Substrate Specificity of the Adenylation Enzyme SgcC1 Involved in the Biosynthesis of the Enediyne Antitumor Antibiotic C-1027*

Received for publication, June 20, 2006, and in revised form, August 1, 2006 Published, JBC Papers in Press, August 1, 2006, DOI 10.1074/jbc.M605887200

Steven G. Van Lanen1, Shuangjun Lin2, Pieter C. Dorrestein3, Neil L. Kelleher4, and Ben Shen1,4,6

From the 1Division of Pharmaceutical Sciences, 2University of Wisconsin National Cooperative Drug Discovery Group, 3Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53705 and 4Department of Chemistry, University of Illinois, Urbana, Illinois 61801

C-1027 is an enediyne antitumor antibiotic composed of a chromophore with four distinct chemical moieties, including an (S)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety that is derived from L-α-tyrosine. SgcC4, a novel aminomutase requiring no added co-factor that catalyzes the formation of the first intermediate (S)-β-tyrosine and subsequently SgcC1 homologous to adenylation domains of nonribosomal peptide synthetases, was identified as specific for the SgcC4 product and did not recognize any α-amino acids. To definitively establish the substrate for SgcC1, a full kinetic characterization of the enzyme was performed using amino acid-dependent ATP-[32P]PPi exchange assay to monitor amino acid activation and electrospray ionization-Fourier transform mass spectroscopy to follow the loading of the activated β-amino acid substrate to the peptidyl carrier protein SgcC2. The data establish (S)-β-tyrosine as the preferred substrate, although SgcC1 shows promiscuous activity toward aromatic β-amino acids such as β-phenylalanine, 3-chloro-β-tyrosine, and 3-hydroxy-β-tyrosine, but all were <50-fold efficient. A putative active site mutant P571A adjacent to the invariant aspartic acid residue of all α-amino acid-specific adenylation domains known to date was prepared as a preliminary attempt to probe the substrate specificity of SgcC1; however, the mutation resulted in a loss of activity with all substrates except (S)-β-tyrosine, which was 142-fold less efficient relative to the wild-type enzyme. In total, SgcC1 is now confirmed to catalyze the second step in the biosynthesis of the (S)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety of C-1027, presenting downstream enzymes with an (S)-β-tyrosyl-S-SgcC2 thioester substrate, and represents the first β-amino acid-specific adenylation enzyme characterized biochemically.

C-1027 is an enediyne antitumor antibiotic isolated from the fermentation broth of Streptomyces globisporus (1, 2). It is produced as a chromoprotein complex consisting of a binding protein (annotated as CagA) and the reactive C-1027 chromophore containing a conjugated enediyne harbored within a nine-membered cyclic ring (see Fig. 1). Upon release from CagA, the enediyne core of the C-1027 chromophore readily undergoes a Bergman cycloaromatization to yield a transient biradical species capable of extracting hydrogen atoms from DNA, which in the presence of molecular oxygen can ultimately lead to single- and double-stranded DNA breaks (3). C-1027 and the entire enediyne family share this mode of action, and as a family, their potent cytotoxicity rivals that of any previously discovered natural product (4).

The biosynthetic gene cluster for C-1027 was previously cloned and sequenced, and analysis of the open reading frames (ORFs)4 provided genetic evidence to adequately propose a mechanism of biosynthesis, transport, resistance, and regulation of C-1027 (5). Comparison of the gene cluster to other members of the enediyne family, including calicheamicin (6) and neocarzinostatin (7), revealed a unified biosynthetic approach among the enediynes and supported a convergent biosynthesis of C-1027 from four components, a deoxy aminosugar, benzoxazolin, β-amino acid, and enediyne core starting from glucose-1-phosphate, chorismic acid, L-α-tyrosine, and acyl-CoAs, respectively (Fig. 1). The pathway to generating the C-1027 deoxy aminosugar was initially targeted for biochemical analysis, and the first reaction catalyzed by SgcA1 has been characterized as a β-α-glucopyranosyl-1-phosphate thymidylyltransferase (8). Of the remaining moieties, only the initial step for the β-amino acid moiety (S)-3-chloro-4,5-dihydroxy-β-phenylalanine has been identified, a novel aminomutase reaction catalyzed by SgcC4 converting L-α-tyrosine to (S)-β-tyrosine (9).

Among the 56 ORFs identified within the boundaries of the C-1027 biosynthetic gene cluster, three ORFs exhibit high homology to nonribosomal peptide synthetase (NRPS) domains. NRPSs are typically large modular proteins that catalyze the synthesis of a wide range of biologically active peptides

---

4 The abbreviations used are: ORF, open reading frame; NRPS, nonribosomal peptide synthetase; PCP, peptidyl carrier protein; ESI-M5, electrospray ionization-mass spectrometry; ESI-FTMS, electrospray ionization-Fourier transform mass spectrometry; PheA, the excised L-Phe-specific adenylation domain of GrsA; ACES, 2-[(L-amino-2-oxoethyl)amino]ethanesulfonic acid.

---
A minimal NRPS module consists of three domains, an adenylation domain that sequentially selects, activates, and loads an amino acid (or carboxylic acid) to the 4'-phosphopantetheine prosthetic group of a peptidyl carrier protein (PCP) domain and a condensation domain that catalyzes the formation of an amide bond. In most examples, a single NRPS protein contains multiple modules, and the primary sequence of the resulting peptide product is correlated to the linear architecture of the modules (hence the so-called “thiotemplate co-linearity” rule). Extra domains embedded in the multimodular NRPS proteins, such as epimerization, methyltransferase, or oxidation domains, can be present to afford final peptides with enriched chemistry. As opposed to large multimodular proteins, singular domains or didomains of NRPS as individual ORFs have also been recently characterized, and this phenomenon has facilitated research toward understanding domain-domain recognition and interaction in NRPS and the molecular mechanism of NRPS-catalyzed nonribosomal peptide biosynthesis.

Within the C-1027 biosynthetic gene cluster, two ORFs upstream and in opposite orientation to sgcC4 were found to encode proteins with sequence similarities to condensation (SgcC5) and PCP (SgcC2) domains (Fig. 2A). An ORF located adjacent to and downstream of sgcC4 was also identified and encodes an 881-amino-acid protein, SgcC1, having sequence homology to adenyltransfer domains. Although SgcC1 has closest homology to adenyltransfer domains that activate L-α-tirosine, preliminary characterization has revealed that SgcC1 recognizes S-β-tirosine as opposed to any of the standard α-amino acids, including either isomer of α-tirosine or α-phenylalanine or analogs (11). Although these results support the findings that SgcC4 is the initial enzyme in the pathway (catalyzing the stereospecific formation of S-β-tirosine from L-α-tirosine) and SgcC1 catalyzes the subsequent step (Fig. 2B), they contradicted the prediction on the basis of the so-called “nonribosomal codes” (12, 13) that SgcC1 would be an α-amino acid-specific adenyltransfer domain (11).

As a follow-up to the initial characterization of SgcC1, we now provide definitive experimental data to assign the substrate specificity for SgcC1. Using amino acid-dependent ATP-[32P]PP exchange assays and electrospray ionization-Fourier transform mass spectrometry (ESI-FTMS), chlorinated and hydroxylated tiroside or phenylalanine analogs were tested revealing that SgcC1, following the SgcC4 aminomutase reaction, catalyzes the second step in the biosynthesis of the S-3-chloro-4,5-dihydroxy-β-phenylalanine moiety from 1-α-tirosine by forming the S-β-tirosine intermediate. Subsequent steps involve halogenation (SgcC3), hydroxylation (SgcC), and incorporation of the fully modified β-amino acid unit into the enediyne core (SgcC5), although the precise timing of the last steps awaits further validation (Fig. 2B). SgcC1 as a naturally occurring, discrete protein could therefore provide a relatively simple platform to address the inherent amino acid specificity-conferring elements of adenyltransfer domains to β-amino acids, of which the current nonribosomal codes were based solely on a single structure of the excised, 1-α-phenylalanine-specific adenylase.
tion domain of GrsA (PheA) (12, 13). As an initial attempt to determine the structural elements of SgcC1 necessary for proper β-amino acid selection, truncated protein and a single point mutation of SgcC1 were generated and the results reported.

**EXPERIMENTAL PROCEDURES**

**Chemicals and Instrumentation**—If not mentioned, chemicals and instruments used were identical to those previously reported (11). Cinnamic acid, p-hydroxybenzoinic acid, and 3-amino-3-phenylpropionic acid (β-phenylalanine) were from Fisher-Acrros (Pittsburgh, PA). (R)-3-Amino-(4-hydroxy-phenyl)-propionic acid ((R)-β-tyrosine) was from PepTech Corporation (Burlington, MA). Restriction enzymes were from New England Biolabs (Beverly, MA) or Invitrogen, and expression vectors were from Novagen (Madison, WI). Unless explicitly stated, the compounds used were racemic mixtures. Protein analysis was performed with position-specific iterated-BLAST and BL2SEQ using the San Diego Supercomputer Center Biology WorkBench, version 3.2, and the adenyl domain active sites were extracted using the NRPS Predictor software from Universität Tübingen.

**Synthesis of Racemic 3-Chloro-β-tyrosine**—Synthesis of 3-chloro-β-tyrosine was achieved by following the method reported by Weaver and co-worker (14). 3-Chlorobenzaldehyde (76 mg) was refluxed with 1 equivalent of malonic acid (51 mg) and 2 equivalents of ammonium acetate (76 mg) in ethanol (5 ml) for 2 days. The reaction mixture was adjusted to an approximate pH of 4 with 1 N HCl and separated by a strong acid cation exchange column (Dowex® 50WX8). The product, 3-chloro-β-tyrosine, was eluted with 1% ammonium hydroxide. The residue was dissolved in distilled water after the solvent was removed through evaporation under reduced pressure and further purified on C-18 reverse phase column (Alltima 10 μm, 5 μm; Grace Davison Discovery Sciences). 1H NMR assignments in D2O containing CF3CO2D at 400 MHz are as follows: δ 6.74 (d, J = 6.8 Hz, H2, 1H), 4.60 (t, J = 8.4 Hz, Hβ, 1H), 2.79 (dd, J = 1.5 Hz, 8.0 Hz, Hα, 1H), and 2.69 (dd, J = 16.0, 6.4 Hz, Hα, 1H). 13C NMR assignments in D2O at 125 MHz are as follows: δ 139.2 (ArC), 128.7 (ArC), 145.6 (ArC), 145.5 (ArC), and 177.9 (CO). ESI-MS gave m/z: [M + H]+, 198.0; [M – H]−, 195.9; calculated 197.0 for C9H10NO3Cl.

**Synthesis of Racemic 3-Chloro-5-hydroxy-β-tyrosine**—3-Chloro-5-hydroxy-β-tyrosine was synthesized by the same procedure as that used for the synthesis of 3-chloro-β-tyrosine beginning from 3-chloro-4,5-dihydroxybenzaldehyde, which was prepared from 3-chloro-4-hydroxyl-5-methoxybenzaldehyde following the method reported by Perchellet and co-workers (15). 1H NMR assignments in D2O at 500 MHz are as follows: δ 7.04 (d, J = 2.5 Hz, 1H), 6.92 (d, J = 2.5 Hz, 1H), 4.60 (t, J = 7.5 Hz, 1H), 3.04 (dd, J = 17.0, 8.0 Hz, 1H), and 2.95 (dd, J = 17.0, 6.5 Hz, 1H). 13C NMR assignments in D2O at 125 MHz are as follows: δ 38.8 (Ca), 51.6 (Cβ), 113.6, 120.2, 121.8, 128.6, 141.8, 146.1 (ArC), and 175.1 (CO). ESI-MS gave m/z: [M + H]+, 232.1; calculated 231.0 for C9H11NO4Cl.

**DNA Manipulations**—Cloning and construction of the sgcC1 (pBS1033) and sgcC2 (pBS1034) expression vectors were previously described (11). A truncated form of SgcC1 was prepared by PCR amplification using pBS1005 (5) as a template and a forward primer of 5′-GGGAATTCATATGGCGCTCTGGCGCTGGG-3′ (Ndel site underlined) and a reverse primer of 5′-GGGAAACCTTGGGGTGGCGGAGGGCC-3′ (HindIII site underlined). The amplified 1629-base-pair fragment was cloned into the same sites of pET29a to yield pBS1037. After confirmation of the DNA sequence fidelity, the Ndel/HindIII fragment was isolated and cloned into pET28a to yield pBS1038. Although the former construct produces SgcC1 as a C-terminal His6-tagged protein, the latter construct produces SgcC1 as a C- and N-terminal His6-tagged protein, both
The SgcC1 (S)-β-Tyrosine-specific Adenylation Enzyme

of which have the 338 amino acids at the N terminus of SgcC1 deleted.

A P571A point mutation of SgcC1 was generated by PCR amplification of the template pbS1033 (11) using the Expand long template PCR system (Roche Applied Science). Reactions were performed using the manufacturer’s provided Buffer 2 with 5% Me2SO, primers of 5'-GTCCTCCCCGAGCAGCAGACGGCGGCTGGCCAGGTGCT-3' and the reverse complement (with the Ala codon underlined), and a PCR program consisting of an initial hold at 94 °C for 2 min followed by 20 cycles of 94 °C for 10 s, 56 °C for 30 s, and 68 °C for 7 min. The template DNA was digested with 10 units of DpnI for 1 h at 37 °C followed by heating to 90 °C for 5 min and cooling at room temperature before transformation. The introduction of the correct point mutation and the fidelity of the entire gene including 250 bp upstream and downstream were confirmed by DNA sequencing to yield pbS1039.

Amino Acid-dependent ATP-[32P]PPi Exchange Assays—Assessment of the SgcC1 adenylation enzyme substrate specificity was performed as described previously (11). A pH profile for SgcC1 was generated using a three-buffer system of 52/52/100 mm Tris-ACES-ethanolamine as described previously (16). Assays were run at 30 °C in 1× buffer with 5 mM MgCl2, 0.1 mM EDTA, 5 mM ATP, 0.9 μM [32P]PPi, 1 mM (S)-β-tyrosine, and 10 mM SgcC1 under initial velocity conditions (<10% isotopic conversion into ATP). Each data point represents a minimum of four duplicate end-point assays. Because the pH profile was utilized only for determining the conditions for optimal activity, the data were fitted to a smooth curve in contrast to a specific rate equation, which would warrant further kinetic analysis at varied pH.

The steady-state kinetic parameters for SgcC1 were determined with activity assays carried out at 30 °C in 100 mM Tris-HCl (pH 9.0), 5 mM MgCl2, 0.1 mM EDTA, 5 mM ATP, 1.0 μM [32P]PPi, and varied co-substrate as follows: 25–1600 μM for 3-chloro-β-tyrosine, 3-hydroxy-β-tyrosine, and β-phenylalanine, 5–370 μM for (R)-β-tyrosine, and 0.5–200 μM for (S)-β-tyrosine. Enzyme concentration for each substrate ranged from 1.9×10–9 to 1.9×10–6 M for (S)-β-tyrosine but not (R)-β-tyrosine (Fig. 2A).

Enzyme concentration for each substrate ranged from 1.9×10–9 to 1.9×10–6 M for (S)-β-tyrosine but not (R)-β-tyrosine (Fig. 2A).

RESULTS

Organization of NRPS within C-1027 Biosynthetic Gene Cluster—The gene cluster for C-1027 biosynthesis contains three ORFs with sequence homology to condensation (SgcC5), adenylation (SgcC1), and PCP (SgcC2) domains found in modular NRPSs. The ORFs are located in close proximity, interrupted only by sgcC4, of which the gene product has already been characterized and was identified to catalyze the initial step in the biosynthesis of the β-amino acid moiety of C-1027 by forming (S)-β-tyrosine (Fig. 2A). Therefore, these three ORFs, which constitute a minimal NRPS module, were hypothesized to process the SgcC4 product to maturation and its subsequent incorporation into the C-1027 chromophore (Fig. 2B).

Substrate Specificity of SgcC1 Using Amino Acid-dependent ATP-PPi Exchange Assay—SgcC1 consists of 881 amino acids, of which 550 residues of the C terminus are similar to AMP-forming adenylation domains of NRPS. SgcC1 contains the 10 core regions that are conserved in all NRPS adenylation domains (10) and overall has closest homology to TycC3 (38/55% identity/similarity from BL2SEQ analysis), NovH (35/46), and SimH (33/43), all β-aminoadipic acid-specific adenylation domains exemplified by Tcp12 (Q70722); δ-glutamic acid-specific adenylation domains that activate the δ-carboxylic acid exemplified by Myca (Q9NR34), and dihydroxyphenylglycine-specific adenylation domains exemplified by Tep12 (Q07AZ6). Given in parentheses in the figure legend are GenBank® accession numbers. The "variable" Aas-235 and Lys-517 (PheA numbering) that interact with the α-amino and carboxylate groups, thereby locking the orientation of the side chain of the α-amino acid in the substrate binding pocket of adenylation domains, are highlighted in bold, and the residues subjected to point mutation are boxed. Aad, δ-aminoadipic acid; dhpg, dihydroxyphenylglycine.

Activity of SgcC1 with SgcC2—Preparation of recombinant SgcC1 and SgcC2 proteins, including overproduction in Escherichia coli, purification, and in vitro 4'-phosphopantetheinylation of apoSgcC2, were as described previously (11). Acylation with (S)-β-tyrosine was analyzed as described previously (11). Briefly, a quenched sample was loaded onto a Jupiter C4 column (Phenomenex, La Jolla, CA) to purify and desalt SgcC2 species. Mass spectrometric analysis was performed with a custom 8.5-T ESI-FTMS spectrometer equipped with a front-end quadrupole. The sample was introduced using a Nanomate 100 for automated nanospray (Advion Biosciences, Ithaca, NY) and typically 500 ms of ion accumulation/scan was used. The instrument was externally calibrated using ubiquitin (Sigma).
The SgcC1 (S)-β-Tyrosine-specific Adenylation Enzyme

FIGURE 4. Substrate specificity of SgcC1. The data are grouped into four different categories: α-amino acids where $R_i$ = respective side chain (A); products of aromatic α-amino acids upon subjecting to ammonia lyase chemistry, cinnamic acid ($R_i = H$) and p-hydroxycinnamic acid ($R_i = OH$) (B); α-tyrosine analogs that could be viewed conceptually as possible biosynthetic intermediates, 3-chloro-α-tyrosine ($R_i = Cl$) and 3-hydroxy-α-tyrosine ($R_i = OH$) (C); and β-tyrosine analogs that could be viewed conceptually as possible biosynthetic intermediates, β-phenylalanine ($R_4 = H$, $R_5 = H$, $R_6 = H$), (S)-β-tyrosine and (R)-β-tyrosine ($R_4 = OH$, $R_5 = H$, $R_6 = H$), 3-hydroxy-β-tyrosine ($R_4 = OH$, $R_5 = OH$, $R_6 = H$), 3-chloro-β-tyrosine ($R_4 = OH$, $R_5 = Cl$, $R_6 = H$), and 3-chloro-5-hydroxy-β-tyrosine ($R_4 = OH$, $R_5 = Cl$, $R_6 = OH$) (D).

FIGURE 5. Effect of pH on the relative activity of SgcC1 as an adenylation enzyme as determined with (S)-β-tyrosine as a substrate by the ATP-[32P]PPi exchange assay.

α-amino acids or β-tyrosine analogs are recognized by SgcC1 (Fig. 4). Among the possible substrates that had little or no activity were cinnamic acid and p-hydroxycinnamic acid, products of ammonia lyase chemistry of the corresponding α-amino acids (20), and 3-chloro-α-tyrosine or 3-hydroxy-α-tyrosine, possible intermediates assuming halogenation or hydroxylation, respectively, were the initial steps in biosynthesis of the β-amino acid moiety. The highest activity was observed with (S)-β-tyrosine followed closely by the R enantiomer (95% relative to (S)-β-tyrosine). Other substrates that were recognized include β-phenylalanine (75%), 3-hydroxy-β-tyrosine (62%), 3-chloro-β-tyrosine (51%), and 3-chloro-5-hydroxy-β-tyrosine (9%).

pH Profile of SgcC1—The optimal pH for SgcC1 activity was analyzed using amino acid-dependent ATP-PPi exchange assay and (S)-β-tyrosine as the substrate. A three-buffer system was employed to maintain a constant ionic strength (16). SgcC1 was active between pH 5 and 11, and the pH profile exhibited two maxima at pH values 7.5 and 9.0 (Fig. 5). When 100 mM Tris-HCl was substituted, no change in activity was observed (data not shown). Because SgcC4 had a pH optimum of ~9.0, further experiments were performed in 100 mM Tris-HCl at pH 9.0.

Single-substrate Kinetics of SgcC1—Kinetic parameters were determined using the five substrates that had significant activity when tested with SgcC1. All of the substrates displayed typical Michaelis-Menten kinetics in the presence of saturating ATP, as exemplified by (S)- or (R)-β-tyrosine (Fig. 6). The best substrate, (S)-β-tyrosine, had a $k_{cat}$ of 2.2 ± 0.5 s$^{-1}$ and $K_m$ of 3.2 ± 0.6 μM, whereas the other substrates had a similar but slightly lower $k_{cat}$ and significantly increased $K_m$ (Table 1). Overall, SgcC1 catalysis was ~50-fold more efficient with (S)-β-tyrosine than any substrate analog tested, excluding the R-enantiomer, which had a 25-fold lower efficiency.

SgcC1-catalyzed Loading of SgcC2—The second half-reaction of SgcC1 was monitored by ESI-FTMS following the mass shift observed upon loading the activated β-tyrosine onto holo-SgcC2 PCP to generate a β-tyrosyl-S-SgcC2 thioester. With either (S)- or (R)-β-tyrosine as the substrate, the appropriate mass shift of $+162.6$ was observed (11). Time course analysis with (S)-β-tyrosine resulted in a rate constant of 2.1 s$^{-1}$ and with a (R)-β-tyrosine rate constant of 1.9 s$^{-1}$, both of which are nearly identical to the $k_{cat}$ calculated from the first half-reaction as determined by the ATP-[32P]PPi exchange assays (Table 1).
The SgcC1 (S)-β-Tyrosine-specific Adenylation Enzyme

**FIGURE 6.** Kinetic characterization of SgcC1 as an adenylation enzyme with (S)-β-tyrosine (A) and (R)-β-tyrosine (B) as a substrate by plotting the initial velocities of ATP-[32P]PP exchange as a function of β-tyrosine concentration.

**TABLE 1**
Michaelis-Menten constants for SgcC1 as an adenylation enzyme toward selected β-amino acids

| Substrate   | $K_m$ (μM) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (μM$^{-1}$ s$^{-1}$) | Relative efficiency |
|-------------|------------|----------------------|----------------------------------|--------------------|
| (S)-β-Tyr   | 3.2 ± 0.6  | 2.2 ± 0.5            | 0.69                             | 1                  |
| (R)-β-Tyr   | 53 ± 12    | 1.5 ± 0.3            | 0.028                            | 0.04               |
| 3-OH-β-Tyr  | 81 ± 16    | 1.2 ± 0.1            | 0.015                            | 0.02               |
| β-Phe       | 1.9 × 10^5 | 1.4 ± 0.4            | 0.0075                           | 0.01               |
| 3-Cl-β-Tyr  | 3.0 × 10^5 | 1.6 ± 0.3            | 0.0053                           | 0.008              |

**TABLE 2**
Michaelis-Menten constants for SgcC1(P571A) as an adenylation enzyme toward (S)-β-tyrosine and comparison with that of the wild-type SgcC1 enzyme

| Enzyme                  | $K_m$ (μM) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (μM$^{-1}$ s$^{-1}$) | Relative efficiency |
|-------------------------|------------|----------------------|----------------------------------|--------------------|
| SgcC1 (wild-type)       | 3.2 ± 0.6  | 2.2 ± 0.5            | 0.69                             | 1                  |
| SgcC1 (P576A)           | 13 ± 4     | 7.0 × 10^{-2} ± 0.02 | 0.0054                           | 0.007              |

**DISCUSSION**

Adenylation domains of NRPS catalyze the activation, at the expense of ATP, and subsequent loading of the activated amino acids or carboxylic acids to a PCP domain to form an aminoceryl-S-PCP thioester. SgcC1 has significant sequence homology to numerous adenylation domains of NRPS, and as a result, has been hypothesized to be involved in the formation, activation, and incorporation of the (S)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety into C-1027, although the precise timing of the steps cannot be predicted a priori (Fig. 2). Sequence analysis predicted L-α-tyrosine as the probable substrate for SgcC1, as deduced from its closest sequence identity to domains that activate L-α-tyrosine, such as TycC3, NovH, and SimH, and comparisons to the nonribosomal codes of NRPS suggested that L-α-tyrosine was the most likely substrate for SgcC1 (Fig. 3) (12, 13). However, our previous results with SgcC4 preferring L-α-tyrosine were inconsistent with this proposal (11), and therefore, a full biochemical characterization of SgcC1 was undertaken.

As previously reported, when L-α-tyrosine was tested as a substrate, SgcC1 was only active in the presence of SgcC4, which converts L-α-tyrosine to (S)-β-tyrosine in situ (11). In contrast, authentic (S)-β-tyrosine, the established product of SgcC4 (9), was recognized and loaded to SgcC2 to form an (S)-β-aminoacyl-S-SgcC2 thioester intermediate. When a full spectrum of substrates was tested, including substituted α- and β-phenylalanine analogs that could hypothetically be generated during biosynthesis of the (S)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety of C-1027, only β-amino acids were activated, all with relatively high rates under the conditions employed, prompting a thorough kinetic analysis. Although the five aromatic β-amino acids examined were all recognized and activated by SgcC1, (S)-β-tyrosine was clearly the preferred substrate by >50-fold. The enantiomer (R)-β-tyrosine is also a substrate for SgcC1 but is 25-fold less efficient, and this non-specific activity is likely to be further compensated by the stereospecificities of SgcC4 and downstream enzymes such as SgcC5. Interestingly, condensation domains have recently been shown to contribute to substrate specificity by discriminating between isomers (21).

Rational engineering of the adenylation domain specificity without adverse effect on the structural integrity of the NRPS biosynthetic machinery could be viewed as one of the most powerful combinatorial biosynthesis strategies for natural product diversity. Central to this strategy is the development of the general rule, known as nonribosomal codes, for substrate recognition in the adenylation domain of an NRPS, which has provided a molecular basis for engineered biosynthesis of novel peptides (12, 13, 22). Based on the structure of PheA (23), 10
Amino acid residues that line the amino acid binding pocket were identified as the general specificity-conferring code of adenylation domains for NRPSs (Fig. 3). According to this model, the Asp residue at position 235 and the Lys residue at 517 interact with the α-amino and the carboxylate groups, respectively, to lock orientation of the L-α-amino acid upon activation. This configuration projects the side chain of the amino acid into the binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).

An important feature of NRPSs is their ability to incorporate nonproteinogenic amino acids, such as β-amino acids, and their presence in the resultant metabolites contributes critically to their biological activities and therapeutic value. Because the current model is based on the structure of PheA, the prediction model is severely limited to L-α-amino acids and cannot be used to predict adenylation domain specificity for β-amino acids or several nonproteinogenic substrates, given the key interaction between the α-amino group and Asp-235, which is fundamental to this model, as this interaction locks the configuration of the amino acid. In β-amino acids, this fundamental interaction is no longer applicable, and thus, the predictive model breaks down. The β-amino acid specificity of SgcC1 reported here unveiled the limitation of the current model, providing an outstanding opportunity to decipher the nonribosomal codes for β-amino acids.

Of the common amino acids, Pro has the uncanny ability to modulate stability, structure, and function of a given protein (24). The presence of Pro generates kinks in α-helix structure, and in particular, the torsional angle inherent in Pro is highly coupled to the preceding residue in the case of SgcC1, the invariant Asp residue, the carboxylate side chain of which is critical in developing the current nonribosomal codes for α-amino acids (12, 13). A survey of L-tyrosine or L-phenylalanine-specific adenylation domains showed that an Ala (or Gly in a couple of cases) occupies this position, and 32 of 1250 α-amino acid residues that line the amino acid binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).

An important feature of NRPSs is their ability to incorporate nonproteinogenic amino acids, such as β-amino acids, and their presence in the resultant metabolites contributes critically to their biological activities and therapeutic value. Because the current model is based on the structure of PheA, the prediction model is severely limited to L-α-amino acids and cannot be used to predict adenylation domain specificity for β-amino acids or several nonproteinogenic substrates, given the key interaction between the α-amino group and Asp-235, which is fundamental to this model, as this interaction locks the configuration of the amino acid. In β-amino acids, this fundamental interaction is no longer applicable, and thus, the predictive model breaks down. The β-amino acid specificity of SgcC1 reported here unveiled the limitation of the current model, providing an outstanding opportunity to decipher the nonribosomal codes for β-amino acids.

Of the common amino acids, Pro has the uncanny ability to modulate stability, structure, and function of a given protein (24). The presence of Pro generates kinks in α-helix structure, and in particular, the torsional angle inherent in Pro is highly coupled to the preceding residue in the case of SgcC1, the invariant Asp residue, the carboxylate side chain of which is critical in developing the current nonribosomal codes for α-amino acids (12, 13). A survey of L-tyrosine or L-phenylalanine-specific adenylation domains showed that an Ala (or Gly in a couple of cases) occupies this position, and 32 of 1250 α-amino acid residues that line the amino acid binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).

An important feature of NRPSs is their ability to incorporate nonproteinogenic amino acids, such as β-amino acids, and their presence in the resultant metabolites contributes critically to their biological activities and therapeutic value. Because the current model is based on the structure of PheA, the prediction model is severely limited to L-α-amino acids and cannot be used to predict adenylation domain specificity for β-amino acids or several nonproteinogenic substrates, given the key interaction between the α-amino group and Asp-235, which is fundamental to this model, as this interaction locks the configuration of the amino acid. In β-amino acids, this fundamental interaction is no longer applicable, and thus, the predictive model breaks down. The β-amino acid specificity of SgcC1 reported here unveiled the limitation of the current model, providing an outstanding opportunity to decipher the nonribosomal codes for β-amino acids.

Of the common amino acids, Pro has the uncanny ability to modulate stability, structure, and function of a given protein (24). The presence of Pro generates kinks in α-helix structure, and in particular, the torsional angle inherent in Pro is highly coupled to the preceding residue in the case of SgcC1, the invariant Asp residue, the carboxylate side chain of which is critical in developing the current nonribosomal codes for α-amino acids (12, 13). A survey of L-tyrosine or L-phenylalanine-specific adenylation domains showed that an Ala (or Gly in a couple of cases) occupies this position, and 32 of 1250 α-amino acid residues that line the amino acid binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).

An important feature of NRPSs is their ability to incorporate nonproteinogenic amino acids, such as β-amino acids, and their presence in the resultant metabolites contributes critically to their biological activities and therapeutic value. Because the current model is based on the structure of PheA, the prediction model is severely limited to L-α-amino acids and cannot be used to predict adenylation domain specificity for β-amino acids or several nonproteinogenic substrates, given the key interaction between the α-amino group and Asp-235, which is fundamental to this model, as this interaction locks the configuration of the amino acid. In β-amino acids, this fundamental interaction is no longer applicable, and thus, the predictive model breaks down. The β-amino acid specificity of SgcC1 reported here unveiled the limitation of the current model, providing an outstanding opportunity to decipher the nonribosomal codes for β-amino acids.

Of the common amino acids, Pro has the uncanny ability to modulate stability, structure, and function of a given protein (24). The presence of Pro generates kinks in α-helix structure, and in particular, the torsional angle inherent in Pro is highly coupled to the preceding residue in the case of SgcC1, the invariant Asp residue, the carboxylate side chain of which is critical in developing the current nonribosomal codes for α-amino acids (12, 13). A survey of L-tyrosine or L-phenylalanine-specific adenylation domains showed that an Ala (or Gly in a couple of cases) occupies this position, and 32 of 1250 α-amino acid residues that line the amino acid binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).

An important feature of NRPSs is their ability to incorporate nonproteinogenic amino acids, such as β-amino acids, and their presence in the resultant metabolites contributes critically to their biological activities and therapeutic value. Because the current model is based on the structure of PheA, the prediction model is severely limited to L-α-amino acids and cannot be used to predict adenylation domain specificity for β-amino acids or several nonproteinogenic substrates, given the key interaction between the α-amino group and Asp-235, which is fundamental to this model, as this interaction locks the configuration of the amino acid. In β-amino acids, this fundamental interaction is no longer applicable, and thus, the predictive model breaks down. The β-amino acid specificity of SgcC1 reported here unveiled the limitation of the current model, providing an outstanding opportunity to decipher the nonribosomal codes for β-amino acids.

Of the common amino acids, Pro has the uncanny ability to modulate stability, structure, and function of a given protein (24). The presence of Pro generates kinks in α-helix structure, and in particular, the torsional angle inherent in Pro is highly coupled to the preceding residue in the case of SgcC1, the invariant Asp residue, the carboxylate side chain of which is critical in developing the current nonribosomal codes for α-amino acids (12, 13). A survey of L-tyrosine or L-phenylalanine-specific adenylation domains showed that an Ala (or Gly in a couple of cases) occupies this position, and 32 of 1250 α-amino acid residues that line the amino acid binding pocket, comprising the 10 amino acid residues within a radius of ~5.5 Å, which in the adenylation domain could then be used to predict the nature of the side chain, and thereby, the amino acid specificity. This model is based on the structural conservation of the binding pocket as reflected by the relatively high sequence similarity of adenylation domains and has been supported by predicting the adenylation domain substrate specificity according to the nonribosomal code followed by experimental verification (12, 13) and re-engineering to alter adenylation domain specificity followed by isolation of the resultant products with the targeted amino acid incorporated (22).
The SgcC1 (S)-β-Tyrosine-specific Adenylation Enzyme

C. M., and Thorson, J. S. (2002) Science 297, 1173–1176
7. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) Chem. Biol. 12, 1–10
8. Murrell, J. M., Liu, W., and Shen, B. (2004) J. Nat. Prod. 67, 206–213
9. Christenson, S. D., Wu, W., Spies, M. A., Shen, B., and Toney, M. D. (2003) Biochemistry 42, 12708–12718
10. Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nat. Prod. Rep. 20, 275–287
11. Van Lanen, S. G., Dorrestein, P. C., Christenson, S. D., Liu, W., Ju, J., Kelleher, N. L., and Shen, B. (2005) J. Am. Chem. Soc. 127, 11594–11595
12. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) Chem. Biol. 6, 493–505
13. Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Chem. Biol. 7, 211–224
14. Tan, C. Y. K., and Weaver, D. F. (2002) Tetrahedron 58, 7449–7461
15. Hua, D. H., Huang, X., Chen, Y., Battina, S. K., Tamura, M. S., Noh, K., Koo, S. I., Namatame, I., Tomoda, H., Perchellet, E. M., and Perchellet, J. P. (2004) J. Org. Chem. 69, 6065–6074
16. Ellis, K. J., and Morrison, J. F. (1982) Methods Enzymol. 87, 405–426
17. Mootz, H. D., and Marahiel, M. A. (1997) J. Bacteriol. 179, 6843–6850
18. Chen, H., and Walsh, C. T. (2001) Chem. Biol. 8, 301–312
19. Galm, U., Schimana, J., Fiedler, H., Schmidt, J., Li, S., and Heide, L. (2002) Arch. Microbiol. 178, 102–114
20. Poppe, L. (2001) Curr. Opin. Struct. Biol. 4, 902–911
21. Belshaw, P. I., Walsh, C. T., and Stachelhaus, T. (1999) Science 284, 486–489
22. Stachelhaus, T., Schneider, A., and Marahiel, M. A. (1995) Science 269, 69–72
23. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) EMBO J. 16, 4174–4183
24. Reiersen, H., and Rees, A. R. (2001) Trends Biochem. Sci. 26, 679–683
25. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D. H. (2005) Nucleic Acids Res. 33, 5799–5808
26. Zhang, J., Wolfe, S., and Demain, A. L. (1992) Biochem. J. 283, 691–698
27. Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B. A. (2000) Chem. Biol. 7, 753–764
28. van Wageningen, A. M. A., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., Jones, M., Jones, S. J. M., and Solenberg, P. J. (1998) Chem. Biol. 5, 155–162
29. Conti, E., Franks, N. P., and Brick, P. (1996) Structure (Lond.) 4, 287–298
30. May, J. J., Kebler, N., Marahiel, M. A., and Stubbs, M. T. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 2509–2514