Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity

Aram Chang^{a,b}, Shanteri Singh^c, Kate E. Helmich^a, Randal D. Goff^c, Craig A. Bingham^{a,b}, Jon S. Thorson^{c,1}, and George N. Phillips, Jr.^{a,b,1}

^aDepartment of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, WI 53706; ^bCenter for Eukaryotic Structural Genomics, University of Wisconsin, 433 Babcock Drive, Madison, WI 53706; and ^cLaboratory for Biosynthetic Chemistry, Pharmaceutical Sciences Division, School of Pharmacy, and National Cooperative Drug Discovery Group Program, University of Wisconsin, 777 Highland Avenue, Madison, WI 53705

Edited by Barbara Imperiali, Massachusetts Institute of Technology, Cambridge, MA, and approved July 25, 2011 (received for review May 26, 2011)

Glycosyltransferases are useful synthetic catalysts for generating natural products with sugar moieties. Although several natural product glycosyltransferase structures have been reported, design principles of glycosyltransferase engineering for the generation of glycodiversified natural products has fallen short of its promise, partly due to a lack of understanding of the relationship between structure and function. Here, we report structures of all four calicheamicin glycosyltransferases (CalG1, CalG2, CalG3, and CalG4), whose catalytic functions are clearly regiospecific. Comparison of these four structures reveals a conserved sugar donor binding motif and the principles of acceptor binding region reshaping. Among them, CalG2 possesses a unique catalytic motif for glycosylation of hydroxylamine. Multiple glycosyltransferase structures in a single natural product biosynthetic pathway are a valuable resource for understanding regiospecific reactions and substrate selectivities and will help future glycosyltransferase engineering.

Natural products with antibiotic and/or anticancer activities are a valuable pharmaceutical resource (1). Sugar moieties in these natural products are often critical to a given metabolite's biological activity and can impact the delivery of the natural product to the target, present high affinity and specificity for a given target, as well as modulate both mechanism and in vivo properties of the natural product (2). Due to these roles, altering the sugar moieties utilizing promiscuous or engineered glycosyltransferases (GTs) represents a prominent method for redesigning natural products for pharmacological applications (3–6). The crystal structures of GTs and, more specifically, an intricate understanding of how GTs achieve regio- and stereospecific reactions, will guide structure-based design and help to interpret the outcomes of directed evolution (7, 8). However, due to the lack of substrate bound GT structures, these engineering methods have thus far been only successful in very limited cases (9, 10).

Calicheamicin γ_1^I (CLM), the flagship member of the naturally occurring 10-membered enediynes, provides a unique model for interrogating the regiochemistry of GTs (11). While an iterative type I polyketide synthase in conjunction with tailoring enzymes provide the novel enediyne core (12–14), four unique GTs are required to complete the biosynthesis of the CLM aryltetrasaccharide, composed of four novel sugar moieties and an orsellinic acid-like moiety (Fig. 1). Some CLM GTs are highly promiscuous and can perform forward, reverse, and exchange reactions, enabling chemoenzymatic methods to generate glycodiversified CLM analogs (15, 16). Based upon biochemical studies, CalG1 and CalG4 were found to be external GTs, acting as a rhamnosyltransferase for sugar moiety D and as an aminopentosyltransferase for sugar moiety E, respectively. Alternatively, CalG2 and CalG3 were characterized as internal GTs, acting as a thiosugar-transferase for sugar moiety B and as a hydroxylaminoglycosyltransferase for sugar moiety A, respectively (Fig. 1). Previously, a CalG3 unliganded structure was reported (16); however, the absence of substrates in the model prevented understanding of

the binding mode of CLM and identification of the origins of regiospecificity.

Here, we report the ligand-bound CalG3, CalG2, CalG1, and unliganded CalG4 structures and complete the GT structure analysis of CLM biosynthetic pathway. The entire set of CLM GT structures reveal a conserved CLM coordination motif among this GT set as well as the key features that dictate the different binding modes of the substrates and the resulting distinct regiospecific reactions. In addition, this comprehensive GT structural study is anticipated help guide future GT engineering efforts.

Results

Overall Structure Description and Donor Molecule Binding in the C-Terminal Domain of CLM GTs. The crystal structure of CalG3 with thymidine diphosphate (TDP) and CLM T_0 (Fig. 1) was solved to a resolution of 1.6 Å (Fig. 2A and Table S1); CalG2 with TDP and CLM T_0 was solved to a resolution of 2.2 Å (Fig. 2B and Table S1); CalG4 in an unliganded form was solved to a resolution of 1.9 Å (Fig. 2C and Table S2); and CalG1 with TDP and CLM α_3^I (Fig. 1) was solved to a resolution of 2.3 Å (Fig. 2D and Tables S1 and S2). Despite their low sequence identities (Fig. S1 A and B), all CLM GTs adopt a conserved GT-B fold, with the N-terminal and C-terminal domains forming a Rossmann fold connected by a linker region. All substrate bound structures adopt a “closed” conformation, while previous CalG3 and CalG4 unliganded structures demonstrate an “open” conformation (Fig. S2). With the exception of some variability in CalG2, the TDP molecule is bound in a highly conserved manner in the C-terminal domain through π -stacking interactions with tryptophan side chain and through hydrogen bonds with nitrogen and oxygen atoms of the polypeptide backbone (Fig. S3). This structural consistency implies that the main causes of regiospecificity among the structures are within the acceptor binding regions of the proteins.

CalG3 Acceptor Binding Mode. CLM T_0 , when bound to CalG3, is located between the N-terminal and the C-terminal domains

Author contributions: A.C., S.S., J.S.T., and G.N.P. designed research; A.C., S.S., and K.E.H. performed research; R.D.G. contributed new reagents/analytic tools; A.C., S.S., K.E.H., C.A.B., J.S.T., and G.N.P. analyzed data; and A.C., J.S.T., and G.N.P. wrote the paper.

The authors declare a conflict of interest (such as defined by PNAS policy). The authors declare competing financial interests. J.S.T. is cofounder of Centrose, Madison, WI.

This article is a PNAS Direct Submission.

Data deposition: The structure factor amplitudes and coordinates of CalG3 with TDP and calicheamicin T_0 , CalG2 with TDP and calicheamicin T_0 , CalG2 with TDP, CalG4, CalG1 with TDP and calicheamicin α_3^I , CalG1 with TDP were deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3OTI, 3R5C, 3IAA, 3IA7, 3OTH, and 3OTG, respectively).

¹To whom correspondence may be addressed. E-mail: phillips@biochem.wisc.edu or jsthorson@pharmacy.wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1108484108/-DCSupplemental.

Fig. 1. Proposed calicheamicin glycosylation pathway. CalG3 mediates an internal glycosylation to the aglycon, while CalG2 mediates an internal glycosylation and CalG4 mediates an external glycosylation to the sugar A. CalG1 operates external glycosylation to the orsellinic acid-like moiety (moiety C). The order of the CalG1 and CalG4 reactions are not characterized in vivo. The names of calicheamicin intermediates are indicated below the structure. The calicheamicin γ_1^1 chemical structure and sugar nomenclature is in the bottom right. The aryltetrasaccharide portion (four sugars and orsellinic acid-like moiety) is colored in blue.

(Fig. 24 and Fig. S44). CLM T_0 is recognized by three specific aromatic residues, which define a distinct CLM recognition motif (17) (Fig. 3A). The planar imidazole side chain of His11, a catalytic residue, is orthogonal to the enediyne plane, and the position of Ne2 of His11 is near the center of the 10-membered ring of CLM T_0 , forming a cation- π interaction. Phe60 is orthogonal to another face of the ring, pointing toward one of the conjugated single bonds of the enediyne, showing a CH- π or edge-to-face interaction. Phe310 forms a π -stacking interaction with the cyclohexenone, although this ring is slightly tilted with respect to the plane. Most of these residues adopt different conformations in the unliganded structure and show evidence of either conformational selection or induced fit (Fig. 3A). The methylated trisulfide

Fig. 2. Overall calicheamicin GT structures and different binding mode. (A) Cartoon representation of the overall structure of CalG3 with TDP and calicheamicin T_0 complex monomer, a closed conformation and bi-domain binding mode. (B) CalG2 with TDP and calicheamicin α_3^1 complex structure, a closed conformation and N-terminal domain cavity binding mode. (C) CalG4 unliganded form, an open conformation. (D) CalG1 with TDP and calicheamicin α_3^1 complex structure, a closed conformation and bi-domain binding mode. All ligands are shown as spheres (TDP: purple, CLM: orange).

Fig. 3. Calicheamicin coordination and catalytic residues in CLM GTs. (A) CalG3 complex structure (green) and unliganded structure (silver) with the key residues that recognize the 10-membered enediyne moiety and cyclohexenone (orange). The side chain of His11 rotates 90° and the Phe60 side chain undergoes a flip upon acceptor binding. The rotation of His11 forms a hydrogen bond between the two catalytic residues to facilitate the glycosyltransfer reaction. (B) CalG2 complex structure of CLM. Phe67, Tyr80, and His77 are utilized for coordination of the enediyne moiety (orange). Thr238 or Asp325 is proposed as a catalytic residue. (C) CalG4 structure (light orange) overlaid with the CLM in CalG2 structure (silver) Tyr82, Trp146, and His79 are proposed to be involved in the coordination of CLM. Phe60, Phe63, or His64 are also proposed to be involved in the coordination of CLM via induced fit. Catalytic residues are His16 and Asp108. (D) CalG1 complex structure (cyan). The aryltetrasaccharide moiety is located in the hydrophobic cleft between the two domains and Phe90 is involved in a π -stacking interaction with moiety C. The small box in the upper left corner in all figures represent the whole structure and the black box indicate the region that is zoomed in. N and C means N-terminal and C-terminal domains, respectively.

group is surrounded by hydrophobic residues (Fig. S44). The Glu/Asp–Gln pair, which has been proposed as a determinant of the donor sugar specificity (18, 19), is not conserved in CalG3. Only Gln311 remains and interacts with sugar A (C2-OH in the sugar A with Ne2 of Gln311, and C3-OH in the sugar A with Oe1) (Fig. S54).

CalG2 Acceptor Binding Mode. Although CalG2 and CalG3 are closely related functionally (the product of CalG3 is the substrate of CalG2), the binding mode of CLM T₀ in CalG2 is clearly distinct, binding within a hydrophobic cavity in the N-terminal domain (Fig. 2B and Fig. S4B). Among three specific aromatic residues that coordinate the CLM enediyne moiety in the CalG3 structure, only two of them are identified in the CalG2 structure (Fig. 3B). Phe67 points toward the center of the 10-membered ring forming a CH- π interaction, and Tyr80 forms a π -stacking interaction with the cyclohexenone, corresponding to His11 and Phe310 of CalG3, respectively. Also, there is a hydrogen bond between a hydroxyl group in the enediyne ring and His77. The methylated trisulfide is again located in the hydrophobic region that is surrounded by the N3 loop and α helix. There is no direct interaction between sugar A and the surrounding CalG2 residues. Asp325 remains in the Glu/Asp–Gln pair; however, its role is not clear due to the lack of a donor sugar moiety in the structure (Fig. S5B).

CalG4 Acceptor Binding Mode. Because of the highly similar conformations of the N3 and N5 regions (Fig. S64), which is the most important determinant of acceptor molecule binding, the CLM binding mode in CalG4 can be predicted from the overlay of the CalG2 structure on the CalG4 structure (Fig. 3C). Tyr82 and His79 of CalG4 are in the same position as Tyr80 and His77 of CalG2. The Phe67 residue of CalG2, involved in a CH- π interaction with the enediyne moiety, is not conserved in CalG4; however, Phe60, Phe63, or His64 might take a similar role via an induced fit upon substrate binding. Besides these residues, Trp146 is proposed to coordinate the enediyne moiety by pointing a conjugated single bond, similar to Phe60 of CalG3. Although the same aglycon binding modes are expected in both CalG2 and CalG4, sugar A needs to be adjusted in CalG4 to bring its O2 reactive group close to the catalytic residue, His16. When sugar A is adjusted in the CalG4 model, not only will O2 be pointing toward the catalytic residue, but also the hydroxylamine of C4 will be pointing toward the cleft between the two domains. This means that the C4 position has the capacity to accommodate an extra moiety and thus explains why the CalG4 reaction is promiscuous for CLM variants at this position (15).

CalG1 Acceptor Binding Mode. In the CalG1 structure, CLM α_3^1 was seen bound in the hydrophobic cleft between the N-terminal and C-terminal domains (Fig. 2D). The electron density for sugar D is missing, presumably removed by the CalG1 reverse reaction (Fig. S4C). Unlike CalG3, CalG2, and possibly CalG4, CalG1 mainly utilizes the aryltetrasacchride of CLM for substrate coordination (Fig. 3D). Phe90 forms a π - π stacking interaction with the C moiety and is considered one of the essential residues for the coordination of that aromatic ring. The C2 OH group in sugar A points outward, which explains why the CalG1 reaction does not discriminate among CLM sugar E variants (15). The enediyne is located at the opening of the cleft in the solvent exposed area and does not have direct interactions with CalG1. The trisulfide is located in the hydrophobic region, generated by the N α 3a and N α 3b helices, similar to other CLM GTs. Again, the Glu/Asp–Gln pair is not conserved in CalG1 (Fig. S5D). Only Asp319 is present in the conserved region, implying possible interactions with the equatorial C4-OH of sugar D, which might provide for a wide range of donor sugar promiscuity.

Active Site Architecture. CalG1, CalG3, and CalG4 utilize a catalytic dyad, histidine and aspartate, located in the cleft between the two domains, which is highly conserved in other GTs (19–22) (Fig. 3A, C, and D and Fig. S6B). The low barrier hydrogen bond formation between Asp and His side chains will facilitate nucleophilic attack on the acceptor hydroxyl group in the CLM via a serine hydrolase-like mechanism (23–25). In the case of CalG2, Leu14 takes the typical position of histidine, whose catalytic activity is missing due to a lack of nucleophilicity, which indicates a different mechanism in CalG2, or a different nucleophile (Fig. 3B and Fig. S6B). Based on the distance from the hydroxylamine group in sugar A to the CalG2 residues, candidates for the catalytic residues of CalG2 are either Thr238 or Asp325 (3.9 Å and 2.4 Å, respectively). However, Asp325 is present in the Glu/Asp–Gln motif, which interacts with the transferring sugar in other CLM GTs (18, 19) and is thus not unique to CalG2.

Discussion

All four CLM GT structures adopt the same GT-B fold and donor molecule binding region and demonstrate good alignment despite quite low sequence identities (Fig. S1A and B). The principles for the coordination of the acceptor molecule are conserved. Enediyne coordination is accomplished via interactions with three aromatic residues (or two in CalG2) (Fig. 3). Also, the residues that accommodate the methyltrisulfide serve to “protect” the methyltrisulfide from reductive activation, thus preventing a premature Bergman cycloaromatization event. Despite these similarities, the acceptor molecule binding region of the CLM GTs displays specialization, demonstrated by the N-terminal domains, most notably by the N3 and N5 regions (α -helices and loops located between strands β 3 and β 4, β 5 and β 6, respectively) (Fig. 4 and Fig. S1), that display strong sequence and structural variation in which, in turn, invokes functional differentiation.

Differentiation of CalG3/CalG1 and CalG2/CalG4 Functions. Based on their acceptor molecule binding modalities, CalG3 and CalG1 can be grouped together as using a “bi-domain” binding mode (22) and CalG2 and CalG4 can be grouped together as using an “N-terminal cavity” binding mode (19–21). The determinant of the binding mode is driven by the presence or absence of a cavity produced by the N3 and N5 regions. In the “bi-domain” binding mode of CalG3 and CalG1, there is only one N α 5 helix, which is very close to the N α 3c helix, contributing to the lack of space between the N3 and N5 regions, in turn requiring a different acceptor molecule binding region (Fig. 4A and D). Meanwhile, CalG2 and CalG4 have multiple, long N α 5 helices, which create a substantial cavity between the N3 and N5 regions for acceptor molecule binding (Fig. 4B and C). This observation implies that the overall GT structure provides a general catalytic platform and that the GT chimeras produced by swapping the N3 and N5 regions might contribute to changes in the acceptor regiospecificity and increased reaction promiscuity. This contention is further supported by prior mutagenesis studies that implicate the N3 and N5 loops as influencing reaction specificity (26–28).

Differentiation of CalG3 and CalG1. CalG3 and CalG1 function as internal and external GTs, respectively. The key residues to build the different binding site architectures and invoke an internal vs. external reaction are Pro95 and Phe152 of CalG3 in the middle of the N3c helix and the N5 helix, respectively, which act as a “helix breaker.” Due to these two residues, CalG3 adopts bent N3c and N5 helices, which contribute to the creation of a “smaller” acceptor binding space (Fig. 4A). On the other hand, CalG1 has linear N3c and N5 helices, which form a straight wall within the cleft between the two domains and coordinate a lengthy substrate (Fig. 4D). Therefore, residues remote from the active sites con-

Fig. 4. Differences in CLM GTs in the N3 and N5 regions and mode of acceptor molecule binding. (A) CalG3 N3 (Asp49-Asp110) and N5 (Arg135-Ala169) regions. (B) CalG2 N3 (Pro53-Asp101) and N5 (Ser128-Leu185) regions. (C) CalG4 N3 (Leu55-Asp103) and N5 (Thr130-Leu187) regions. (D) CalG1 N3 (Ala52-Asp112) and N5 (His137-Pro177) regions. Bound CLM is shown as spheres. In A, the CLM sugar A moiety in the model was deleted to display a substrate, not a product structure. The small box in the upper middle corner in all figures represent the whole structure and the black box indicates the region of interest.

tribute to the different architectures of substrate binding and also influence the regiospecificity of the reactions. Electrostatic properties are another determinant of the differential binding mode (Fig. S6 C and D). CalG1 has slightly negatively charged residues in the CalG3 trisulfide moiety binding region, which is governed by hydrophobic residues. This feature prevents CalG1 from possible CalG3 substrate (calicheamicone) binding. The N-terminal domain cavities of other natural product GTs are also dominated by hydrophobic residues.

Differentiation of CalG2 and CalG4. Due to the expected similarity of the acceptor molecule binding modes in CalG2 and CalG4, catalytic residue relocation in CalG2 compared to CalG4 is utilized to achieve the regiospecificity (Fig. 3 B and C and Fig. S6). The nucleophile on the acceptor of CalG2 is a hydroxylamine, which is more reactive than the typical hydroxyl group (pKa of 13.7 vs. 15 ~ 16). Therefore, CalG2 appears not to need the usual catalytic dyad, and Thr238 or Asp325 may mediate the reaction.

Phylogenetic Origins of CLM GTs. All CLM GTs have been assigned to the GT-1 family in the CAZy database (29). Phylogenetic analysis of the bacterial GT-1 family suggests that while most GTs in the same pathway are highly related, CLM GTs might have been derived from distant ancestor genes (Fig. S1C). CalG2 and CalG4 likely originate from a relatively recent common ancestor sequence, as expected from their sequential and structural similarity. However, CalG3 and CalG1 likely come from a much more distant phylogenetic origin than CalG2 and CalG4. An attempt to predict different binding modes or to identify “helix breaker” residues from the phylogenetic tree, alignment of sequences, or predicted secondary structure elements failed to produce recognizable patterns.

Conclusion

CLM GTs are prime examples of how structurally homologous enzymes achieve their regiospecific reactions and thereby contribute to diverse chemical reactivities. The set of GT structures in the CLM biosynthetic pathway possess the conserved CLM coordination signature (Fig. 3); CalG3, CalG2, and CalG4 utilize three (or two) aromatic residues for the enediyne coordination through cation- π and/or CH- π interaction and π stacking interaction. The dispositions of these residues in each GT are different in

order to accommodate different acceptor molecule positions and regiospecific reactions. CalG1 is distinguished from other CLM GTs because there is no direct interaction with the enediyne core. In this report, we show that fundamental determinants of acceptor molecule binding are localized in the N3 and N5 regions (CalG1, CalG3 vs. CalG2, CalG4), which suggest that mutating and exchanging these regions would be best place to focus engineering. Also, two “helix breaker” residues of CalG3 (Pro95 and Phe152), electrostatic charges (CalG3 vs. CalG1) and catalytic residue reorientation (CalG2 vs. CalG4) are able to contribute to the further regiospecific functional differentiation among the four CLM GTs (Fig. 5). The lesson from the CLM GT structures explains not only the common principle of enzymes in natural product biosynthesis pathway but also provides various possible methods for the rational design of the alteration of GT specificities.

Fig. 5. Principles of CLM GTs regiospecificity. Simplified phylogenetic tree showing pairs of GTs and their specified adaptations. CalG3 and CalG1 share 28% sequence identity and have their acceptor bound between the two domains, and CalG2 and CalG4 share 49% sequence identity and have their acceptor bound internally. In CalG3, the N3c and N5 helices are bent by two helix breaker residues. In CalG2, catalytic residues are altered for the hydroxylamine glycosidic bond linkage.

Methods

Sample Preparation. CLM α_3^1 was provided by Pfizer. CLM T₀ was prepared as previously described (16). CalG3, CalG2, and CalG1 with TDP samples were prepared by mixing 10 mg/mL of CalG3 or 20 mg/mL of CalG2 or CalG1 protein samples with 25 mM TDP. For preparing CalG3, CalG2, or CalG1 with TDP and CLM T₀ or α_3^1 , approximately 0.1 mg of CLM powder were dissolved in 5 μ L 100% methanol, then added to 20 μ L of CalG3, CalG2 or CalG1 protein with TDP sample prepared above, before methanol evaporated. Samples were centrifuged at max speed for 10 s to remove precipitated CLM and make fully saturated CalG3, CalG2, or CalG1 with TDP and CLM T₀ or α_3^1 solutions. Supernatants were taken out and clear but tint red color was observed. All crystal screens are set up with these supernatants.

X-ray Crystallography. Initial screens were performed with a local screen UW192, IndexHT, and SaltHT (Hampton research) utilizing a Mosquito® dispenser (TTP labTech) by the sitting drop method. Crystal growth was monitored by Bruker Nonius Crystal Farms at 20 °C and 4 °C.

CalG3 with TDP and CLM T₀ crystals were grown by mixing 1 μ L of sample solution and 1 μ L of reservoir solution, 28% MEPEG 2K, 160 mM Na₃Citrate, and 100 mM NaAcetate pH 4.5 at 20 °C using hanging drop method. CalG2 with TDP and CLM T₀ crystals were grown by mixing 1 μ L of sample solution and 1 μ L of reservoir solution, 0.5% MEPEG 5K, 800 mM Na K-tartrate, and 100 mM Tris pH 8.5 at 20 °C using hanging drop method. CalG2 with TDP crystals were grown by mixing 10 μ L of sample solution and 10 μ L of reservoir solution, 800 mM Na₃Citrate and 100 mM BisTris pH 6.5 at 20 °C using batch method. CalG4 crystals were grown by mixing 1 μ L of sample solution and 1 μ L of reservoir solution, 20% PEG 4K, 80 mM CaCl₂, 100 mM Arg-Glu, and 100 mM CHES pH 9.5 at 4 °C using hanging drop method. Streak seeding was utilized to provide diffraction-quality crystals. CalG1 with TDP and CLM α_3^1 crystals were grown by mixing 1 μ L of sample solution and 1 μ L of reservoir solution, 16% MEPEG 5 K, 160 mM CaCl₂, and 100 mM MES/Acetate pH 5.5 at 20 °C using hanging drop method. CalG1 with TDP crystals were grown by mixing 1 μ L of sample solution and 1 μ L of reservoir solution, 20% PEG3350, 0.2 M LiSO₄, 100 mM BisTris pH 6.5 at 20 °C using hanging drop method. All crystals were cryoprotected with reservoir solution and 20% ethylene glycol except CalG2 with TDP and CLM T₀ crystal, which were protected by foblin, and were flash frozen in liquid nitrogen. Cryosolutions of CalG2 with TDP and CalG1 with TDP require an additional 10 mM TDP.

Data Collection. X-ray diffraction data were collected at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) with X-ray wavelength of 0.9794 Å (CalG2/TDP), 0.9794 Å and 0.9642 Å (CalG4 and CalG1/TDP, peak and remote) and at the Life Science Collaborative Access Team (LS-CAT) with X-ray wavelength of 0.9794 Å (CalG3/TDP/CLM, CalG2/

TDP/CLM, and CalG1/TDP/CLM) at the Advanced Photon Source at Argonne National Laboratory.

Datasets were indexed and scaled using HKL2000 (30). CalG2/TDP/CLM dataset displays a lattice translocation disorder and requires special treatment (31, 32) (*SI Text* and *Fig. S7*). For phasing experiments (CalG4, CalG1/TDP), phenix.HySS (33) and ShelxD (34) were utilized for determining the selenium substructures, autoSHARP for phasing (35), DM for density modification (36), and phenix.autobuild for automatic model building (33). For CalG3 with bound TDP and CLM T₀ structures, molecular replacement was used with a separated N-terminal domain (1–200) and C-terminal domain (201–375) using the previously determined CalG3 structure (PDB ID code 3D0R) as a starting model. For the CalG2 with bound TDP and CLM T₀ structure, molecular replacement was used with the CalG2/TDP structure (PDB ID code 3IAA) as a starting model. For the CalG2 with bound TDP structure, molecular replacement was used with a separated N-terminal domain (1–200) and C-terminal domain (201–375) of the CalG4 structure (PDB ID code 3IA7) as a starting model. For the CalG1 with bound TDP and CLM α_3^1 , molecular replacement was used starting with the CalG1/TDP structure (PDB ID code 3OTG). phenix-AutoMR and phenix.AutoBuild were utilized for molecular replacement and model rebuilding (33). The structures were completed with alternating rounds of manual model building with COOT (37) and refinement with phenix.refine (33). The final rounds of CalG1 and TDP structure refinement included eight TLS groups (38). Structure quality was assessed by Procheck (39) and Molprobit (40). All figures in this paper were generated by PyMOL (41).

ACKNOWLEDGMENTS. We thank Dr. Christopher M. Bianchetti for helpful discussion; Younghee Shin for the help with confirming the calicheamicin α_3^1 compound with NMR measurements; and Dr. Atilla Sit for the help with programming that handled CalG2/TDP/CLM lattice translational defect problem. We thank Pfizer for graciously providing calicheamicins. This research was supported in part by National Institutes of Health (NIH) Grant CA84374 (J.S.T.), U54 GM074901 (G.N.P.), U01 GM098248 (G.N.P.), and NIH Molecular Biophysics Training Grant GM08293 (A.C.). J.S.T. is a University of Wisconsin HI Romnes Fellow and holds the Laura and Edward Kremers Chair in Natural Products. The General Medicine and Cancer Institute Collaborative Access Team (GM/CA-CAT) has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). The Life Sciences Collaborative Access Team (LS-CAT) has been supported by Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-102-ENG-38.

- Walsh CT, Fischbach MA (2010) Natural products version 2.0: Connecting genes to molecules. *J Am Chem Soc* 132:2469–2493.
- Weymouth-Wilson AC (1997) The role of carbohydrates in biologically active natural products. *Nat Prod Rep* 14:99–110.
- Thibodeaux CJ, Melancon CE, Liu HW (2007) Unusual sugar biosynthesis and natural product glycodiversification. *Nature* 446:1008–1016.
- Williams GJ, Gantt RW, Thorson JS (2008) The impact of enzyme engineering upon natural product glycodiversification. *Curr Opin Chem Biol* 12:556–564.
- Griffith BR, Langenhan JM, Thorson JS (2005) ‘Sweetening’ natural products via glycorandomization. *Curr Opin Biotechnol* 16:622–630.
- Blanchard S, Thorson JS (2006) Enzymatic tools for engineering natural product glycosylation. *Curr Opin Chem Biol* 10:263–271.
- Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem* 77:521–555.
- Williams GJ, Thorson JS (2009) Natural product glycosyltransferases: properties and applications. *Adv Enzymol Relat Areas Mol Biol* 76:55–119.
- Palcic MM (2011) Glycosyltransferases as biocatalysts. *Curr Opin Chem Biol* 15:226–233.
- Chang A, Singh S, Phillips GN, Thorson JS (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol*, 10.1016/j.copbio.2011.04.013.
- Thorson JS, et al. (2000) Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research. *Curr Pharm Des* 6:1841–1879.
- Ahlert J, et al. (2002) The calicheamicin gene cluster and its iterative type I enediynes PKS. *Science* 297:1173–1176.
- Liu W, Christenson SD, Standage S, Shen B (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* 297:1170–1173.
- Horsman GP, Chen Y, Thorson JS, Shen B (2010) Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes. *Proc Natl Acad Sci USA* 107:11331–11335.
- Zhang C, et al. (2006) Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* 313:1291–1294.
- Zhang C, et al. (2008) Biochemical and structural insights of the early glycosylation steps in calicheamicin biosynthesis. *Chem Biol* 15:842–853.
- Kim KH, Kwon BM, Myers AG, Rees DC (1993) Crystal structure of neocarzinostatin, an antitumor protein-chromophore complex. *Science* 262:1042–1046.
- Hu Y, et al. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci USA* 100:845–849.
- Bolam DN, et al. (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc Natl Acad Sci USA* 104:5336–5341.
- Mulichak AM, et al. (2003) Structure of the TDP-epi-vancosaminyltransferase GtFA from the chloroeremomycin biosynthetic pathway. *Proc Natl Acad Sci USA* 100:9238–9243.
- Mulichak AM, Lu W, Losey HC, Walsh CT, Garavito RM (2004) Crystal structure of vancosaminyltransferase GtFd from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands. *Biochemistry* 43:5170–5180.
- Offen W, et al. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J* 25:1396–1405.
- Frey PA, Whitt SA, Tobin JB (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* 264:1927–1930.
- Cleland WW, Kreevoy MM (1994) Low-barrier hydrogen bonds and enzymic catalysis. *Science* 264:1887–1890.
- Cleland WW, Frey PA, Gerlt JA (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 273:25529–25532.
- Hoffmeister D, Ichinose K, Bechthold A (2001) Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. *Chem Biol* 8:557–567.
- Hoffmeister D, et al. (2002) Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity. *Chem Biol* 9:287–295.
- Williams GJ, Zhang C, Thorson JS (2007) Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat Chem Biol* 3:657–662.
- Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZY): An expert resource for Glycogenomics. *Nucleic Acids Res* 37:D233–238.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.

31. Wang J, Kamtekar S, Berman AJ, Steitz TA (2005) Correction of X-ray intensities from single crystals containing lattice-translocation defects. *Acta Crystallogr D Biol Crystallogr* 61:67–74.
32. Hare S, Cherepanov P, Wang J (2009) Application of general formulas for the correction of a lattice-translocation defect in crystals of a lentiviral integrase in complex with LEDGF. *Acta Crystallogr D Biol Crystallogr* 65:966–973.
33. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
34. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
35. de la Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* 276:472–494.
36. Cowtan KD, Main P (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D Biol Crystallogr* 52:43–48.
37. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
38. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr* 39:109–111.
39. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
40. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–383.
41. Delano WL (2002) *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA).