Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases

David E Ehmann, John W Trauger, Torsten Stachelhaus* and Christopher T Walsh

Background: Nonribosomal peptide synthetases (NRPSs) are large multidomain proteins that catalyze the formation of a wide range of biologically active natural products. These megasynthetases contain condensation (C) domains that catalyze peptide bond formation and chain elongation. The natural substrates for C domains are biosynthetic intermediates that are covalently tethered to thiolation (T) domains within the synthetase by thioester linkages. Characterizing C domain substrate specificity is important for the engineered biosynthesis of new compounds.

Results: We synthesized a series of aminoacyl-*N*-acetylcysteamine thioesters (aminoacyl-SNACs) and show that they are small-molecule substrates for NRPS C domains. Comparison of rates of peptide bond formation catalyzed by the C domain from enterobactin synthetase with various aminoacyl-SNACs as downstream (acceptor) substrates revealed high selectivity for the natural substrate analog L-Ser-SNAC. Comparing L- and D-Phe-SNACs as upstream (donor) substrates for the first C domain from tyrocidine synthetase revealed clear D- versus L-selectivity.

Conclusions: Aminoacyl-SNACs are substrates for NRPS C domains and are useful for characterizing the substrate specificity of C domain-catalyzed peptide bond formation.

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

*Present address: Biochemie/Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein Strasse, D-35032 Marburg, Germany.

Correspondence: Christopher T Walsh
E-mail: walsh@walsh.med.harvard.edu

Keywords: Condensation domain; Nonribosomal peptide synthetase; Substrate specificity

Received: **26 May 2000**
Revisions requested: **3 July 2000**
Revisions received: **17 July 2000**
Accepted: **27 July 2000**

Published: **18 August 2000**

Chemistry & Biology 2000, **7**:765–772

1074-5521/00/$ – see front matter
© 2000 Elsevier Science Ltd. All rights reserved.
PII: S1074-5521(00)00022-3

Introduction

A variety of peptide natural products, including the antibiotics penicillin and vancomycin, the immunosuppressant cyclosporin A and the bacterial siderophores yersiniabactin, mycobactin and enterobactin, are assembled by nonribosomal peptide synthetases (NRPSs) [1–9]. NRPSs are very large, multidomain proteins that contain sets of functional domains termed modules. The sequence of the peptide produced by a NRPS is determined by the order of modules. Each module contains a thiolation (T) domain (~80 residues) that is posttranslationally modified with phosphopantetheine (P-pant) in a reaction catalyzed by a phosphopantetheinyl transferase [10]. Chain initiation involves each module loading a specific amino acid onto its T domain, a process that involves amino acid activation catalyzed by an adenylation (A) domain (~550 residues) followed by reaction of the resulting aminoacyl adenylate with the P-pant thiol group to form an aminoacyl thioester (aminoacyl-S-T). Subsequent chain elongation steps are catalyzed by peptide bond-forming condensation (C) domains (~450 residues) (Figure 1). In addition to C, A and T domains, which comprise a core elongation module, modules may contain optional modification domains including L- to D-epimerization (E) domains and *N*-methylation domains.

Determination of the substrate specificity of C domains, which is important for the engineered biosynthesis of novel peptides, is challenging because the substrates and products of the reaction are covalently attached to and specified by the synthetase. Reconstitution of one peptide bond-forming reaction to produce a dipeptidyl-S-T domain product requires two T domains primed with phosphopantetheine and loaded with an amino acid. Only recently have such two module reconstitutions been achieved with purified proteins [11,12]. In the absence of downstream domains the dipeptidyl-S-T domain condensation product remains covalently tethered to the enzyme (except in special cases) so turnover cannot occur. Probing substrate specificity is also difficult because the upstream donor and downstream acceptor substrates are defined by the A domains [13,14].

To directly evaluate C domain specificity, we set out to synthesize small-molecule surrogates for aminoacyl-S-T domains, one of the natural substrates for C domains. Precedent from work with polyketide synthases suggested that aminoacyl-S-T domains could be replaced by aminoacyl *N*-acetylcysteamine (NAC) thioesters (aminoacyl-SNACs) [15,16]. The NAC moiety mimics the terminal part of P-pant (Figure 2). We report here the synthesis of a series of aminoacyl-SNACs and the ability of some of these to serve as downstream acceptor substrates for the C domain

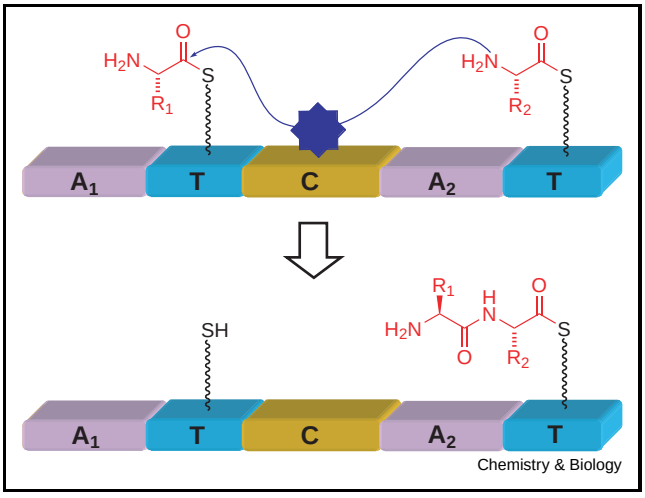

Figure 1. C domains in NRPSs catalyze peptide bond formation between intermediates that are covalently bound to T domains by thioester linkages. Peptide bond formation results in transfer of the upstream acyl donor to the downstream aminoacyl acceptor. The wavy line represents phosphopantetheine.

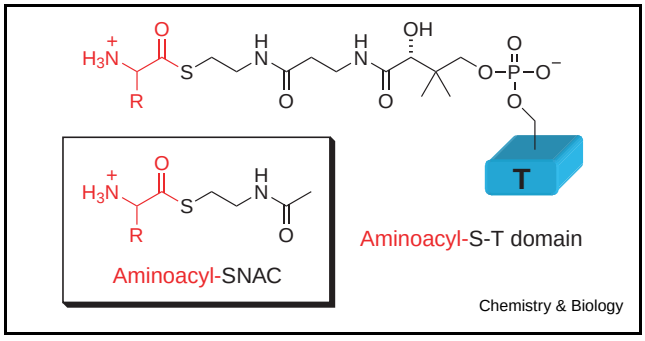

Figure 2. Aminoacyl-SNACs are small-molecule mimics of aminoacyl-S-T domains, one of the natural protein-bound substrates for C domains. NAC is identical to the terminal portion of phosphopantetheine.

in the EntF subunit of enterobactin synthetase [6] (Figure 3a) and as upstream donor substrates for the first C domain of tyrocidine synthetase (Figure 3b) [11]. Peptide bond formation rates for different aminoacyl-SNAC substrates were determined to provide a readout of the C domains' specificities.

Results

Synthesis and characterization of aminoacyl-SNACs

A series of aminoacyl-SNACs were synthesized as outlined in Figure 4. Coupling of NAC to a Boc-protected amino acid was effected with DCC/HOBt and K_2CO_3. Deprotection of the resulting Boc-aminoacyl-SNAC followed by ether precipitation afforded the aminoacyl-SNAC as the trifluoroacetic acid (TFA) salt. L- and D-isomers of Ser-,

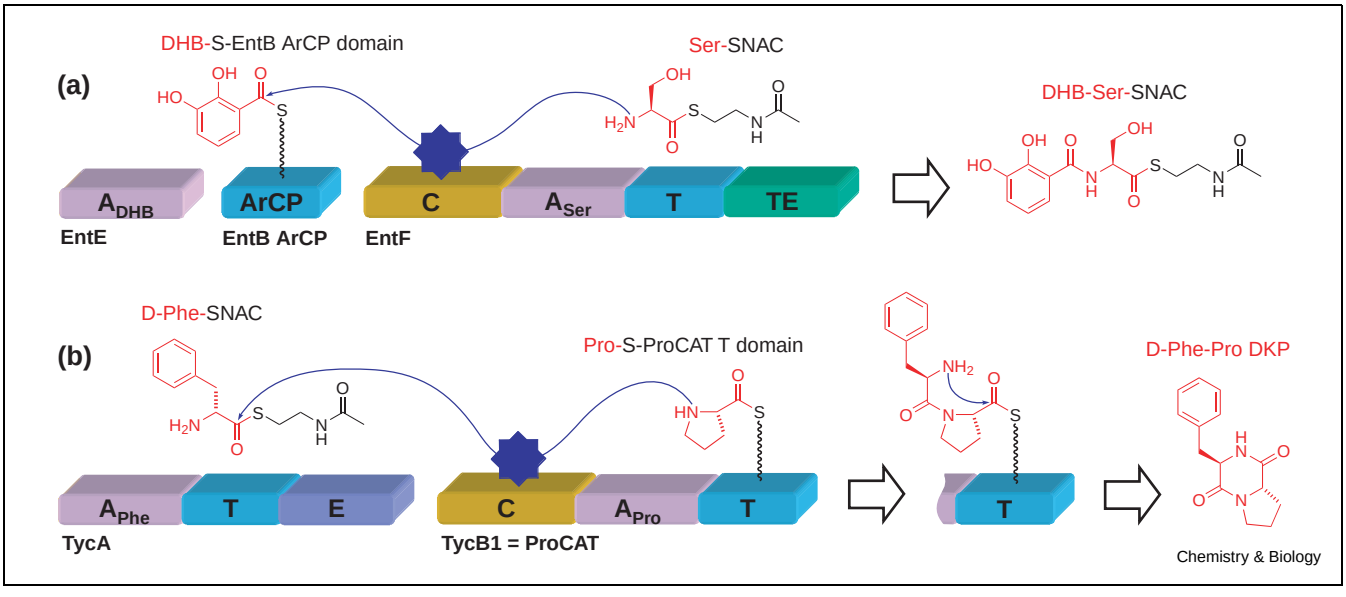

Figure 3. Replacement of aminoacyl-S-T domains with aminoacyl-SNACs. **(a)** For the C domain in the EntF component of enterobactin synthetase, Ser-SNAC mimics Ser-S-EntF T domain, the C domain's natural downstream (acceptor) substrate. **(b)** For the first C domain from the TycB component of tyrocidine synthetase, D-Phe-SNAC mimics D-Phe-S-TycA T domain, the C domain's natural upstream (donor) substrate. Experiments in this study use the protein ProCAT, which consists of the first three domains of TycB. The enzyme-bound dipeptide product D-Phe-Pro-S-ProCAT T domain undergoes non-catalyzed intramolecular cyclization to form D-Phe-Pro diketopiperazine (DKP).

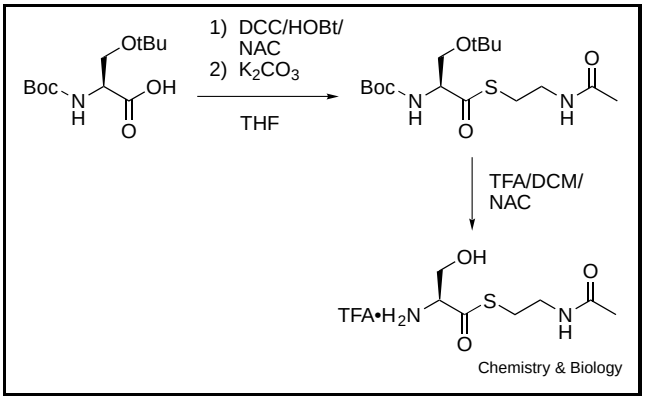

Figure 4. Synthesis of aminoacyl-SNACs. The synthesis of L-Ser-SNAC is shown as a representative example.

Ala-, Thr-, Phe-, Tyr-, Pro- and Leu-SNAC were synthesized. The identity and purity of the compounds was verified by MALDI-TOF mass spectrometry and ^{1}H NMR.

To be useful as C domain substrates, the aminoacyl-SNACs must have sufficient stability toward non-enzymatic hydrolysis and dimerization. We determined half-lives for non-enzymatic hydrolysis/dimerization for the L-aminoacyl-SNACs (75 mM Tris–HCl, pH 7.5, 37°C) by detection of the free thiol of NAC using the DTNB colorimetric assay [17]. With the exception of L-Pro-SNAC, all of the aminoacyl-SNACs displayed half-lives between 50 and 130 min. The L-Pro-SNAC exhibited a shorter half-life of only 15 min, probably due to formation of Pro-Pro diketopiperazine. Since all of the enzymatic reactions described in this study use 2–15 min reaction times, the aminoacyl-SNACs are sufficiently stable to be useful as C domain substrates.

L-Ser-SNAC as downstream (acceptor) substrate for the EntF C domain

The enterobactin NRPS consists of three subunits, EntE, EntB and EntF. The EntF subunit (140 kDa, four domains) has a single C domain that uses as upstream (donor) substrate 2,3-dihydroxybenzoic acid (DHB) tethered to the aryl carrier protein (ArCP) domain of EntB (DHB-S-EntB ArCP) and as downstream (acceptor) substrate serine tethered to the T domain of EntF (Ser-S-EntF T domain) [6]. We set out to determine if Ser-SNAC could replace Ser-S-EntF T domain as acceptor substrate for the EntF C domain (Figure 3a). If so, then the condensation product would be released into solution rather than remaining tethered to the EntF T domain, allowing one to monitor multiple turnovers of the C domain.

In a reaction containing EntF, the donor substrate DHB-S-EntB ArCP domain (10 kDa) and the acceptor substrate L-Ser-SNAC, we observed formation of the condensation

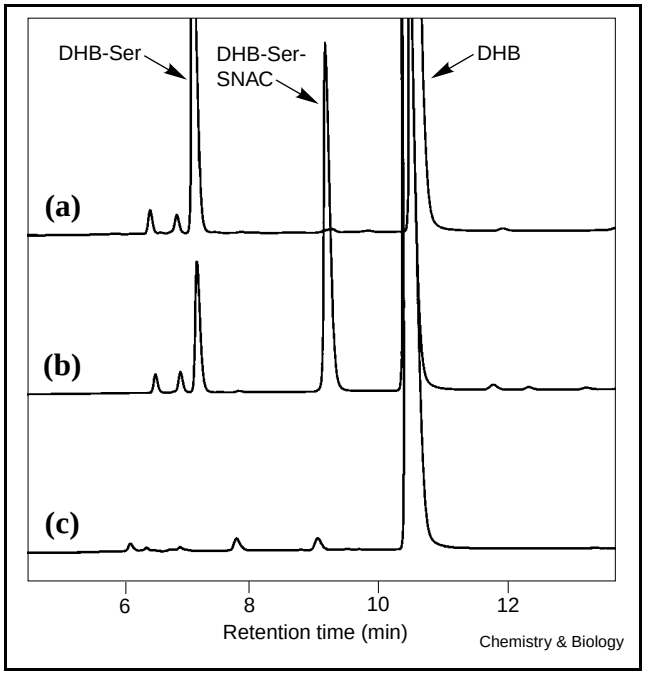

Figure 5. L-Ser-SNAC is a downstream (acceptor) substrate for the EntF C domain. HPLC traces of reactions that contained EntB ArCP loaded with DHB and 5 mM L-Ser-SNAC in the presence or absence of EntF are shown: **(a)** reaction with 1 μM EntF followed by addition of potassium hydroxide, **(b)** reaction with 1 μM EntF, and **(c)** control reaction without EntF. All reactions were carried out as described in Materials and methods.

product DHB-Ser-SNAC and DHB-Ser (which results from thioester hydrolysis during the reaction and work-up) (Figure 5). The identity of the products was determined by MALDI-TOF mass spectrometry and, for DHB-Ser, by HPLC co-elution with a synthetic standard. No product formation was detectable in the absence of EntF. In subsequent assays, the condensation product DHB-Ser-SNAC was completely hydrolyzed to DHB-Ser by briefly heating with potassium hydroxide before HPLC analysis to simplify product quantitation. A reaction with 2 mM L-Ser-SNAC and 500 nM EntF (4000:1 substrate:enzyme) showed linear DHB-Ser formation over 15 min. The concentration dependence of the rate of peptide bond formation demonstrates that the reaction follows Michaelis–Menten kinetics, with a K_m of 7 mM and a robust k_{cat} of 230 min^{-1} (Figure 6). Thus, the EntF C domain can utilize L-Ser-SNAC as a downstream (acceptor) substrate in place of the normal substrate L-Ser-S-EntF T domain.

Peptide bond formation catalyzed by truncated EntF proteins

EntF contains four domains, an N-terminal C domain (50 kDa) followed by a serine-specific adenylation (A) domain [18], a T domain (10 kDa) and a C-terminal thioesterase domain (30 kDa). We next prepared the truncated EntF

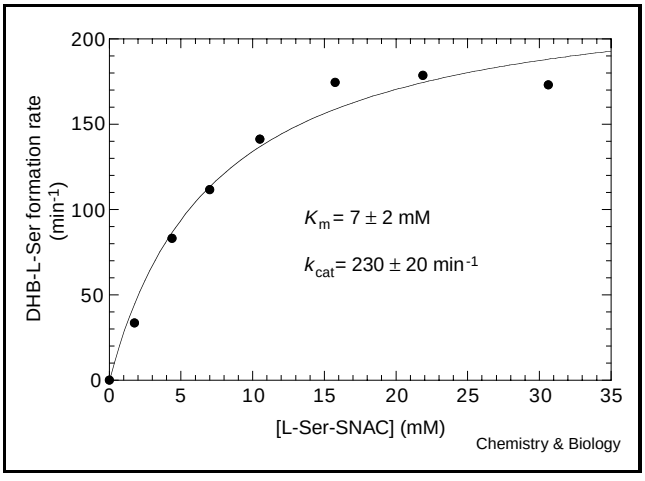

Figure 6. Plot of EntF-catalyzed DHB-Ser formation rate versus Ser-SNAC concentration. Reactions contained EntB ArCP loaded with DHB, Ser-SNAC and 250 nM EntF and were incubated at 37°C for 4 min.

proteins EntF C-A-T (110 kDa) and EntF C-A (100 kDa), which delete the last one or two domains of EntF. Both EntF fragments retain C domain activity with DHB-S-EntB ArCP as donor and L-Ser-SNAC as acceptor. In reactions with 2 mM L-Ser-SNAC, full-length EntF shows a k_{obs} of 30 min^{-1}, while EntF C-A-T and EntF C-A show k_{obs} values of 66 min^{-1} and 64 min^{-1}, respectively (Figure 7). No DHB-Ser formation was observed in the absence of the EntF fragment. We also observed that serine could not replace Ser-SNAC as acceptor substrate (data not shown).

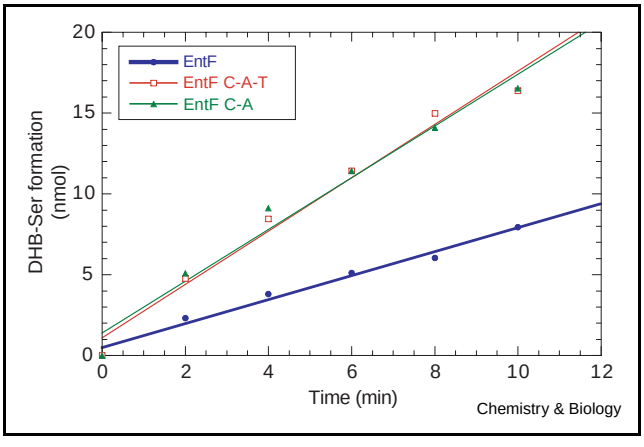

Figure 7. DHB-Ser formation rates with truncated EntF proteins. Reactions contained EntB ArCP loaded with DHB, 2 mM Ser-SNAC and 500 nM EntF, EntF C-A-T or EntF C-A. Full-length EntF displays a k_{obs} for DHB-Ser formation of 30 min^{-1}, while the truncated EntF proteins EntF C-A-T and EntF C-A display k_{obs} values of 66 min^{-1} and 64 min^{-1}, respectively.

With the two-domain protein EntF C-A, which contains no T domain, there is no possibility that peptide bond formation results from reaction of Ser-SNAC with the P-pant group of the EntF T domain to form Ser-S-EntF T domain. Thus, the C domain activity observed for EntF C-A confirms that Ser-SNAC is used directly as a substrate by the C domain.

EntF C domain recognition of side chain size and L- vs. D-configuration

We next tested a series of aminoacyl-SNACs that vary the side chain size and L- versus D-configuration as acceptor substrates for the EntF C domain. All assays were carried out using 500 nM EntF and 2 mM aminoacyl-SNAC. When D-Ser-SNAC was used, a rate of 0.7 min^{-1} was observed compared to a rate of 30 min^{-1} observed with L-Ser-SNAC (Table 1). Thus, the EntF C domain is 40-fold selective for L- versus D-Ser-SNAC. Further assays with Ala-, Thr-, Tyr-, Phe-, Leu- and Pro-SNAC revealed that the EntF C domain is selective for its natural substrate analog Ser-SNAC (Table 1). Compared with Ser-SNAC, the reaction rates with L-Ala-SNAC, L-Thr-SNAC, L-Tyr-SNAC and L-Phe-SNAC were diminished 2-fold, 30-fold, 40-fold and 50-fold, respectively, and no peptide bond formation was detectable with Leu-SNAC and Pro-SNAC. In addition, none of the D-aminoacyl-SNACs D-Ala-, D-Thr-, D-Tyr-, D-Phe-, D-Leu- and D-Pro-SNAC were substrates. Thus, the EntF C domain recognizes both the side chain and L- vs. D-configuration of its downstream (acceptor) substrate.

D-Phe-SNAC as upstream (donor) substrate for the first C domain of tyrocidine synthetase

Tyrocidine synthetase consists of three subunits, TycA, TycB and TycC. The first C domain in this synthetase is the first domain of the TycB subunit, and normally uses D-Phe loaded onto the T domain of TycA as upstream (donor) substrate and Pro loaded onto the first T domain of TycB as downstream (acceptor) substrate. We next examined whether this first C domain could recognize D-Phe-

Table 1
Rates (k_{obs}, min^{-1}) of EntF C domain-catalyzed DHB-amino acid formation with EntB ArCP loaded with DHB as upstream (donor) substrate and an aminoacyl-SNAC as downstream (acceptor) substrate.

Aminoacyl-SNAC	Side chain configuration	
	L	D
Ser-SNAC	30	0.7
Ala-SNAC	16	< 0.1
Thr-SNAC	1.1	< 0.1
Tyr-SNAC	0.7	< 0.1
Phe-SNAC	0.6	< 0.1
Leu-SNAC	< 0.1	< 0.1
Pro-SNAC	< 0.1	< 0.1

Reactions were carried out with 500 nM EntF and 2 mM aminoacyl SNAC as described in Materials and methods.

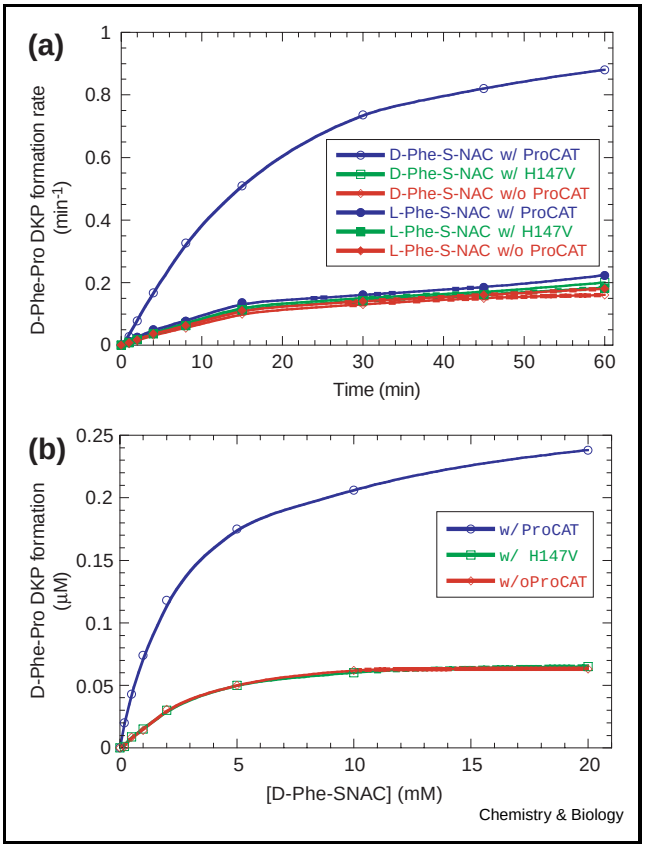

Figure 8. D-Phe-SNAC is an upstream (donor) substrate for the C domain of ProCAT. **(a)** Amount of Phe-Pro DKP condensation product formed as a function of time in reactions that contained either 500 nM holo-ProCAT, 500 nM holo-ProCAT(H147V) or no ProCAT, 1.3 μM [^{3}H]Pro and either 5 mM D-Phe-SNAC or 5 mM L-Phe-SNAC. The Pro-CAT proteins were converted to holo-form using the phosphopantetheinyl transferase Sfp. **(b)** ProCAT-catalyzed D-Phe-Pro DKP formation rate versus D-Phe-SNAC concentration. Reactions contained 500 nM holo-ProCAT, 1.3 μM [^{3}H]Pro and 5 mM D-Phe-SNAC and were incubated at 37°C for 8 min.

SNAC as an upstream (donor) substrate in place of the natural substrate D-Phe-S-T. Prior efforts [11] have established that first module of gramicidin S synthetase (GrsA, which is functionally equivalent to TycA) and the isolated first module of TycB (referred to as ProCAT) function together in vitro to form a peptide bond, yielding D-Phe-Pro-S-ProCAT T domain. In the absence of downstream modules, the enzyme-bound dipeptide undergoes intramolecular cyclization, resulting in release of D-Phe-Pro diketopiperazine (DKP) at the slow rate of 0.5 min^{-1} [11,13,19].

When ProCAT (500 nM) was incubated with [^{3}H]Pro, ATP and D-Phe-SNAC (5 mM), we observed time-dependent production of radiolabeled D-Phe-Pro DKP, reaching about 1.7 equivalents relative to enzyme over 1 h (Figure 8a). The identity of the product D-Phe-Pro DKP was verified by HPLC co-elution with a synthetic standard and by MALDI-TOF mass spectrometry. The reaction showed saturation kinetics with a K_m of about 2.5 mM (Figure 8b). When ProCAT was either omitted or replaced by the catalytically inactive C domain mutant Pro-CAT(H147V), no D-Phe-Pro DKP formation above background was observed (Figure 8a). When L-Phe-SNAC was used in place of D-Phe-SNAC, no DKP formation above the non-enzymatic background rate was observed. This result indicates that the ProCAT C domain is selective for the D- versus L-configuration of its natural upstream (donor) substrate Phe-S-TycA T domain.

Non-catalytic DKP formation at a rate comparable to the enzyme-catalyzed rate was observed in control reactions without ProCAT (Figure 8a). The high background rate is probably due to slow, rate-limiting product release by DKP formation (0.5 min^{-1}). With D-Leu-SNAC as donor substrate for the ProCAT C domain, we observed no DKP formation above the non-enzymatic background rate. However, because the background rate was comparable to the ProCAT-catalyzed rate with D-Phe-SNAC (the analog of the natural substrate), we cannot conclude whether this result reflects actual selectivity of the C domain for D-Phe-SNAC versus D-Leu-SNAC. Thus, while we are able to demonstrate D- versus L-donor selectivity for the Pro-CAT C domain, the GrsA/ProCAT system is not suited to systematically probing donor selectivity.

The GrsA/ProCAT system provided an opportunity to analyze whether ProCAT can use Pro-SNAC as downstream (acceptor) substrate. However, in experiments using GrsA loaded with D-Phe as donor and Pro-SNAC as acceptor no DKP formation above background was observed (data not shown). Also, when D-Phe-SNAC and L-Pro-SNAC were tested together with ProCAT to evaluate if one could use aminoacyl-SNACs as both donor and acceptor substrates at the same time, no DKP formation over the background level was observed.

Discussion

In studies of the substrate specificity of C domains in NRPSs, the use of small-molecule substrate analogs to mimic both the upstream (donor) substrate aminoacyl- or peptidyl-S-T domain and the downstream (acceptor) substrate aminoacyl-S-T domain is potentially very useful. Our approach of using aminoacyl-SNACs as small-molecule substrate analogs builds on precedent from work with polyketide synthases (PKSs), where acyl-S-T domain donors have been replaced by acyl-SNAC donors to change chain initiation units and to bypass inactivated chain elongation domains [15,16]. In addition, work with the actinomycin NRPS demonstrated that an aryl thioester could serve as a C domain donor substrate [20].

We showed here that the C domain in the EntF subunit of enterobactin synthetase can utilize L-Ser-SNAC as a downstream (acceptor) substrate, allowing us to measure for the first time the value of k_{cat} (230 min^{-1}) for an NRPS C domain. Our results here indicate that the EntF C domain recognizes the CH_2OH side chain of its natural substrate (L-Ser-S-EntF T domain) since L-Ala-, L-Thr-, L-Phe-, L-Tyr-, L-Leu- and L-Pro-SNAC are poorer substrates. The observed 40:1 preference of the EntF C domain for L-Ser-SNAC versus D-Ser-SNAC and the fact that none of the other D-aminoacyl-SNACs tested are substrates demonstrates that this C domain is selective for L- versus D-aminoacyl acceptor substrates. This result is in accord with earlier work in which aminoacyl CoAs were used to generate aminoacyl-S-T domain acceptors [13]. With a convenient assay for EntF C domain activity in hand, we were then able to show that the EntF C-A fragment retained full peptide bond-forming activity. In future studies, further truncation of the EntF C-A protein and point mutations within the C domain will allow definition of a minimal C domain and illuminate the role of conserved residues.

We also showed here that the first C domain of tyrocidine synthetase (i.e., the ProCAT C domain) can utilize D-Phe-SNAC as an upstream (donor) substrate. The observed preference for D-Phe-SNAC versus L-Phe-SNAC donors shows that this C domain is selective for aminoacyl-S-T domain donors having the D-configuration. This result agrees with earlier work using aminoacyl CoAs [13] and with an experiment that used a GrsA mutant that has an inactivated epimerization domain together with ProCAT, in which no peptide bond formation was observed despite formation of the potential donor substrate L-Phe-S-GrsA T domain [19]. We were unable due to the slow turnover and high non-enzymatic background rate of this system to systematically probe specificity of the ProCAT C domain for different D-aminoacyl-SNAC donors.

The C domain donor and acceptor substrate specificities observed here are likely to be general for other C domains. Thus, it is likely that C domains in general will specifically recognize the side chain and L- versus D-configuration of their downstream (acceptor) aminoacyl substrates and will recognize the L- versus D-configuration of their upstream (donor) substrates. These results are useful guidelines for the engineered biosynthesis of new compounds. The approach described here of using aminoacyl-SNACs to probe C domain specificity could be broadened to the use of peptidyl-SNACs [21], which in principle would allow analysis of the donor specificity of internal C domains, as opposed to first C domains which use aminoacyl thioesters as substrates. Finally, aminoacyl- and peptidyl-SNACs should be useful small-molecule substrates for other NRPS domains including the heterocyclizing condensation domains [22], epimerization domains [19] and *N*- and *C*-methylation domains. Indeed, peptidyl-SNACs have recently been used to probe the substrate specificity of a chain-terminating thioesterase domain [21].

Significance

Nonribosomal peptide synthetases (NRPSs) catalyze the formation of a diverse array of natural products. In these large, multidomain proteins, substrate specificity is well-documented for adenylation domains, which catalyze amino acid activation, but less is known about the selectivity of condensation (C) domains, which catalyze peptide bond formation. In this study, we used aminoacyl *N*-acetylcysteamine thioesters (SNACs) as small-molecule substrates for C domains to assess substrate specificity. Studies of one C domain using aminoacyl-SNACs as downstream (acceptor) substrates revealed high selectivity for the side chain size and L- versus D-configuration of the substrate. Experiments with a second C domain from a different NRPS using aminoacyl-SNACs as upstream (donor) substrates revealed clear D- versus L-selectivity. This study demonstrates a new approach to characterizing C domain specificity, which is important for the engineered biosynthesis of novel compounds.

Materials and methods

General procedure for synthesis of aminoacyl-SNAC substrates

^1H NMR spectra were recorded on either a Varian Unity NMR spectrometer at 500 MHz or a Varian M200 spectrometer at 200 MHz. MALDI-TOF mass spectrometry was carried out using a Perseptive Biosystems Voyager-DE STR mass spectrometer. Boc-protected amino acids were obtained either from Calbiochem or Bachem. To the Boc-protected amino acid (1 eq.) in THF was added a mixture of DCC (1 eq.) and HOBt (1 eq.) in THF followed by *N*-acetylcysteamine (1 eq., except for Tyr derivatives, for which 10 eq. was used). After stirring the resulting mixture for 45 min at 24°C, K_2CO_3 (0.5 eq.) was added and the reaction stirred for 3 h at 24°C. The reaction was then filtered and concentrated by rotary evaporation. The residue was taken up in ethyl acetate and washed once with one volume of 10% aqueous $NaHCO_3$. The organic layer was dried (Na_2SO_4) and concentrated. The crude product was purified by flash chromatography (3–5% MeOH in $CHCl_3$) and concentrated to afford the Boc-aminoacyl-SNAC. Yields were in the range of 50–80%. Deprotection was carried out by dissolving the Boc-aminoacyl-SNAC in 50% TFA/CH_2Cl_2 (for Ser and Thr derivatives, which have a *t*-butyl side chain protecting group, the deprotection solution also contained 5% (v/v) *N*-acetylcysteamine). After allowing the solution to stand at 24°C for 1 h, the reaction was concentrated by rotary evaporation. The residue was twice taken up in CH_2Cl_2 and concentrated to remove TFA. The resulting residue was taken up in a minimal volume of CH_2Cl_2 and precipitated with diethyl ether. The resulting solid was washed twice with ether and dried to afford the aminoacyl-SNAC in 30–80% yield.

Characterization of aminoacyl-SNAC substrates

L- and D-Boc-Ser(tBu)-SNAC: ^1H NMR (DMSO-d_6): δ 7.97 (t, 1H), 7.25 (d, 1H, J=8.0 Hz), 4.08 (m, 1H), 3.47 (d, 2H, J=5.4 Hz), 3.06 (m, 2H), 2.79 (t, 2H, J=7.0 Hz), 1.72 (s, 3), 1.35 (s, 9H), 1.04 (s, 9H).

L- and D-Ser-SNAC: ^1H NMR (DMSO-d_6): δ 8.39 (br s, 3H), 8.05 (m, 1H), 5.80 (br s, 1H), 4.33 (t, 1H, J=4.0 Hz), 3.81 (d, 2H, J=4.0 Hz), 3.19 (m, 2H), 3.04 (m, 2H), 1.78 (s, 3H); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 207.5 (207.1 calc.).

L-Boc-Thr(tBu)-SNAC: ^1H NMR (DMSO-d_6): δ 7.99 (t, 1H), 6.35 (d, 1H,

J=8.6 Hz), 4.02 (m, 2H), 3.06 (m, 2H), 2.80 (m, 2H), 1.72 (s, 3H), 1.36 (s, 9H), 1.01 (m, 12H).

D-Boc-Thr(tBu)-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.00 (t, 1H), 6.36 (d, 1H), 4.05 (m, 2H), 3.12 (m, 2H), 2.84 (m, 2H), 1.76 (s, 3H), 1.40 (s, 9H), 1.05 (m, 12H).

L-Thr-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.33 (br s, 1H), 8.07 (t, 1H), 5.63 (br s, 1H), 4.14 (m, 2H), 3.20 (m, 2H), 3.05 (m, 2H), 1.78 (s, 3H), 1.20 (d, 3H, J=6.6 Hz); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 221.5 (221.1 calc.).

D-Thr-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.33 (br s, 3H), 8.07 (t, 3H), 5.63 (br s, 1H), 4.14 (m, 2H), 3.20 (m, 2H), 3.02 (m, 2H), 1.78 (s, 3H), 1.20 (d, 3H, J=6.2 Hz); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 221.5 (221.1 calc.).

L- and D-Boc-Ala-SNAC: ^{1}H NMR (DMSO-d_6): δ 7.99 (s, 1H), 7.60 (d, 1H, J=7.5 Hz), 4.06 (q, 1H, J=7.0 Hz), 3.12 (m, 2H), 2.83 (m, 2H), 1.78 (s, 3H), 1.39 (s, 9H), 1.20 (d, 3H, J=7.5 Hz).

L- and D-Ala-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.34 (br s, 1H), 8.06 (s, 1H), 4.31 (q, 1H, J=7.5 Hz), 3.21 (m, 2H), 3.05 (m, 2H), 1.79 (s, 3H), 1.42 (d, 3H, J=7.5 Hz). MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 191.1 (191.1 calc.).

L- and D-Boc-Phe-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.01 (s, 1H), 7.64 (d, 2H, J=8.5 Hz), 7.25 (m, 5H), 4.23 (m, 1H), 3.15 (m, 2H), 3.03 (m, 1H), 2.85 (m, 2H), 2.78 (m, 1H), 1.78 (s, 3H), 1.31 (s, 9H).

L- and D-Phe-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.49 (br s, 3H), 8.01 (t, 1H), 7.28 (m, 5H), 4.51 (t, 1H, J=6.6), 3.13 (m, 4H), 2.95 (m, 2H), 1.78 (s, 3H); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 267.2 (267.1 calc.).

L- and D-Boc-Tyr-SNAC: ^{1}H NMR (DMSO-d_6): δ 9.16 (br s, 1H), 7.98 (t, 1H, J=5.4 Hz), 7.56 (d, 1H, J=8.0 Hz), 6.97 (d, 2H, J=8.8 Hz), 6.59 (d, 2H, J=8.7 Hz), 4.06 (m, 1H), 3.07 (m, 2H), 2.80 (m, 4H), 1.73 (s, 3H), 1.28 (s, 9H).

L- and D-Tyr-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.44 (br s, 3H), 8.01 (t, 1H), 7.03 (d, 2H, J=8.5 Hz), 6.70 (d, 2H, J=8.4 Hz), 4.42 (t, 1H, J=6.6 Hz), 3.14 (m, 2H), 2.98 (m, 4H), 1.78 (s, 3H); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 283.1 (283.1 calc.).

L- and D-Boc-Leu-SNAC: ^{1}H NMR (DMSO-d_6): δ 7.99 (s, 1H), 7.57 (d, 1H, J=8.0 Hz), 4.03 (m, 1H), 3.11 (m, 2H), 2.83 (m, 2H), 1.77 (s, 3H), 1.62 (m, 1H), 1.41 (m, 2H), 1.40 (s, 9H), 0.86 (d, 3H, J=6.0 Hz), 0.82 (d, 3H, J=6.0 Hz).

L- and D-Leu-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.46 (br s, 3H), 8.07 (s, 1H), 4.20 (t, 1H), 3.20 (m, 2H), 3.07 (m, 2H), 1.78 (s, 3H), 1.76 (m, 1H), 1.61 (m, 2H), 0.91 (d, 6H, J=6.5 Hz); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 233.2 (233.1 calc.).

L- and D-Boc-Pro-SNAC: ^{1}H NMR (DMSO-d_6): δ 8.02 (d, 1H), 4.30 (t, 1H, J=7.5 Hz), 3.33 (m, 2H), 3.13 (m, 2H), 2.87 (m, 2H), 2.19 (m, 2H), 1.83 (m, 2H), 1.76 (s, 3H), 1.32 (s, 9H).

L- and D-Pro-SNAC: ^{1}H NMR (DMSO-d_6): δ 9.22 (br s, 2H), 8.03 (m, 1H), 4.59 (t, 1H, J=7.0 Hz), 3.36 (m, 2H), 3.15 (m, 2H), 2.96 (m, 2H), 2.22 (m, 2H), 1.82 (m, 2H), 1.65 (s, 3H); MALDI-TOF MS (monoisotopic, L-isomer) [M+H]: 217.2 (217.1 calc.).

Determination of aminoacyl-SNAC hydrolysis/dimerization rates

Aminoacyl-SNAC hydrolysis/dimerization rates were determined by measuring formation of *N*-acetylcysteamine using 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) [17]. Reactions (200 μl) contained 3 mM DTNB, 75 mM Tris–HCl (pH 7.5) and 5 mM aminoacyl-SNAC at 37°C. Rates were determined by measuring the increase in absorbance at 412 nm (due to formation of the 2-nitro-5-thiobenzoate anion, extinction coefficient = 14 140 M^{-1} cm^{-1}) on a Perkin-Elmer Lambda 6 spectrophotometer over 8 min.

Protein expression and purification

EntD, EntE, EntB ArCP, full-length EntF and EntF C-A proteins were purified as previously described [9,23,24]. To generate the EntF C-A-T protein, the T domain coding sequence was amplified by PCR from pPMS22 [18] using the primers EntF426-F (5′-GAATTCCATATGCTGCTGTGCGGCGATGTCGAT-3′) and EntF1046-R (5′-TGACCGCTCGAGTTCTTCAGCATCAATAATCG-3′). The PCR product was digested with *Nsi*I and *Xho*I and ligated into *Nsi*I/*Xho*I-digested pET22b-EntF-(1–974-His_6) [24] to create plasmid pET22b-EntF(1–1046-His_6). *Escherichia coli* BL21(DE3) cells harboring this plasmid were grown at 37°C in 2 l LB medium with 100 μg/l ampicillin to an optical density at 600 nm of 0.6. The cells were then induced with 1 mM IPTG and grown at 20°C for 8 h. After lysing the cells using a French press, the EntF C-A-T protein was purified by Ni^{2+} affinity chromatography. The eluted protein was dialyzed against 25 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 2 mM dithiothreitol and 5% glycerol and stored at −80°C. Protein concentration was determined by the Bradford dye assay and by using the calculated extinction coefficient at 280 nm of 138 180 M^{-1} cm^{-1} [25]. The overall yield was 18 mg of purified protein per liter of medium. EntB was primed with P-pant using the phosphopantetheinyl transferase EntD as described previously [9].

Cloning of the peptide synthetase fragments *grsA*, *proCAT* and *proCAT(H147V)* and purification of the expressed His_6-tagged proteins was carried out as described previously [11]. All proteins were purified to apparent homogeneity as judged by SDS–PAGE using single-step Ni^{2+} affinity chromatography. Fractions containing the recombinant protein were pooled and dialyzed against assay buffer (20 mM MES, 100 mM NaCl, 10 mM $MgCl_2$ and 1 mM EDTA, pH 8.0) and stored at −80°C. Protein concentrations were determined using the calculated extinction coefficients at 280 nm: 138 690 M^{-1} cm^{-1} for PheATE and 92 230 M^{-1} cm^{-1} for ProCAT and ProCAT(H147V). Priming of the expressed proteins with P-pant was achieved using the phosphopantetheinyl transferase Sfp as described previously [26].

HPLC assay for EntF condensation domain activity

Reactions (50 μl) contained 75 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM tris-(2-carboxyethyl)phosphine, 500 μM coenzyme A, 550 nM EntD, 1 μM EntB ArCP and were preincubated for 30 min at 37°C to allow phosphopantetheinylation of EntB ArCP [23]. Then were added 650 nM EntE, 2 mM DHB, 5 mM ATP, EntF to the indicated concentration, and 2 mM aminoacyl-SNAC. All aminoacyl-SNACs were initially dissolved in dilute TFA, pH 3, and stored at −80°C to minimize hydrolysis. After incubation at 37°C, reactions were quenched with 50 μl 0.1 N KOH, heated to 60°C for 4 min, cooled on ice and extracted with 0.5 ml ethyl acetate. After removal of ethyl acetate by centrifugation under reduced pressure, samples were dissolved in 200 μl 0.1% TFA/water, filtered and analyzed by HPLC. Analytical HPLC was carried out on a Beckman Gold Nouveau system with a Vydac protein and peptide C18 column using a gradient of 10–40% acetonitrile in 0.1% TFA/water over 22 min at a flow rate of 1 ml/min. Samples were monitored at 254 nm and absorbance area correlated with a synthetic DHB-Ser standard [9].

Radioassay for ProCAT condensation domain activity

Reaction mixtures (400 μl) in assay buffer containing 2 mM ATP, 0.5 μM holo-ProCAT and 1.3 μM L-[5-^{3}H]Pro (30 mCi/mmol) were incubated at 37°C for 3 min. The condensation reaction was then initiated by adding either D- or L-Phe-SNAC to a final concentration of 5 mM. At various time points 50 μl samples were quenched by adding 0.5 ml 10% TCA (w/v). The reaction mixtures were extracted with 0.5 ml 4:1 (v/v) butanol/

chloroform, the organic layers washed once with 0.5 ml 0.1 M NaCl and the amount of extractable label (DKP) quantified by liquid scintillation counting. For determination of kinetic constants, the assay conditions were the same except for the following minor modifications: the reaction volume was 100 μl, the reaction time was 8 min and the concentration of aminoacyl-SNAC was 0–20 mM. All reactions were performed in triplicate.

Acknowledgements

This work was supported by NIH Grant GM 20011 to C.T.W. D.E.E. is a National Science Foundation Graduate Research Fellow. J.W.T. is a NIH postdoctoral fellow. T.S. is a fellow of the European Molecular Biology Organization.

References

1. Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chem. Biol.* **4**, 561–567.
2. Konz, D. & Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? *Chem. Biol.* **6**, R39–R48.
3. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2673.
4. Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R. & Morris, H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.* **271**, 15248–15435.
5. Kallow, W., von Dohren, H. & Kleinkauf, H. (1998). Penicillin biosynthesis: energy requirement for tripeptide precursor formation by delta-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Acremonium chrysogenum*. *Biochemistry* **37**, 5947–5952.
6. van Wageningen, A.M.A., Kirkpatrick, P.N., Williams, D.H., Harris, B.R., Kershaw, J.K., Lennard, N.J., Jones, S.J. & Solenberg, P.J. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* **5**, 155–162.
7. Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., Walsh, C.T. & Perry, R. (1998). Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573–586.
8. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. (1998). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631–645.
9. Gehring, A.M., Mori, I. & Walsh, C.T. (1998). Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648–2659.
10. Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C. & Walsh, C.T. (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936.
11. Stachelhaus, T., Mootz, H.D., Bergendahl, V. & Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* **273**, 22773–22781.
12. Doekel, S. & Marahiel, M.A. (2000). Dipeptide formation on engineered hybrid peptide synthetases. *Chem. Biol.* **7**, 373–384.
13. Belshaw, P.J., Walsh, C.T. & Stachelhaus, T. (1999). Aminoacyl-CoA's as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489.
14. Stachelhaus, T., Mootz, H.D. & Marahiel, M. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505.
15. Holzbaur, I.E., Harris, R.C., Bycroft, M., Cortes, J., Bisang, C., Staunton, J., Rudd, B.A. & Leadlay, P.F. (1999). Molecular basis of Celmer's rules: the role of two ketoreductase domains in the control of chirality by the erythromycin modular polyketide synthase. *Chem. Biol.* **6**, 189–195.
16. Gokhale, R.S., Hunziger, D., Cane, D.E. & Khosla, C. (1999). Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **6**, 117–125.
17. Riddles, P.W., Blakeley, R.L. & Zerner, B. (1983). Reassessment of Ellman's reagent. *Methods Enzymol.* **91**, 49–60.
18. Reichert, J., Sakaitani, M. & Walsh, C.T. (1992). Characterization of EntF as a serine-activating enzyme. *Protein Sci.* **1**, 549–556.
19. Stachelhaus, T. & Walsh, C.T. (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**, 5775–5787.
20. Stindl, A. & Keller, U. (1993). The initiation of peptide bond formation in the biosynthesis of actinomycin. *J. Biol. Chem.* **268**, 10612–10620.
21. Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. & Walsh, C.T. (2000). Peptide cyclization catalyzed by the thioesterase domain of tyrocidine synthetase. *Nature*, in press.
22. Gehring, A.M., Mori, I., Perry, R. & Walsh, C.T. (1998). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637–11650.
23. Gehring, A.M., Bradley, K.A. & Walsh, C.T. (1997). Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2, 3-dihydroxybenzoate. *Biochemistry* **36**, 8495–8503.
24. Ehmann, D.E., Shaw-Reid, C.A., Losey, H.C. & Walsh, C.T. (2000). The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl. Acad. Sci. USA* **97**, 2509–2514.
25. Gill, S.C. & von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
26. Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. & Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595.