Insights into Channel Architecture and Substrate Specificity from Crystal Structures of Two Macrocycle-Forming Thioesterases of Modular Polyketide Synthases[†,‡]

Shiou-Chuan Tsai,[§] Hongxiang Lu,[||] David E. Cane,[||] Chaitan Khosla,[*,§,⊥,@] and Robert M. Stroud[#]

Departments of Chemical Engineering, Chemistry, and Biochemistry, Stanford University, Stanford, California 94305-5025, Department of Chemistry, Brown University, Box H, Providence, Rhode Island 02912-9108, and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received April 24, 2002; Revised Manuscript Received July 22, 2002

ABSTRACT: Modular polyketide synthases (PKSs) synthesize the polyketide cores of pharmacologically important natural products such as erythromycin and picromycin. Understanding PKSs at high resolution could present new opportunities for chemoenzymatic synthesis of complex molecules. The crystal structures of macrocycle-forming thioesterase (TE) domains from the picromycin synthase (PICS) and 6-deoxyerythronolide B synthase (DEBS) were determined to 1.8−3.0 Å with an R_{crys} of 19.2−24.4%, including three structures of PICS TE (crystallized at pH 7.6, 8.0, and 8.4) and a second crystal form of DEBS TE. As predicted by the previous work on DEBS TE [Tsai, S. C., et al. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 14808−14813], PICS TE contains an open substrate channel and a hydrophobic dimer interface. Notwithstanding their similarity, the dimer interfaces and substrate channels of DEBS TE and PICS TE reveal key differences. The structural basis for the divergent substrate specificities of DEBS TE and PICS TE is analyzed. The size of the substrate channel increases with increasing pH, presumably due to electrostatic repulsion in the channel at elevated pH. Together, these structures support previous predictions that macrocycle-forming thioesterases from PKSs share the same protein fold, an open substrate channel, a similar catalytic mechanism, and a hydrophobic dimer interface. They also provide a basis for the design of enzymes capable of catalyzing regioselective macrocyclization of natural or synthetic substrates. A series of high-resolution snapshots of a protein channel at different pHs is presented alongside analysis of channel residues, which could help in the redesign of the protein channel architecture.

Modular polyketide synthases (PKSs)[1] are a family of multienzyme complexes that synthesize the polyketide cores of biologically active compounds, including pharmaceutically important natural products such as erythromycin, rifamycin, FK506, rapamycin, and avermectin (*1*). Their remarkable combination of substrate tolerance and selectivity is largely due to their modular architecture, in which different catalytic domains are combined into "modules" such that each module contains several enzymes and adds one additional building block to a growing polyketide chain. Two examples of related but distinct modular PKSs are the 6-deoxyerythronolide B synthase (DEBS) (*2*, *3*) and the picromycin synthase (PICS) (*4*, *5*), which synthesize the 14-membered macrolactone cores of the related antibiotics erythromycin and picromycin (Figure 1). The picromycin synthase also synthesizes a truncated 12-membered macrolactone product, 10-deoxymethynolide, which is subsequently processed into the antibiotic methymycin (*6*). Both PKSs contain ca. 30 distinct domains that are organized into a loading module, six extension modules, and a terminal thioesterase (TE).

The TE domains of DEBS and PICS catalyze the cyclization and concomitant release of highly functionalized heptaketide chains via lactonization (Figure 1). In each case, this domain is covalently linked to the acyl carrier protein (ACP) domain of the last module. Prior to cyclization, the acyclic polyketide substrate is transferred from the phosphopantethienyl arm of the ACP to the TE. Cyclization occurs via a covalent mechanism involving an acyl−enzyme intermediate at a universally conserved serine residue in the active site of the TE. Regioselective displacement of the enzyme-bound intermediate by a secondary alcohol at the distal end of the chain results in formation of the macrocycle product. At least 12 other modular PKSs are known to harbor homologous terminal TE domains, with a level of sequence identity of 29−75% with respect to DEBS TE. Although the natural product of DEBS TE is a 14-membered ring, this enzyme has been shown to support the formation of alternatively functionalized 6-, 8-, 12-, 14-, and 16-membered ring systems (*7*, *8*). At the same time, different TE domains appear to have different specificity profiles for acyl chains with alternative functionality and stereochemistry (*9*, *10*). Given the importance of macrocycle formation in synthetic

[†] This work was supported by grants from the National Institutes of Health to R.M.S. (CA 63081), C.K. (CA 66736), and D.E.C. (GM 22172).

[‡] The atomic coordinates have been deposited in the Protein Data Bank (entries 1MN6, 1MNA, 1MNQ, and 1MO2).

* To whom correspondence should be addressed: C.K.: e-mail, ck@chemeng.stanford.edu. R.M.S.: e-mail, stroud@msg.ucsf.edu.

[§] Department of Chemical Engineering, Stanford University.

[||] Brown University.

[⊥] Department of Chemistry, Stanford University.

[@] Department of Biochemistry, Stanford University.

[#] University of California.

[1] Abbreviations: PKS, polyketide synthase; DEBS, 6-deoxyerythronolide B synthase; PICS, picromycin synthase; ACP, acyl carrier protein; TE, thioesterase.

10.1021/bi0260177 CCC: $22.00
Published on Web 09/27/2002

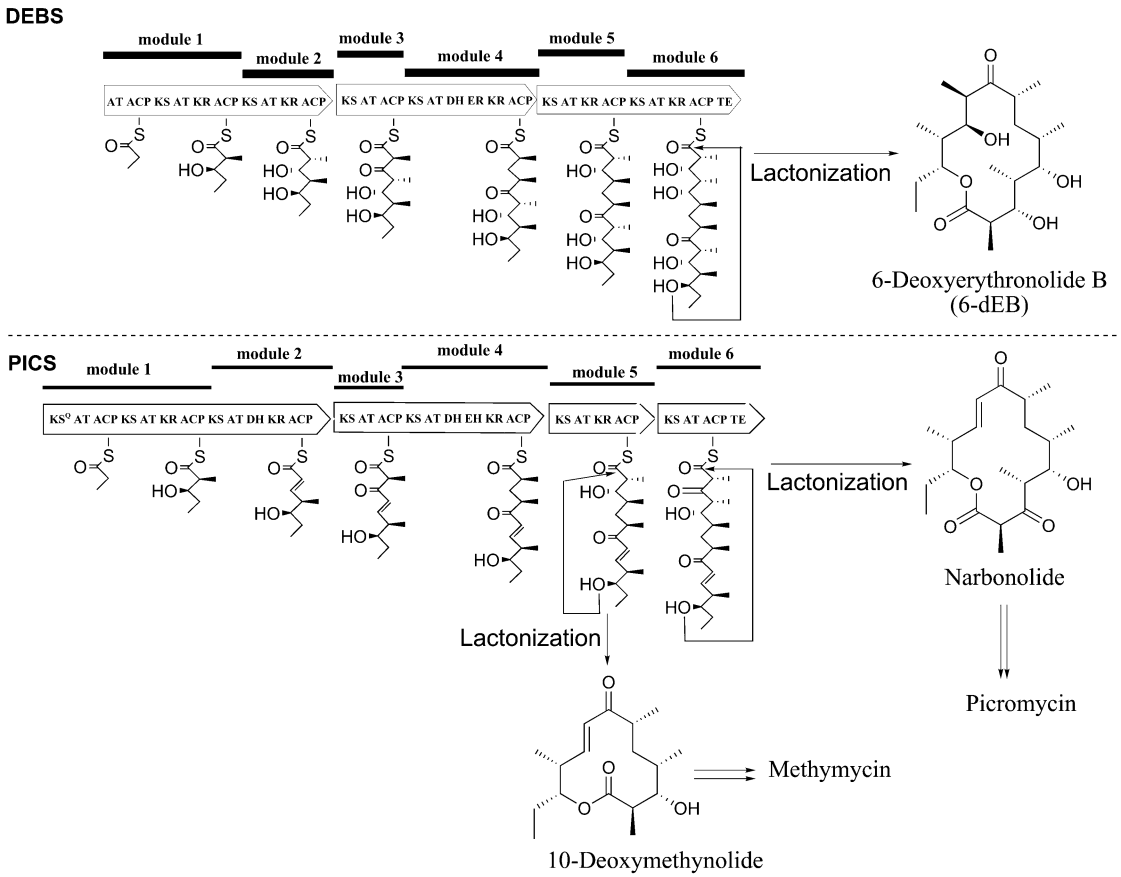

FIGURE 1: Biosynthesis of 6-deoxyerythronolide (6-dEB) by 6-deoxyerythronolide synthase (DEBS) and narbonolide by picromycin synthase (PICS). Abbreviations for individual catalytic domains: KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase.

chemistry (and natural product chemistry in particular), understanding the mechanism and substrate specificity of this class of enzymes could present new opportunities for chemoenzymatic synthesis of complex molecules.

Recently, we reported an X-ray crystal structure of the DEBS TE (*11*). A particularly noteworthy feature of this member of the hydrolase family of enzymes was a substrate channel that passed through the entire protein. Modeling suggested that the active site could accommodate and orient the 6-deoxyerythronolide B precursor uniquely. Moreover, the geometry and organization of functional groups was consistent with the observed substrate specificity of this TE. To gain a better understanding of how PKS-derived TE domains control macrocycle formation, a second related but distinct TE domain was targeted for X-ray crystallographic analysis. The PICS TE was chosen because it synthesizes a related 14-membered macrocycle (*4*), although its substrate specificity is distinct from that of the DEBS TE. In particular, as might be anticipated from the structural differences between 6-deoxyerythronolide B and narbonolide (Figure 1), kinetic analysis of DEBS TE and PICS TE revealed significant difference in their substrate specificity (*12*). Here, we report the structures from three crystal forms of the PICS TE, as well as the structure from a second crystal form of DEBS TE. A comparative analysis of these five crystal structures provides new insights not only into the properties of the individual TE domains but also into the general properties of macrocycle-forming TE domains. In particular, the influence of pH on the structure of the active site channel is analyzed. Comparison of these five macrocycle-forming TE structures with other α,β-hydrolases, including TEs of fatty acid biosynthesis, offers insight into the structural basis for the difference in substrate channel shape. This study also offers a series of high-resolution snapshots of an active site channel at different pHs.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of DEBS TE. Recombinant DEBS TE was expressed in *Escherichia coli* and purified as described previously (*9*). DEBS TE eluted from a Ni−agarose column was further purified using a Poros HQ column (Perceptive Biosystems) and eluted at 0.8 M NaCl.

Cloning, Expression, and Purification of PICS TE. Recombinant PICS TE was expressed in *E. coli* (*12*) and purified on a Ni−agarose column. Approximately 80 mg of protein was produced per liter of LB medium. This one-step purification gave 98% pure protein. The buffer was changed to assay buffer [50 mM phosphate buffer (pH 8.0)] or crystallization sample buffer [10 mM HEPES (pH 7.0)] in a PD-10 column. For crystallization, the protein sample was further concentrated with a Centricon YM-10 concentrator to 10 mg/mL.

Crystallization of DEBS TE. Crystals of DEBS TE were grown in hanging drops at 4 °C by vapor diffusion. The well buffer contained 30% PEG 400 and 100 mM HEPES (pH 7.0−7.4) or 100 mM Bicine (pH 8.5), 2 mM DTT, and 100 mM $MgCl_2$. The protein concentration was 10 mg/mL in 20 mM HEPES (pH 7.5) and 2 mM dithiothreitol (DTT). Drops

Table 1: Crystallographic Parameters for DEBS TE and PICS TE

	PICS TE, pH 7.6 form	PICS TE, pH 8.0 form	PICS TE, pH 8.4 form	DEBS TE, pH >8.5 form
(A) data collection				
source	SSRL 7-1	SSRL 9-2	SSRL 9-2	ALS 5.0.2
space group	*C*2	$p2_12_12$	$p2_12_12_1$	$p2_12_12_1$
cell dimensions	$a = 117.1$ Å, $b = 58.3$ Å, $c = 90.9$ Å, $\alpha = \gamma = 90.0°$, $\beta = 96.2°$	$a = 108.1$ Å, $b = 130.6$ Å, $c = 56.8$ Å, $\alpha = \beta = \gamma = 90°$	$a = 58.9$ Å, $b = 106.1$ Å, $c = 114.2$ Å, $\alpha = \beta = \gamma = 90.0°$	$a = 80.5$ Å, $b = 102.5$ Å, $c = 156.5$ Å, $\alpha = \beta = \gamma = 90°$
resolution (Å)	2.2	1.8	2.2	3.0
mosaicity (deg)	0.4	0.2	0.8	1.1
no. of observations	203580	690823	418911	350978
no. of unique reflections	31231	75591	35346	26440
redundancy	6.5	9.1	13.8	13.2
completeness (%) (last shell)	100.0 (100.0)	99.4 (91.5)	97.4 (99.6)	99.5 (100.0)
$I/\sigma(I)$ (last shell)	13.3 (3.1)	31.3 (2.3)	14.6 (2.2)	9.7 (1.7)
R_{merge} (%) (last shell)	9.8 (35.7)	6.1 (58.7)	10.3 (64.8)	10.3 (51.0)
(B) structural refinement				
resolution (Å)	2.2	1.8	2.2	3.0
no. of reflections	30136	70844	30344	22595
data cutoff (σ)	0.0	0.0	0.0	0.0
no. of protein atoms	4193	4193	4185	3776
no. of waters	452	321	372	61
R_{crys} (%)	19.2	23.7	22.1	24.4
R_{free} (%)	24.1	26.3	24.8	28.7
(C) geometry				
rmsd for bonds (Å)	0.005	0.005	0.017	0.006
rmsd for angles (deg)	1.22	1.22	1.74	1.67
rmsd for *B* main chain	1.6	1.6	0.8	1.6
rmsd for *B* side chain	2.2	2.2	2.8	1.9
Ramachandran plot (%)				
most favored	85.1	88.3	88.8	77.7
favored	12.7	9.7	9.6	15.8
generously allowed	1.7	1.3	1.1	6.0
disallowed	0.5	0.7	0.5	0.5

were generated by mixing 4 μL of the purified protein solution with 4 μL of well buffer. The crystal form grown at pH 7.2 fell in space group $p3_121$, and crystals grew in 1 week (*11*). The pH 8.5 crystal form, in space group $p2_12_12_1$ (Table 1), grew in 4 months.

Crystallization of PICS TE. Crystals of PICS TE were grown in hanging drops at room temperature by vapor diffusion. The well buffers for the three crystal forms contained 30% PEG 8000, 100 mM HEPES (pH 7.6), 2 mM DTT, and 100 mM $MgCl_2$ (form 1, space group *C*2, $a = 117.1$ Å, $b = 58.3$ Å, $c = 90.9$ Å, $\alpha = \gamma = 90.0°$, and $\beta = 96.2°$), 1 M $LiSO_4$, 100 mM Tris-HCl (pH 8.0), and 2 mM DTT (form 2, space group $p2_12_12$, $a = 108.1$ Å, $b = 130.6$ Å, $c = 56.8$ Å, and $\alpha = \beta = \gamma = 90°$), and 30% PEG 4000, 100 mM Tris (pH 8.4), 2 mM DTT, and 100 mM $LiSO_4$ (form 3, space group $p2_12_12_1$, $a = 58.9$ Å, $b = 106.1$ Å, $c = 114.2$ Å, and $\alpha = \beta = \gamma = 90°$) (Table 1). In all three crystallization conditions, the protein concentration was 10 mg/mL in 10 mM HEPES (pH 7.5) and 2 mM DTT. Drops were generated by mixing 4 μL of the purified protein solution with 4 μL of well buffer. All three crystal forms grew in 1 week.

Data Collection. X-ray diffraction data from the four crystal forms of DEBS TE and PICS TE were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) and Advanced Light Source (ALS), with the highest resolution ranging from 1.9 to 3.0 Å (Table 1). Crystals were frozen in mother liquor for the corresponding crystallization conditions. The crystal space groups and cell dimensions are listed in Table 1. Diffraction intensities were integrated, reduced using the program DENZO, and scaled using SCALEPACK (*13*). A summary of the crystallographic data is shown in Table 1.

Molecular Replacement and Refinement. Initial phases were determined by molecular replacement using the pH 7.2 crystal form of dimeric DEBS TE (Protein Data Bank entry 1KEZ) as the search model, with residues 170−190 (helix 6) removed. A cross-rotational search followed by translational search was performed utilizing the program EPMR (*14*). After the structure had been rebuilt using Quanta, further refinement was performed using CNS (*15*). The placement of the two noncrystallographically related monomers treated as rigid bodies was refined using CNS to give an initial R_{crys} of 35−52%. A preliminary round of refinement using torsion angle simulated annealing followed by energy minimization and positional and individual *B*-factor refinement reduced R_{crys} to 31−41%. Subsequent rounds of model building and refinement were carried out using the maximum likelihood-based approach implemented within CNS using all data to the highest resolution. Strict noncrystallographic symmetry restraints were applied for the first two rounds of refinement. Refinement was continued to an R_{crys} of <25% (R_{free} < 28%). Table 1 lists the components of the final model for each crystal form.

RESULTS AND DISCUSSION

Overall Structure of PICS TE. Three crystal forms (pH 7.6, 8.0, and 8.4) of PICS TE were grown, the structures of which were determined by molecular replacement using the pH 7.2 DEBS TE dimer with the deletion of helix 6 as the search model. Two noteworthy structural features, an open substrate channel and a hydrophobic dimer interface (Figures 2 and 3), were observed in both the crystal structures of PICS and DEBS TE. Both these features were predicted from the crystal structure of DEBS TE and the sequence alignment

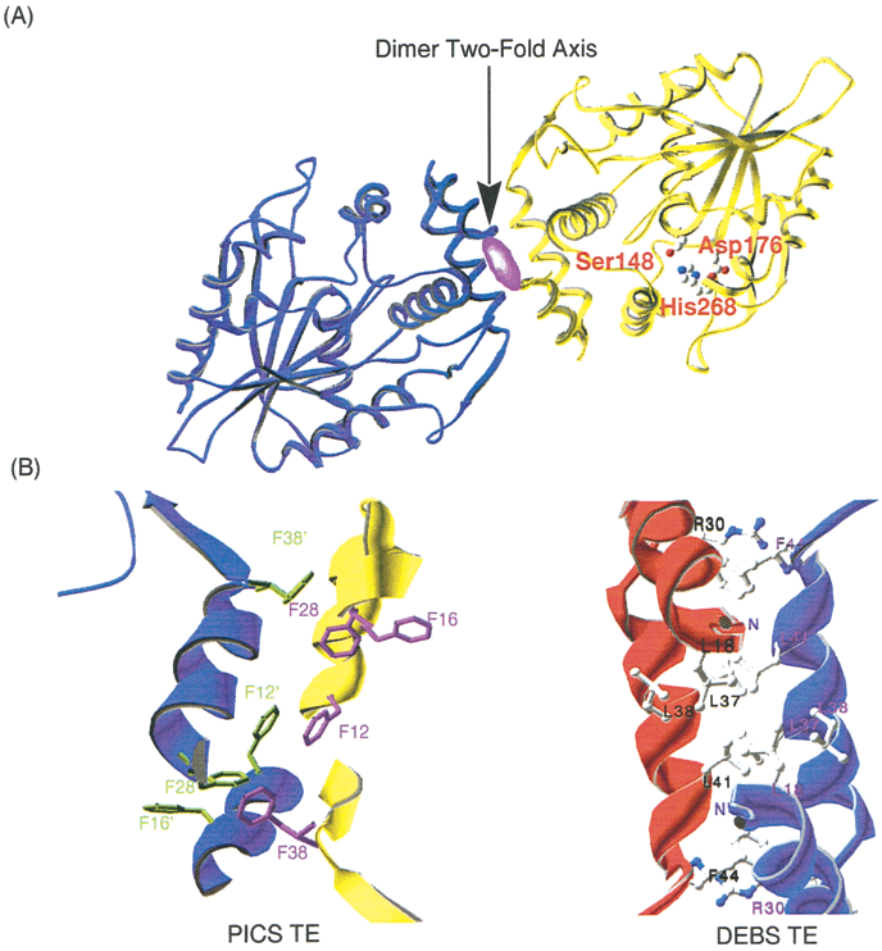

FIGURE 2: (A) Overall structure of the PICS TE (pH 7.6 form) dimer, with a hydrophobic dimer interface and an open substrate channel. The active site triad (Asp176-His268-Ser148) is located in the middle of the substrate channel. (B) Hydrophobic dimer interface of PICS TE (left panel in blue and yellow, F12, F16, F28, and F38), in comparison with DEBS TE (right panel in blue and red, L17, L18, L37, and L41). This figure was generated using Swiss PDB Viewer rendered with POV-Ray (*30*).

(Figure 4) to be a general property of the macrocycle-forming thioesterases (*11*).

As predicted by the structure of DEBS TE (*11*), the crystal structures of PICS TE revealed that it is a dimer. Each monomer has an α,β-hydrolase fold (*16*, *17*) consisting of a central seven-stranded β-sheet with the second strand (β_2) antiparallel to the remaining strands (Figure 2A). The β-pleated sheets from all the DEBS TE and PICS TE structures can be superimposed on each other, and are flanked on either side by α-helices, two on one side and four on the other. Further, as predicted by the structure of DEBS TE (*11*), two additional α-helices at the N-terminal end of the PICS TE also make up a hydrophobic dimer interface (Figure 2B). Two loop insertions in PICS TE (residues 64−68 and 112−118; disordered in all three crystal forms) do not significantly influence any of these structural features. An analogue of the arginine-rich electropositive groove on DEBS TE, proposed from docking simulations to be the binding site for the upstream acyl carrier protein domain (*11*), is also present in PICS TE (Figure 3). Most remarkably, the open substrate channel observed in DEBS TE is also observed in PICS TE, and passes through the entire protein (Figure 3). These observations reinforce earlier predictions that the macrocycle-forming thioesterases from polyketide synthases are a closely related family of enzymes (*11*) with a similar fold, a hydrophobic dimer interface, and an open substrate channel.

Active Site of PICS TE. The substrate channel runs through the protein over a distance of ca. 20 Å, with openings on both the side containing the N-terminus (termed the N-side) and the side containing the C-terminus (termed the C-side). The active site triad of PICS TE (Ser148-His268-Asp176) is located in the middle (7.5 Å from the N-side surface) of the substrate channel. Ser148 is the site of covalent attachment of the acyl chain, and is located in the nucleophilic elbow between β_6 and α_5, a feature that is characteristic of α/β-hydrolases (*16*, *17*). This unique spatial requirement results in the placement of Ser148 in the disallowed region of the Ramachandran plot (*16*, *17*).

Besides the active site triad (Ser148-His268-Asp176), the substrate channel is composed of six conserved hydrophobic residues (Ala78, Pro180, Ile186, Leu193, Met213, and Phe269), five semiconserved or nonconserved hydrophobic residues (Leu29, Leu152, Leu197, Ala217, and Ala221), six conserved hydrophilic residues (Glu85, His147, Tyr178, Glu184, Ser190, and Asp267), and five nonconserved hydrophilic residues (Tyr25, Thr77, Asn80, Glu187, and Gln192). As in the substrate channel of DEBS TE (*11*), there are many more acidic residues (Asp and Glu) than basic ones (only His147 and His268) in the substrate channel of PICS TE. The acidic nature of the substrate channel is evident from the calculated negative electrostatic potential that is mapped on the molecular surface (Figure 3, red in the substrate channel).

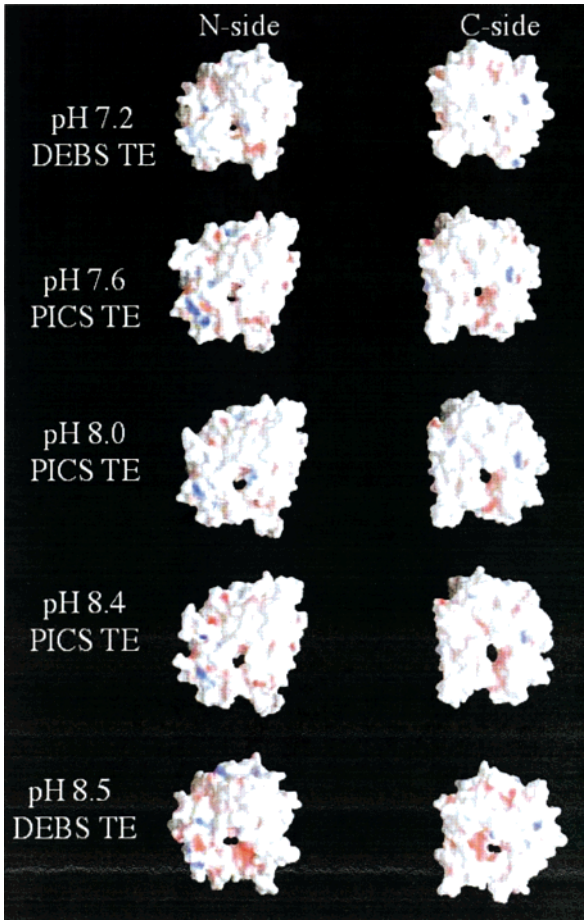

FIGURE 3: Views of the N-side and C-side of the TE molecular surfaces, showing that the substrate channel size increases with crystallization pH. Surfaces with electrostatic potential (positive charge in blue and negative charge in red) were generated using GRASP (*31*).

Ser148, His268, and Asp176 form the hydrogen-bonded catalytic triad. This catalytic triad is consistent with the mechanism for this family of thioesterases, in which His268, stabilized by Asp176, acts as the catalytic base which deprotonates Ser148 as it attacks the ACP6-bound thioester substrate to yield an acyl−*O*-serine intermediate. The NH group of Gly149 may serve as the oxyanion-stabilizing functional group. The enzyme-bound intermediate is subsequently attacked either by water (hydrolysis) or by a hydroxyl group on the polyketide chain to complete the cyclization reaction (Figure 5A). Notably, there is a near-perfect overlap of the catalytic triad and oxyanion group among DEBS TE and all three (pH 7.6, 8.0, and 8.5) PICS TE structures (Figure 5B). This suggests that DEBS TE and PICS TE have the same mechanism throughout this pH range. The above analysis of the substrate channel suggests that the catalytic mechanism and the negatively charged nature of the substrate channel may be conserved among macrocycle-forming TEs.

Comparison between DEBS TE and PICS TE. Notwithstanding their similarity, the dimer interfaces and substrate channels of DEBS TE and PICS TE reveal key differences. Although both dimer interfaces are hydrophobic in nature, the interface is comprised of phenylalanine residues (F12, F16, F28, and F38) in PICS TE versus leucine residues (L17, L37, L38, and L41) in DEBS TE (*11*) (Figure 2B). This results in a slight tilt of the PICS TE dimer interface relative to that of DEBS TE. The region that forms the dimer interface, composed of helices α_1 and α_2, is relatively unconserved among macrocycle-forming thioesterases (Figure 4). It is possible that by varying these residues on helices α_1 and α_2, TEs adapt the angle of their monomer−monomer interfaces to fine-tune features such as channel geometries or docking sites with donor acyl carrier protein domains.

The biggest difference between the pH 7.2 structure of DEBS TE and the three PICS TE structures reported here has to do with the size of the substrate channel (Figure 3), which appears to be controlled by the conformation of helices 6 and 7 in PICS TE [the corresponding region of DEBS TE has a single unbroken helix, designated helix 6 previously (*11*)]. Indeed, since the crystal structures of PICS TE could not be determined from the DEBS TE structure without first deleting helix 6, it was recognized early during model building that there were likely to be significant differences in the size and shape of the substrate channels of DEBS TE and PICS TE. Moreover, a comparison of the structures of PICS TE at different pH values also revealed that the size of the substrate channel increases with increasing pH (Figure 3); on the N-side and the C-side, the pH 7.2 DEBS TE has approximate mouth diameters of 8.0 and 5.5 Å, respectively, which become 8.8 and 9.2 Å in the pH 7.6 PICS TE structure, 9.6 and 10.2 Å in the pH 8.0 PICS TE structure, and 9.5 and 11.4 Å in the pH 8.4 PICS TE structure, respectively.

What can be the structural basis for the increasing channel size with increasing pH? Structure comparison among the DEBS TE (pH 7.2) and PICS TE (pH 7.6, 8.0, and 8.4) structures indicates that there is no significant global protein conformational change. Except for helices 6 and 7 of PICS TE (or helix 6 in DEBS TE), these TE structures, albeit crystallized at different pH values, can be overlapped nearly perfectly. Thus, it appears that only a specific region in TE responds to the increase in pH, namely the region of helices 6 and 7 of PICS TE (helix 6 in DEBS TE). The presence of a proline residue in this helical region of PICS TE (but not DEBS TE) might give rise to conformational differences between PICS TE and DEBS TE in this region, resulting in a larger channel size in PICS TE. However, this cannot account for the variation in channel size among PICS TE structures at different pHs. Structure comparison shows that most of the structure variation among PICS TE structures occur at charged residues of helices 6 and 7 (Figure 5C), suggesting altered ionic interactions at different pHs. The acidic residues Asp and Glu ($pK_a = 4.5$) are likely to be predominantly in the anionic (basic) form at pH 7.2−8.4, contributing one negative charge per residue throughout this pH range. In contrast, the ionic state of the two basic residues in the substrate channel, His147 and His268 ($pK_a = 6.5-7.5$), is more likely to be altered between pH 7.2 and 8.4. At higher pH values, the two His residues will be neutral, resulting in an elevated negative potential in the substrate channel. It should be noted that the side chain of the active site base His268 forms a hydrogen bond with the side chain of the active site triad Asp176, while the side chain of His147 does not hydrogen bond with any ionic residues. The pH effect of the substrate channel geometry does not appear to be caused by specific side chain interactions but rather by an overall increase in the number of negative charges developed in the substrate channel as pH increases. Therefore, from structure comparison, a possible explanation for the variability in the size of the substrate channel may lie in

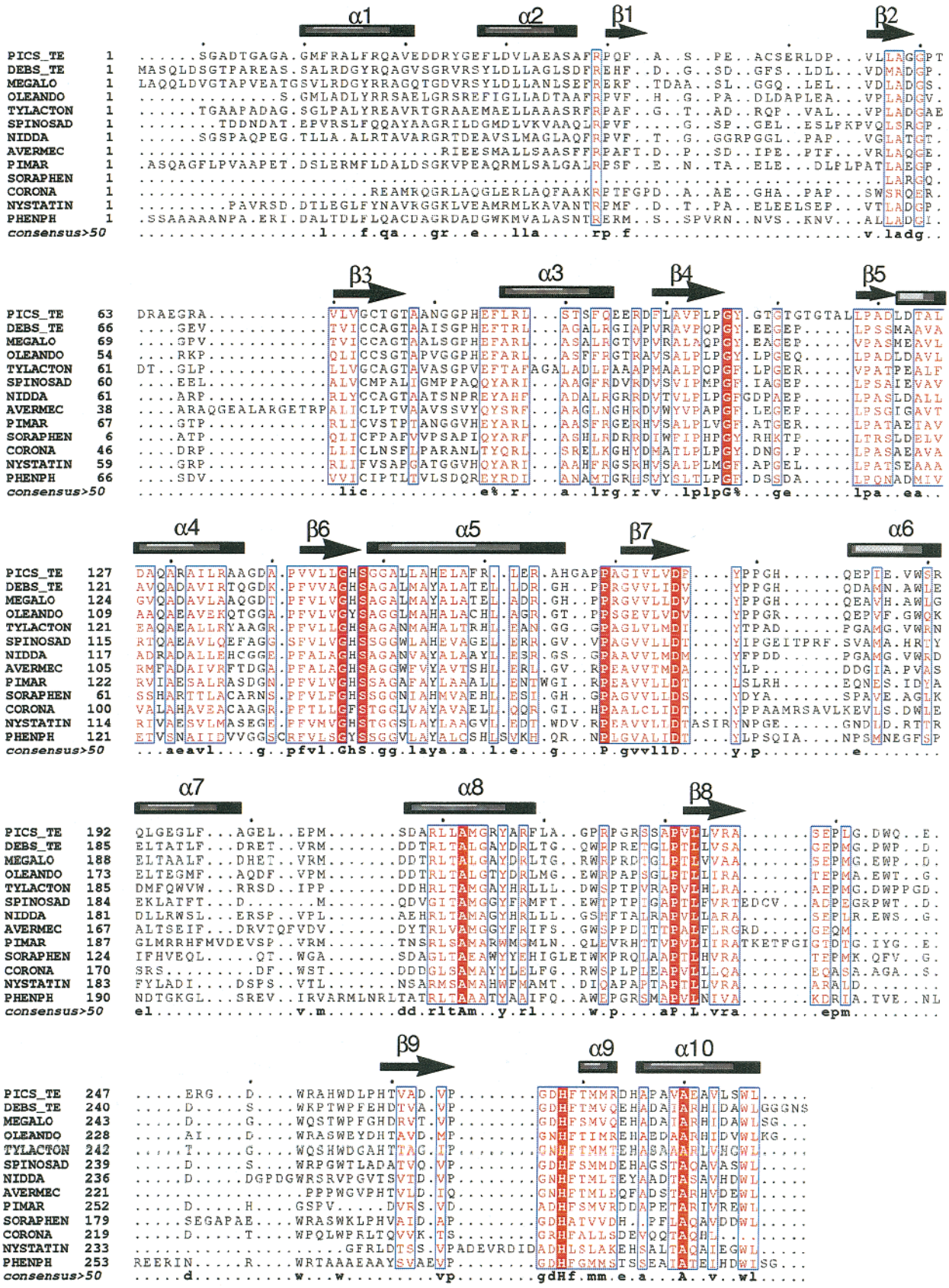

FIGURE 4: Sequence alignments of PKS TEs. Symbols denote the final macrolide product for the corresponding modular PKSs: PICS_TE, methymycin, neomethymycin, narbomycin, and picromycin; DEBS_TE, 6-deoxyerythronolide; MEGALO, megalomycin; OLEANDO, oleandomycin; TYLACTON, tylactone (precursor of tylosin); SPINOSAD, spinosyn aglycone; NIDDA, niddamycin; AVERMEC, avermectin; PIMAR, pimaricin; SORAPHEN, soraphen; CORONA, coronafacic acid; NYSTATIN, nystatin; PHENPH, phenylphthiocerol.

the highly acidic nature of the substrate channel. Effectively, channel geometry would be controlled by electrostatic repulsion, which would be enhanced at higher pH due to the deprotonation of the His and/or Asp/Glu residues.

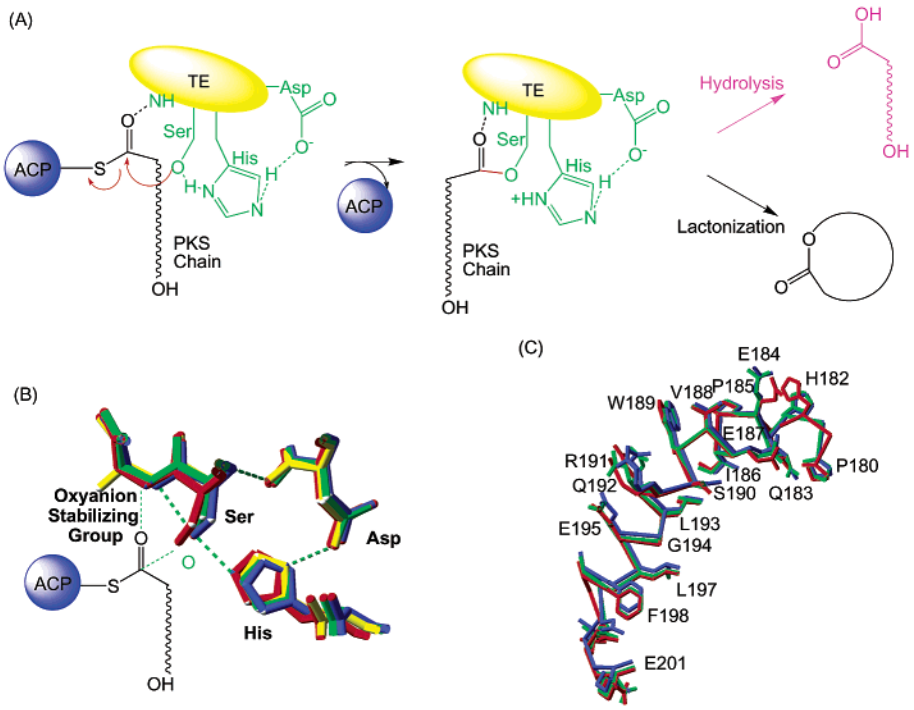

FIGURE 5: (A) Mechanism of the thioesterase-catalyzed reaction. His268, stabilized by Asp176, acts as the catalytic base which accepts the proton from Ser148, as it attacks the ACP6-bound thioester substrate, to yield an acyl−*O*-serine intermediate. The negative charge developed on the thioester during the nucleophilic attack may be stabilized by the tentative oxyanion-stabilizing group (Gly149). The acyl−enzyme intermediate is subsequently attacked either by water (hydrolysis) or by a hydroxyl group on the polyketide chain (cyclization). (B) Near-perfect overlap of the catalytic triad and oxyanion-stabilizing group among DEBS TE (pH 7.2 in yellow) and PICS TE (pH 7.6 in green, pH 8.0 in blue, and pH 8.4 in red), suggesting a similar catalytic mechanism for DEBS TE and PICS TE. (C) An overlap of the region of helices 6 and 7 of PICS TE reveals that most of the structural variance came from ionic residues (pH 7.6 in green, pH 8.0 in blue, and pH 8.4 in red). This figure was generated using Swiss PDB Viewer rendered with POV-Ray (*30*).

Alternatively, the helical region may assume multiple conformations, and the observed variability in channel geometry as a function of pH may reflect the pH-dependent shift in equilibrium among different conformers.

The effect of pH and the resulting variable substrate channel could influence catalysis in multiple ways. Our observation that the positions of the catalytic triad and tentative oxyanion group remain unchanged among DEBS TE and PICS TE (Figure 5B) supports the assumption that the catalytic mechanism does not vary significantly at different pHs or in different TEs. Under such circumstances, a wider channel might allow for improved access of water and substrate thioesters to the active site, as well as improved release of the product. This may account for the 1.8−7.0-fold higher k_{cat}/K_M of PICS TE versus that of DEBS TE at pH 8.0 for the hydrolysis of artificial substrates (*12*). Further, on the basis of a combinatorial effect of catalysis, substrate binding, and product release, one would also predict that the hydrolytic activity of PICS TE should increase with increasing pH due to the effect of a wider channel and increasing nucleophilicity of the active site serine and/or histidine residues. Indeed, as demonstrated in the companion paper (*12*), the initial hydrolysis rate of a representative diketide substrate increases with increasing pH.

In addition to affecting hydrolysis, the variable channel geometry could also influence macrocyclization. To investigate this possibility, we docked the 14-membered macrolides, narbonolide and 6-deoxyerythronolide B (Figure 1), and a 16-membered macrolide, tylactone, into the substrate channels of the different structures of PICS TE, using DOCK (*18*) and using manual docking followed by energy minimization. In contrast to our earlier results with the DEBS TE, where the docked 6-dEB assumed a unique orientation in the active site, each of these macrolides was found to fit in the active sites of all three PICS TE structures loosely. The unique orientation of docked 6-dEB in the active site of DEBS TE arose from multiple hydrogen bonds between the macrocycle and the protein, which also shielded the active site completely from water. In contrast, only one or two residues in the PICS TE structures were predicted to form hydrogen bonds with one or two functional groups in the docked polyketides. Notwithstanding this predicted loose fit, PICS TE is an effective cyclase for its cognate hexaketide and heptaketide substrates (Figure 1). Perhaps this ability to catalyze macrocycle formation arises from the flexibility in the region of helices 6 and 7, which has the highest *B* factor of all the PICS TE structures. This flexibility may allow the protein to wrap around polyketide substrates of alternative sizes (an "induced fit"), thereby shielding the acyl chain from bulk water.

Sequence differences between DEBS TE and PICS TE in this helical region could also contribute to differences in substrate specificity. For example, Asn180 and Glu184 in DEBS TE, which were predicted to interact with the C-3 OH and C-5 OH groups of 6-dEB, respectively, are replaced with Glu187 and Arg191 in PICS TE. It was found that both DEBS TE and PICS TE prefer a 3-keto thioester substrate over similar 2-methyl-3-hydroxyl thioesters; however, PICS TE has a relatively higher affinity for the 3-keto thioester than DEBS TE (*12*). Further studies on the hydrolytic and cyclization specificities of wild-type and mutant TEs should be useful in elucidating the roles of these residues (*12*).

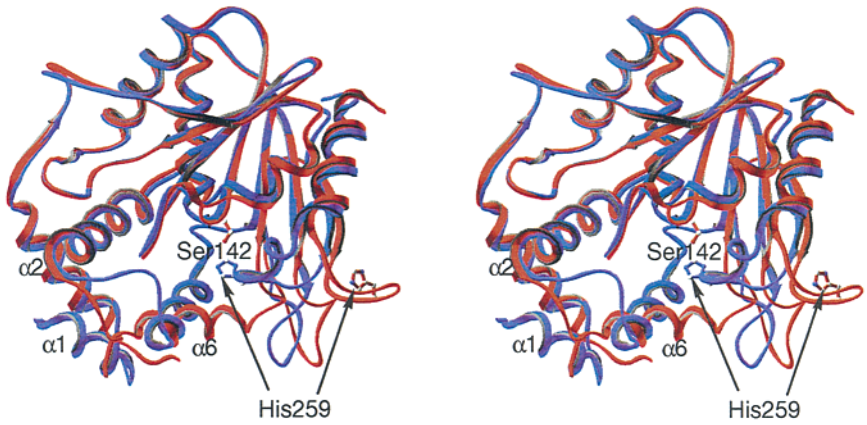

FIGURE 6: Comparison of the pH 7.2 (in blue) and pH 8.5 (in red) structures of DEBS TE. In the pH 8.5 form, α-helix 6 swung aside due to negative charge repulsion, resulting in the displacement of one loop of residues 233−244, which in turn displaces the loop of residues 256−267 that contains the active site His259. The enzyme is presumably inactive due to the misplacement of the active site base (His259). Additionally, the loop of residues 256−267 displaces helices α_1 and α_2, resulting in the disorder of α_1 and weakening of the dimer interface, which may decrease the stability of TE. This figure was generated using Swiss PDB Viewer rendered with POV-Ray (*30*).

The physiological pH for modular polyketide synthases and thioesterases should be close to pH 7.0. Presumably, at this pH, the substrate channel geometry will be closer to that of the smaller channel observed for DEBS TE (pH 7.2). However, the macrocycle-forming thioesterases could be synthetically useful catalysts for macrocycle formation; in such situations, pH is an experimental parameter that could be altered for optimal catalytic activity. Our observation of the pH effect on the substrate channel should therefore be taken into consideration when evaluating the utility of the macrocycle-forming thioesterases as self-standing biocatalysts.

A Second Form of DEBS TE Is Likely To Be Inactive. A new crystal form of DEBS TE was grown at pH 8.5 in 4 months. The structure of this crystal form was determined by molecular replacement of the pH 7.2 DEBS TE dimer with the deletion of α-helix 6. In this structure, the active site triad residues, Asp169 and His259, are no longer hydrogen-bonded to the active site Ser142 (Figure 6); therefore, this crystal form is most likely of an inactive state. Like the pH 7.2 form, the pH 8.5 form of DEBS TE is also a dimer with an open substrate channel. However, the substrate channel is much more open in the pH 8.5 form (Figure 3).

A detailed comparison between the structures of the two forms of DEBS TE has shed light on how the enzyme is inactivated over time at elevated pH. In this crystal form, α-helix 6 has swung aside, resulting in displacement of the loop of residues 233−244, which in turn displaces the loop of residues 256−267 containing the active site His259 (Figure 6). The enzyme therefore appears to be inactivated due to the misplacement of the active site base His259. Additionally, movement of α-helix 6 also appears to induce disorder in helices α_1 and α_2, a feature that might weaken the dimer interface. As shown in Figure 3, when calculated surface charges are mapped onto the molecular surface of the inactive DEBS TE, the substrate channel is highly electronegative. As suggested above for PICS TE, perhaps the enhanced electrostatic repulsion in the protein is responsible for pushing helix 6 aside. Interestingly, kinetic analysis of DEBS TE at higher pH using a diketide substrate yielded k_{cat}/K_M values of 5.2 and 3.8 M^{-1} s^{-1} at pH 8.0 and 9.0, respectively (*12*), indicating that DEBS TE is active at pH 9.0. Therefore, the observed inactive crystal structure of DEBS TE at pH 8.5 is presumably due to a combined effect of pH and time.

Structural Elements that Affect the Substrate Channel Geometry of Thioesterases. To date, seven structures of thioesterases have been reported. Two of these belong to the 4-hydroxybenzoyl CoA TE from *Pseudomonas* (*19*) and thioesterase II from *E. coli* (*20*). They share a distinct fold and catalyze hydrolysis via a mechanism different from that of the macrocycle-forming TEs. The remaining five TE structures include the myristoyl acyl carrier protein thioesterase from *Vibrio harveyi* (*21*), the palmitoyl protein thioesterase (*22*), the human acyl protein thioesterase (*23*), and recently the macrocycle-forming thioesterases from polyketide [DEBS TE (*11*)] and nonribosomal peptide [srf TE (*24*)] biosynthetic pathways. Although their sequences are less than 10% identical, these five TEs have the same α,*β*-hydrolase fold with an rmsd of 1.5−3.7 Å in the core region. Further, their catalytic triad can be overlapped, indicating that a similar catalytic mechanism is shared among the five TEs. However, their substrate specificity varies significantly, ranging from aliphatic linear chains (fatty acid) (*22*) to nonribosomal peptides (surfactin) (*24*). An understanding of the structural basis for divergent substrate specificity among these TEs is of fundamental and biocatalytic interest.

By overlapping the crystal structures of the five TEs described above, it was found that the substrate binding pocket of TEs is defined by the "cap" region (*16*) that corresponds to α-helices 6 and 7 of DEBS TE. Insertion of an extra loop in this region creates a tighter binding pocket for myristoyl acyl carrier protein thioesterase and palmitoyl protein thioesterase, resulting in a preference for linear aliphatic chains. In fact, it was found that the aromatic serine-protease inhibitor phenylmethanesulfonyl fluoride (PMSF) cannot inhibit palmitoyl TE due to the narrowness of its substrate pocket (*25*). The alteration of channel geometry via the cap region (helices 6 and 7 of DEBS TE) is even more significant in the case of the macrocycle-forming srfTE of the nonribosomal peptide synthesis pathway, the crystal structure of which contains two monomers per asymmetric unit. These monomers represent open and closed conformations of the enzyme, but differ only in one helix that corresponds to helix 6 of DEBS TE. Instead of forming a

channel, the substrate pocket of srfTE is sealed into a bowl shape by several hydrophobic residues. The above comparison demonstrates the versatility of the region of helices 6 and 7 in the macrocycle-forming TEs. Via alteration of this region, it should be possible to engineer the substrate specificity and regiospecificity for cyclization as well as the partitioning between hydrolysis and cyclization.

pH Dependence of Channel Proteins. In this study, we are able to observe the effect of pH on the substrate channel of a new class of enzymes of biosynthetic significance, the macrocycle-forming thioesterases of DEBS and PICS. As shown in the cases of the potassium channel (*26*), the diphtheria toxin T domain (*27*), the proton pump of heme-copper oxidases (*28*), and a calcium channel Annexin XII E105K (*29*), the substrate channels of these TE domains were shown to be altered by pH. Like those of TEs, the residues that contribute to the pH dependence of the active site architectures of these other proteins are charged (*26*). As a result, different protonation states of these ionic residues at different pHs result in different conformations in the corresponding channel regions. In particular, repulsive interactions between acidic residues play an important role in conformational changes. This is observed not only for TE but also for the potassium channel (*26*), the diphtheria toxin T domain (*27*), and the calcium channel Annexin XII E105K (*29*). Mutations of these charged residues in the active sites of these channel proteins should enable the alteration of channel geometry; alternatively, these ionic interactions can be utilized as anchors for the design of drugs with an affinity for these protein channels.

The five crystal structures of DEBS TE and PICS TE confirmed our earlier prediction that macrocycle-forming thioesterases in general have an open substrate channel, similar catalytic mechanisms, and a hydrophobic dimer interface. It also offered insight into how electrostatic interactions might gauge the geometry of the active site, from a conformation where the substrate is tightly anchored (crystallized at pH 7.2) to a conformation that appeared to be inactive (crystallized at pH 8.5 in 4 months). In addition to providing a structural basis for altering substrate specificity, these structural biological advances provide an improved basis for predicting the properties of a growing number of functionally uncharacterized proteins from this family of thioesterases that continue to appear in genetic databases.

ACKNOWLEDGMENT

Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, the National Center for Research Resources, the Biomedical Technology Program, and the National Institute of General Medical Sciences. Portions of this research were carried out at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

REFERENCES

1. Cane, D. E., Walsh, C. T., and Khosla, C. (1998) *Science 282*, 63−68.
2. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) *Science 252*, 675−679.
3. Cortes, J., Wiesmann, K. E. H., Roberts, G. A., Brown, M. J., Staunton, J., and Leadlay, P. F. (1995) *Science 268*, 1487−1489.
4. Xue, Y., and Sherman, D. H. (2000) *Nature 403*, 571−575.
5. Xue, Y., and Sherman, D. H. (2001) *Metab. Eng. 3*, 15−26.
6. Chen, S., Xue, Y., Sherman, D. H., and Reynolds, K. A. (2000) *Chem. Biol. 7*, 907−918.
7. Kao, C. M., Luo, G., Katz, L., Cane, D. E., and Khosla, C. (1995) *J. Am. Chem. Soc. 117*, 9105−9106.
8. Jacobsen, J. R., Hutchinson, C. R., Cane, D. E., and Khosla, C. (1997) *Science 277*, 367−369.
9. Gokhale, R. S., Hunziker, D., Cane, D. E., and Khosla, C. (1999) *Chem. Biol. 6*, 117−125.
10. Rowe, C. J., Bohm, I. U., Thomas, I. P., Wilkinson, B., Rudd, B. A., Foster, G., Blackaby, A. P., Sidebottom, P. J., Roddis, Y., Buss, A. D., Staunton, J., and Leadlay, P. F. (2001) *Chem. Biol. 8*, 475−485.
11. Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C., and Stroud, R. M. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 14808−14813.
12. Lu, H., Tsai, S. C., Khosla, C., and Cane, D. E. (2002) *Biochemistry 41*, 12590−12597.
13. Otwinoswski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
14. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) *Acta Crystallogr. D55*, 484−491.
15. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
16. Nardini, M., and Dijkstra, B. W. (1999) *Curr. Opin. Struct. Biol. 9*, 732−737.
17. Schrag, J. D., and Cygler, M. (1997) *Methods Enzymol. 284*, 85−107.
18. Ewing, T. J., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001) *J. Comput.-Aided Mol. Des. 15*, 411−428.
19. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) *J. Biol. Chem. 273*, 33572−33579.
20. Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z. S. (2000) *Nat. Struct. Biol. 7*, 555−559.
21. 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.
22. Bellizzi, J. J., III, Widom, J., Kemp, C., Lu, J. Y., Das, A. K., Hofmann, S. L., and Clardy, J. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 4573−4578.
23. Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L., and Derewenda, Z. S. (2000) *Struct. Folding Des. 8*, 1137−1146.
24. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., and Stubbs, M. T. (2002) *Struct. Folding Des. 10*, 301−310.
25. Das, A. K., Bellizzi, J. J., III, Tandel, S., Biehl, E., Clardy, J., and Hofmann, S. L. (2000) *J. Biol. Chem. 275*, 23847−23851.
26. Meuser, D., Splitt, H., Wagner, R., and Schrempf, H. (1999) *FEBS Lett. 462*, 447−452.
27. Mindell, J. A., Silverman, J. A., Collier, R. J., and Finkelstein, A. (1994) *J. Membr. Biol. 137*, 29−44.
28. Das, T. K., Tomson, F. L., Gennis, R. B., Gordon, M., and Rousseau, D. L. (2001) *Biophys. J. 80*, 2039−2045.
29. Cartailler, J. P., Haigler, H. T., and Luecke, H. (2000) *Biochemistry 39*, 2475−2483.
30. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714−2723.
31. Nicholls, A., and Honig, B. (1991) *J. Comput. Chem. 12*, 435−445.

BI0260177