_m (data not shown). The 20-fold difference that

Table 4: Use of Various Noncognate ArCPs by the Adenylation Domains EntE and VibE

ArCP	V_{max}/E_TK_m (min^{-1} μM^{-1})[a]	
	EntE	VibE
(●) VibB	0.15 ± 0.01	**5.8 ± 0.2**
(○) PchE ArCP	10 ± 1	6.8 ± 0.5
(□) AngB	1.5 ± 0.1	4.1 ± 0.3
(■) EntB	**100 ± 10**	0.65 ± 0.01
(△) HMWP2 ArCP	3.2 ± 0.2	0.26 ± 0.01

[a] Native substrate values in bold.

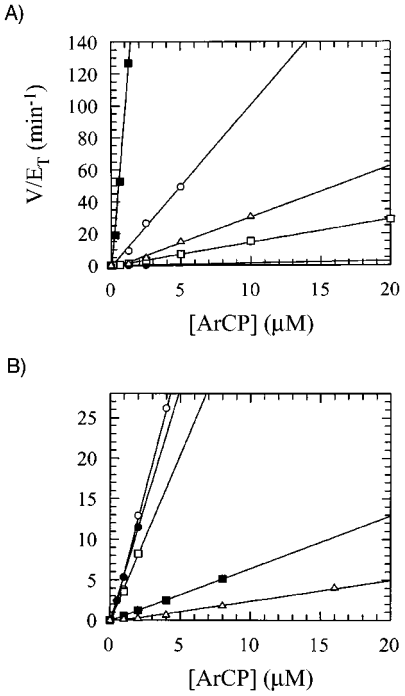

FIGURE 3: Use of ArCPs from various siderophore NRPS systems by (A) EntE and (B) VibE, including VibB (●), PchE ArCP (○), AngB (□), EntB (■), and HMWP2 ArCP (△).

can be attributed to a decrease in V_{max}/E_T with PchD brings the K_m effect in line with the other two A domains, at about 4 log units. This is a potent demonstration of the importance of these carrier proteins in A domain substrate recognition and is consistent with the effect of using carrier protein-free substrates with other NRPS domains (*21*, *22*), as well as in analogous polyketide synthase domains (*26*).

Use of Various Noncognate ArCPs by EntE and VibE. While the significant loss in efficiency observed in carrier protein-free substrates implicates the protein scaffold in contributing important binding energy during catalysis, it does not address the ability of A domains to discriminate cognate from noncognate carrier proteins. We tested the ability of EntE and VibE to utilize ArCPs from a variety of different siderophore NRPS systems (Figure 2B). The five carrier proteins tested ranged from 20% to 36% in sequence alignment similarity. Due to difficulties in providing some ArCPs at concentrations that saturate enzyme activity, as well as complications arising from substrate inhibition, we restricted analysis to the first-order part of the Michaelis−Menten curve, in which slope is equal to V_{max}/E_TK_m. It can be seen from Table 4 and Figure 3 that all five of the ArCPs tested were used by both A domains. As expected, EntB was the best substrate for EntE. The same was not true for VibE, however, with VibB edged slightly by PchE ArCP. The range from best to worst substrate was much larger for EntE (670-fold) than for VibE (26-fold). Furthermore, the substrates were more evenly distributed across this range in EntE, with two distinct groups of "good substrates" and "bad substrates" evident with VibE. Although both A domains used the other's native substrate quite poorly, a reciprocal pattern of usage was not observed. Taken together, these findings suggest that ArCP recognition in A domains is complex, with multiple determinants, as might be expected from a substrate with a large surface area for binding. The differences observed in EntE and VibE may be a reflection of this, in which the additive effect of multiple contributions, even with identical binding mechanisms, could lead to a complex pattern of substrate recognition. Alternatively, the binding mechanism of EntE and VibE could be different.

Amino Acid Alignment of ArCPs. To try to identify some of the ArCP recognition determinants that would allow us to begin mapping the interaction between A domain and ArCP, we took advantage of the clustering of "good" and "bad" ArCP substrates observed with VibE. Figure 4 shows a selected region of a CLUSTAL X alignment of the three good substrates VibB, PchE ArCP, and AngB with the worst substrate HMWP2 ArCP, using the guideline that carrier proteins extend about 37 amino acids on either side of the posttranslationally modified serine. Our objective was to mutate residues in HMWP2 ArCP to the amino acid at the corresponding position in VibB, turning a bad substrate into a good one. There was no immediately apparent cluster of residues exclusively present in the good substrates. We identified eight positions where either (a) residues were conserved in some physical property in the good substrates but not in HMWP2 ArCP (potential contributors to binding) or (b) a residue in HMWP2 ArCP was not similar to residues of any of the good substrates at that position (a potential detractor to binding). The most likely candidate for a binding detractor was R78, and the R78A mutant was made first. To reduce the likelihood that R78 would mask any improvements obtained by further mutation, all subsequent mutants were made on the R78A background.

Mutagenesis and Purification of HMWP2 ArCP. All mutants were made using SOE mutagenesis (*20*). The R78A mutation was made using the plasmid pHMWP2ArCP as template, double mutants were made with R78A template, and the triple mutant was made using a H66E/R78A template. All mutants were sequenced to confirm their identities. The expression vector used appended a C-terminal $(His)_6$ tag to the protein, allowing purification by using nickel affinity chromatography. All mutants were obtained in yield and purity comparable to those of the wild-type protein.

Assay of HMWP2 ArCP Mutants with VibE and VibH. One of the concerns of making ArCP mutants is whether the mutants will be competent for in vitro posttranslational modification with pPant by the transferase enzyme Sfp. All of the mutants were acylated to the same stoichiometric end point as wild-type ArCP, corresponding to 65% of the total protein as determined by Bradford assay. The efficiencies of utilization of the mutants by VibE are reported in Table 5 and Figure 5A. R78A incurred a negligible increase in efficiency. W64L, W67K, and R74D mutations had no effect, while P42V and H46D each produced a modest increase of 1.3-fold. The most significant increases were produced with

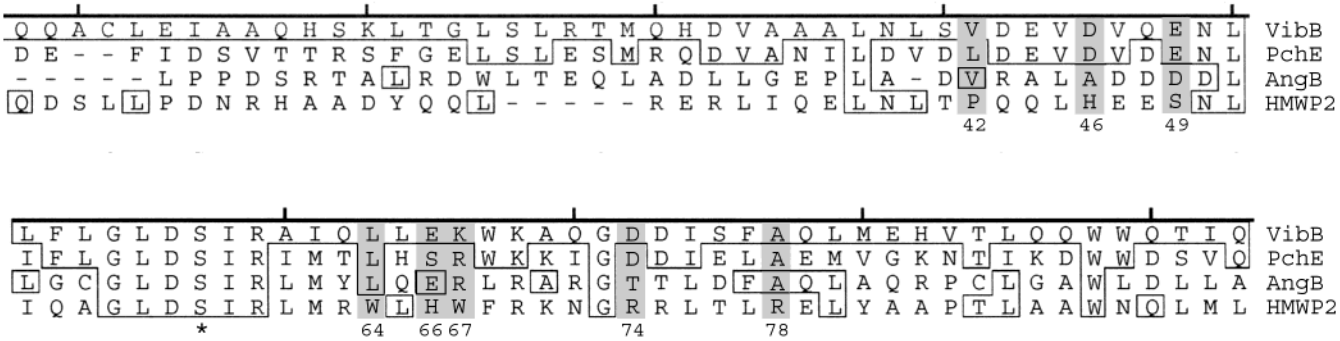

FIGURE 4: Section of an amino acid alignment of selected ArCPs. Boxed residues match the VibB residue. The serine that is posttranslationally modified with pPant is indicated with an asterisk. Shaded positions indicate where HMWP2 ArCP residues were mutated to VibB residues.

Table 5: Use of HMWP2 ArCP Mutants by the A Domain VibE and the C Domain VibH

		V_{max}/E_TK_m ($min^{-1}\,\mu M^{-1}$)	
	ArCP	VibE	VibH
(△)	HMWP2 ArCP	0.26 ± 0.01	5.3 ± 0.2
	R78A	0.27 ± 0.01	8.0 ± 0.3
	P42V/R78A	0.33 ± 0.02	ND
	H46D/R78A	0.34 ± 0.03	ND
(■)	S49E/R78A	0.91 ± 0.09	14 ± 2
	W64L/R78A	0.24 ± 0.01	ND
(○)	H66E/R78A	1.0 ± 0.1	20 ± 1
	W67K/R78A	0.26 ± 0.01	ND
(▲)	R74D/R78A	0.23 ± 0.01	ND
	S49E/H66E/R78A	1.6 ± 0.1	31 ± 1
(●)	VibB	5.8 ± 0.2	110 ± 4

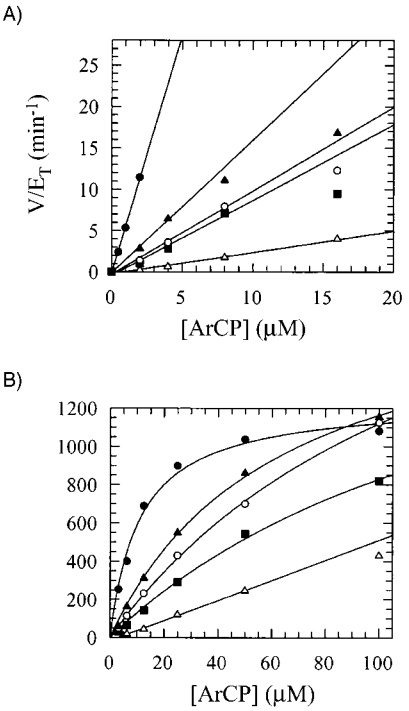

FIGURE 5: Use of HMWP2 ArCP mutants by (A) VibE and (B) VibH, including wild-type HMWP2 ArCP (△), S49E/R78A (■), H66E/R78A (○), S49E/H66E/R78A (▲), and VibB (●). Where possible, data were fit to a Michealis−Menten equation; otherwise the data were fit to a linear equation using selected data points.

S49E and H66E, which boosted the signal 3.5-fold and 3.8-fold over wild type, respectively. The triple mutant S49E/H66E/R78A behaved as the sum of its components, 6.2-fold better than wild type and 3.6 times worse than the native substrate VibB. This suggested that the two enhancing mutations were not totally independent of one another, in which case a 13-fold improvement would be expected. Thus, we identified three residues that have no positive or negative effects, two that have slightly positive effects, and two that have a significantly positive effect on productive substrate utilization by the VibE adenylation domain.

We wanted to determine what effect the two most potent mutations might have on the next downstream enzyme to act on the ArCP in the NRPS assembly line. As illustrated in Figure 1B, C domains use the activated acyl-*S*-carrier protein as substrate in condensation reactions with a downstream amine (often an amino acid itself tethered to a carrier protein). VibH is a free-standing C domain that uses an untethered amine (norspermidine) as a nucleophile and is the natural downstream partner of VibE and VibB (*23*). We used excess VibE to maintain the ArCP in an acylated state and measured how efficiently VibH used R78A, S49E/R78A, H66E/R78A, and the triple mutant, compared to wild-type HMWP2 ArCP and VibB (Table 5, Figure 5B). Similar to the case with VibE, the range of activity spanned between HMWP2 ArCP and VibB was 21-fold. Unlike with VibE, however, R78A had a significant, albeit modest, effect, increasing activity 1.5-fold. This increase in background means that less of the 2.6-fold and 3.8-fold increases experienced by S49E/R78A and H66E/R78A were attributable to the second mutation; however, overall these increases are very much in keeping with the observed effects on VibE. Accordingly, the triple mutant was up 5.8-fold, weighing in at 3.5-fold worse than VibB. Clearly, the S49E and H66E mutations had very similar effects on VibH as they did on VibE, suggesting that these two enzymes recognize similar determinants on the carrier protein surface.

To determine whether the lessons learned about carrier protein recognition by the VibE and VibH enzymes could translate to other systems, we made mutations at the S49 and H66 positions of HMWP2 ArCP to make it more "EntB-like" and measured the effect on catalysis by EntE. To allow direct comparison to the VibE system, the S49D and H66A mutants were made on the R78A background. It can be seen in Table 6 that, as with VibE, R78A alone had no effect. Unlike with VibE, however, mutations at S49 and H66 toward EntB did not significantly increase activity but rather induced mildly negative and positive effects, respectively. This suggests that EntE recognizes its carrier protein differently than does VibE and is consistent with the difference

Table 6: Use of HMWP2 ArCP Mutants by the A Domain EntE

ArCP	V_{max}/E_TK_m ($min^{-1}\mu M^{-1}$)
HMWP2 ArCP	3.2 ± 0.2
R78A	3.3 ± 0.2
S49D/R78A	5.2 ± 0.6
H66A/R78A	2.0 ± 0.2

between these two enzymes in their recognition of pPant derivatives.

DISCUSSION

Nonribosomal peptide synthetases are large assemblies of covalently linked enzymatic domains interspersed with substrate-bearing carrier proteins, clustered as modules and arranged in a very specific order which determines the order in which amino acid monomers are strung together into peptides. The relationship between module order and the order of peptide assembly suggests that physical arrangement may be more important than substrate specificity in the *inter*module translocation of peptide intermediates. The importance of specificity in *intra*modular processes has been established in the characterization of A domain acyl group recognition (*24*) and to a lesser extent the ability of C domains to utilize various aminoacyl-*N*-acetylcysteamine thioesters (*25*). An intrinsic part of the substrates utilized by NRPS enzymatic domains, however, is the cofactor−carrier protein component, and little is known about the extent to which these acyl-bearing protein scaffolds contribute to substrate recognition and to substrate discrimination. Understanding this contribution and how it may differ in intramodule vs intermodule catalysis would be an important step forward in our ability to engineer hybrid NRPS systems. This study examined the nature of carrier protein recognition by A and C domains in siderophore systems, in which carrier proteins are often naturally separate from the enzymes that work upon them, setting the stage for comparison of carrier protein recognition in different natural contexts.

In the first series of experiments, "carrier protein-free" substrates, in the form of pPant derivatives of varying lengths, were all found to be detectable but poor substrates for the three A domains tested. In addition to establishing the importance of the protein component of the cofactor−carrier protein substrate, the data indicate significant differences in how pPant is recognized by these homologous A domains, responding differently to the trimming back of the CoA-derived cofactor. This was not expected as there would be minimal evolutionary pressure on a binding pocket for an invariant component of substrate. This suggests that A domains may have to undergo significant rearrangements to respond to changes in substrate specificity, rather than simply change a residue here or there. While not very good substrates when compared to the native substrate, turnover rates for some of these surrogate substrates were sufficiently fast to validate them as readily available (and fairly generic) aminoacyl activating groups. Thus this methodology can serve to prepare molecules that enable the analysis of downstream NRPS processes.

The differences between A domains were further evident in the abilities of EntE and VibE to use noncognate carrier proteins. Here differences were reflected in the range of efficiency (from worst to best substrate), as well as in the distribution of efficiencies, with EntE demonstrating a much stronger discrimination against noncognate substrates on both counts. It is important to note that this is not necessarily a reflection of the A domain's ability to discriminate cognate from noncognate carrier proteins that it may encounter in its natural environment, as none of the noncognate substrates used would ever be present in this context. The differences between EntE and VibE suggest that carrier protein recognition is complex and dependent on multiple recognition elements, as is observed in the analogous recognition of transfer RNA by aminoacyl-tRNA synthetases (*27*−*29*). In tRNA, recognition elements are a subset of identity elements, which include nucleotides responsible for discrimination by noncognate synthetases. It is likely that carrier protein recognition by A domains occurs via a similar strategy, and a remaining challenge is the identification and classification of these elements.

While EntE expressed a greater range of substrate preferences and thus allowed for a greater potential effect in mutagenesis analysis, the clustered distribution of substrate efficiencies generated by VibE allowed for identification of residues potentially important in substrate recognition/discrimination. Eight positions were identified in which the amino acid residue might either contribute to or detract from binding and the appropriate mutations made in the worst substrate HMWP2 ArCP to the corresponding residue in the native substrate, VibB. From this analysis, two positions, S49 and H66 in HMWP2, corresponding to E239 and E256 in VibB, were found to influence carrier protein recognition by VibE, and the combined effect improved activity over 6-fold. Given that even the best carrier protein-free substrates lost 3 orders of magnitude in efficiency, it is clear that the residues at these positions are not large contributors to the overall energy released upon carrier protein binding. It is likely that the region around the posttranslationally modified Ser, which is invariant at several positions regardless of the quality of the substrate, contributes significantly to binding. However, this is also the region that is likely recognized by the pPant-transferase enzyme, which must faithfully identify all of the carrier proteins in the system if they are to be functionally modified with the pPant cofactor. Therefore, NRPS enzymes must use other parts of the carrier protein to distinguish cognate from noncognate substrates, and these interactions can be significantly less potent than those around the core region. In the transition from peptidoglycan terminating in D-Ala to D-lactate in clinically observed vancomycin resistance, *Enterococcus faecium* relies on an enzyme which prefers D-lactate to D-Ala by only 10-fold (*30*). Residues that are responsible for the discrimination achieved by these more subtle binding interactions will be the targets of NRPS engineers. In the recognition of VibB by VibE, VibB residues E239 and E256 were identified as components of this subtle recognition system.

The next stage in the vibriobactin assembly line requires the transfer of the activated acyl group from the cofactor−carrier protein to the symmetric triamine norspermidine, catalyzed by the free-standing C domain VibH (Figure 1). That VibH also benefited from the mutations at the S49 and H66 position, and to essentially the same degree as did VibE, reinforces the recognition role of these two positions on the carrier protein and also indicates that the recognition elements utilized by these two NRPS enzymes have significant

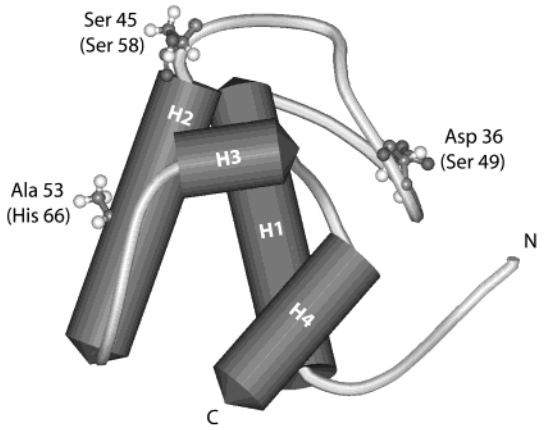

FIGURE 6: Schematic of the NMR-solved structure of a PCP from the tyrocidine biosynthetic protein TyrC (*31*). The posttranslationally modified Ser 45 is labeled, as are the residues that are sequentially equivalent to HMWP2 ArCP S49 and H66. N- and C-termini are indicated.

overlap. By using a highly homologous recognition pattern, NRPS enzymes that operate on the same carrier protein appear to make efficient use of recognition determinants and avoid undesired cross-talk with foreign substrates. Enzymes which have "committed" evolutionarily to a carrier protein may therefore have a relationship with not only the carrier protein but all of the other enzymes that have a similar commitment, and it might be expected that these groups travel together in the natural evolution of novel NRPS pathways. It will be interesting to see if this is a more general communication phenomenon in the NRPS world and, if so, how it helps to define the evolutionary links between NRPS domains fused together in a given synthetase.

To extend the analysis further, the S49 and H66 positions of HMWP2 ArCP were mutated in the direction of EntB and tested with EntE, and the effects observed were minor. This suggests that, as was observed with their use of pPant derivatives, EntE and VibE bind their substrates differently. The broader implication is that there is diversity in the positions of the recognition determinants utilized by carrier proteins. This could be problematic for NRPS engineers who seek to use combinatorial methods; however, it remains to be seen how far-reaching this phenomenon is.

The NMR structure of an apo-PCP from the tyrocidine synthetase TycC has been solved (*31*), and the 74-residue peptide folds into a four-helix bundle, in a manner analogous to the ACP of fatty acid synthase systems. The posttranslationally modified Ser 45 sits atop of helix 2 following a 17-residue loop that has a reasonable amount of conformational flexibility (Figure 6). The structure of an ArCP has not been solved but is predicted to look very similar. When S49 and H66 are mapped onto this PCP, they are found on opposite faces of the protein, about 18 Å apart. This suggests that the binding sites of VibE and VibH utilize the larger surface area of the carrier protein substrate in recognition. This is also observed in tRNA recognition by its synthetase and allows for a more complex matrix of identity elements. While the crystal structures of two A domains have been solved, PheA of *B. brevis* (*32*) and DhbE of *Bacillus subtilis* (M. Marahiel, personal communication), the binding site for a carrier protein has yet to be identified.

In a recent study, the entire helix 2 of a PCP from the TycC (structured A-PCP-E) was replaced with the corresponding helix of an ACP from the same organism (*33*). While the hybrid was competent for aminoacylation by the A domain, it was not a substrate for the downstream E domain. This argues against a model in which enzymes acting on the same carrier protein recognize identical recognition determinants. The *in cis* nature of the enzymes with their carrier protein substrate limited the study's kinetic analysis, which may have masked discrimination against the hybrid PCP. Nevertheless, the refusal by the E domain is absolute and may reflect a special case for these enzymes. Indeed, as the authors point out, carrier proteins upstream of E domains have very distinct conserved residues in sequence alignments, particularly in helix 2.

The carrier proteins used in this study are acted upon intermolecularly (*in trans*) by their native A domains. Thus, selective pressures have evolved these A domains to "seek out" their substrate in the cell as most enzymes do. However, most carrier proteins are acted on *in cis*, by their A domains as well other NRPS enzymes. In these cases, physical associations that favor or disfavor potential substrates may play a greater role than recognition determinants on the carrier proteins themselves. It should be interesting to see how an equivalent study on such an *in cis* system compares with the findings generated here. Additionally, it would be worthwhile to compare C domain recognition of carrier proteins at both the donor and acceptor sites. While A-PCP interactions are always intramodular, there is some question as to the modular organization of C domains (C-A-PCP vs A-PCP-C). Pairing recognition determinants of both C domain sites with neighboring A domains would be a strong argument toward one model and would be an important step toward combinatorial biosynthetic engineering strategies.

REFERENCES

1. Kleinkauf, H., and von Dohren, H. (1995) *Antonie Van Leeuwenhoek 67*, 229−242.
2. Mootz, H. D., and Marahiel, M. A. (1997) *Curr. Opin. Chem. Biol. 1*, 543−551.
3. Keating, T. A., and Walsh, C. T. (1999) *Curr. Opin. Chem. Biol. 3*, 598−606.
4. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) *Curr. Opin. Chem. Biol. 5*, 525−534.
5. Kleinkauf, H., and Von Dohren, H. (1996) *Eur. J. Biochem. 236*, 335−351.
6. Stein, T., Vater, J., Kruft, V., Wittmann-Liebold, B., Franke, P., Panico, M., Mc Dowell, R., and Morris, H. R. (1994) *FEBS Lett. 340*, 39−44.
7. 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.
8. Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M. A., Kruft, V., and Wittmann-Liebold, B. (1991) *J. Biol. Chem. 266*, 23135−23141.
9. Vater, J., Stein, T., Vollenbroich, D., Kruft, V., Wittmann-Liebold, B., Franke, P., Liu, L., and Zuber, P. (1997) *J. Protein Chem. 16*, 557−564.
10. Actis, L. A., Fish, W., Crosa, J. H., Kellerman, K., Ellenberger, S. R., Hauser, F. M., and Sanders-Loehr, J. (1986) *J. Bacteriol. 167*, 57−65.
11. Haag, H., Hantke, K., Drechsel, H., Stojiljkovic, I., Jung, G., and Zahner, H. (1993) *J. Gen. Microbiol. 139*, 2159−2165.
12. Neilands, J. B. (1976) *Ciba Found. Symp.*, 107−24.
13. Griffiths, G. L., Sigel, S. P., Payne, S. M., and Neilands, J. B. (1984) *J. Biol. Chem. 259*, 383−385.
14. Ankenbauer, R. G., Toyokuni, T., Staley, A., Rinehart, K. L., Jr., and Cox, C. D. (1988) *J. Bacteriol. 170*, 5344−5351.

15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
16. Roblot, G., Wylde, R., Martin, A., and Parello, J. (1993) *Tetrahedron 49*, 6381−6398.
17. Quadri, L. E., Keating, T. A., Patel, H. M., and Walsh, C. T. (1999) *Biochemistry 38*, 14941−14954.
18. Gehring, A. M., Bradley, K. A., and Walsh, C. T. (1997) *Biochemistry 36*, 8495−8503.
19. Keating, T. A., Marshall, C. G., and Walsh, C. T. (2000) *Biochemistry 39*, 15513−15521.
20. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene 77*, 51−59.
21. Ehmann, D. E., Trauger, J. W., Stachelhaus, T., and Walsh, C. T. (2000) *Chem. Biol. 7*, 765−772.
22. Marshall, C. G., Burkart, M. D., Keating, T. A., and Walsh, C. T. (2001) *Biochemistry 40*, 10655−10663.
23. Keating, T. A., Marshall, C. G., and Walsh, C. T. (2000) *Biochemistry 39*, 15522−15530.
24. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) *Chem. Biol. 6*, 493−505.
25. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) *Science 284*, 486−489.
26. Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) *J. Am. Chem. Soc. 123*, 6465−6474.
27. Schulman, L. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 23−87.
28. Normanly, J., and Abelson, J. (1989) *Annu. Rev. Biochem. 58*, 1029−1049.
29. Beuning, P. J., and Musier-Forsyth, K. (1999) *Biopolymers 52*, 1−28.
30. Bugg, T. D., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P., and Walsh, C. T. (1991) *Biochemistry 30*, 10408−10415.
31. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) *Struct. Folding Des. 8*, 407−418.
32. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) *EMBO J. 16*, 4174−4183.
33. Mofid, M. R., Finking, R., and Marahiel, M. A. (2002) *J. Biol. Chem. 26*, 26.

BI0202575