Generality of Peptide Cyclization Catalyzed by Isolated Thioesterase Domains of Nonribosomal Peptide Synthetases†

Rahul M. Kohli,‡ John W. Trauger,‡ Dirk Schwarzer,§ Mohamed A. Marahiel,§ and Christopher T. Walsh*,‡

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, and Biochemie/Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, 35032 Marburg, Germany

Received January 4, 2001; Revised Manuscript Received March 19, 2001

ABSTRACT: The C-terminal thioesterase (TE) domains from nonribosomal peptide synthetases (NRPSs) catalyze the final step in the biosynthesis of diverse biologically active molecules. In many systems, the thioesterase domain is involved in macrocyclization of a linear precursor presented as an acyl-S-enzyme intermediate. The excised thioesterase domain from the tyrocidine NRPS has been shown to catalyze the cyclization of a peptide thioester substrate which mimics its natural acyl-S-enzyme substrate. In this work we explore the generality of cyclization catalyzed by isolated TE domains. Using synthetic peptide thioester substrates from 6 to 14 residues in length, we show that the excised TE domain from the tyrocidine NRPS can be used to generate an array of sizes of cyclic peptides with comparable kinetic efficiency. We also studied the excised TE domains from the NRPSs which biosynthesize the symmetric cyclic decapeptide gramicidin S and the cyclic lipoheptapeptide surfactin A. Both TE domains exhibit expected cyclization activity: the TE domain from the gramicidin S NRPS catalyzes head-to-tail cyclization of a decapeptide thioester to form gramicidin S, and the TE domain from the surfactin NRPS catalyzes stereospecific cyclization to form a macrolactone analogue of surfactin. With an eye toward generating libraries of cyclic molecules by TE catalysis, we report the solid-phase synthesis and TE-mediated cyclization of a small pool of linear peptide thioesters. These studies provide evidence for the general utility of TE catalysis as a means to synthesize a wide range of macrocyclic compounds.

Nonribosomal peptide synthetases (NRPSs),[1] polyketide synthetases (PKSs), and hybrid NRPS/PKS systems produce an incredibly diverse set of biologically active molecules (Figure 1) (*1*, *2*). Many of these molecules have macrocyclic structures, including the immunosupressant cyclosporin, the antibiotics erythromycin and daptomycin, and the anticancer agent epothilone. Macrocyclic structure decreases the conformational flexibility of a molecule compared to its linear analogue, which can constrain it to a biologically active conformation.

The modular organization of nonribosomal peptide synthetases facilitates biosynthesis of diverse structures. Each NRPS module contains several semiautonomous enzymatic domains (Figure 2A) (*1*−*4*). An adenylation (A) domain activates a specific amino acid as an aminoacyl adenylate, which subsequently becomes covalently tethered by a thioester linkage to a phosphopantethiene containing peptidyl carrier protein (PCP). At this stage, the tethered amino acid may be modified by additional domains, including epimerization, *N*-methylation, and heterocyclization domains. Chain elongation is catalyzed by condensation (C) domains which couple the growing chain from an upstream PCP domain to the downstream aminoacyl-S-PCP, culminating in the generation of the complete linear peptidyl precursor tethered to the most C-terminal PCP domain (peptide-S-PCP).

Release of the final product is usually catalyzed by a C-terminal thioesterase (TE) domain, which catalyzes deacylation of the most C-terminal PCP (Figure 2B). The mechanism of TE catalysis involves transfer of the linear peptide from the terminal PCP domain to the active site serine residue in the TE domain to form a peptide-O-TE intermediate (*5*, *6*). Deacylation of the intermediate involves either hydrolysis to release a linear peptide or, in the case of cyclic products, reaction of an intramolecular nucleophile. For cyclizing TE domains, the TE provides a source of diversity and complexity as a variety of groups can be the nucleophile in the cyclization reaction including the N-terminal amino group (resulting in head-to-tail cyclization) as in the peptide antibiotics tyrocidine A (*7*) and gramicidin S (*8*), a side chain nucleophile (resulting in a branched cyclic molecule) as in the antibiotics bacitracin A (*9*) and daptomycin (*10*), and the β-hydroxyl group of a β-hydroxy fatty acid as in surfactin A (*11*) (Figure 1).

† This work was supported by National Institutes of Health Grant GM-20011 (C.T.W.) and the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie (M.A.M.). J.W.T. was supported by a NIH postdoctoral fellowship.

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

‡ Harvard Medical School.

§ Philipps-Universität Marburg.

[1] Abbreviations: NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; PCP, peptidyl carrier protein domain; TE, thioesterase domain; NAC, *N*-acetylcysteamine; Orn, ornithine; SNAC, *N*-acetylcysteamine thioester; TycC TE, TE domain from the tyrocidine NRPS; GrsB TE, TE domain from the gramicidin S NRPS; SrfA-C TE, TE domain from the surfactin NRPS; NMP, *N*-methyl-2-pyrrolidinone; DIEA, diisopropylethylamine; GLP, gramicidin linear precursor; SLP, surfactin linear precursor; Ac, acetyl.

10.1021/bi010036j CCC: $20.00
Published on Web 05/19/2001

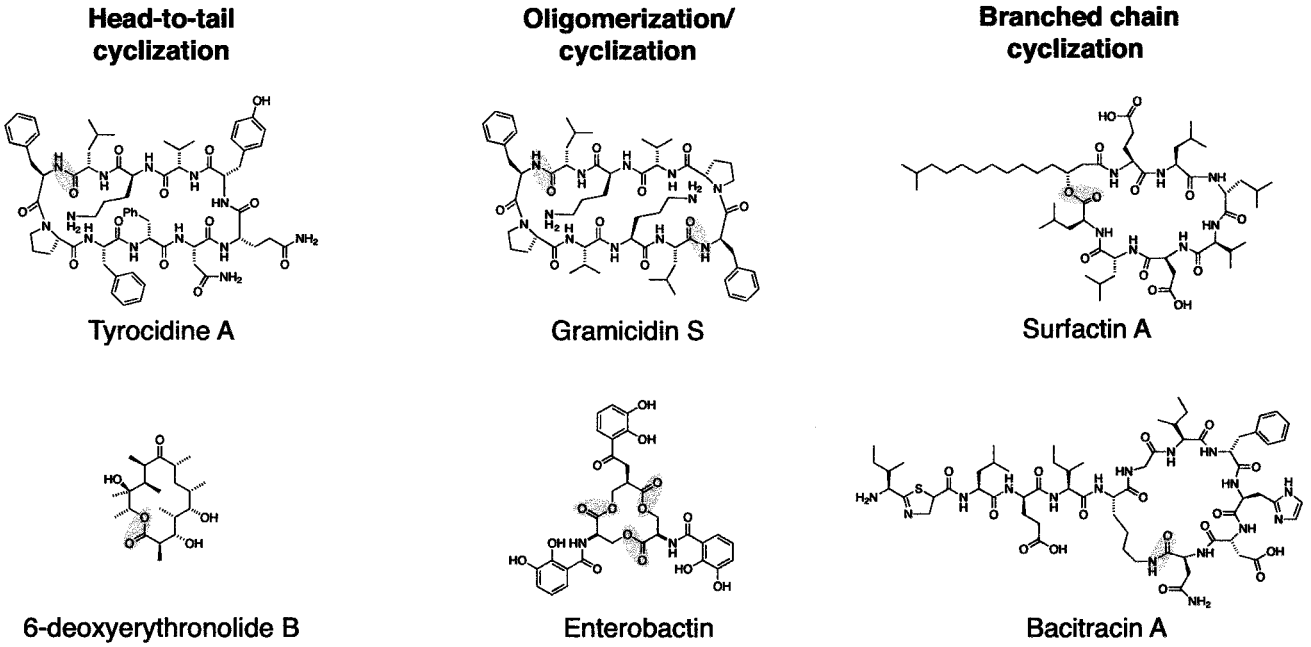

FIGURE 1: Examples of macrocyclic structures produced by NRPS, PKS, and hybrid NRPS/PKS systems, organized according to the type of cyclization linkage formed by TE catalysis: head-to-tail cyclization, oligomerization/cyclization, or branched chain cyclization. The bonds formed by TE catalysis during biosynthesis are highlighted by shading.

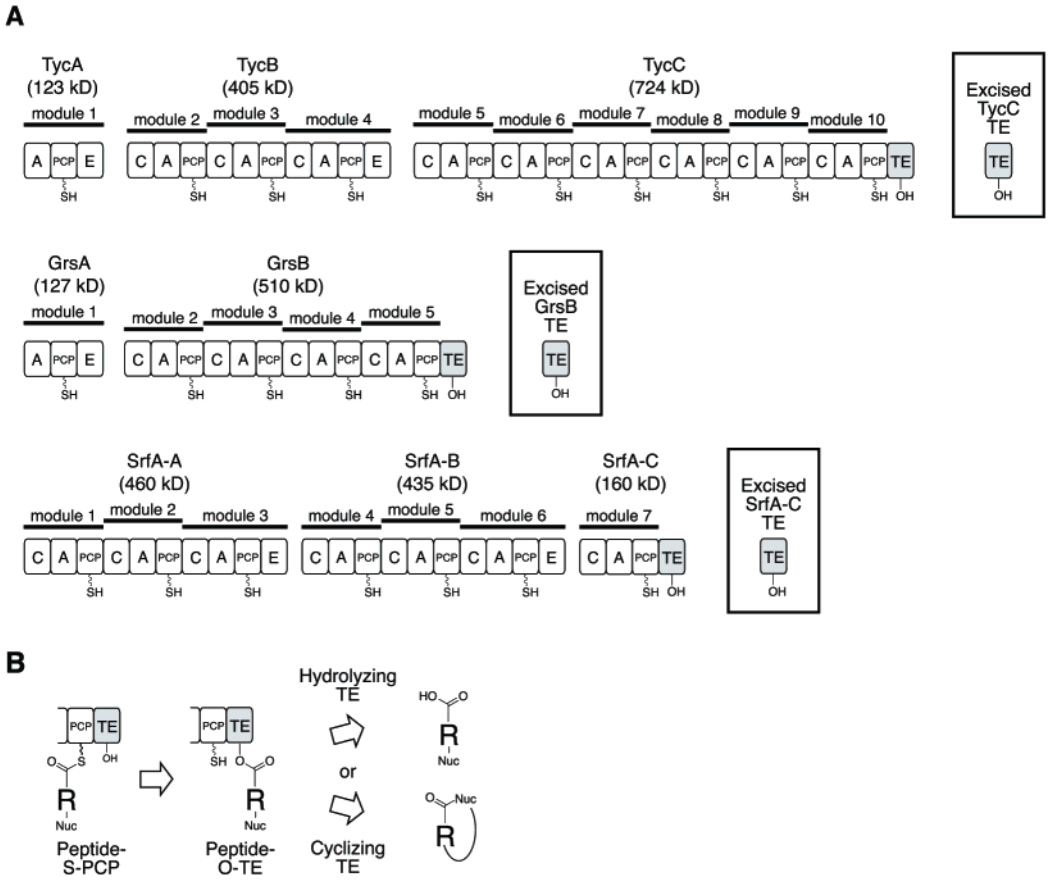

FIGURE 2: (A) Schematic diagram of the tyrocidine, gramicidin S, and surfactin nonribosomal peptide synthetases. Boxes represent individual protein domains: A, adenylation; PCP, peptidyl carrier protein; E, epimerization; C, condensation;TE, thioesterase (shaded). SH represents phosphopantetheine and OH the TE active site serine. In this paper, the TE from each synthetase was excised for study by cloning, expression, and purification of the isolated domain. (B) Mechanism of thioesterase domain catalysis: a peptide-O-TE acyl-enzyme intermediate is formed by transfer of the peptidyl chain from the phosphopantethiene of the terminal peptidyl carrier protein (PCP) to the active site serine of the TE domain. For hydrolyzing TE domains, the intermediate is captured by water, generating the linear peptide; for cyclizing TE domains, an intramolecular nucleophile captures the intermediate, resulting in a cyclic product.

TE domains that have been produced as isolated domains can retain cyclization activity, providing a possible enzyme-catalyzed approach to the synthesis of macrocyclic compounds. We recently cloned the TE domain (TycC TE) from

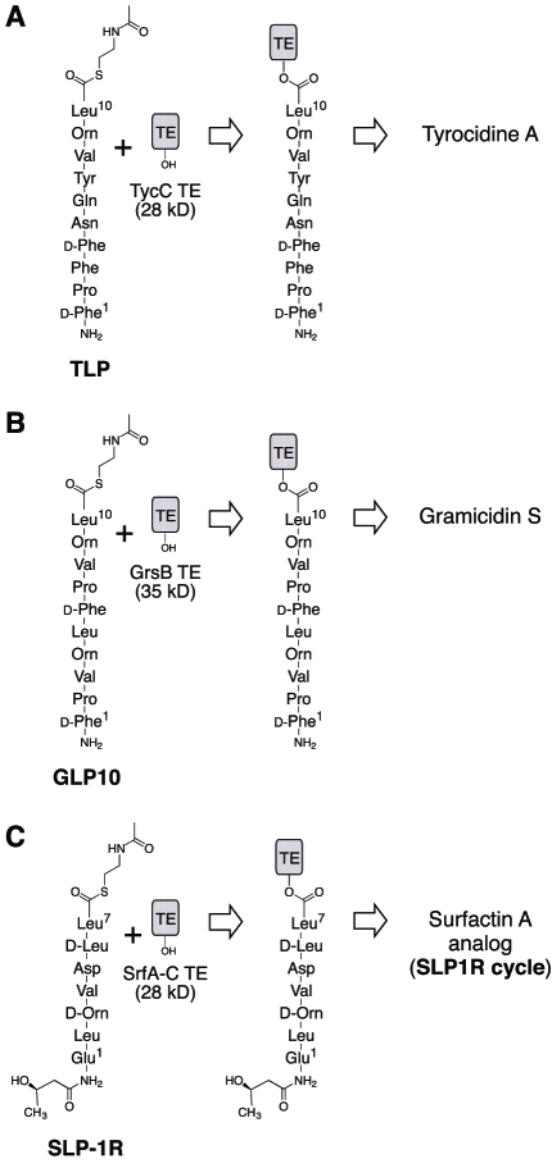

FIGURE 3: Cyclization of peptide thioester substrates catalyzed by excised thioesterase domains from (A) the tyrocidine NRPS, (B) the gramicidin NRPS, and (C) the surfactin NRPS. The peptide thioester substrates mimic peptide-S-PCP, the natural substrate for the TE domains.

the tyrocidine NRPS subunit TycC (724 kDa) and expressed it as an isolated 28 kDa domain (*12*). A peptide *N*-acetylcysteamine thioester (peptide-SNAC) was synthesized as a mimic of the full-length linear peptide linked to the phosphopantethiene of the most C-terminal PCP (peptide-S-PCP). The isolated TycC TE efficiently catalyzed head-to-tail cyclization of the peptide-SNAC to form the cyclic antibiotic tyrocidine A (Figure 3A). Varying the substrate residues revealed that the TE was quite permissive in its ability to cyclize substrate analogues. Only two residues, one near each end of the decapeptide substrate, proved to be required for efficient cyclization: the N-terminal D-Phe provided an absolute stereochemical and side chain requirement for cyclization, and changing an Orn adjacent to the C-terminal residue to Glu drastically reduced the cyclization rate (*12*). Further probes into peptide backbone requirements for cyclization indicated that some backbone amide bonds near the cyclization junction are required for efficient cyclization (*13*). The model that has emerged from these studies suggests that various substrates which meet these minimal requirements will be cyclized by TE catalysis (*13*).

With an eye toward exploring the general synthetic utility of isolated TE domains, we set out to examine three different TE domains, excised from the tyrocidine, gramicidin S, and surfactin synthetases. Having established that TycC TE can catalyze cyclization of decapeptide thioester substrates, here we investigate its ability to catalyze cyclization to form cyclic peptides ranging in size from 6 to 14 residues. Further, we clone, express, and purify isolated TE domains from the gramicidin S NRPS (GrsB TE) and from the surfactin NRPS (SrfA-C TE) and assess their ability to catalyze cyclization to form gramicidin S and a lipopeptide analogue of surfactin A, respectively (Figure 3). Finally, we report the solid-phase synthesis of a small pool of peptide thioesters and test the ability of TycC TE to convert it to a pool of corresponding cyclic peptides. These studies provide evidence for the general utility of this approach to the synthesis of macrocyclic compounds.

EXPERIMENTAL PROCEDURES

Cloning and Expression of TycC TE, GrsB TE, and SrfA-C TE Domains. TycC TE was cloned, expressed, and purified as previously described (*12*). For GrsB TE and SrfA-C TE, gene fragments were amplified by PCR from chromosomal DNA using Vent polymerase (New England Biolabs). GrsB TE (containing GrsB residues 4163−4452) was amplified using the oligonucleotides 5′-TTT CCA TGG TTA CAC ATA AAG AAT CAG AA-3′ and 5′-ATA GGA TCC TTT TAC TAC AAA TGT CCC TTG-3′ from chromosomal DNA of *Bacillus brevis* ATCC 9999, digested with *Nco*I and *Bam*HI (Amersham), and ligated into *Nco*I/*Bgl*II digested pQE60 (Qiagen), which appends a C-terminal His_6 tag to the expressed protein. The SrfA-C TE coding sequence (containing SrfA-C residues 1041−1274) was amplified using the oligonucleotides 5′-ATA AGA TCT AAC GGG GGC TCT GAT GG-3′ and 5′-TAT GGA TCC TGA AAC CGT TAC GGT TTG TG-3′ from chromosomal DNA of *Bacillus subtilis* JH642, digested with *Bam*HI and *Bgl*II, and ligated into pQE60 previously digested with *Bgl*II and dephosphorylated using CIP (New England Biolabs), yielding the vector pQE60-TE(SrfA-C). Standard procedures were used for all DNA manipulations (*14*). Cloning products were confirmed by sequencing with an ABI prism 310 genetic analyzer. For expression, the recombinant plasmids were transformed into *Escherichia coli* M15/pREP4 (Qiagen). Cells were grown to OD = 0.5 (600 nm), induced with 0.2 mM IPTG, and then grown at 24 °C for pQE60-TE(GrsB) and at 30 °C for pQE-TE(SrfA-C) for 6 h. The expressed proteins were purified by Ni−NTA affininty chromatography (Figure 4) and dialyzed into 25 mM HEPES, 50 mM NaCl, and 10% glycerol, pH 7.0.

Standard Peptide-SNAC Synthesis. Peptide thioester substrates, excluding the library of peptide thioesters **LIB1** (detailed below), were synthesized by a two-step process involving solid-phase peptide synthesis followed by solution-phase thioester formation. Peptide synthesis was carried out on a Perseptive Biosystems 9050 synthesizer (0.3 mmol scale) using diisopropylcarbodiimide (DIPCDI)/hydroxybenzotriazole (HOBt) chemistry and 2-chlorotrityl resin derivatized with L-Leu. Couplings were carried out with Fmoc-

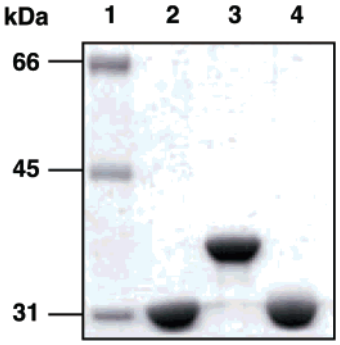

FIGURE 4: Purified excised thioesterase domains used in this study. Lanes: 1, molecular weight standards with masses listed; 2, TycC TE; 3, GrsB TE; 4, SrfA-C TE. Enzymes were overproduced in *E. coli* and purified by Ni−NTA affinity chromatography. Protein samples were resolved by SDS−PAGE (10%) and stained with Coomassie blue.

protected amino acids (Novabiochem), except for the N-terminal D-Phe of **GLP** substrates, which was Boc protected. Side chain protecting groups were Boc for Orn and Trp, *tert*-butyl for Tyr, Thr, Asp, and Glu, and trityl for Asn and Gln. For the final coupling in the synthesis of substrates **SLP-1R** and **SLP-1S**, respectively, *R*- and *S*-3-hydroxybutyric acids were used (Aldrich). A mixture of 1:1:3 acetic acid/trifluoroethanol/CH_2Cl_2 was used to cleave the peptide from the resin (3 h, 24 °C). The resin was then removed by filtration, the peptide precipitated with *n*-hexane, and the solvent removed by rotary evaporation. The protected peptide was dissolved in THF and thioester formation initiated by addition of DCC (1.2 equiv), HOBt (1.2 equiv), and NAC (3 equiv). After being stirred for 1 h (24 °C), K_2CO_3 (0.6 equiv) was added and the reaction stirred for an additional 2 h. The reaction was then filtered and solvent removed by rotary evaporation. Deprotection of the peptide-SNAC was effected using 16:3:1 TFA/CH_2Cl_2/NAC (3 h, 24 °C). The reaction was precipitated with cold ether (4 °C) and the precipitate washed twice with ether. The reaction product was purified by preparative HPLC with a reversed-phase (C_{18}) column, using a gradient from 20% to 60% acetonitrile in 0.1%TFA/water over 40 min. Lyophilization afforded the peptide-SNAC as a white powder. The identity and purity of the peptide-SNACs were assessed by analytical HPLC and MALDI-TOF mass spectrometry. The substrates were generated in >95% purity with the exception of **GLP14**, which contained ~50% of the free acid as estimated by MALDI-TOF MS. For **GLP14**, 50% purity was assumed for kinetic analysis.

Solid-Phase Peptide-SNAC Synthesis. For solid-phase synthesis of a small pool of four peptide thioesters, **LIB1**, an acylsulfonamide safety-catch linker was utilized (*15*). 4-Sulfamylbutyryl resin (NovaBiochem) was loaded with Fmoc-L-Leu using the method of Ingenito and co-workers (*16*). The resin (13.5 g, 1 equiv) was washed with CH_2Cl_2 and suspended in 50 mL of CH_2Cl_2. Diisopropylethylamine (10 equiv) and Fmoc-L-Leu (5 equiv) were added, and the mixture was stirred for 20 min at 24 °C. The mixture was then cooled to −20 °C and PyBOP (5 equiv) added. After 8 h, the reaction was filtered and washed with CH_2Cl_2. The coupling procedure was repeated (total of three coupling cycles). The resin loading was 0.1 mmol/g as assessed by treatment with 20% piperidine and quantitation of the released Fmoc group. Subsequent amino acid couplings were carried out using automated DIPCDI/HOBt couplings with protected amino acids as above. The peptide mixture was synthesized with the wild-type tyrocidine sequence diversified at positions 3 and 4 by utilizing 1.5 equiv each of L-Phe and L-Ala (for position 3) and 1.5 equiv each of D-Phe and D-Trp (for position 4). Activation and cleavage from the resin were carried out as per the method of Shin et al. (*17*). Briefly, the resin was activated by addition of NMP (4 mL) and DIEA (200 μL), followed by iodoacetonitrile (180 μL) that was prefiltered through basic alumina. After agitation for 24 h at 24 °C, the resin was washed with NMP followed by THF. Thioester formation was effected by agitation of the resin with 8% *N*-acetylcysteamine in THF (24 h, 24 °C). The resin was removed by filtration and the solvent removed by rotary evaporation. Deprotection of the peptide was carried out with a 16:3:1 solution of TFA/CH_2Cl_2/NAC (3 h, 24 °C). The pool of peptide-SNACs was then precipitated with cold ether (4 °C), dissolved in 20% CH_3CN/water, and lyophilized, providing **LIB1** as a white solid.

Assays. Reactions were conducted in 400 μL volume, at 24 °C, with 25 mM MOPS, pH 7.0, 1 mM NaCl, and variable amounts of glycerol (0.02−1%). Stock solutions of the enzyme were diluted to 10× desired concentration in 10 mM NaCl and 10 mM MOPS, pH 7.0. Reactions were initiated by addition of enzyme and quenched by addition of 25 μL of 1.7% TFA/water, flash frozen with liquid N_2, and stored at −80 °C. Time zero reactions were conducted by addition of 25 μL of 0.1% TFA/water to the reaction mixture prior to initiation of reaction. The reactions were thawed, 85 μL of acetonitrile was added, and samples were analyzed by HPLC (Beckman System Gold) with a reversed-phase (C_{18}) column (Vydac) with a gradient from 20% to 100% 0.1% TFA/acetonitrile in 0.1% TFA/water over 35 min. Experimentally determined exinction coefficients (ϵ) at 220 nm for peptide-SNACs were used to determine concentrations of peptide-SNACs, hydrolysis, and cyclic products in reaction mixtures, assuming equal ϵ at 220 nm. Product formation with TycC TE with the wild-type substrate **TLP** and GrsB TE with the wild-type substrate **GLP10** was seen to be linear to 2 min (data not shown). Determination of kinetics of cyclization for all **GLP** substrates was conducted by the method of initial rates, using 1 min time points. For SrfA-C TE, under steady-state turnover conditions, product formation is linear to 40 min (data not shown). Kinetics for reactions of **SLP-1R** with SrfA-C TE were determined by the method of initial rates, using 10 min time points. No uncatalyzed product formation was detected after 24 h for any of the substrates examined (data not shown).

Product Characterization. Assignment of peaks on analytical HPLC was conducted by MALDI-TOF mass spectrometry using a Perseptive Biosystems Voyager DE instrument at the Dana Farber Cancer Institute, Molecular Biology Core Facility. Calculated and observed masses are given in Table 1. Identification of the peptide-SNACs in **LIB1** and the corresponding cyclic products and HPLC MS-MS analysis of **SLP-1R cycle** were carried out using a Hewlett-Packard 1100 HPLC coupled to a Finnigan-MAT LCQ ion trap mass spectrometer at the Biopolymers Facility of the Howard Hughes Medical Institute/Harvard Medical School.

Table 1: Characterization of Substrates and Reaction Products by MS

compound	species	mass calculated	mass observed
by MALDI-TOF MS			
GLP6	$[M + H]^+$	868.5	868.5
GLP8	$[M + H]^+$	1062.6	1062.8
GLP12	$[M + H]^+$	1486.9	1486.8
GLP14	$[M + H]^+$	1685.1	1685.1
GLP6 cycle	$[M + H]^+$	749.5	749.5
GLP8 cycle	$[M + H]^+$	943.6	643.6
GLP8 dimer	$[M + H]^+$	2005.2	2005.3
GLP8 dimer cycle	$[M + H]^+$	1886.2	1886.2
GLP12 cycle	$[M + H]^+$	1367.8	1368.0
GLP14 cycle	$[M + H]^+$	1566.0	1566.1
SLP1-R	$[M + H]^+$	1002.6	1002.5
SLP1-R hydrolyzed	$[M + H]^+$	901.5	901.6
SLP1-R glycerolysis	$[M + H]^+$	975.6	975.6
SLP1-R cycle	$[M + H]^+$	883.6	883.6
SLP1-S	$[M + H]^+$	1002.6	1002.6
SLP2	$[M + H]^+$	1059.6	1059.6
by ESI−ion trap MS			
L1	$[M + 2H]^{2+}$	676.9	677.0
L2	$[M + 2H]^{2+}$	657.4	657.4
L3	$[M + 2H]^{2+}$	714.9	714.8
L4	$[M + 2H]^{2+}$	695.4	695.4
L1 cycle	$[M + H]^+$	1233.7	1233.7
L2 cycle	$[M + H]^+$	1194.7	1194.7
L3 cycle	$[M + H]^+$	1309.7	1309.7
L4 cycle	$[M + H]^+$	1270.7	1270.7

RESULTS

TycC TE Catalyzes Cyclization To Form 6−14 Residue Cyclic Peptides. To examine the ability of TycC TE to catalyze cyclization to generate various sizes of cyclic molecules, we synthesized a series of peptide thioester substrates which were 6, 8, 12, or 14 residues in length (Figure 5A). The sequences were based on the gramicidin S linear precursor sequence (**GLP10**), which we have previously synthesized and shown to be a substrate for cyclization catalyzed by TycC TE (*12*). The peptides were designed to be symmetric following the gramicidin model and to retain the key side chain recognition elements revealed by our previous study, including an N-terminal D-Phe residue and an Orn adjacent to the C-terminal residue. Assaying for TycC TE catalyzed cyclization (35 μM peptide-SNAC, 200 nM TycC TE, 25 mM MOPS, pH 7.0) revealed that all of the peptide-SNACs are cyclized (Figure 5B). The identity of the cyclic products was confirmed by MALDI-TOF MS. With **GLP6**, the substrate was completely consumed, producing the hexapeptide cycle as well as the hydrolysis product. This result with a hexapeptide substrate is in contrast to a previous result with a pentapeptide-SNAC which undergoes TycC TE catalyzed dimerization but not cyclization (*12*). With **GLP8** several products were formed: the expected octapeptide cycle, as well as a small amount of the 16-residue dimer-SNAC which in turn was converted to the 16-residue dimer cycle.

We next examined the kinetics of cyclization of the **GLP** series of substrates (1 min reactions). Cyclization reactions follow Michaelis−Menten kinetics with the K_m and k_{cat} values reported in Table 2. Remarkably, the cyclization rates (k_{cat}) are all within a factor of 3 of the 10-mer **GLP10** (6−30 turnovers/min) for all substrates, and K_m values (3−6 μM) are similar for all substrates. The cyclization efficiencies (k_{cat}/

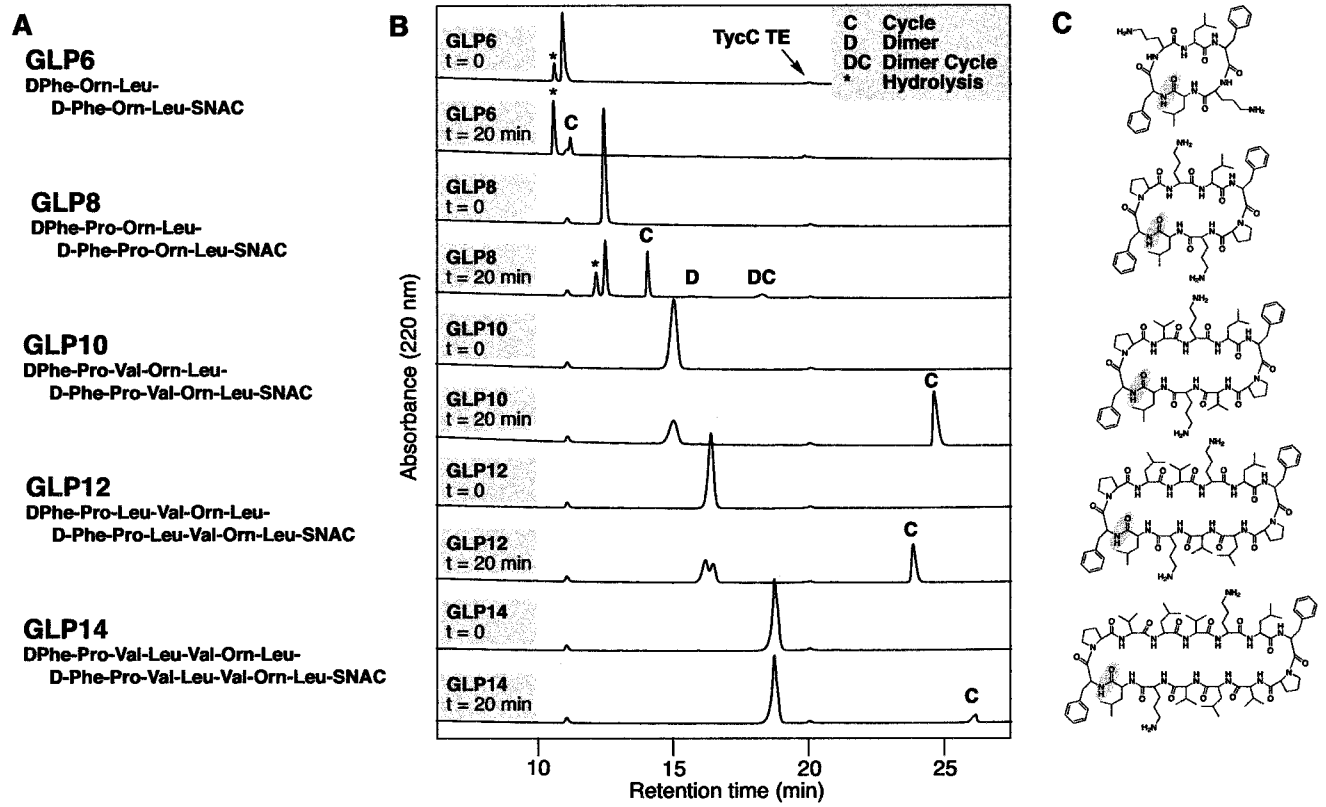

FIGURE 5: Formation of 6−14 residue cyclic peptides catalyzed by the TE domain from the tyrocidine NRPS, TycC TE. (A) Sequences of synthetic peptide thioester substrates (SNAC denotes *N*-acetylcysteamine thioester). (B) HPLC traces of reactions that initially contained 35 μM peptide-SNAC substrate, 200 nM TycC TE, and 25 mM MOPS, pH 7.0, 24 °C (time = 0 or 20 min). The products were characterized by MALDI-TOF mass spectrometry and assigned as cyclic product (C), dimer (D), dimer cycle (DC), or hydrolysis product (*). (C) Structures of the cyclic peptide products. The bond formed upon cyclization is highlighted by shading.

Table 2: Kinetics of Cyclization of Increasing Size Substrates by TycC TE[a]

substrate	k_{cat} (turnover/min)	K_m (μM)	k_{cat}/K_m
GLP6	30	4	8
GLP8	17	3	6
GLP10[b]	12	5	2
GLP12	22	6	4
GLP14	6	5	1

[a] Standard deviation ±15% (see ref *12*). [b] From ref *12*.

K_m) over the range of peptide sizes lie within an order of magnitude.

GrsB TE Catalyzes Head-to-Tail Cyclization To Generate Gramicidin S. Gramicidin S is a symmetrical cyclic decapeptide generated by *B. brevis* by a nonribosomal peptide synthetase (Figures 1 and 2A). The NRPS consists of two large proteins: GrsA, which encodes one amino acid activating module, and GrsB, which encodes four additional amino acid activating/elongating modules as well as a C-terminal thioesterase domain (GrsB TE) (*8*, *18*). The structure of the NRPS suggests a mechanism in which GrsB TE serves as a holding bay for a gramicidin linear pentapeptide (*19*). After the buildup of a second pentapeptide, TE catalysis generates a decapeptide which is subsequently acted upon again by the TE domain to generate the cyclized product. This mechanism was supported by our finding that TycC TE can catalyze dimerization of the gramicidin S pentapeptide-SNAC followed by cyclization of the resulting decapeptide-SNAC (*12*). To examine the generality of TE-catalyzed cyclization of peptide thioesters, we cloned and expressed the GrsB TE domain from the gramicidin S NRPS. Cloning of GrsB TE was modeled on previous success in isolating the TE domain of TycC, using a start site 38 amino acids downstream of the core serine of the adjacent PCP domain. Cloning of the GrsB TE domain using this start site generates a protein with a 26-residue N-terminal extension relative to TycC TE. The gene encoding GrsB TE was amplified from *B. brevis* genomic DNA and incorporated into a pQE60 expression vector with a hexahistidine C-terminal tag to facilitate purification. GrsB TE (35 kDa) was soluble when expressed in *E. coli* at 24 °C and was purified by Ni−NTA affinity chromatography yielding 2 mg/L of culture (Figure 4, lane 3).

GrsB TE was tested for its ability to catalyze head-to-tail cyclization of the gramicidin decapeptide linear precursor, **GLP10** (Figure 3B). Upon incubation with GrsB TE, **GLP10** was cyclized to gramicidin S (Figure 6). The identity of the cyclic product was confirmed by MALDI-TOF MS and by coelution with authentic gramicidin S. TycC TE has previously been shown to catalyze cyclization of **GLP10**, with a k_{cat} of 12 min^{-1} and K_m of 5 μM (*12*). We characterized the kinetics of cyclization of **GLP10** by GrsB TE and discovered that cyclization was an order of magnitude slower than observed with TycC TE, with a k_{cat} of 1.4 min^{-1} and a K_m of 13 μM. We also examined the ability of GrsB TE to catalyze the dimerization of the pentapeptide thioester **GLP5** (DPhe-Pro-Val-Orn-Leu-SNAC). In contrast to TycC TE, which efficiently catalyzes the dimerization of **GLP5** to form **GLP10**, GrsB TE showed barely detectable dimerization activity (data not shown).

SrfA-C TE Catalyzes Stereospecific Cyclization to a Lipopeptide Analogue of Surfactin A. TE catalysis generates

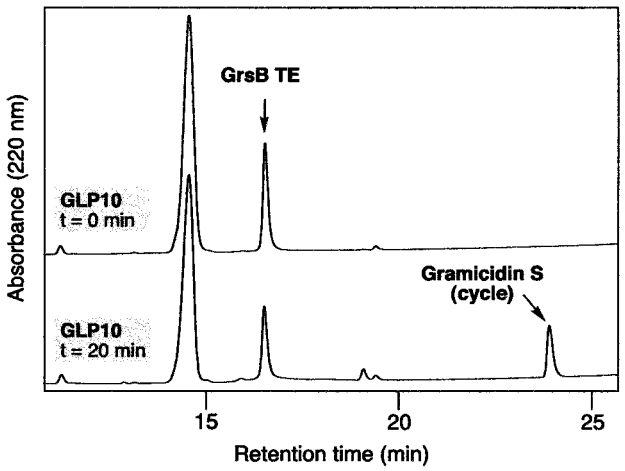

FIGURE 6: Head-to-tail cyclization of the gramicidin linear decapeptide thioester (**GLP10**) to form gramicidin S catalyzed by the isolated thioesterase domain from the gramicidin S NRPS (GrsB TE). HPLC traces of reactions that initially contained 35 μM **GLP10**, 500 nM GrsB TE, and 25 mM MOPS, pH 7.0, 24 °C (time = 0 or 20 min). The peak assigned to the cyclic product gramicidin S was characterized by MALDI-TOF mass spectrometry and shown to coelute with authentic standard.

a variety of linkages including branched chain cyclic peptides such as surfactin A and bacitracin A (Figure 1). Surfactin A, which is generated by *B. subtilis*, is a powerful biosurfactant, with antimicrobial and antimitotic activities (*20*, *21*). The structure of surfactin consists of a lipoheptapeptide having the sequence β-HFA-Glu-Leu-DLeu-Val-Asp-DLeu-Leu (β-HFA = β-hydroxy fatty acid) cyclized via an ester bond connecting the C-terminus to the β-oxygen of a β-hydroxy $C_{13}-C_{15}$ fatty acid (*22*). A recent study has revealed that surfactin A has *R* stereochemistry at the C_β carbon of the fatty acid (*23*). The NRPS responsible for the production of surfactin consists of three subunits, SrfA-A, SrfA-B, and SrfA-C (Figure 2A) (*24*). The third subunit, SrfA-C, contains an activating module for the C-terminal leucine residue of the linear lipoheptapeptide as well as the TE domain (SrfA-C TE) (*25*). The cyclization reaction to form surfactin A is distinct from the reactions to form tyrocidine and gramicidin S in several regards: (1) the nucleophile is a hydroxyl instead of an amine, (2) the nucleophile is derived from a fatty acid molecule rather than an amino acid, and (3) the cycle produced has a branched cyclic structure rather than a head-to-tail cyclic structure. We examined if the isolated SrfA-C TE domain could independently catalyze cyclization to form a lipopeptide analogue of surfactin A.

The isolated SrfA-C TE domain was cloned in an analogous manner to GrsB TE and TycC TE. It was soluble when expressed at 30 °C in *E. coli* and contains a C-terminal hexahistidine tag to facilitate purification by Ni−NTA affinity chromatography. The purified SrfA-C TE was produced with a yield of 12 mg/L (Figure 4, lane 4). The natural substrate for SrfA-C TE is a lipoheptapeptide linked to the phosphopantethiene of the most C-terminal PCP of SrfA-C. The substrate we synthesized, **SLP-1R**, had two modifications to the wild-type surfactin linear precursor to increase the solubility of the substrate: (1) in the heptapeptide sequence D-Leu3 is altered to D-Orn and (2) instead of conjugating a β-hydroxy $C_{13}-C_{15}$ fatty acid to the N-terminal

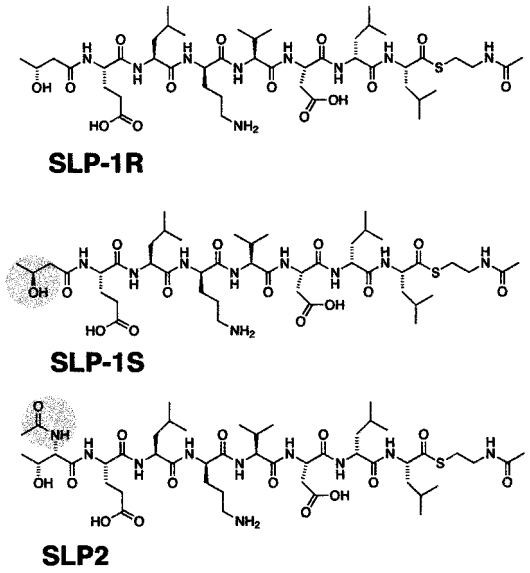

FIGURE 7: Substrates synthesized for study of the surfactin thioesterase domain, SrfA-C TE. **SLP-1R** retains the natural *R* stereochemistry at C_β. Regions highlighted by shading represent differences in substrates **SLP-1S** and **SLP2** from **SLP-1R**. **SLP-1S** inverts stereochemistry at C_β, while **SLP2** differs by substitution of *N*-acetylthreonine for the β-hydroxy acyl chain.

Glu, we used the shorter analogue *R*-3-hydroxybutyric acid (Figure 7).

The reaction of **SLP-1R** (200 μM, 1% glycerol, 25 mM MOPS, pH 7.0) with SrfA-C TE (1 μM) generated three products in a 1:1:1 ratio (Figure 8A). The expected cyclic product formed, along with the products of hydrolysis and glycerolysis of the thioester. All products were identified by MALDI-TOF MS. We isolated the cyclic product and analyzed it by MS-MS fragmentation. Several of the observed fragments included the Leu7 to 3-hydroxybutyric acid linkage formed by the expected cyclization reaction, confirming the structure of the cyclic product (Figure 8C). We characterized the kinetics of product formation with the substrate **SLP-1R** (Table 3), revealing similar k_{cat} (15−22 min^{-1}) and K_m (4.0−5.6 mM) values for the three products formed. No uncatalyzed cyclization was seen after incubation at pH 7.0 for 48 h in the absence of enzyme, indicating a rate enhancement of greater than 2×10^8. The K_m values seen for **SLP-1R** are several orders of magnitude higher than those seen with substrates for TycC TE. Possible sources of the increase in K_m include the side chain alteration of D-Leu3 from the natural substrate, the need for a mimic longer than *N*-acetylcysteamine for phosphopantetheine, the need for a longer mimic of the fatty acid chain, or an inherent difference in the reactivity of the SrfA-C TE.

We also addressed two other questions with regard to the catalytic potential of SrfA-C TE. To test the stereochemical requirements of the nucleophile for SrfA-C TE, we altered the stereochemistry at C_β by coupling *S*-3-hydroxybutyric acid to the N-terminal Glu to generate the peptide thioester **SLP-1S** (Figure 7). Additionally, to test the possibility that Ac-L-Thr could serve as an analogue of *R*-3-hydroxybutyric acid, as it differs only by the addition of an acetylamino group at C_α, we synthesized **SLP2** (Figure 7). When assayed with SrfA-C TE, no cyclization of **SLP-1S** and **SLP2** was observed, but hydrolysis and glycerolysis products were detected for both (Figure 8A). The fact that catalyzed

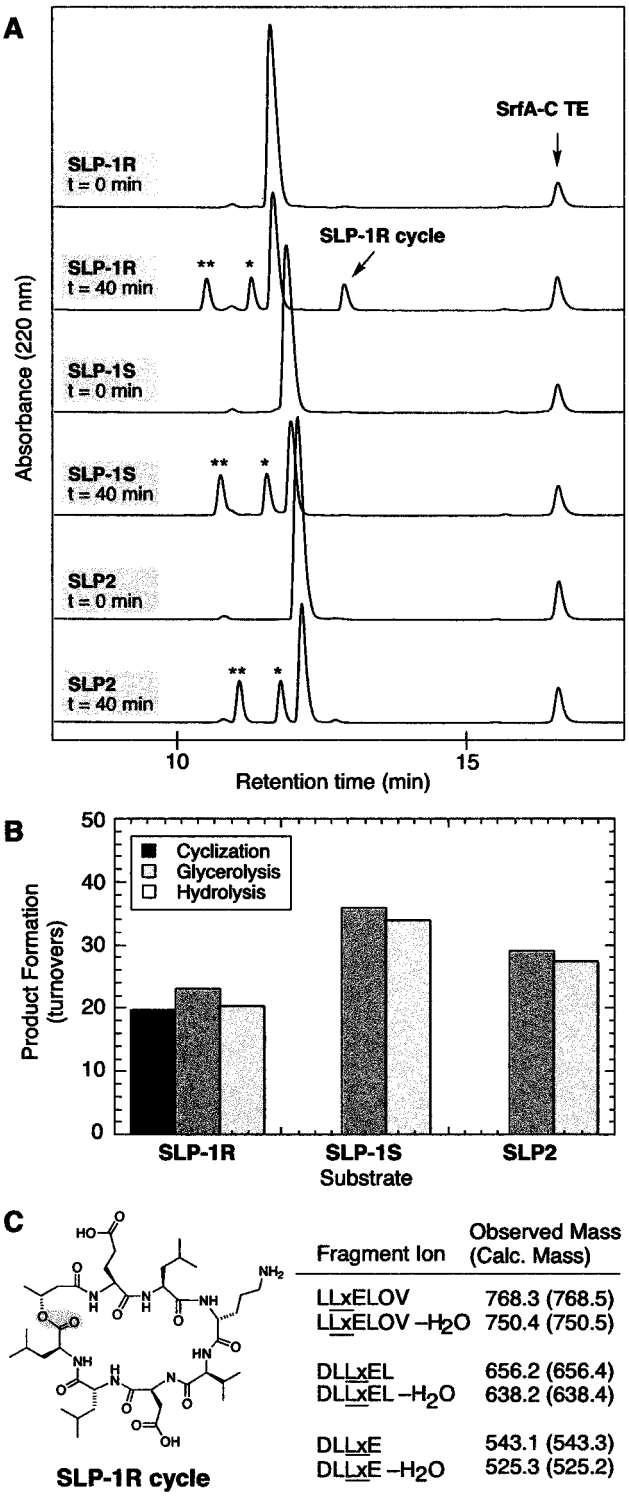

FIGURE 8: (A) HPLC traces of reactions of **SLP** substrates catalyzed by SrfA-C TE. Reactions initially contained 200 μM peptide-SNAC, 1 μM SrfA-C TE, 1% glycerol, and 25 mM MOPS, pH 7.0, 24 °C (time = 0 or 40 min). The cyclic product is labeled, and products of hydrolysis and glycerolysis are noted by * and **, respectively. All products were characterized by MALDI-TOF mass spectrometry. (B) Observed number of SrfA-C TE catalyzed turnovers of **SLP** substrates to cyclization, hydrolysis, and glycerolysis products determined by quantitation of analytical HPLC runs shown in (A). (C) Characterization of the cyclization product of **SLP-1R** by ESI− ion trap mass spectrometry. The calculated and observed masses for selected fragment ions are listed (x denotes 3-hydroxybutyric acid; O denotes ornithine). The underlined sequence represents the 3-hydroxybutyric acid to Leu linkage formed in the cyclization reactions.

Table 3: Kinetics of Product Formation with **SLP-1R** with SrfA-C TE[a]

product	k_{cat} (turnover/min)	K_m (mM)
cyclization	15	4
hydrolysis	20	6
glycerolysis	22	4

[a] Standard deviation ±15% (see ref *12*).

hydrolysis and glycerolysis proceeded at rates in excess of that seen with **SLP-1R** (Figure 8B) suggested that peptide-O-TE formation proceeded efficiently, but the lack of the proper intramolecular nucleophile precluded cyclization, leading instead to reaction with an external nucleophile.

TE Catalysis Can Generate a Small Pool of Cyclic Peptides. Having observed that TE domains from several NRPS systems can catalyze the cyclization of linear thioesters, we began to examine the synthetic utility of TE-catalyzed cyclization by assessing TycC TE-catalyzed cyclization to generate a small pool of four cyclic molecules. On the basis of our previous studies of the side chain specificity of TycC TE, we chose to alter positions 3 and 4 in the linear tyrocidine sequence, generating a pool that contained the combinations of either Phe or Ala at position 3 and either D-Phe or D-Trp at position 4 (Figure 9A). To demonstrate that the process will be amenable to synthesis of large numbers of compounds, we pursued a solid-phase synthesis of the peptide thioester substrates. Such a method would eliminate (1) the need for preparative HPLC purification to separate the free acid from the peptide thioester, (2) the need for the side chain protected substrate to be soluble in organic solvent as required with solution-phase thioester synthesis, and (3) two solvent removal steps that would hamper parallel synthesis efforts. The peptide was synthesized using a safety-catch sulfonamide linker, the linker activated by alkylation with iodoacetonitrile, and cleavage from the resin was carried out with *N*-acetylcysteamine, affording the mixture of four peptide thioesters (**LIB1**) (*15*−*17*). We evaluated conversion of the pool of linear peptide thioesters into a pool of cyclic peptide thioesters by TycC TE catalysis (Figure 9B). Following a 1.5 min reaction, the four peptide-SNACs were largely consumed, and the four expected cyclic products formed, as confirmed by LC-MS analysis. The relative amounts of the cyclic peptide products match the relative amounts of starting peptide-SNACs, suggesting that, as expected, each of the molecules is cyclized with comparable efficiency.

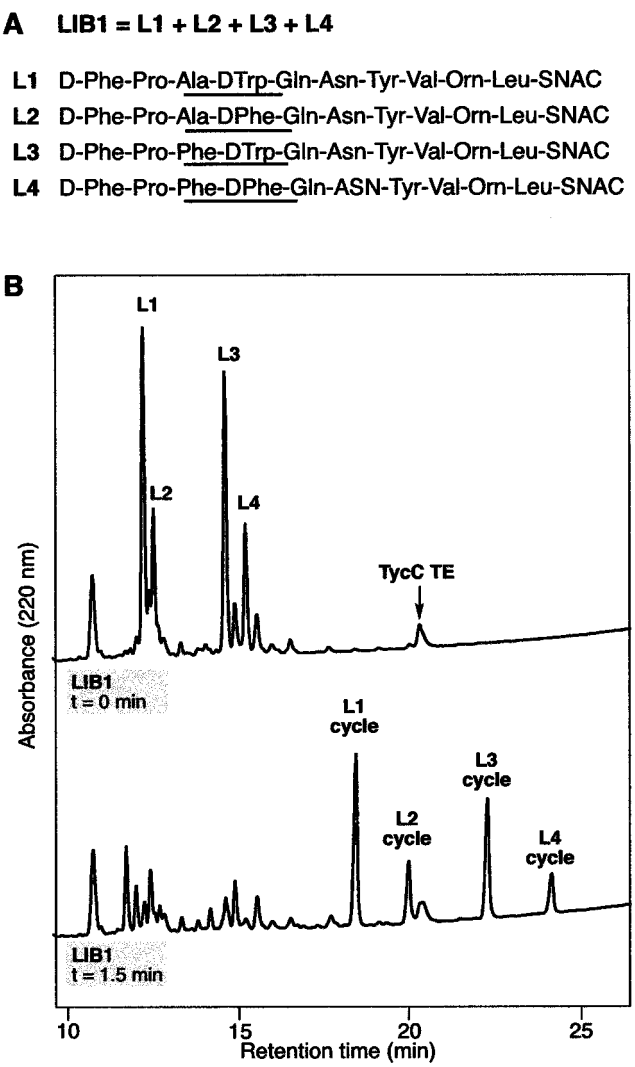

FIGURE 9: Cyclization of a pool of four peptide thioesters catalyzed by TycC TE. (A) **LIB1** contains four peptides with variations in positions 3 and 4 (underlined) in the linear tyrocidine peptide sequence. (B) HPLC trace of reactions initially containing 8 μM total peptide-SNAC, 100 nM TycC TE, and 25 mM MOPS, pH 7.0, incubated at 24 °C for 0 or 1.5 min showing conversion of **LIB1** to a pool of four cyclic peptides. The identity of the peptide-SNACs and cyclic products was confirmed by ESI−ion trap mass spectrometry.

DISCUSSION

Nonribosomal peptide synthetases often contain a C-terminal TE domain which can catalyze hydrolysis or cyclization to release the final product. For systems which contain cyclizing TE domains, a variety of cyclization linkages are possible, including head-to-tail cyclization, oligomerization/cyclization, and branched chain cyclization, which facilitates formation of a diverse array of cyclic peptides, polyketides, hybrid polypeptide/polyketides, and lipopeptides (Figure 1). Previously, we showed that the isolated TycC TE can catalyze the cyclization of the linear tyrocidine decapeptide thioester to generate the antibiotic tyrocidine A (*12*). The observation that TycC TE was quite liberal in its side chain requirements for cyclization (*12*) suggested that isolated TE domains could be utilized for enzymatic synthesis of macrocycles. In this paper, we explore the generality of this approach by several means: (1) we expand on previous results with TycC TE by probing the ability to generate variously sized cyclic molecules, (2) we express and purify the TE domain from a closely related system, the gramicidin S NRPS, and demonstrate that it also catalyzes head-to-tail cyclization, and (3) we express and purify the TE domain from the surfactin NRPS and demonstrate that it can catalyze cyclization via macrolactone formation.

Our previous study of TycC TE revealed that, in addition to generating tyrocidine A, modified substrates in which a single amino acid was deleted or added to the middle of the tyrocidine decapeptide thioester sequence can also be cyclized to generate 9- and 11-membered cyclic peptides (*12*). To test the range of sizes of cyclic peptides that can be formed by TycC TE, we synthesized an array of peptide thioester substrates 6−14 residues in length. The sequences of the peptides are based on that of the symmetric cyclic

peptide gramicidin S. Remarkably, TycC TE catalyzed the cyclization of the various length peptide thioesters with comparable kinetic efficiency (Figure 5, Table 2). These results suggest that TycC TE is quite permissive in its size requirements for cyclization. Additionally, dimerization was not observed for peptides longer than 8 residues (**GLP8**). Chemical approaches to macrocyclization face the significant challenge of controlling for cyclization versus oligomerization. Enzymatic macrocyclization could serve as an alternate means to control for desired product formation.

Recently, the structures of 6−16 residue cyclic peptides, modeled on gramicidin S, have been characterized by NMR (*26*). The study suggested that products with $4n + 2$ residues have high antiparallel β-sheet content, while those with $4n$ residues have low β-sheet content. The accompanying paper (*13*) provided evidence suggesting that substrate preorganization by β-sheet hydrogen bonds contributes to TycC TE-catalyzed cyclization (*13*). Our results with the array of substrate lengths suggest that, assuming β-sheet content is lower in the $4n$ (8- and 12-residue) cyclic products versus the $4n + 2$ (6-, 10-, and 14-residue) cyclic products, the ability to form transient β-sheet hydrogen bonds is sufficient for efficient cyclization.

We further explored the generality of TE-catalyzed cyclization by evaluating isolated TE domains from two other NRPS systems, gramicidin S and surfactin. The excision of isolated GrsB TE and SrfA-C TE domains from the synthetase proteins was modeled on the successful cloning of the TycC TE. The isolated TE domains were cloned, expressed with C-terminal hexahistidine tags, and purified (Figure 4, lane 3). GrsB TE was active as a cyclization catalyst when tested with the decapeptide thioester substrate **GLP10** (Figure 6). The activity observed with this substrate was an order of magnitude slower than that seen with TycC TE with the same substrate (*12*). The decrease in activity for GrsB TE compared with TycC TE could be due to a number of factors, including the need for a PCP-TE fusion to stabilize the domain fold, the need to delete the 26-residue N-terminal extension, or an inherent difference in the TE domains. SrfA-C TE was tested with an acylheptapeptide-SNAC mimic of its natural lipoheptapeptide-S-PCP substrate. This substrate, **SLP-1R**, substituted 3-hydroxybutyric acid for a β-hydroxy C_{13}−C_{15} fatty acid and included a D-Leu3 to D-Orn change to increase substrate solubility. SrfA-C TE catalyzed the cyclization of **SLP-1R**, as well as its hydrolysis and glycerolysis. It is notable that, despite the presence of only 1% glycerol in the reaction, the flux to glycerolysis is similar to that to hydrolysis (Figure 8B). The cyclization was stereospecific, as the substrate **SLP-1S** with altered stereochemistry at C_β was not cyclized by SrfA-C TE. Further, when tested with Ac-L-Thr as a substitute for the 3-hydroxybutyric acid (**SLP2**), no cyclization was observed.

The cyclization activity of GrsB TE and SrfA-C TE demonstrates that peptide cyclization catalyzed by an isolated TE domain, first observed with TycC TE, is a generalizable phenomenon. Some mechanistic lessons can be drawn from our results. TycC TE has previously been shown to exclusively catalyze cyclization of substrates with an N-terminal D-Phe residue, as substrates with substitution of D-Phe1 by either L-Phe or D-Ala are not cyclized (*12*). The results with the SrfA-C TE domain, where **SLP-1R** but not **SLP-1S** or **SLP2** undergo cyclization, provide a second example of stringent stereoselective recognition of the part of the substrate near the cyclizing nucleophile. Further, despite the presence of alternate nucleophiles in the side chains (e.g., Orn in **GLP** and **SLP** substrates) which could lead to alternate cyclic molecules, no such alternative products have been observed. Thus, in the examples investigated so far, TE domains are both stereo- and regiospecific cyclization catalysts. From alanine scanning mutagenesis, we have previously shown that only two residues, one near each end of the tyrocidine decapeptide-SNAC substrate, are critical for cyclization by TycC TE. Similarly, the SrfA-C TE domain was tolerant of mutation of D-Leu3 to D-Orn. In the examples investigated thus far, TE domains appear to be tolerant of side chain substitutions that are not immediately adjacent to the cyclizing nucleophile or the C-terminus of the linear peptide. Finally, in addition to internal TE domains, NRPSs commonly contain an external thioesterase, whose function has remained speculative (*27, 28*). We have shown in three cases that, when taken as an isolated domain, the internal TE domain, independent of the external TE, has cyclization activity.

Chemical approaches to peptide cyclization have been utilized to generate libraries of modest sized cyclic peptides (*29, 30*). These efforts have met significant challenges with regard to the need for strict protecting group control, the need for solubility of the protected substrates, unreliable efficiency of cyclization, and undesired multimerization. TE-catalyzed cyclization may provide an alternative approach for generating libraries of cyclic peptides and a new opportunity for accessing diverse structures such as cyclic lipopeptides, hybrid polypeptide/polyketides, and branched chain cyclic peptide libraries. As a demonstration of the synthetic utility of TE-catalyzed cyclization in the generation of a library of cyclic molecules, we synthesized a small pool of cyclic peptides by TycC TE catalysis. A solid-phase peptide thioester synthesis strategy (*15*−*17*) was used to generate a pool of four peptides which vary at positions 3 and 4 of the linear tyrocidine sequence. When subjected to TycC TE catalysis, the linear peptide thioester pool was efficiently converted into a corresponding pool of cyclic peptides. The approach presented here should be amenable to synthesis of larger libraries. Further, this approach to macrocycle synthesis could complement the engineered biosynthetic approach. A cyclic molecule, initially produced on a small scale by TE-catalyzed cyclization of a synthetic acyl thioester substrate and once identified as an interesting target, in principle could subsequently be produced en masse by engineered synthetases (*31*−*34*).

Our results suggest that cyclization of synthetic acyl thioester substrates catalyzed by excised thioesterase domains from NRPSs may be a generally useful approach to the synthesis of diverse cyclic molecules. Though the approach may not be ideally suited to the generation of large amounts of a single cyclic peptide, the liberal side chain, backbone, and length requirements should allow for the generation of diverse libraries of cyclic molecules which could be probed for activity (*12, 13*). The diversity of molecules potentially accessible by this approach is further amplified by the observation that naturally occurring TE domains from different NRPSs can catalyze diverse ring closure reactions, including head-to-tail cyclization, oligomerization/cyclization, and branched chain cyclization.

ACKNOWLEDGMENT

We gratefully acknowledge Eric Spooner of the Howard Hughes Medical Institute/Harvard Medical School Biopolymers Facility for assistance with HPLC/MS.

REFERENCES

1. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) *Chem. Rev. 97*, 2651−2674.
2. Cane, D. E., Walsh, C. T., and Khosla, C. (1998) *Science 282*, 63−68.
3. Marahiel, M. A. (1997) *Chem. Biol. 4*, 561−567.
4. Zocher, R., and Keller, U. (1997) *Adv. Microb. Physiol. 38*, 85−131.
5. Lawson, D. M., Derewenda, U., Serre, L., Ferri, S., Szittner, R., Wei, Y., Meighen, E. A., and Derewenda, Z. S. (1994) *Biochemistry 33*, 9382−9388.
6. Li, J., Szittner, R., Derewenda, Z. S., and Meighen, E. A. (1996) *Biochemistry 35*, 9967−9973.
7. Mootz, H. D., and Marahiel, M. A. (1997) *J. Bacteriol. 179*, 6843−6850.
8. Krätzschmar, J., Krause, M., and Marahiel, M. A. (1989) *J. Bacteriol. 171*, 5422−5429.
9. Konz, D., Klens, A., Schörgendorfer, K., and Marahiel, M. A. (1997) *Chem. Biol. 4*, 927−937.
10. Mchenney, M. A., Hosted, T. J., Dehoff, B. S., Rosteck, P. R., Jr., and Baltz, R. H. (1998) *J. Bacteriol. 180*, 143−151.
11. Cosmina, P., Rodriguez, F., de Ferra, F., Perego, M., Venema, G., and van Sinderen, D. (1993) *Mol. Microbiol. 8*, 821−831.
12. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A., and Walsh, C. T. (2000) *Nature 407*, 215−218.
13. Trauger, J. W., Kohli, R. M., and Walsh, C. T. (2001) *Biochemistry 40*, 7092−7098.
14. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989*) Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
15. Backes, B. J., and Ellman, J. A. (1999) *J. Org. Chem. 64*, 2322−2330.
16. Ingenito, R., Bianchi, E., Fattori, D., and Pessi, A. (1999) *J. Am. Chem. Soc. 121*, 11369−11374.
17. Shin, Y., Winans, K. A., Backes, B. J., Kent, S. B. H., Ellman, J. A., and Bertozzi, C. (1999) *J. Am. Chem. Soc. 121*, 11684−11689.
18. Turgay, K., Krause, M., and Marahiel, M. A. (1992) *Mol. Microbiol. 6*, 529−546.
19. Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C., and Walsh, C. T. (1999) *Chem. Biol. 6*, 385−400.
20. Rosenberg, E., and Ron, E. Z. (1999) *Appl. Microbiol. Biotechnol. 52*, 154−162.
21. Peypoux, F., Bonmatin, J. M., and Wallach, J. (1999) *Appl. Microbiol. Biotechnol. 51*, 553−563.
22. Bonmatin, J. M., Genest, M., Labbé, H., and Ptak, M. (1994) *Biopolymers 34*, 975−986.
23. Nagi, S., Okimura, K., Kaizawa, K., Ohki, N. K., and Kanatomo, S. (1996) *Chem. Pharm. Bull. (Tokyo) 44*, 5−10.
24. Ullrich, C., Klige, B., Palacz, Z., and Vater, J. (1991) *Biochemistry 30*, 6503−6508.
25. de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C., and Grandi, G. (1997) *J. Biol. Chem. 272*, 25304−25309.
26. Gibbs, A. C., Kondejewski, L. H., Gronwald, W., Nip, A. M., Hodges, R. S., Skyes, B. D., and Wishart, D. S. (1998) *Nat. Struct. Biol. 5*, 284−288.
27. Schneider, A., and Marahiel, M. A. (1998) *Arch. Microbiol. 169*, 404−410.
28. Xue, Y., and Sherman, D. H. (2000) *Nature 403*, 571−575.
29. Spatola, A. F., and Romanovskis, P. (1996) in *Combinatorial Peptide and Nonpeptide Libraries: A Handbook* (Jung, G., Ed.) pp 327−347, VCH Publishers, New York.
30. Hruby, V., Ahn, J.-M., and Liao, S. (1997) *Curr. Opin. Chem. Biol. 1*, 114−119.
31. Stachelhaus, T., Schneider, A., and Marahiel, M. A. (1995) *Science 269*, 69−72.
32. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5848−5853.
33. Jacobsen, J. R., Hutchinson, C. R., Cane, D. E., and Khosla, C. (1997) *Science 277* 367−369.
34. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 1846−1851.

BI010036J