Structural and Biochemical Characterization of the Salicylaldehyde-acyltransferase SsfX3 from a Tetracycline Biosynthetic Pathway*§^

Received for publication, September 1, 2011, and in revised form, September 28, 2011 Published, JBC Papers in Press, September 30, 2011, DOI 10.1074/jbc.M111.299859

Lauren B. Pickens†§, Michael R. Sawaya†§, Huma Rasool†, Inna Pashkov§, Todd O. Yeates§, and Yi Tang†§

From the Departments of †Chemical and Biomolecular Engineering and §Chemistry and Biochemistry, ‡Molecular Biology Institute, *Howard Hughes Medical Institute, and UCLA-DOE Institute for Genomics and Proteomics, UCLA, Los Angeles, California 90095

Background: SsfX3 is an acyltransferase that can acylate tetracycline-like molecules at C-4.

Results: Crystal structure showed SsfX3 contains a structural N-terminal β-sandwich domain and a catalytic C-terminal hydrolase domain. Mutagenesis revealed both are essential for activity.

Conclusion: The N-terminal domain is recruited to bind the large “small molecule” substrate.

Significance: New structural knowledge enables engineering of the acyltransferase toward synthesis of tetracycline analogs.

SsfX3 is a GDSL family acyltransferase that transfers salicylate to the C-4 hydroxyl of a tetracycline intermediate in the penultimate step during biosynthesis of the anticancer natural product SF2575. The C-4 salicylate takes the place of the more common C-4 dimethylamine functionality, making SsfX3 the first acyltransferase identified to act on a tetracycline substrate. The crystal structure of SsfX3 was determined at 2.5 Å, revealing two distinct domains as follows: an N-terminal β-sandwich domain that resembles a carbohydrate-binding module, and a C-terminal catalytic domain that contains the atypical α/β-hydrolase fold found in the GDSL hydrolase family of enzymes. The active site lies at one end of a large open binding pocket, which is spatially defined by structural elements from both the N- and C-terminal domains. Mutational analysis in the putative substrate binding pocket identified residues from both domains that are important for binding the acyl donor and acceptor. Furthermore, removal of the N-terminal carbohydrate-binding module-like domain rendered the stand-alone α/β-hydrolase domain inactive. The additional noncatalytic module is therefore proposed to be required to define the binding pocket and provide sufficient interactions with the spatially extended tetracycline substrate. SsfX3 was also demonstrated to accept a variety of non-native acyl groups. This relaxed substrate specificity toward the acyl donor allowed the chemoenzymatic biosynthesis of C-4-modified analogs of the immediate precursor to the bioactive SF2575; these were used to assay the structure activity relationships at the C-4 position.

Natural products display an incredibly large spectrum of chemical diversity, in large part due to the many combinations of tailoring enzymes that decorate their scaffolds. Group transferases such as glycosyltransferases, methyltransferases, acyltransferases, aminotransferases, and prenyltransferases play key roles in the biosynthesis of many important natural products, such as doxorubicin and erythromycin, and have been successfully used for the combinatorial biosynthesis of unnatural analogs (1). The acyl transfer modification has in many cases been shown to be vital to arming the final products with bioactivity, as seen with the biosynthesis of teicoplanin (2, 3), chromomycin (4), phascolactomin (5), and lovastatin (6). Acyltransferases frequently show broad substrate flexibility with regard to the donor acyl group, and therefore they can be valuable tools to diversify natural product scaffolds at positions important for bioactivity.

SF2575 is a tetracycline natural product produced by Streptomyces sp. SF2575. It has potent anticancer activity (7, 8). SF2575 is structurally unique among tetracyclines due to its extensive tailoring modifications, including glycosylation with D-olivose at C-9, acylation of the C-4 hydroxyl with salicylate, and acylation of the C-4′ of D-olivose with angelate. The sf gene cluster encoding these activities was identified recently, enabling the biosynthetic pathway for SF2575 to be articulated (Fig. 1) (9). Transfer of the salicylic and the angolic groups to the aglycon of SF2575 are both important for the anticancer activity of SF2575 (8). Understanding the enzymatic basis for these transfer group additions offers the possibility of structure-activity relationship studies and subsequent fine-tuning of the pharmacological properties of SF2575. Toward this end, we previously identified SsfX3 as the acyltransferase responsible for transfer of salicylate from salicylate-CoA 5 to the C-4 hydroxyl of the key intermediate 4 to produce the C-4-(R)-salicylated 6 (Fig. 1) (9). The salicylic moiety transforms the inactive 4 into the moderately potent cytotoxic compound 6.

Engineered biosynthesis of new tetracycline compounds has been limited in part by the dearth of specific modifying enzymes. SsfX3 therefore represents a possible tool to specifically decorate the C-4 position of the tetracycline aglycon. This
prompted us to further study the structural and mechanistic features of this enzyme.

A protein family database search revealed that the C-terminal domain of SsfX3 has low but recognizable sequence similarity to the GDSL hydrolase family of enzymes (with an E value of 0.071). This family was first suggested by Upton and Buckley (10), who identified a conserved active site motif at the N terminus of the enzyme (rather than the middle as observed for typical α/β-hydrolases), as well as identifying five conserved sequence blocks. This family of enzymes contains the conserved GDS(S)L(S) motif around the active serine, lacks a well defined nucleophilic elbow or sharp turn identified by the canonical GXGXG motif (11), and has four invariantly conserved residues making up the oxanion hole, Ser, Gly, Asn, and His; these features allowed the additional protein to be classified as members of the SGNH hydrolase family of enzymes (12).

One of the best characterized GDSL hydrolases is *Escherichia coli* thioesterase I (TAP), a versatile enzyme known to function as a thioesterase, esterase, arylesterase, protease, and lysophospholipase (13). Its promiscuity led TAP to be identified as three different proteins, thioesterase I (TesA) (14), protease I (Apel) (15), and lysophospholipase L1 (PlcC) (16) before Nojima and co-workers (17) found that they were identical molecules encoded by the same gene. The protein family database currently identifies 6817 sequences belonging to the GDSL hydrolase family (18). Several of these have been structurally characterized, including *E. coli* TAP (13, 19, 20), *Streptomyces scabies* SsEst (21), rhamnogalacturonan acylesterase from *Aspergillus aculeatus* (22), and esterase EstA from *Pseudomonas aeruginosa* (23). GDSL hydrolases are also observed to be promiscuous with regard to substrate.

Asler and co-workers (24) tested the activities of several known GDSL hydrolases toward a variety of esterase, lipase, thioesterase, phospholipase, and protease substrates and found each of the hydrolases tested to have a unique activity profile for a variety of substrates. Huang and co-workers (25, 26) performed NMR spectroscopy on the TAP enzyme and showed that the active site is highly flexible, which could contribute to the broad substrate specificity.

The GDSL hydrolase consensus sequence mapped only to the C-terminal half of the SsfX3 protein, leaving the function of the N-terminal half unknown. A number of homologs of SsfX3 were identified based on amino acid sequence similarity using BLAST, most of them with uncharacterized functions. Only four SsfX3 homologs from verified natural product gene clusters have been identified to date. The closest homolog in the NCBI database is AviX9, a putative GDSL hydrolase with unknown function from the avilamycin gene cluster (27). Recently, GDSL hydrolases have also been identified in the caprazamycin (28), related liposidomycin (29), and A-90289 (30) gene clusters. These enzymes have been proposed to be involved in the transfer of long chain fatty acyl groups to form liponucleoside antibiotics. Sequence similarity of these acyltransferases to SsfX3 is evident across both the novel N-terminal and the GDSL-like C-terminal domains. To gain further insight into the role of these enzymes as atypical acyltransferases in natural product biosynthesis, we determined the x-ray crystal structure of SsfX3 and performed structurally guided mutagenesis to probe key residues around the active site and to identify residues important for catalysis and substrate binding. The three-dimensional structure confirmed that SsfX3 is indeed composed of two distinct domains separated by a long linker. The N-terminal domain bears resemblance to a carbohydrate-binding module (CBM) domain and is critical for substrate binding, whereas the catalytic C-terminal domain

---

**FIGURE 1. Abbreviated biosynthetic pathway of SF2575 showing key intermediates and tailoring steps.** The upstream sequence of reactions is proposed based on the structures of intermediates 1–3.
displays the α/β-hydrolase fold typical of the GDSL hydrolase family.

**EXPERIMENTAL PROCEDURES**

**SsfX3 and Mutant Plasmid Construction**—The gene encoding SsfX3 was amplified by PCR from cosmID 5F15 (9) and ligated to pCR-Blunt vector (Invitrogen) to generate plasmid pLP59. The gene was excised from pCR-Blunt and cloned into pET28a (Invitrogen) using Ndel and EcoRI restriction sites. SsfX3 mutants were prepared by ligating pLP59 as a template. Full-length PCR products were ligated into pCR-Blunt vector and then excised using Ndel and EcoRI and ligated into pET28a. All genes were confirmed by DNA sequencing (Laragen) prior to cloning into pET28a. Primers used for amplification and splicing by overlap extension are given in [supplemental Table S1](#). Expression plasmids were then transformed into *E. coli* BL21(DE3) for protein expression and purification.

**Heterologous Expression and Purification of Enzymes**—SsfX3, SsfL1, and SsfX3 mutants used for kinetic assays and substrate preparation were expressed from *E. coli* BL21(DE3) as fusion proteins with an N-terminal hexahistidine tag. The C-terminal domain of SsfX3 was expressed from pET24 as a fusion protein with a C-terminal hexahistidine tag. *E. coli* BL21(DE3) transformants containing the plasmid of interest were used to inoculate a 5-ml culture containing LB medium supplemented with 35 mg/liter kanamycin. The seed culture was grown overnight at 37 °C and added to a fresh 2-liter flask containing 500 ml of LB medium with 35 mg/liter kanamycin. The seed culture was grown at 80 °C. Selenomethionyl SsfX3 was expressed by the inhibition of the methionine biosynthetic pathway in minimal media supplemented with selenomethionine as described by Doublie (31). Protein purification was carried out at 4 °C. The cell pellets were harvested by centrifugation and resuspended in Buffer A (50 mM Tris-HCl, pH 7.9, 10 mM imidazole, and 50 mM NaCl). Cell membranes were disrupted by sonication. The cell lysate was centrifuged at 16,000 rpm, and the soluble fraction was collected and incubated with nickel-nitrilotriacetic acid resin (GE Healthcare) for 30 min with gentle rotation. Protein/resin mixture was added to a gravity flow column, and buffers containing increasing concentrations of imidazole were applied stepwise. The target proteins were eluted in Buffer A containing 250 mM imidazole. An Amicon filtration column (Millipore) was used for buffer exchange and concentration of the protein solution. Purified enzymes were stored in Buffer B (50 mM Tris-HCl, pH 7.9, 2 mM EDTA, 2 mM DTT) with 10% glycerol for cryopreservation. Proteins were aliquoted at 0.5 mM concentration, flash-frozen on dry ice, and stored at −80 °C. Protein concentrations were measured by a Bradford assay using bovine serum albumin (BSA) as a standard (32).

**Crystallization of SsfX3**—Purified SsfX3 migrated as two bands when loaded onto an SDS-polyacrylamide gel, possibly due to modification of the surface cysteine. To facilitate crystallization, the nonconserved Cys-68, which resides in the N-terminal domain, was mutated to histidine. The C68H mutant protein migrated as a single band on SDS-PAGE and subsequently afforded the highest quality crystals for x-ray data collection. Kinetic assays verified that the mutation had no impact on the enzyme function. Protein crystals were grown by the hanging drop vapor diffusion method at room temperature. Native crystals were grown from a reservoir containing 3.3 M sodium formate, pH 7.0, with an additive solution giving a final concentration of 0.01 M reduced l-glutathione and 0.01 M oxidized l-glutathione in the drop. Selenomethionine crystals were grown from a reservoir of 3.3 M sodium formate, pH 7, and a drop containing 2.0 M NDSB-201 (3-(1-pyridinio)-1-propane sulfonate) from the Hampton Research additive screen. Crystal trays were set up with a 1.0-μl drop volume by a TTP Labtech Mosquito robot. The crystals appeared as rectangular plates with the longest dimension less than 80 μm. The crystals required approximately 2 weeks to reach their full size.

**Data Collection**—X-ray diffraction data from the native crystal were collected at the Advanced Light Source, beamline 8.2.1, using an ADSC Quantum 315 3X3 CCD array. The crystal was cooled to 100 K in a cryogenic nitrogen stream for data collection. No additional cryo-protectant was included. Each 5-s exposure covered 1° of crystal rotation. The incident beam was not attenuated. A total of 120 images was used in the reported data set, which extended to 2.5 Å resolution (Table 1). Diffraction data from two selenomethionine derivatives were collected at the Advanced Photon Source, beamline 24-ID-C, using the same model detector. Both selenomethionine derivative crystals were cryo-protected by 10-s swipes through a solution containing 7.0 μl of reservoir solution and 3.0 μl of 100% glycerol and then cooled to 100 K in a cryogenic nitrogen stream for data collection. Each exposure was collected for 1 s and covered 1° of crystal rotation. Data from the first selenomethionine crystal were collected using a 70-μm diameter aperture and 5% transmission. Data from the second selenomethionine crystal were collected using a 30-μm aperture and 10% transmission. A total of 100 images was used for the data set at the selenium anomalous inflection wavelength, and 82 images were used in the high remote wavelength data set. The inverse beam mode was not used for either data set. Data reduction and scaling for all crystals were performed using DENZO/SCALEPACK (33). Data collection statistics are reported in Table 1.

**Structure Determination and Refinement**—In a multiwavelength anomalous dispersion (MAD) experiment, three data sets at three different wavelengths were customarily collected from a single crystal to ensure isomorphism. However, because of their small sizes, none of the crystals could withstand the radiation-induced damage caused by a three-wavelength experiment. An attempt to make do with a single wavelength data set by calculating phases using the single wavelength anomalous dispersion method did not result in an interpretable electron density map. Instead, a MAD data set was assembled from two crystals as follows: a data set collected at the peak wavelength using one crystal, in combination with data collected at the inflection, and high remote wavelengths of a second crystal (Table 1). The two selenomethionine crystals were sufficiently isomorphous as indicated by the correlation coefficients between signed anomalous differences reported by SHELXC in the program HKL2MAP (34).
The correlation ranged between 0.55 and 0.25 over a resolution range from 20.0 to 3.5 Å. Coordinates for 7 of 14 possible selenium sites were obtained by SHELXD using both dispersive and anomalous differences from the three data sets (35).

Phases were calculated using MLPHARE and modified by solvent flattening and histogram matching using DM from the CCP4 suite of programs (36). The inflection data set was used as the reference wavelength. Three additional selenium sites could be identified in self-anomalous difference Fourier maps. These were included in another round of phasing. The resulting map was of sufficient quality to allow the manual placement of several β-strands and α-helices with the program COOT (37). At this point, it was possible to recognize a structural similarity with the carbohydrate esterase from Clostridium thermocellum (PDB code 2WAB) (38). Swiss-Model (39) had suggested its use as the basis for building a homology model of SsfX3 before phases could be obtained. Coordinates of PDB code 2WAB had been used in attempts to solve the structure by the molecular replacement method but without success. In retrospect, the failure was probably due to the low sequence identity between SsfX3 and 2WAB (less than 16%). With newly acquired phases, two molecules of 2WAB could be placed in the asymmetric unit using the phased rotation and translation function employed in the program MOLRE (40). Phases were further improved by 2-fold symmetry averaging in DM using the coordinates of the properly placed 2WAB molecules to construct a molecular mask using the program MAMA (41). The initial NCS matrix was obtained by least squares superpositioning of NCS-related selenium atom coordinates using the program O (42) and then improved using Kleywegt’s program IMP (43). Statistics for experimental phases calculated to 2.9 Å are reported in Table 1.

The model for SsfX3 was initially refined with the PHENIX (44)-simulated annealing algorithm and with REFMAC5 using tight 2-fold noncrystallographic symmetry restraints (45). Later, TLS parameterization of domain disorder was employed (46). After each refinement step, the model was visually inspected in Coot, using both 2Fo − Fe and Fo − Fe difference maps. All hydrogen atoms connected to carbon atoms and backbone nitrogen atoms were included at their geometrically calculated positions and refined using a riding model. Finally, the model was refined using Buster (47).

All models were validated with the following structure validation tools: PROCHECK (48), ERRAT (49), and VERIFY3D (50). PROCHECK reports that 90.6% of the residues are in the most favored region of the Ramachandran plot; 9.1% of the residues are in additionally allowed regions; Ala-207 from chain B is in a generously allowed region, and Asp-173 is in a disallowed region. Asp-173 appears in clear electron density. Its unfavorable Ramachandran angles appear to be stabilized by eight hydrogen bonds with neighboring residues. It immediately precedes the active site Ser-174 and might be involved in catalysis. ERRAT reported that 96.9% of the residues were within the 95% certainty limit for rejection. The coordinates of the final model and the merged structure factors have been deposited with the Protein Data Bank. The corresponding PDB code is 3SKV.

Size Exclusion Chromatography—Size exclusion chromatography analysis of SsfX3 was performed on a Superdex 75 (10/300 GL) (GE Healthcare) at 4 °C using a running buffer consisting of 50 mM sodium phosphate, pH 7.0, and 150 mM NaCl.

Preparation of Tetracycline Substrates—Compound 4 was obtained by incubation of a 10 mg/ml solution of SF2575 in 1.0 M NaOH for 15 h. To isolate compound 4, the reaction mixture was adjusted to pH 7 by addition of 6.2 M HCl, and the products were extracted with ethyl acetate/acetate (99:1%). Following evaporation of the solvent, the product was redissolved in methanol and purified by HPLC (Alltech Altima reverse phase column 5 μm, 10 × 250 mm) with an isocratic gradient of 78% CH3CN in water (0.1% TFA). SF2575 intermediates 1 and 2 were generously provided by Peng Wang (UCLA).

Preparation of Aryl-CoA Substrates—Aryl-CoA substrates were prepared enzymatically using salicylyl-CoA ligase SsfL1. Reactions were set up and contain the following: 50 mM HEPES, pH 7.9, 10 mM MgCl2, 5 mM coenzyme-A, 6 mM acid substrates, 5 mM ATP, and 20 μM SsfL1. Reactions were incubated for at least 2 h at room temperature. The reaction mix was then filtered to remove precipitated protein and directly purified by HPLC (Alltech Altima reverse phase column 5 μm, 10 × 250 mm) with a gradient of CH3CN in water (0.1% TFA). The gradient varied depending on the reaction as some products were more hydrophilic than others, requiring adjustment to the method. Butyryl-CoA lithium salt was purchased from Sigma and used directly without further purification.

Kinetic Assays—The assays were performed at room temperature in 50 mM HEPES, pH 7.9, with 10 mM MgCl2. To determine the Km value of compound 4, the concentration of 4 was varied from 5 to 300 μM keeping concentration of compound 5 constant at 300 μM or at least 8-fold higher than observed Km value for compound 5. To determine the Km value of compound 5, the concentration of 5 was varied from 5 μM to 2 mM keeping the concentration of compound 4 constant at 50 μM or 10-fold higher than observed Km value for compound 4. Concentration of SsfX3 mutants were optimized for each to be between 10 and 100 μM depending on the rate of reaction. A 1.2-ml reaction volume was set up, and aliquots of 240 μl were removed and quenched by extraction with 300 μl of ethyl acetate (1% acetic acid) at the time points (30, 60, 120, and 180 s). Extracts were separated by centrifugation, dried, and analyzed by HPLC. The amount of product in each aliquot was determined by comparing the integrated HPLC peak area to a standard curve prepared with known product concentrations. The resulting initial velocity data were fit to the Michaelis-Menten equation to compute the resulting kinetic parameters, kcat and Km.

Chemoenzymatic Preparation of SF2575 Analogs—Analogs of compound 6 were prepared by enzymatic reaction. Six 1-ml reactions were prepared as follows: 50 mM HEPES, pH 7.9, 10 mM MgCl2, 0.5 mM compound 4, 2 mM free coenzyme-A, 2 mM ATP, 2 mM acid substrate, 15 μM SsfL1, and 5 μM SsfX3. Reactions were incubated overnight and extracted with ethyl acetate containing 1% acetic acid. Organic extract was dried, and the final product was purified from the redissolved extract by HPLC.
TABLE 1

Statistics of data collection, phasing, and refinement

The highest resolution shell is shown in parenthesis.

| Data collection | Selenomethionine derivative 1 | Selenomethionine derivative 2 |
|-----------------|-------------------------------|-------------------------------|
| Space group     | P2_12_12                      | P2_12_12                      |
| Cell dimensions |                               |                               |
| a, b, c (Å)     | 66.2, 84.4, 127.9             | 65.7, 83.8, 127.0             |
| a, b, c (°)     | 90.0, 90.0, 90.0              | 90.0, 90.0, 90.0              |
| Resolution (Å)  | 1.000                         | 0.979                         |
| Rmerge (%)      | 0.068 (0.540)                 | 0.089 (0.431)                 |
| Completeness (%)| 96.7 (100.0)                  | 98.5 (100.0)                  |
| Redundancy      | 4.6 (4.8)                     | 4.7 (4.8)                     |
| Phasing         |                               |                               |
| Cells Rmerge    | 0.94                          | 0.96                          |
| Cells Rmap     | 0.040 (0.91)                  | 0.070 (0.95)                  |
| Phasing power  | 0.68 (0.80)                   | 0.40 (0.34)                   |
| Refinement      |                               |                               |
| Resolution (Å)  | 2.5                           | 2.5                           |
| No. reflections | 25,411                        | 25,411                        |
| Rmerge (%)      | 0.170 (0.227)                 | 0.170 (0.227)                 |
| No. atoms      | 5,555                         | 5,555                         |
| Protein        | 12                             | 12                             |
| Water           | 184                            | 184                            |
| I-B factors    | 43.8                           | 43.8                           |
| Protein        | 53.5                           | 53.5                           |
| Water           | 40.7                           | 40.7                           |
| R.m.s deviations | 0.019                        | 0.019                        |
| Bond lengths (Å) | 1.18                        | 1.18                        |
| Bond angles (°) |                               |                               |

Bioactivity Assays—Cell proliferation was determined by 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-(4-sulphophenyl)-2H-tetrazolium assay for adherent HeLa and M249 cells and suspension of Nalm-6 and Jurkat cells. All cells were cultured according to ATCC or DSMZ standards. Jurkat and Nalm-6 cells were seeded at 8 x 10^4 cells/well, and HeLa cells were seeded at 4 x 10^5 cells/well. Compounds of interest dissolved in DMSO were added to wells after 12 h to a final concentration ranging from 10 nM to 10 μM. Cell proliferation was assayed 72 h after treatment using CellTiter 96® AQueous OneSolution cell proliferation assay reagent (Promega) and measured with a PowerWave XS 96-well plate reader at 490 nm. Values were compared with control cells plus 0.1% DMSO.

RESULTS AND DISCUSSION

Structure Determination and Model Quality—The crystal structure of a selenomethionine-containing SsfX3 mutant C68H was solved using the MAD method. The C68H mutant was prepared to eliminate nonspecific oxidation of the free thiol. Size exclusion chromatography was used to determine that SsfX3 was a monomer in solution. Because of severe radiation sensitivity of the crystals, two isomorphous crystals were used for x-ray data collection, one at the peak wavelength and the other at the inflection and remote wavelengths (see under “Experimental Procedures”). Statistics for experimental phases calculated to 2.9 Å are reported in Table 1. The map was of sufficient quality to allow manual placement of several β-strands and α-helices, from which it was possible to recognize structural similarity with the family 2 carbohydrate esterase from Clostridium thermocellum (PDB code 2WAB) (38). With newly acquired phases, two molecules of the esterase could be placed in the asymmetric unit using the phased rotation and translation function. Phases were further improved by 2-fold symmetry averaging, and the final model was refined against native diffraction data to 2.5 Å resolution. The model is mostly complete; of the 384 residues that comprise SsfX3, it was possible to model residues 8–93 and 97–364 in chain A, and residues 12–85, 97–272, and 283–364 in chain B. Refinement statistics are reported in Table 1.

Overall Architecture of SsfX3—From the structure, shown in Figs. 2A and 4A, it is clear that SsfX3 is indeed composed of two domains. A flexible linker divides the N-terminal β-sandwich domain and the C-terminal GDSL hydrolase region as predicted by sequence comparison to the consensus GDSL hydrolase sequence. It also contains a large open binding pocket, which is bounded by the aromatic residues Phe-276, Trp-277, and Tyr-335 from the C-terminal domain at one end and the N-terminal β-sandwich domain on the other end (Fig. 2 and 4). The N-terminal region consists of two short β-strands β1 and β2 followed by a 10-stranded β-sandwich or “jelly roll” fold. Between β4 and β5 in the β-sandwich domain is a loop containing two antiparallel 3_10 helices A and B. The C-terminal domain displays the atypical α/β-hydrolase fold characteristic of GDSL hydrolases (12). The β-strands β13–17 are arranged in a parallel β-sheet with a slight twist. This β-sheet formation is flanked by five α-helices, two on the concave surface (αC and αK) and three on the convex surface (αE, αG, and αI). There is also a short helix αF in the loop connecting β15 and αG, which borders the active site. Another short helix αJ directly follows β17 and precedes the 12-residue loop ending at the general base His-338 in αK.

A structural similarity search using the Dali server (51) revealed five structures having three-dimensional similarity over both the β-sandwich and the α/β-hydrolase-fold domains of SsfX3. All of these were from the CE2 family of carbohydrate esterases, which includes CcCCE2 (38). Despite their low sequence identity (less than 16% for CcCCE2), they were well aligned across both domains with a root mean square deviation (r.m.s.d.) of 2.9 Å and 282/327 residues aligned. These structures were all reported by Montanier et al. (38) as acetyltransferases acting on noncellulosic plant polysaccharides. Additional enzymes were identified that independently mapped to either the β-sandwich or the α/β-hydrolase domains. The C-terminal region, as expected, shares structural similarity with other GDSL hydrolases such as TAP (13) with r.m.s.d. of 2.4 Å and 115/137 residues structurally aligned. Proteins with structural similarity to the N-terminal jelly roll domain of SsfX3 include members from the CBM6 (52) and CBM53 (53) families of type B CBMs (54). CBM53 (Pel-CBM35, PDB code 2W3) (53) is structurally similar to the N-terminal domain of SsfX3 with a 2.5 Å r.m.s.d. and 115/137 residues aligned. CBM modules are often found appended to catalytic domains to promote substrate interaction; an example is Pel-CBM35, which together with a separate catalytic pectin lyase domain composes the multidomain enzyme Pel10 (53).

There is a large loop region between β4 and β5 of the N-terminal domain (residues 41–61) that contains two short 3_10 helices αA and αB. This insertion in the β-sandwich motif is not found in either the Pel-CBM35 or the CcCCE2 structures, as shown in Fig. 2C. Interestingly, this region appears to be con-
erved among proteins with higher sequence similarity, as shown in Fig. 2D, including AviX9 and Cpz23 from other natural product pathways, yet is missing from CtCE2. Instead, in the same spatial region (residues 679–693) of CtCE2 are two short β-strands inserted between αC and β15 of the α/β-hydrolase domain (Fig. 2, B–D). Additional loop regions that are structurally distinct between SsfX3 and CtCE2 are shown in Fig. 2, C and D. CtCE2 contains a large insertion following the GDSL active site motif, which results in a much larger loop that was not found in SsfX3. This larger loop effectively blocks off part of the active site, resulting in a much narrower binding pocket in CtCE2 compared with SsfX3, and it also reduces the contact area of the N-terminal domain with the binding pocket. This larger loop in CtCE2 and the aforementioned β-strand insertion (residues 679–693) results in a binding pocket that is defined almost entirely by the C-terminal α/β-hydrolase domain. In contrast, the SsfX3 binding pocket is outlined by numerous residues from both N- and C-terminal domains, in particular αA and αB, which extend into the active site, thereby increasing the contribution of the N-terminal domain to the active site surface area (Figs. 2 and 4).

Identification of Catalytic Residues—The active site of canonical α/β-hydrolases includes a highly conserved catalytic triad consisting of a nucleophile, an acid, and a general base histidine (11). For the GDSL hydrolase subfamily, this typically manifests as a Ser-Asp-His catalytic triad (12), which was predicted to be Ser-174, Asp-333, and His-338 in SsfX3 based on structural alignment of SsfX3 and SANK 60405 (accession number BAJ05896), Cpz23 from Streptomyces viridochromogenes (accession number AA831711), a putative GDSL family lipolytic protein from Streptomyces bingchenguensis (accession number ADI04849), a putative GDSL family lipolytic protein from Paenibacillus sp. Y412MC10 (accession number YP.000324352), LipT from Streptomyces sp. SANK 60405 (accession number BAJ05896), Cpz23 from Streptomyces sp. MK730–62F2 (accession number ACQ63631), and CcCE2 from C. thermocellum (accession number 2WAB_A).

The D333A mutant, however, retained catalytic activity with a $k_{cat}$ of approximately 1% of the wild type enzyme (Table 2). Once the structure was solved, it became clear that Asp-333 was not positioned with respect to Ser-174 and His-338 in an orientation compatible with catalysis (see Fig. 2A). Asp-333 resides on the far end of the loop between β17 and αK with its side chain extending away from, rather than pointing into, the active site. The carboxylate chain of Glu-330, however, extends toward the active site in an orientation similar to that observed for other known GDSL hydrolase structures as shown in Fig. 3A. The 2.8-Å distance between β17 and αK with its side chain extending away is too low to obtain accurate $K_m$ values.

Interestingly, these catalytic residues lie significantly further apart from each other in SsfX3 than they do in homologous esterases (Fig. 3A). With no substrate bound, the distance between Ser-174 and His-338 is 8.6 Å. The enzyme must therefore undergo a significant conformational change upon substrate binding to optimally position the general base within hydrogen bonding distance to the nucleophilic hydroxyl (55). Compared with other GDSL hydrolases, this distance between the Ser-174 and His-338 is larger than in the E. coli TAP (19, 20) and the three known CE2 esterases (38); those structures show distances in the range of 2.7–3.8 Å for the unbound structures. As expected for catalytic triads in esterases, the His-338 proton is within hydrogen bonding distance (2.7 Å) to the carboxylate of Glu-330 (55).

| Mutant    | $K_m$ of salicyl-CoA 5 $\mu M$ | $K_m$ of tetracycline 4 $\mu M$ | $k_{cat}$ $\text{min}^{-1}$ |
|-----------|-----------------------------|---------------------------------|-----------------------------|
| SsfX3 WT  | 41 ± 11                     | ≤5$^a$                          | 46 ± 5.8                    |
| S174A     | $b$                         | $b$                             | $b$                         |
| H338A     | $b$                         | $b$                             | $b$                         |
| D333A     | 27 ± 2.5                    | ≤5$^a$                          | 24 ± 0.7                    |
| E330A     | $c$                         | $c$                             | ≤0.1                        |
| C176A     | $c$                         | $c$                             | ≤0.1                        |
| C176S     | $c$                         | $c$                             | ≤0.1                        |
| I175N     | $c$                         | $c$                             | ≤0.1                        |
| E276A     | 12 ± 4.1                    | 32 ± 9.1                        | 11 ± 1.2                    |
| W277A     | 35 ± 1.3                    | 18 ± 6.0                        | 9.7 ± 0.1                   |
| Y335A     | 12 ± 4.6                    | 11 ± 0.8                        | 30 ± 3.1                    |
| L142N     | 42 ± 2.6                    | 24 ± 7.6                        | 5.3 ± 0.1                   |
| Q61A      | 215 ± 55                    | ≤5$^a$                          | 3.4 ± 0.4                   |
| R98L      | 941 ± 363                   | ≤5$^a$                          | 2.6 ± 0.4                   |
| P85W      | 91 ± 8.0                    | ≤5$^a$                          | 46 ± 2.1                    |
| Q93L      | 15 ± 3.4                    | ≤5$^a$                          | 32 ± 1.0                    |
| E97A      | 40 ± 8.6                    | ≤5$^a$                          | 19 ± 1.3                    |

$^a$ Lower concentrations were below the assay detection limit.

$^b$ No activity was observed.

$^c$ Conversion was too low to obtain accurate $K_m$ values.

FIGURE 2. A, overall structure of SsfX3 highlighting residues discussed in the text. B, structure of CtCE2 complexed with cellohexaose (PDB code 2WAO) (38). C, structural alignment of SsfX3 and CcCE2. Where the structures diverge, green ribbons indicate the backbone trace of CtCE2 and purple ribbons indicate the backbone trace of SsfX3. D, sequence alignment of SsfX3 with CtCE2 and other homologs identified based on sequence similarity. Homologs shown are AviX9 from Streptomyces viridochromogenes (accession number AA831711), a putative GDSL family lipolytic protein from Streptomyces bingchenguensis (accession number ADI04849), a putative GDSL family lipolytic protein from Paenibacillus sp. Y412MC10 (accession number YP.000324352), LipT from Streptomyces sp. SANK 60405 (accession number BAJ05896), Cpz23 from Streptomyces sp. MK730–62F2 (accession number ACQ63631), and CcCE2 from C. thermocellum (accession number 2WAB_A).
site, however, is the number of residues separating the aspartate and histidine. In nearly all GDSL hydrolases, the carboxylate residue (usually aspartate) and the histidine are separated by only one or two residues (12). In the case of SsfX3, residues Glu-330 and His-338 are separated by a seven-residue flexible loop. Glu-330 resides at the start of the loop following a short \( \alpha \) helix \( \alpha \)H, and His-338 is positioned at the end of the loop immediately before \( \alpha \)K. This large loop may allow both Glu-330 and His-338 more flexibility to undergo large scale conformational changes required to form the required hydrogen bond network with Ser-174 upon substrate binding. The flexibility of this active site loop is supported by the structural variations observed in this region between the two SsfX3 molecules in the asymmetric unit. In the other SsfX3 molecule, the Ser-174 O\(^{\gamma} \) and His-338 N\(^{\varepsilon_{2}} \) distance is 12.9 Å, whereas the Glu-330 O\(^{\varepsilon_{2}} \) and His-338 N\(^{\varepsilon_{3}} \) distance is 4.17 Å. The variation between these two independent molecules is mostly localized to the active site, which further points to flexibility in this region when no substrate is bound. The two molecules were much closer in conformation over the remainder of their structures, with an overall r.m.s.d. of 1.2 Å.

Residues Ile-175 and Cys-176, which immediately follow the catalytic serine (GDSI) and line the bottom of the binding pocket, were mutated to determine their effects on kinetic parameters (Fig. 2A). Because it is unusual for GDSL hydrolases to contain an isoleucine following the nucleophilic serine, an I175N mutant was prepared to determine whether a nonpolar residue is a requirement at this position. The resulting mutant had no observable acyltransferase activity, suggesting that a hydrophobic residue is indeed required there for enzyme activity. In the early years following the discovery of this class of \( \alpha/\beta \)-hydrolases, the conserved active site motif was defined as GDSLS, because a serine is frequently found in this fifth position. A S31A mutation in EstA arylolesterase from \textit{Vibrio mimicus} decreased the \( K_{m} \) by a factor of 2 compared with the wild type enzyme and had little effect on the catalytic rate (56). In
contrast, mutation of S18G or S18V in GCAT from *Aeromonas hydrophila* resulted in complete inactivation of the enzyme (57). Unlike the homologous GDSL enzymes shown in Fig. 2D, SsfX3 has a cysteine residue in this position. The mutant C176A resulted in reduction of *k* <sub>cat</sub> to ≤0.1 min<sup>−1</sup>. Interestingly, despite the common occurrence of a serine following the GDSL motif, the C176S mutant also showed greatly reduced activity (Table 2). Therefore, the cysteine thiol is highly important for SsfX3 catalysis; its specific role is presently unknown, but it may be involved in substrate interaction or in somehow maintaining the correct active site configuration. As a note, CtCE2 also contains a cysteine of unknown function, Cys-616, in close proximity to the active serine (Fig. 2D).

In addition to the catalytic triad, GDSL hydrolases are identified by four invariant residues (Ser, Gly, Asn, and His) making up the oxyanion hole. These residues lie in conserved blocks I–III and V and lead to the alternative designation of “SGNH hydrolases” for this family of enzymes (12). Because of the low sequence similarity of SsfX3 to other GDSL hydrolases, only blocks I and III are conserved in SsfX3 and only weakly. Based on sequence alignment (Fig. 2D), residues that make up the oxyanion hole were initially predicted to be Ser-174, Gly-209, Asn-236, and His-338 (9). A pairwise structure comparison using the DaliLite server with TAP and CtCE2 revealed that Gly-209 is distant from the active site and that the backbone amide of Ala-207 is more likely to contribute to the oxyanion hole. The spatial arrangement of Ala-207 is also nearly identical to that of Gly-44 of TAP (19) and Gly-658 of CtCE2 (38). The peptide backbone of Asn-236 does align closely with the aforementioned structures. However, the side chain extends away rather than into the active site, and rotation of the side chain amide to accommodate the oxyanion appears unlikely. The side chain of the neighboring residue Ser-235 is facing the active site and may be a more likely candidate to serve as a hydrogen bond donor to the oxyanion. A sequence alignment with homologs of SsfX3 shows that Ser-235 is aligned with asparagine residues in four of the six homologs examined (Fig. 2D). Additionally, homologs Czp23 and LipT contain nonpolar residues at this position, suggesting that the oxyanion hole residues may not be strictly conserved.

**Site-directed Mutagenesis of Binding Pocket Residues**—Unfortunately, we were unable to obtain a co-crystal structure with either compounds 4 or 6 to identify the exact residues and dimensions of the binding pocket. Attempts to model substrate binding using computational docking software were also unsuccessful, likely due to the large conformational changes predicted to occur upon substrate binding. To map regions that are potentially important to substrate binding, we instead utilized site-directed mutagenesis to experimentally probe the role of individual regions in and around the putative binding pocket. All of the mutants were solubly expressed and purified from *E. coli* at levels similar to that of the wild type SsfX3 (≈60 mg/liter culture).

Because of the large size of the tetracycline substrate 4, which spans ~16 Å from end to end (~20 Å for the product 6), the substrate is expected to extend from the active site pocket of the α/β-hydrolase domain to the interface of the β-sheet domain (Fig. 4, B and C). During biochemical analysis of CtCE2, a previous investigation found that aromatic residues lining the top and bottom of the binding pocket were important for substrate interactions (38). In the active site pocket of SsfX3, three aromatic residues (Phe-276, Trp-277, and Tyr-335) form one end of the binding pocket as seen in Figs. 2A and 4B. The aromatic side chains of Tyr-335 and Trp-277 line the top and bottom of the cavity, similar to Trp-790 and Trp-746 in CtCE2, respectively (Fig. 2). Phe-276 forms a wall at the end of the binding pocket, thereby making it an enclosed cavity. The helix αH where Phe-276 is found in SsfX3 is missing in CtCE2. As a result, the binding pocket of CtCE2 is more like an open-ended channel (compare Fig. 2, A and B). These aromatic residues in SsfX3 were mutated to alanine. The resulting kinetic parameters are shown in Table 2. F276A and W277A both resulted in an ~4-fold decrease in *k* <sub>cat</sub>. As neither of those residues is predicted to be involved in catalysis, the diminished activity is likely due to effects on the conformation of nearby residues caused by removing a bulky side chain from the active site. All three mutants F276A, W277A, and Y335A resulted in an increase in the *K* <sub>M</sub> value for compound 4. On the contrary, Y335A and F276A resulted in an ~4-fold lower *K* <sub>M</sub> value toward the acyl donor 5. *K* <sub>M</sub> values for compound 5 remained relatively unchanged for W277A. These results indicate that the aromatic residues F276A, W277A and Y335A play important roles in binding the polycyclic tetracycline substrate 4, but may not be involved in the binding of compound 5.

**Role of the N-terminal β-Sandwich Domain**—As GDSL hydrolases are often isolated as single domain enzymes without fusion to the β-sandwich domain, we were interested to determine whether the C-terminal hydrolase domain had any activity as a stand-alone protein or whether it required the additional β-sandwich domain for which no catalytic activity was predicted. The excised C-terminal hydrolase domain, starting from residue Thr-162 in the linker region, was expressed in *E. coli* at yields comparable with wild type SsfX3 (~60 mg/liter). Neither the forward transcylation nor the reverse hydrolysis activities were detected with the stand-alone enzyme, demonstrating that the N-terminal domain is vital to the activity of SsfX3. The additional domain may be needed to provide sufficient binding energy for large substrates such as compound 4. The canonical, single domain acyltransferases may not have sufficient binding interactions to confer substrates specificity for the large substrate.

To further examine the role of the N-terminal domain, residues that reside at the interface of the two domains, and in close proximity to the putative binding pocket (Fig. 4), were mutated to determine whether they had an impact on reaction kinetics or substrate affinity. During analysis of CBM35, Montanier *et al.* (53) found that binding of carbohydrates to CBM35 is facilitated by a combination of hydrophobic interactions involving several aromatic residues and hydrogen bonding from polar side chains. In the β-sandwich domain of SsfX3, however, no aromatic residues were positioned in close proximity to the likely substrate binding pocket. Instead, the side chain of Leu-142 positioned at the beginning of β11 inserts into the inter-domain binding pocket and may be oriented for contact with the hydrophobic portions of the tetracycline substrate 4, such
as the aromatic D-ring or D-olivose. As expected, mutating Leu-142 to the polar residue asparagine resulted in a significantly increased $K_M$ for compound 4 and a 10-fold reduction in $k_{cat}$. The L142N mutation did not affect the $K_M$ value for compound 5, suggesting that this residue is not involved in binding the acyl donor. The side chains of Arg-58, Gln-61, Gln-93, and Glu-97 protrude into the putative binding pocket and may form hydrogen bonds with either substrate 4 or 5. These residues were mutated to nonpolar amino acids, and the kinetic parameters were evaluated (Table 2). Neither E97A nor Q93L mutants, which reside on the large flexible loop between $\beta_6$ and $\beta_7$, showed a significantly negative effect on enzyme activity. The
E97A mutant had no measurable effect on the $K_M$ value of either substrate but resulted in an ~2-fold decrease in $k_{cat}$. The Q93L mutant had a slightly decreased $k_{cat}$ and an ~3-fold reduction in $K_M$ for compound 5.

Arg-58 and Gln-61 both lie in the large loop region between $\beta_4$ and $\beta_5$, which contains the two $\beta_3$ helices that line the bottom of the putative binding pocket (Fig. 4B). Mutation of these residues had the largest effect on kinetic parameters. The Q61A mutant showed a reduction in $k_{cat}$ of ~10-fold and a decrease in affinity for compound 5 with a ~4-fold increase in $K_M$, although the $K_M$ value for compound 4 remained below detection limits. The R58L mutant had a similar reduction in $k_{cat}$ and an even more pronounced increase in $K_M$ for compound 5 of greater than 20-fold. Pro-85 is also in this region at the interface of the N-terminal domain and putative substrate binding pocket. The proline residue was mutated to a tryptophan to determine whether reducing the size of the binding pocket might influence the kinetic parameters. The $k_{cat}$ and $K_M$ values for compound 4 remained unchanged, whereas the $K_M$ value for compound 5 was twice the wild type value. These mutagenesis results hint at a possible difference in binding locations for compounds 4 and 5. Pro-85 lies adjacent in space to Gln-61 and Arg-58 and may further define the binding region for the acyl donor.

These results indicate that the N-terminal domain has an important role in substrate binding for both the tetracycline substrate 4 and acyl donor 5. In particular, we noticed that the surface area of the N-terminal domain in the putative binding pocket was enlarged by the aforementioned $\beta_3$ helices $\alpha A$ and $\alpha B$. Although this structural feature was not observed for carbohydrate esterase CtcE2, the region containing the two $\beta_3$ helices appears to be present in AviX9 and Cpz23. Interestingly, all of these bidomain GDSL hydrolases are predicted or known to bind large extended substrates (although all very different from compound 4). Hence, recruitment of the additional $\beta$-sandwich domain by the GDSL catalytic domain may be a general strategy used to define the binding pocket for the “large” small molecule substrates such as compound 4, to provide sufficient interactions and binding energies required for forming the enzyme-substrate complexes.

**Substrate Specificity of the Acyl Group and Hydroxyl Donor**—Broad substrate specificity with regard to the acyl donor is commonly observed for acyltransferases (12), and this flexibility has been exploited in several cases to generate analogs of natural products with varied acyl substituents (2, 3, 5, 58–60). SsfX3 has been shown previously to accept a variety of acyl-CoA substrates, including those containing small substitutions around the aromatic ring, such as 2-chlorobenzoyl-CoA, 2,3-dihydroxybenzoyl-CoA, and 2,5-dihydroxybenzoyl-CoA. Larger groups such as 4-aminobenzoyl-CoA were shown not to be tolerated, resulting in no detectable analog of compound 6 (9). The kinetic parameters of SsfX3 with respect to several of the different acyl-CoA substrates were determined (Table 3). Although the $k_{cat}$ value for each of the analogs was at least 20-fold lower than that observed for compound 5, the $K_M$ values varied. Substitution at the ortho position is clearly preferred, as the $K_M$ value for both 2-chlorobenzoic acid and compound 5 were lower than that of benzoyl-CoA. Additional substitutions around the ring led to unfavorable binding to SsfX3, as noted by the much higher $K_M$ value for 2,3-dihydroxybenzoyl-CoA. This indicated that the SsfX3 binding pocket affords a tight fit to the aryl functionality. Surprisingly, however, SsfX3 displayed measurable activity toward the butyryl group, as the C-4-butyryl analog of 6 was detected in the transacylation assay using butyryl-CoA. Despite the significantly reduced reaction velocity, this result demonstrated that the aromatic moiety is not an absolute requirement for SsfX3. Unfortunately, in the absence of a co-crystal structure, the spatial arrangement of the binding pocket cannot be mapped in detail.

The substrate specificities of natural product acyltransferases with respect to the acyl acceptor are assayed less frequently, which may be due in part to difficulties in preparing or isolating analogs of the natural substrate to test. In the case of ansamitocin biosynthesis, Asm19 showed broad substrate specificity toward various acyl substrates but was unable to acylate the C-3 hydroxyl of related substrate maytansinol (58). N-Acyltransferases tAtf and aAtf, which acylate the sugar amino groups during the biosynthesis of glycopeptides teicoplanin and A-40,926 respectively, demonstrated flexibility for both the acyl substrates and acyl acceptors (3). To examine the substrate specificity of SsfX3 toward the aryl acceptor, we assayed the transacylation reaction using upstream intermediates 1 and 2 (Fig. 1) in the presence of 5. Compound 2 differs from 4 in that the C-ring has not been oxidized and is an analog of anhydrotetracycline instead of tetracycline. Compound 2 also lacks the C-6 O-methoxy substitution present in compound 4 and is therefore nearly isosteric to compound 4 but with important physical differences. Compound 1 is the precursor of compound 2 and lacks the C-9 d-olivose group. From active site mutations such as L142N, we reasoned that the distal end of compound 4, which includes the C-glycosylated d-olivose, makes van der Waals contacts with the N-terminal CBM-like domain. Therefore, we anticipated that the smaller compound 1 would be a poor substrate for SsfX3. Product assays were performed with 0.1 mM compounds 1, 2, or 4, 0.3 mM compound 5 and 1 mM SsfX3. Reactions were incubated at room temperature for 1 h, extracted, and analyzed by LCMS. Both compounds 1 and 2 were salicylated by SsfX3, albeit at a much slower rate than the native substrate 4. As expected, compound 1 had the lowest conversion as follows: only 1.6% in 1 h compared with 17% for compound 2 and 90% for native substrate 4 (Supplemental Fig. S1). Although not tested in vitro, trace amounts of SF2575 variants lacking either the C12a or C-6 O-methylation have been identified in the fermentation extract of strains producing SF2575, indicating that these demethyl analogs of compound 4 are also accepted as substrates of SsfX3.

### Table 3

| Substrate                      | $K_M$  | $k_{cat}$ | $k_{cat}/K_M$ |
|-------------------------------|--------|-----------|---------------|
| Salicyl-CoA compound 5        | 41 ± 11| 46 ± 5.8  | 1.0           |
| Benzoyl-CoA                   | 75 ± 18| 2.7 ± 0.2 | 0.032         |
| 2-Chlorobenzoyl-CoA           | 11 ± 1.4| 2.0 ± 0.08| 0.16          |
| 2,3-Dihydroxybenzoyl-CoA      | 357 ± 39| 1.4 ± 0.005| 0.0035        |
| Butyryl-CoA                   | 922 ± 194| 0.54 ± 0.04| 0.0005        |
These studies demonstrate that SsfX3 has somewhat relaxed substrate specificity for both the acyl donors and acyl acceptors. However, the substrate specificity observed for SsfX3, although somewhat tolerant to closely related substrates, seems to be much more stringent when compared with the extremely broad range of substrates accepted by GDSL hydrolases reported by Asler et al. (24). According to the crystal structure of SsfX3, this increased substrate specificity is likely influenced significantly by the putative N-terminal binding module, as discussed previously.

**Bioactivity of Chemoenzymatically Prepared Analogs**—As SsfX3 had sufficient flexibility with regard to the acyl donor, we were able to chemoenzymatically prepare analogs of compound 6 to determine the effect of various substituents around the aromatic ring of the C-4 salicylic acid ester on the bioactivity. A 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium cell proliferation assay was used to determine the anti-proliferative effects of the analogs against various types of cancer cell lines (Table 4). The most surprising result was that all three dihydroxybenzoate analogs had greatly reduced activity against all cell lines tested. Removal of the ortho-hydroxyl, or substitution with chlorine, however, had little effect on the bioactivity, suggesting that this position is not critical to the interaction with the molecular target, which is suggested to be DNA topoiso­merase I for related compounds TAN-1518A and TAN-1518B (61).

**Conclusions**—SsfX3 is a critical enzyme in the biosynthesis of the potent anticancer compound SF2575, being responsible for installing the C-4 salicylate group vital for bioactivity. Additionally, it is the first acyltransferase identified to specifically act on a tetracycline substrate. To further understand the mechanism of this enzyme, we crystallized SsfX3 and elucidated the structure. The structure of SsfX3 revealed a bidomain architecture consisting of an N-terminal β-sandwich domain and a C-terminal α/β-hydrolase domain. Comparison of the structure with known enzymes revealed that SsfX3 is remotely homologous to the CE2 family of carbohydrate esterases, a relationship that could not be determined previously due to low sequence identity. With the structure in hand, we were able to probe the binding pocket and active site residues to identify those important for binding and catalysis. Interestingly, although the N-terminal domain was not predicted to have any catalytic activity (i.e. it does not harbor any of the catalytic residues), it was determined to be vital to the activity of the enzyme, and several of the residues at the interface of the two domains were shown to affect the kinetic parameters dramatically. SsfX3 was also demonstrated to be flexible with regard to the acyl donor, which enabled the chemoenzymatic prepara-

---

**REFERENCES**

1. Olano, C., Méndez, C., and Salas, J. A. (2010) *Nat. Prod. Rep.* 27, 571–616
2. Li, T. L., Huang, F., Haydock, S. F., Mironenko, T., Leadlay, P. F., and Spencer, J. B. (2004) *Chem. Biol.* 11, 107–119
3. Kruger, R. G., Lu, W., Oberthür, M., Tao, J., Kahne, D., and Walsh, C. T. (2005) *Chem. Biol.* 12, 131–140
4. Menéndez, N., Braña, A. F., Salas, J. A., and Méndez, C. (2007) *Microbiology* 153, 3061–3070
5. Ghate, M. S., Palaniappan, N., Alhamadsheh, M. M., DiBari, J., and Reynolds, K. A. (2009) *Appl. Environ. Microbiol.* 75, 3469–3476
6. Xie, X., Meehan, M. J., Xu, W., Dorrsett, P. C., and Tang, Y. (2009) *Am. Chem. Soc.* 131, 8388–8389
7. Hatsu, M., Sasaki, T., Gomi, S., Kodama, Y., Sezaki, M., Inouye, S., and Kondo, S. (1992) *J. Antibiot.* 45, 325–330
8. Hatsu, M., Sasaki, T., Watabe, H., Miyadoh, S., Nagasawa, M., Shomura, T., Sezaki, M., Inouye, S., and Kondo, S. (1992) *J. Antibiot.* 45, 320–324
9. Pickens, L. B., Kim, W., Wang, P., Zhou, H., Watanabe, K., Gomi, S., and Tang, Y. (2009) *J. Am. Chem. Soc.* 131, 17677–17689
10. Upton, C., and Buckley, J. T. (1995) *Trends Biochem. Sci.* 20, 178–179
11. Nardin, M., and Dijkstra, B. W. (1999) *Curr. Opin. Struct. Biol.* 9, 732–737
12. Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004) *Prog. Lipid Res.* 43, 534–552
13. Lo, Y. C., Lin, S. C., Shaw, J. F., and Liaw, Y. C. (2003) *J. Mol. Biol.* 330, 539–551
14. Cho, H., and Cronan, J. E., Jr. (1993) *J. Biol. Chem.* 268, 9238–9245
15. Pautau, M., and Uriel, J. (1997) *Eur. J. Biochem.* 243, 435–442
16. Karasawa, K., Kudo, I., Kobayashi, T., Homma, H., Chiba, N., Mizushima, H., Inoue, K., and Nojima, S. (1991) *J. Biochem.* 109, 288–293
17. Karasawa, K., Yokoyama, K., Setaka, M., and Nojima, S. (1999) *J. Biochem.* 126, 445–448
18. Finn, R. D., Tate, I., Murray, I., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) *Nucleic Acids Res.* 36, D281–D288
19. Lee, L. C., Lee, Y. L., Lee, R. J., and Shaw, J. F. (2006) *Biochem. J.* 397, 69–76
20. Lo, Y. C., Lee, Y. L., Shaw, J. F., and Liaw, Y. C. (2000) *Acta Crystallogr. D Biol. Crystallogr.* 56, 756–757
21. Wei, Y., Schottel, J. L., Derewenda, U., Swenson, L., Patkar, S., and Dere­wendza, Z. S. (1995) *Nat. Struct. Biol.* 2, 218–223
22. Molgaard, A., Kauppinen, S., and Larsen, S. (2000) *Structure* 8, 373–383
23. van den Berg, B. (2010) *J. Mol. Biol.* 396, 627–633
24. Lešèi Asler, I., Iviæ, N., Kovaèiæ, F., Schell, S., Knorr, J., Krauss, U., Wilhelm, S., Kojiaè-Prodiaè, B., and Jaeger, K. E. (2010) *ChemBioChem* 11, 2158–2167
