BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

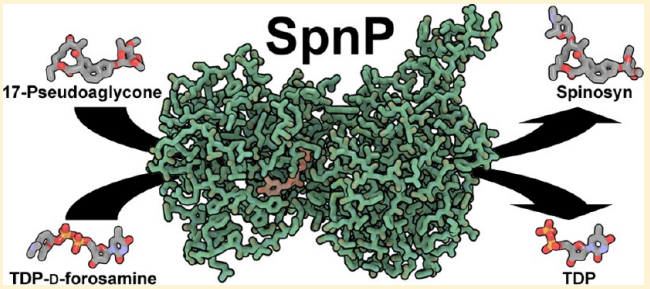
Structural Studies of the Spinosyn Forosaminyltransferase, SpnP

Eta A. Isiorho,[†,‡] Byung-Sun Jeon,[†,‡] Nam Ho Kim,[†] Hung-wen Liu,*[,†,‡] and Adrian T. Keatinge-Clay*[,‡]

[†]Division of Medicinal Chemistry, College of Pharmacy, and [‡]Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712, United States

Supporting Information

ABSTRACT: Spinosyns A and D (spinosad) are complex polyketide natural products biosynthesized through the cooperation of a modular polyketide synthase and several tailoring enzymes. SpnP catalyzes the final tailoring step, transferring forosamine from a TDP-D-forosamine donor substrate to a spinosyn pseudoaglycone acceptor substrate. Sequence analysis indicated that SpnP belongs to a small group of glycosyltransferases (GTs) that require an auxiliary protein for activation. However, unlike other GTs in this subgroup, no putative auxiliary protein gene could be located in the biosynthetic gene cluster. To learn more about SpnP, the structures of SpnP and its complex with TDP were determined to 2.50 and 3.15 Å resolution, respectively. Binding of TDP causes the reordering of several residues in the donor substrate pocket. SpnP possesses a structural feature that has only been previously observed in the related glycosyltransferase EryCIII, in which it mediates association with the auxiliary protein EryCII. This motif, H-X-R-X_5-D-X_5-R-X_{12-20}-D-P-X_3-W-L-X_{12-18}-E-X_4-G, may be predictive of glycosyltransferases that interact with an auxiliary protein. A reverse glycosyl transfer assay demonstrated that SpnP possesses measurable activity in the absence of an auxiliary protein. Our data suggest that SpnP can bind its donor substrate by itself but that the glycosyl transfer reaction is facilitated by an auxiliary protein that aids in the correct folding of a flexible loop surrounding the pseudoaglycone acceptor substrate-binding pocket.



Spinosyns A and D (**9** and **10**, respectively; termed spinosad) derived from *Saccharopolyspora spinosa* are the active ingredients of commercial insecticides for the treatment of lice, fleas, and agricultural pests.[1,2] Spinosyns are biosynthesized through the action of a modular polyketide synthase (PKS) and several tailoring enzymes, two of which are glycosyltransferases (GTs), SpnG and SpnP (Figure 1).[3−6] They aid in spinosyn biosynthesis by attaching rhamnose and forosamine sugars, critical for the biological activity of spinosad, to the aglycone intermediates **3** and **4** and pseudoaglycone intermediates **7** and **8**, respectively.[7,8]

GTs generally deliver sugars from activated nucleotide diphosphate sugar donor substrates to acceptor substrates.[9] These enzymes can be classified into more than 90 families, of which family 1 (GT1) is the most common in the glycosylation of secondary metabolites.[10,11] SpnG, whose crystal structure was recently determined, belongs to the GT1 family.[12,13] Other members of GT1 family include the calicheamicin GT CalG1, the erythromycin GT EryCIII, the oleandomycin GT OleD, the pikromycin GT DesVII, the tylosin GT TylM2, the urdamycin GT UrdGT2, and the vancomycin GT GtfA.[14−20] Each of these is an inverting GT-B enzyme ("inverting" signifies that the stereochemistry at the anomeric position of the product is the opposite of that of the TDP substrate; "GT-B" denotes one of two principle glycosyltransferase folds) that transfers a sugar from a UDP- or TDP-containing donor substrate to an aglycone acceptor substrate in the biosynthesis of a natural product.[10]

Sequence analysis reveals that SpnP is also a member of the GT1 family. It catalyzes the attachment of forosamine (biosynthesized from TDP-4-keto-6-deoxy-D-glucose by SpnN, SpnO, SpnQ, SpnR, and SpnS) to 17-pseudoaglycone (**7** and **8**).[21] The *in vivo* activity of SpnP has been established by gene deletion experiments; however, its *in vitro* activity has not yet been demonstrated.[4] Interestingly, SpnG and SpnP (25% identical sequences) add sugars to similar substrates but display nearly opposite regiospecificities. SpnG transfers rhamnose to the 9-OH of **3** and **4**, while SpnP transfers forosamine to the 17-OH of **7** and **8**. Thus, SpnG and SpnP must differentially orient similar spinosyn carbon skeletons within their acceptor substrate-binding sites.[16]

A subset of GT1 enzymes, including the GTs AknS, DesVII, EryCIII, MycB, and TylM2, are known to require auxiliary proteins for their catalytic activities.[17,22−26] These GTs often transfer deoxyaminosugars and are encoded by a gene neighboring that of its auxiliary protein in the biosynthetic cluster. A typical auxiliary protein possesses a P450 fold but lacks the conserved cysteine residue that serves as a ligand to the heme iron. These auxiliary proteins have a long, N-terminal

Received: March 25, 2014
Revised: May 24, 2014
Published: June 19, 2014

   dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

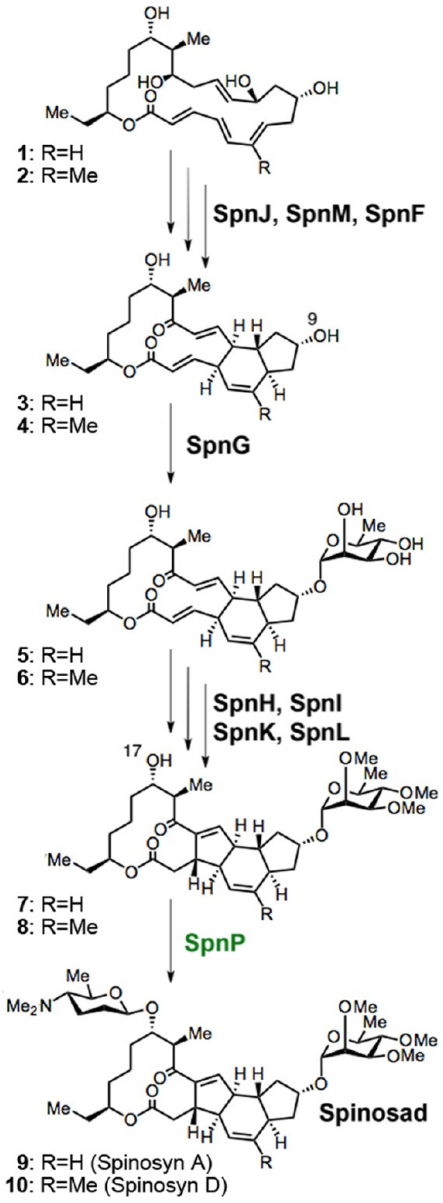


Figure 1. Spinosad biosynthetic pathway. SpnP is proposed to forosaminylate intermediates **7** and **8** at 17-OH.

helix and are believed to be critical for the activation of their cognate GTs. Thus far, the only available crystal structure of a GT requiring an auxiliary protein is the binary complex of EryCIII with its auxiliary protein, EryCII.[15]

To complete the characterization of GTs involved in spinosyn biosynthesis, we purified SpnP and demonstrated its *in vitro* activity through a reverse glycosyl transfer assay. Crystal structures of SpnP in its unliganded form and bound to the substrate donor analogue TDP were also determined. On the basis of sequence and structure comparisons with EryCIII and other GTs aided by auxiliary proteins, we hypothesize SpnP requires an auxiliary protein for activation. Although the identity of the auxiliary protein remains elusive, the SpnP structures reported herein provide insight into a GT that is poised for pairing with an auxiliary protein partner. These data add to the characterization of the spinosyn biosynthetic pathway and illuminate the role of auxiliary proteins in the activation of GTs.

MATERIALS AND METHODS

Preparation of Spinosyn A. Spinosad (**9** and **10**) was prepared from water/ethyl acetate (1:1) extractions of Dow AgroSciences Tracer Naturalyte solutions. The pooled organic fractions were washed with brine and concentrated.

Preparation of 17-Pseudoaglycone. The 17-pseudoaglycone **7** was prepared as previously described.[27] Briefly, 1 N H_2SO_4 was added to spinosyn A (**9**) and the mixture heated to 80 °C for 2 h. The reaction mixture was cooled to room temperature, and the precipitate was filtered, washed with 1 N H_2SO_4, and dissolved in dichloromethane. The resulting solution was washed four times with brine, dried over $MgSO_4$, and concentrated.

Cloning, Expression, and Purification of SpnP. The *spnP* gene was amplified from genomic DNA of *S. spinosa* strain NRLL18537 and cloned into a pET28b(+) vector at the NdeI and XhoI restriction sites. The forward primer 5′-AATAAG-GC**CATATG**CGTGTCCTGTTCACC-3′ and reverse primer 5′-AACT**CTCGAG***TCA*CGGATGGCCATCAGAC-3′ were used for PCR amplification (restriction sites in bold and start and stop codons in italics).

The expression plasmid was then transformed into *Escherichia coli* BL21(DE3). Overnight cultures (LB with 50 μg/mL kanamycin, 5 mL) of the transformed expression hosts were grown and used to inoculate 6 × 1 L cultures (LB with 50 mg/L kanamycin). The cultures were incubated while being shaken at 37 °C until the OD_{600} reached 0.6, after which isopropyl β-D-thiogalactopyranoside (final concentration of 0.5 mM) was added. The temperature was decreased to 15 °C, and the cultures were grown for an additional 18 h. Cells were pelleted via centrifugation (5000*g* for 10 min) and resuspended in lysis buffer [5% (v/v) glycerol, 500 mM NaCl, and 30 mM HEPES (pH 7.5)] for sonication. Lysed cells were centrifuged (30000*g* for 30 min), and the lysate was loaded onto a Ni-NTA column (Qiagen). Fifteen column volumes of 20 mM imidazole in lysis buffer was used to wash the column before SpnP was eluted with 150 mM imidazole in lysis buffer. SpnP was further purified over a gel filtration column (Superdex 200, GE Healthcare Life Sciences) equilibrated with 5% (v/v) glycerol, 150 mM NaCl, and 10 mM HEPES (pH 7.5). A protein concentrator (Amicon, YM10 membrane) was used to achieve a concentration of 20 mg/mL. Aliquots were flash-frozen in liquid nitrogen and stored at −78 °C until they were used further.

Crystallization, Data Processing, and Structure Determination. Crystals of SpnP were grown using the sitting-drop vapor-diffusion method at 22 °C. The crystallization buffer consisted of 7 mM 17-pseudoaglycone **7**, 46% (w/v) PEG 200, 0.6−3.0% (w/v) dextran sulfate M_R 5000, 0.2 M sodium chloride, and 0.1 M phosphate citrate buffer (pH 4.24). Each drop consisted of 2 μL of the SpnP protein solution (20 mg/mL) and 1 μL of the crystallization solution with 500 μL of crystallization buffer in the reagent reservoir. Crystals were flash-frozen in liquid nitrogen prior to data collection. Crystals of SpnP complexed with TDP were obtained by supplying 20 mM TDP (final concentration) to the crystallization drop and incubating it for 1 h prior to flash-freezing. Data for SpnP and the SpnP−TDP complex were collected at Advanced Light Source beamlines 5.0.2 and 5.0.3, respectively, and processed with HKL2000.[28] Molecular replacement was utilized to determine the unliganded SpnP structure, using the EryCIII monomer [Protein Data Bank (PDB) entry 2YJN] as the

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

Table 1. Data Collection and Refinement Statistics for SpnP and the SpnP−TDP Complex

	SpnP	SpnP−TDP
Crystallization Data		
resolution (Å)	46.93−2.50 (2.56−2.50)	114.81−3.15 (3.23−3.15)
wavelength (Å)	0.976	0.976
space group	$P4_32_12$	$P4_32_12$
a, *b*, *c* (Å)	162.4, 162.4, 81.2	162.4, 162.4, 81.3
no. of molecules per asymmetric unit	2	2
no. of measured intensities	502778	274295
no. of unique reflections	71988	19214
completeness (%)	99.7 (97.3)	99.0 (95.3)
R_{sym}	0.110	0.17
redundancy	7.0 (6.4)	14.3 (10.8)
average $I/\sigma I$	49.3 (3.1)	16.7 (2.4)
Refinement		
no. of reflections	36239	18175
R_{cryst}/R_{free}	0.203/0.257	0.209/0.314
no. of protein atoms	5772	5827
no. of ligand atoms	–	50
no. of solvent atoms	82	67
average *B* factor (Å^2)		
monomer A	63.1	62.7
monomer B	59.5	59.4
water	54.2	47.9
TDP	–	71.64
root-mean-square deviation		
bond lengths (Å)	0.016	0.012
bond angles (deg)	1.956	1.814
Ramachandran plot (%)		
preferred regions	93.94	87.70
allowed regions	3.17	7.38
outliers	2.89	4.92

search model in Phaser.[29,30] The model was initially refined with ARP/wARP and built through several rounds of refinement with Coot and Refmac.[31−33] ARP/wARP was used to model in waters, which were then evaluated manually. The TDP complex structure was determined by molecular replacement using the unliganded SpnP structure.

Reverse Glycosyl Transfer Assay. The SpnP reverse glycosyl transfer reaction mixture consisted of 100 μM spinosad (**9** and **10**), 1.0% (v/v) DMSO, 2.5 mM $MgCl_2$, 2 mM TDP, 10 μM SpnP, and 100 mM HEPES (pH 8.0) (final volume of 200 μL). Reaction mixtures were incubated for 16 h at 30 °C, after which 200 μL of ethanol was added. The resulting mixtures were centrifuged (10000*g* for 5 min) to remove debris, and samples were analyzed by high-performance liquid chromatography (HPLC) with a C_{18} column (Varian): 80 to 90% B over 30 min [solvent A consisting of water and 20 mM ammonium acetate and solvent B consisting of an acetonitrile/methanol mixture (50:50) and 20 mM ammonium acetate] at a 1 mL/min flow rate. Formation of 17-pseudoaglycone **7** was monitored at 254 nm and confirmed by mass spectrometry.

RESULTS

Expression and Purification of SpnP. Attempts to express and purify SpnP were problematic. Re-examination of the sequence of the *spnP* gene and the flanking regions revealed that its start codon likely resides at a region more downstream than that originally assigned. The corrected open reading frame (ORF) is 48 bp shorter, corresponding to 16 fewer N-terminal residues. The truncated construct begins with residues MRVL, more typical for GT-B enzymes. Expression of this new construct yielded soluble protein. The numbering of the residues in SpnP described here is based on the MRVL starting sequence.

Crystallization and Structure Determination. In the presence of 17-pseudoaglycone **7**, crystals of SpnP were observed within 2 weeks. In the absence of **7**, crystals appeared after 3 weeks but were smaller and did not diffract well. The addition of dextran sulfate M_R 5000 increased the crystal size and quality. The EryCIII monomer (PDB entry 2YJN) was used as the molecular replacement search model.[15] Unliganded SpnP was refined to a resolution of 2.50 Å, and the SpnP−TDP complex was refined to a resolution of 3.15 Å (Table 1).

Overall Structure of SpnP. SpnP is homodimeric with W22 from each monomer forming a π-stacking interaction across the 2-fold axis, a structural feature also observed in SpnG, CalG3, and SsfS6[13,14,34] (Figure 2A). In addition, H414 makes a second stacking interaction across the 2-fold axis of the SpnP dimer. The structure of the SpnP monomer is consistent with the structural fold possessed by GT-B enzymes, with two Rossmann-like domains, each containing a parallel β-sheet with 321456 topology surrounded by α-helices.[9] The N-terminal domain consists of residues 1−224, as well as residues 412−426 from the C-terminal helix, while the C-terminal domain is comprised of residues 247−410. The domains are linked by a long loop (residues 225−246). Two disordered regions are present in each monomer, residues 63−98 and 261−271 in monomer A and residues 60−98 and 258−277 in monomer B. The equivalent flexible loops of the nogalamycin GT SnogD

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

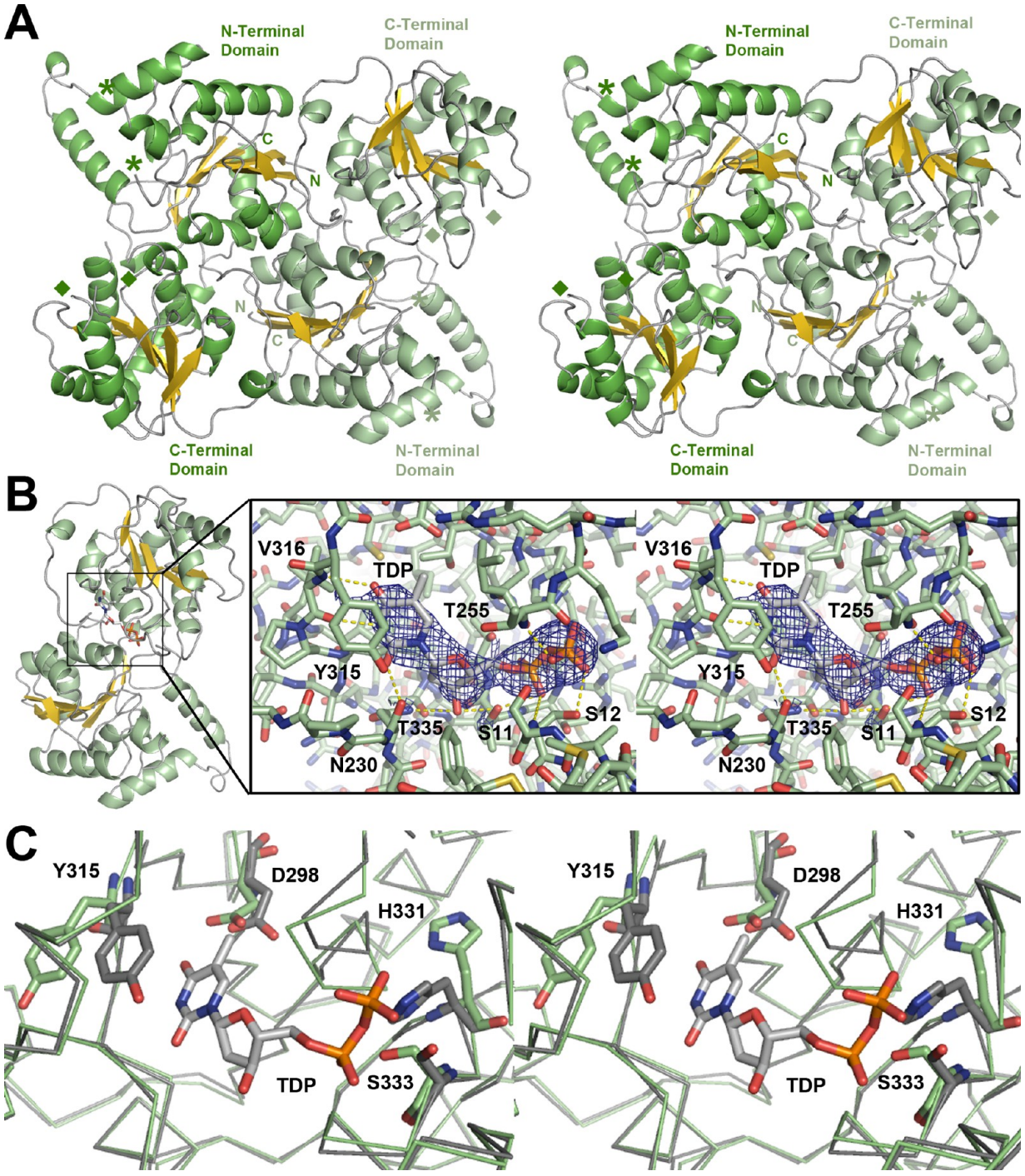


Figure 2. SpnP homodimer. (A) Stereodiagram showing the asymmetric unit. Asterisks mark the beginning and end of the first disordered loop (FL1, residues 60−98). Diamonds mark the beginning and end of the second disordered loop (FL2, residues 258−277). (B) Stereodiagram of the donor substrate-binding site in the SpnP−TDP complex showing the $F_o - F_c$ omit map (contoured at 2.5 root-mean-square deviations). The thymine base, deoxyribose, and the pyrophosphate of TDP contact the labeled residues in the donor-binding site. (C) Stereodiagram showing the SpnP active site in the absence of TDP (green) and in the presence of TDP (gray). The TDP α-phosphate stabilizes one conformation of the pyrophosphate-binding loop (H-X_3-G-T motif), causing several residues to shift, notably H331 and S333. Y315 also reorients to stack with the thymine base.

have been termed FL1 and FL2, respectively; thus, this nomenclature is utilized here.[35] The active site is located in a cleft between the N- and C-terminal domains with H13 likely serving as the catalytic base (by analogy with other GT1 enzymes).

Donor Substrate-Binding Site. Similar to other GT1 enzymes, the donor substrate, TDP-D-forosamine, is expected to bind to the C-terminal domain of SpnP. An $\alpha-\beta-\alpha$ motif (residues 315−343) constitutes the major portion of the donor substrate-binding pocket (Figure 2B,C). In the complex of SpnP with the donor substrate analogue TDP, the thymine forms hydrophobic interactions with L254, Y315, and L318. Several known GT1 enzymes that select TDP-linked sugars over UDP-linked sugars (e.g., SpnG and SsfS6) contain an

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

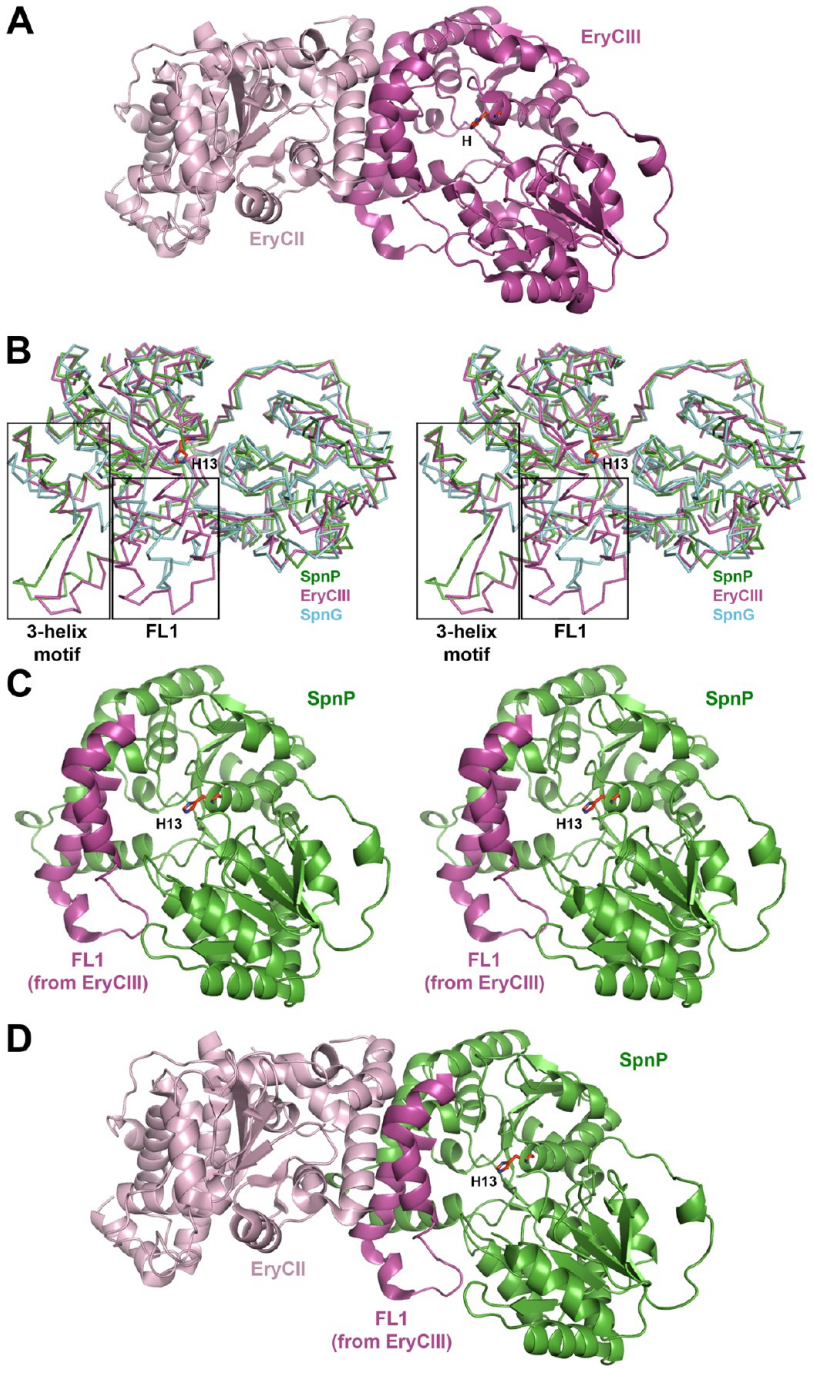


Figure 3. Auxiliary protein contact. (A) Complex of the GT EryCIII (fuchsia) with its auxiliary protein EryCII (light pink). (B) Structural alignment of SpnP (green), SpnG (blue), and EryCIII (fuchsia). The putative catalytic base of SpnP, H13, is labeled. EryCIII and SpnP contain a three-helix motif as well as a long FL1 (however, in the SpnP structure, FL1 is disordered). (C) SpnP modeled with the EryCIII FL1. (D) Modeled dimeric interface between SpnP and an auxiliary protein.

asparagine/threonine pair, of which the asparagine NH_2 forms a hydrogen bond with the 3′-OH of the deoxyribose but sterically excludes UDP by clashing with its 2′-OH.[13,34] Likewise, N230 and T335 are present in SpnP, with the NH_2 of N230 forming a hydrogen bond with the 3′-OH of TDP. As with other structurally characterized GT-B enzymes, an H-X_3-G-T loop (H331−T336 in SpnP) contacts the pyrophosphate group of TDP through interactions with the backbone NH groups (T335 and T336) and side chains (S11, S12, H331, S333, T335, and T336). The binding of TDP causes several residues to adopt conformations different from those in unliganded SpnP. For example, Y315 reorders over the thymine to π-stack with it and forms a hydrogen bond with the N230 backbone carbonyl (similar interactions have been observed in other GTs)[14,16,20,36] (Figure 2C). In addition, the H331 side chain reorients to form a charged hydrogen bond with the pyrophosphate group. The sugar-binding motif in most GT1 enzymes, D/E-Q, which forms hydrogen bonds with the sugar substituents, is replaced with D-E in SpnP (residues 356 and 357).[37]

Acceptor Substrate-Binding Site. The binding site for the 17-pseudoaglycone acceptor substrate (**7** and **8**) is

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

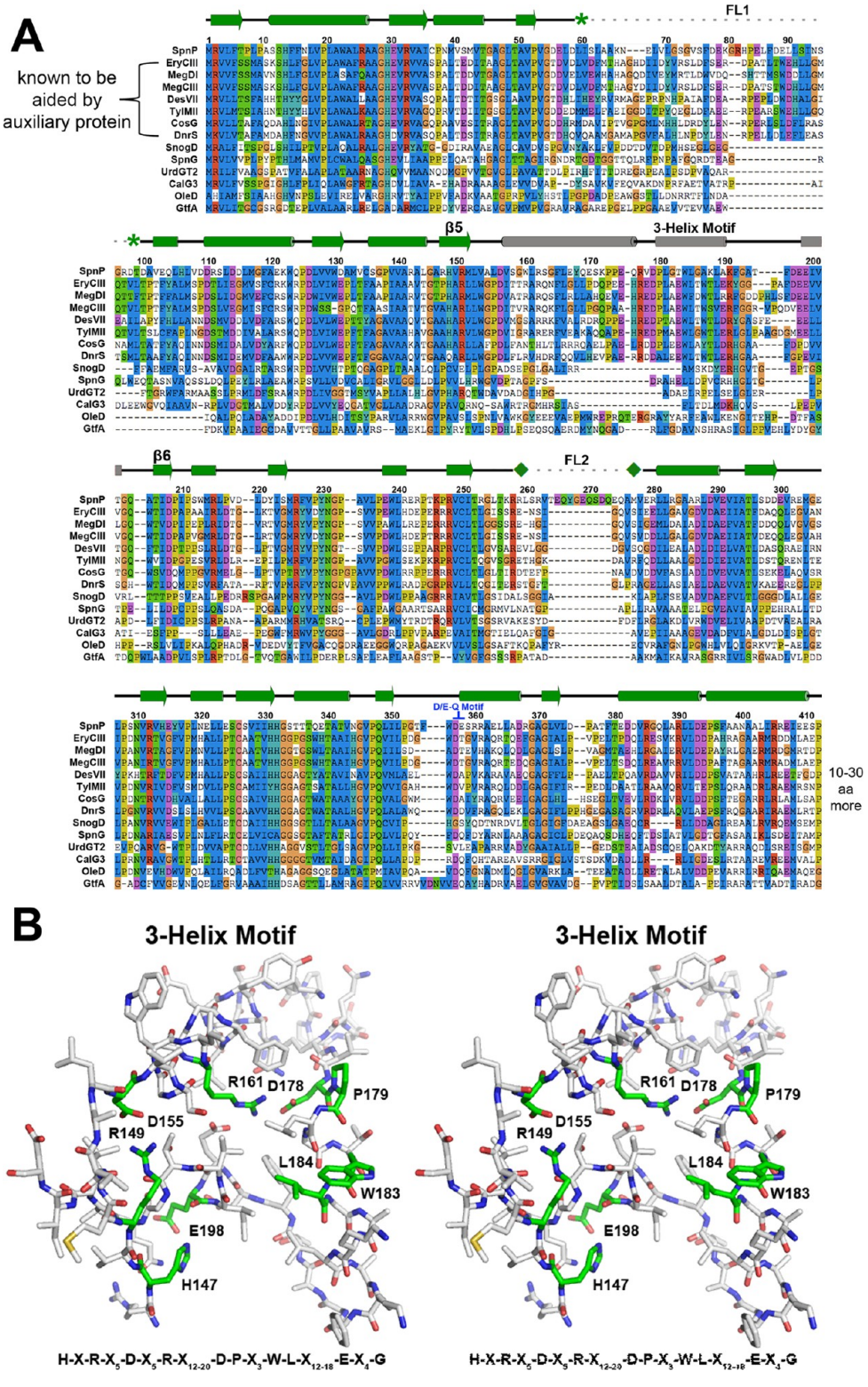


Figure 4. GT1 sequence alignment and three-helix motif. (A) A long FL1 and the three-helix motif (between the fifth and sixth β-strands of the N-terminal domain) distinguish GT1s aided by auxiliary proteins. FL1 (*) and FL2 (ϕ) are the longest disordered regions of SpnP. The D/E-Q motif, known to help bind the sugar of the donor substrate, is often substituted with other residues. The secondary structure is from SpnP. Accession numbers: SpnP, AAG23277; EryCIII, YP_001102993; MegCIII, CAC37820; DesVII, AAC68677; TylMII, CAA57472; CosG, ABC00729; DnrS, AAD15267; SnogD, AAF01811; SpnG, AAG23268; UrdGT2, AAF00209; CalG3, AAM94798; OleD, ABA42119; GtfA, AAB49292. (B) Stereodiagram of the three-helix motif showing the locations of its most highly conserved residues (green).

principally formed by the N-terminal domain of SpnP (Figure 2). Neither cocrystallization nor soaking experiments with the 17-pseudoaglycone acceptor substrate **7** yielded the anticipated density. Soaking with spinosad (**9** and **10**) also did not yield new density. FL1 (also termed the "specificity loop") was invisible in the electron density maps.[16] This region has been

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

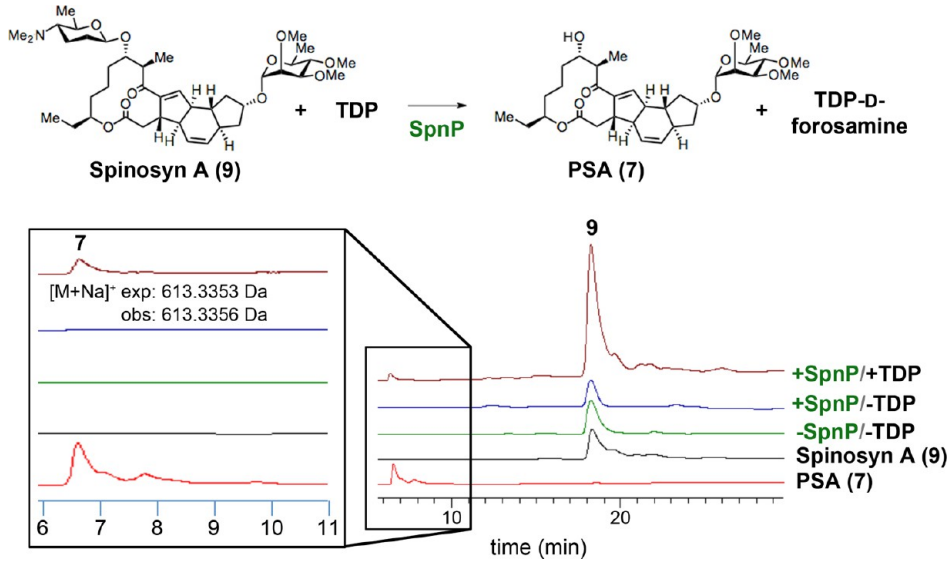


Figure 5. Reverse glycosyl transfer assay. SpnP-mediated formation of 17-pseudoaglycone **7** from spinosyn A (**9**) and superstoichiometric TDP was observed by reversed-phase HPLC (λ = 254 nm). Only when SpnP and TDP were added to **9** was **7** produced [confirmed by HRMS (see Figure S1 of the Supporting Information)].

observed to be structured in homologous GTs upon binding an acceptor substrate and forms a large portion of the acceptor substrate-binding pocket; however, it is often unstructured in the absence of an acceptor substrate.[13,14,16,35,38] Interestingly, most of FL1 in EryCIII forms helices that make substantial contact with the auxiliary protein EryCII. FL2 (residues 261–278), from the C-terminal domain, is also in position to close over a bound acceptor substrate as observed for CalG3 (PDB entry 3OTI).[14]

Three-Helix Motif. SpnP possesses a structural motif that has been observed in only one other GT structure, EryCIII (Figure 3). This motif is comprised of three helices between the fifth and sixth β-strands of the N-terminal domain. In EryCIII, these helices, along with the helices of the structured FL1, form the interface with the bound auxiliary protein EryCII. Significantly, no structurally characterized GT1 apart from SpnP and EryCIII possesses this motif. Because no other structurally characterized GTs require an auxiliary protein for activity, the presence of this motif appears to be a unique trait for a GT that pairs with an auxiliary protein. Indeed, a sequence alignment shows that this motif exists in GT1s known to require auxiliary proteins but is absent in GTs that act alone to catalyze glycosyl transfer reactions (Figure 4). The alignment of the top 100 BLAST hits for SpnP reveals a strong H-X-R-X$_5$-D-X$_5$-R-X$_{12-20}$-D-P-X$_3$-W-L-X$_{12-18}$-E-X$_4$-G motif that begins in the fifth β-strand and ends before the sixth β-strand of the N-terminal domain. Some of these highly conserved hydrophobic residues (e.g., P179 and W183) are surface-exposed. Other surface-exposed, hydrophobic residues in the motif include W159, L160, F164, Y167, and F191. These residues are likely at the interface formed between SpnP and its putative auxiliary protein. Several highly conserved neighboring residues (e.g., W129 and T205 in SpnP) provide structural support for the motif.

Activity Assays. The SpnP donor substrate, TDP-D-forosamine, is very unstable, readily losing its TDP substituent. Because of the difficulty of preparing sufficient amounts of TDP-D-forosamine to assay SpnP-catalyzed forosaminylation of 17-pseudoaglycone **7** directly, the activity of SpnP was analyzed in the reverse direction.[39] Formation of 17-pseudoaglycone **7** was monitored in the assay mixture containing SpnP, superstoichiometric TDP, and spinosad (primarily **9**) under the reverse glycosyl transfer conditions (Figure 5). A small quantity of **7** could be detected only when SpnP was present in the reaction conditions. Addition of auxiliary proteins for other GTs, such as EryCII and DesVIII, had no effect on the reaction (see the Supporting Information).

DISCUSSION

Among the growing number of GTs characterized *in vitro*, at least five have been demonstrated to be aided by an auxiliary protein: EryCIII, CosG, DnrS, MegCIII, and DesVII. The presented structural and bioinformatics analyses of forosaminyltransferase SpnP, which catalyzes the final transformation in the biosynthesis of spinosad (spinosyns A and D, **9** and **10**, respectively), indicate that SpnP is likely a new member of this subset of GTs requiring an auxiliary protein for optimal activity. The only other GT in this subset that has been structurally characterized is EryCIII, which was observed to bind to its auxiliary protein EryCII in the absence of substrates.

A comparison of the structures of SpnP and EryCIII clearly reveals the resemblance of these two proteins. Half of the EryCIII–EryCII interface is formed by helices located between the fifth and sixth β-strands of the N-terminal subdomain. A similar three-helix structural motif is also present in SpnP. A sequence alignment of the top 100 BLAST hits for SpnP shows a high degree of conservation for this motif with the consensus sequence H-X-R-X$_5$-D-X$_5$-R-X$_{12-20}$-D-P-X$_3$-W-L-X$_{12-18}$-E-X$_4$-G.[40] This motif is present in GTs known to be aided by auxiliary proteins and may be a prognostic trait of such GTs. Interestingly, SpnP can bind TDP by itself, suggesting that the SpnP homodimer can bind the donor substrate TDP-D-forosamine without the aid of its auxiliary protein. TDP is bound by SpnP in much the same way as it is bound by SpnG, with the diphosphate-binding site reconfiguring.[13] In view of its ability to form a complex with TDP and its N230/T339 motif, SpnP is expected to be selective for TDP-linked sugars. A glutamate replaces a glutamine in the second position of the D/

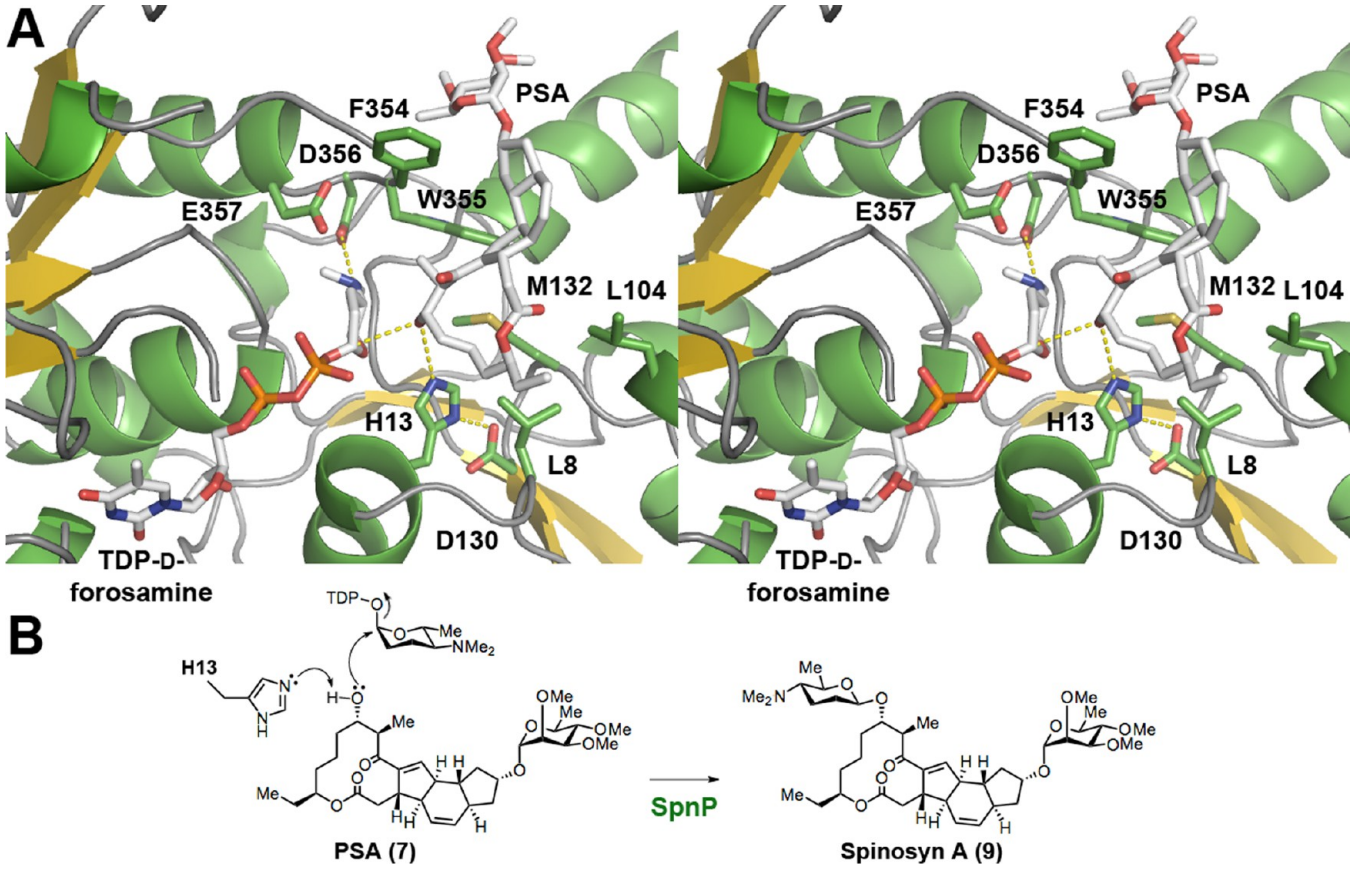


Figure 6. Model of the SpnP ternary complex. (A) TDP-D-forosamine and 17-pseudoaglycone **7** (PSA) were placed using the SpnP−TDP complex structure. The dimethylamine moiety of the forosamine sugar is in the proximity of the D356 and E357 carboxylates (the altered "D/E-Q motif"). PSA is positioned to make hydrophobic interactions with residues such as F354 and W355. The permethylated rhamnose is pointed away from the active site, while 17-OH is adjacent to the catalytic base, H13, and the electrophilic forosamine C1. (B) SpnP forosaminylation. H13 is proposed to abstract a proton from 17-OH, which nucleophilically attacks C1 of TDP-D-forosamine to displace TDP and form spinosyn A (**9**).

E-Q motif observed in most GT-B enzymes; the negatively charged residues, D356 and E357, may form ionic interactions with the positively charged tertiary amine of the forosamine sugar (Figure 6).

Although the reverse glycosyl transfer assay presented here demonstrates that SpnP possesses measurable activity in the absence of an auxiliary protein, it is relatively weak. Binding with its auxiliary protein may facilitate the proper folding of SpnP and thereby improve its catalytic efficiency. In the EryCIII−EryCII complex, FL1 is structured into helices that make significant contact with EryCII. Similarly, binding with an auxiliary protein may help order FL1 in SpnP. The closed, structured loop would then be more organized for binding the 17-pseudoaglycone acceptor substrate. Indeed, when the 17-pseudoaglycone is docked into the SpnP structure, the rhamnose sugar is positioned by FL1. An interaction between rhamnose and FL1 may help orient the substrate so that the 17-OH is placed next to the active site base H13.

Genes encoding auxiliary proteins have been found only in gene clusters encoding aminosugar-bearing metabolite biosynthesis and are typically located directly upstream of the gene of the GT responsible for the transfer of the amino sugar. However, the spinosyn biosynthetic gene cluster does not contain a gene encoding an auxiliary protein homologous to EryCII, AknT, DesVIII, TylM3, or MydC, for SpnP. Several GTs requiring auxiliary proteins are known to be activated by noncognate auxiliary proteins,[41] so SpnP could be activated by an auxiliary protein encoded by a gene in a different region in the *S. spinosa* genome. However, the P450-related proteins identified through a BLAST search of the *S. spinosa* proteome using EryCII as the query all possess the cysteine residue that is important for heme coordination and do not contain an unusually long N-terminal extension. Lacking both signature features of GT-activating auxiliary proteins, these P450-related proteins are unlikely candidates to be the auxiliary protein for SpnP.

In summary, the structure of SpnP, an aminosugar GT in the spinosyn biosynthetic pathway, was determined. The resemblance of its structure to that of EryCIII provides a strong argument for the participation of an auxiliary protein in SpnP-catalyzed glycosylation. If so, the structure of SpnP represents the second example of GTs whose catalysis is aided by an auxiliary protein. Several key hydrophobic residues, highly conserved in the three-helix motif observed in GTs aided by auxiliary proteins, are surface-exposed in SpnP and may play roles similar to those of their counterparts within EryCIII in binding an auxiliary protein. While the identity and function of the putative auxiliary protein for SpnP remain to be determined, the results presented here will impact the characterization and application of this group of GTs important in the tailoring of biologically active secondary metabolites.

dx.doi.org/10.1021/bi5003629 | *Biochemistry* 2014, 53, 4292−4301

■ ASSOCIATED CONTENT

Supporting Information

Cloning, expression, purification, and assay conditions of auxiliary proteins DesVIII and EryCII and HRMS spectra of compound 7 (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Atomic coordinates for the apo form of SpnP and its binary complex with TDP have been deposited in the Protein Data Bank as entries 4LDP and 4LEI, respectively.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: h.w.liu@mail.utexas.edu. Phone: (512) 232-7811.
*E-mail: adriankc@utexas.edu. Phone: (512) 471-2977.

Funding

This work is supported by grants provided by the National Institutes of Health (GM35906 to H.-w.L. and GM106112 to A.T.K.-C.) and the Robert A. Welch Foundation (F-1511 to H.-w.L. and F-1712 to A.T.K.-C.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Yasushi Ogasawara and Allen Yu for cloning the *spnP*, *eryCIII*, and *desVIII* genes and Steven Mansoorabadi for discussions and advice. Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology of The University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract DE-AC02-05CH11231.

■ ABBREVIATIONS

GT, glycosyltransferase; GT1, glycosyltransferase family 1; HRMS, high-resolution mass spectrometry; LB, Luria-Bertani broth; PCR, polymerase chain reaction; PKS, polyketide synthase; TDP, thymidine diphosphate; UDP, uridine diphosphate.

■ REFERENCES

(1) Kirst, H. (1991) A83543A-D, unique fermentation-derived tetracyclic macrolides. *Tetrahedron Lett. 32*, 4839−4842.

(2) Sparks, T., Crouse, G., and Durst, G. (2001) Natural products as insecticides: The biology, biochemistry and quantitative structure-activity relationships of spinosyns and spinosoids. *Pest Manage. Sci. 57*, 896−905.

(3) Waldron, C., Madduri, K., Crawford, K., Merlo, D., Treadway, P., Broughton, M. C., and Baltz, R. (2000) A cluster of genes for the biosynthesis of spinosyns, novel macrolide insect control agents produced by *Saccharopolyspora spinosa*. *Antonie van Leeuwenhoek 78*, 385−390.

(4) Waldron, C., Matsushima, P., Rosteck, P. R., Broughton, M. C., Turner, J., Madduri, K., Crawford, K., Merlo, D., and Baltz, R. (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem. Biol. 8*, 487−499.

(5) Kim, H. J., Pongdee, R., Wu, Q., Hong, L., and Liu, H.-w. (2007) The biosynthesis of spinosyn in *Saccharopolyspora spinosa*: Synthesis of the cross-bridging precursor and identification of the function of SpnJ. *J. Am. Chem. Soc. 129*, 14582−14584.

(6) Kim, H., Ruszczycky, M., Choi, S.-H., Liu, Y.-N., and Liu, H.-w. (2011) Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature 473*, 109−112.

(7) Creemer, L., Kirst, H., Paschal, J., and Worden, T. (2000) Synthesis and insecticidal activity of spinosyn analogs functionally altered at the 2′-, 3′- and 4′-positions of the rhamnose moiety. *J. Antibiot. 53*, 171−178.

(8) Sparks, T., Thompson, G., Kirst, H., Hertlein, M., Larson, L., Worden, T., and Thibault, S. (1998) Biological activity of the spinosyns, new fermentation derived insect control agents, on tobacco budworm (Lepidopter: Noctuidae) larvae. *J. Econ. Entomol. 91*, 1277−1283.

(9) Lairson, L., Henrissat, B., Davies, G., and Withers, S. (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem. 77*, 521−555.

(10) Coutinho, P., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol. 328*, 307−317.

(11) Cantarel, B., Coutinho, P., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res. 37*, D233−D238.

(12) Chen, Y.-L., Chen, Y.-H., Lin, Y.-C., Tsai, K.-C., and Chiu, H.-T. (2009) Functional characterization and substrate specificity of spinosyn rhamnosyltransferase by *in vitro* reconstitution of spinosyn biosynthetic enzymes. *J. Biol. Chem. 284*, 7352−7363.

(13) Isiorho, E., Liu, H.-w., and Keatinge-Clay, A. (2012) Structural studies of the spinosyn rhamnosyltransferase, SpnG. *Biochemistry 51*, 1213−1222.

(14) Chang, A., Singh, S., Helmich, K., Goff, R., Bingman, C., Thorson, J., and Phillips, G., Jr. (2011) Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity. *Proc. Natl. Acad. Sci. U.S.A. 108*, 17649−17654.

(15) Moncrieffe, M., Fernandez, M., Spiteller, D., Matsumura, H., Gay, N., Luisi, B., and Leadlay, P. (2012) Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *J. Mol. Biol. 415*, 92−101.

(16) Bolam, D., Roberts, S., Proctor, M., Turkenburg, J., Dodson, E., Martinez-Fleites, C., Yang, M., Davis, B., Davies, G., and Gilbert, H. (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc. Natl. Acad. Sci. U.S.A. 204*, 5336−5341.

(17) Borisova, S., Zhao, L., Melancon, C., Kao, C.-L., and Liu, H.-w. (2004) Characterization of the glycosyltransferase activity of DesVII: Analysis of and implication for the biosynthesis of macrolide antibiotics. *J. Am. Chem. Soc. 126*, 6534−6535.

(18) Gandecha, A. R., Large, S. L., and Cundliffe, E. (1997) Analysis of four tylosin biosynthetic genes from the tylLM region of the *Streptomyces fradiae* genome. *Gene 184*, 197−203.

(19) Mittler, M., Bechthold, A., and Schulz, G. E. (2007) Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. *J. Mol. Biol. 372*, 67−76.

(20) Mulichak, A., Losey, H. C., Lu, W., Wawrzak, Z., Walsh, C., and Garavito, R. (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U.S.A. 100*, 9238−9243.

(21) Hong, L., Zhao, Z., Melançon, C. E., III, Zhang, H., and Liu, H.-w. (2008) *In vivo* characterization of the enzymes involved in TDP-D-forosamine biosynthesis in the spinosyn pathway of *Saccharopolyspora spinosa*. *J. Am. Chem. Soc. 130*, 4954−4967.

(22) Lu, W., Leimkuhler, C., Gatto, G., Kruger, R., Oberthür, M., Kahne, D., and Walsh, C. (2005) AknT is an activating protein for the

glycosyltransferase AknS in L-aminodeoxysugar transfer to the aglycone of aclacinomycin A. *Chem. Biol. 12*, 527−534.
(23) Borisova, S., Zhang, C., Takahashi, H., Zhang, H., Wong, A., Thorson, J., and Liu, H.-w. (2006) Substrate specificity of the macrolide-glycosylating enzyme pair DesVII/DesVIII: Opportunities, limitations, and mechanistic hypotheses. *Angew. Chem., Int. Ed. 45*, 2748−2753.
(24) Borisova, S., and Liu, H.-w. (2010) Characterization of glycosyltransferase DesVII and its auxiliary partner protein DesVIII in the methymycin/picromycin biosynthetic pathway. *Biochemistry 49*, 8071−8084.
(25) Yuan, Y., Chung, H., Leimkuhler, C., Walsh, C., Kahne, D., and Walker, S. (2005) *In vitro* reconstitution of EryCIII activity for the preparation of unnatural macrolides. *J. Am. Chem. Soc. 127*, 14128−14129.
(26) Melançon, C., Takahashi, H., and Liu, H.-w. (2004) Characterization of tylM3/tylM2 and mydC/mycB pairs required for efficient glycosyltransfer in macrolide antibiotic biosynthesis. *J. Am. Chem. Soc. 126*, 16726−16727.
(27) Creemer, L., Kirst, H., and Paschal, J. (1998) Conversion of spinosyn A and spinosyn D to their respective 9- and 17-pseudoaglycones and their aglycones. *J. Antibiot. 51*, 795−800.
(28) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.
(29) Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D59*, 1131−1137.
(30) McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., and Read, R. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.
(31) Cohen, S., Morris, R., Fernandez, F., Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V., Kleywegt, G., and Perrakis, A. (2004) Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D60*, 2222−2229.
(32) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.
(33) Vagin, A., Steiner, R., Lebedev, A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D60*, 2184−2195.
(34) Wang, F., Zhou, M., Singh, S., Yennamalli, R., Bingman, C., Thorson, J., and Phillips, G., Jr. (2013) Crystal structure of SsfS6, the putative C-glycosyltransferase involved in SF2575 biosynthesis. *Proteins 81*, 1277−1282.
(35) Claesson, M., Siitonen, V., Dobritzsch, D., Metsa-Ketela, M., and Schneider, G. (2012) Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater*. *FEBS J. 279*, 3251−3263.
(36) Mulichak, A., Lu, W., Losey, H., Walsh, C., and Garavito, R. (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands. *Biochemistry 43*, 5170−5180.
(37) Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B., Tarling, C., Ford, C., Bowles, D., and Davies, G. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J. 25*, 1396−1405.
(38) Mulichak, A., Losey, H., Walsh, C., and Garavito, R. (2001) Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure 9*, 547−557.
(39) Zhang, C., Griffith, B., Fu, Q., Albermann, C., Fu, X., Lee, I.-K., Li, L., and Thorson, J. (2006) Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science 313*, 1291−1294.
(40) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403−410.
(41) Hong, J., Park, S., Parajuli, N., Park, S., Koh, H., Jung, W., Choi, C., and Yoon, Y. (2007) Functional analysis of desVIII homologues involved in glycosylation of macrolide antibiotics by interspecies complementation. *Gene 386*, 123−130.