Chemistry & Biology
Article

# Biochemical and Structural Insights of the Early Glycosylation Steps in Calicheamicin Biosynthesis

Changsheng Zhang,[1] Eduard Bitto,[2] Randal D. Goff,[3] Shanteri Singh,[3] Craig A. Bingman,[2] Byron R. Griffith,[3] Christoph Albermann,[3] George N. Phillips, Jr.,[2,*] and Jon S. Thorson[3,*]
[1]Guangdong Key Laboratory of Marine Materia Medica, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 164 Xingang West Road, Guangzhou 510301, China
[2]Department of Biochemistry
[3]Laboratory for Biosynthetic Chemistry, Pharmaceutical Sciences Division, School of Pharmacy, University of Wisconsin-National Cooperative Drug Discovery Group Program
University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI 53706-1544, USA
*Correspondence: phillips@biochem.wisc.edu (G.N.P.), jsthorson@pharmacy.wisc.edu (J.S.T.)
DOI 10.1016/j.chembiol.2008.06.011

## SUMMARY

**The enediyne antibiotic calicheamicin (CLM) $\gamma_1^I$ is a prominent antitumor agent that is targeted to DNA by a novel aryltetrasaccharide comprised of an aromatic unit and four unusual carbohydrates. Herein we report the heterologous expression and the biochemical characterization of the two "internal" glycosyltransferases CalG3 and CalG2 and the structural elucidation of an enediyne glycosyltransferase (CalG3). In conjunction with the previous characterization of the "external" CLM GTs CalG1 and CalG4, this study completes the functional assignment of all four CLM GTs, extends the utility of enediyne GT-catalyzed reaction reversibility, and presents conclusive evidence of a sequential glycosylation pathway in CLM biosynthesis. This work also reveals the common GT-B structural fold can now be extended to include enediyne GTs.**

## INTRODUCTION

Calicheamicin (CLM) $\gamma_1^I$ (Figure 1, **1**) from *Micromonospora echinospora* spp. *calichensis* is a prominent member of the enediyne family because of its unprecedented molecular architecture, remarkable mechanism of action, and clinical utility (Galm et al., 2005; Thorson et al., 2000; Van Lanen and Shen, 2008). Structurally, CLM is a member of the 10-membered enediynes, which all share a signature bicyclo[7.3.1]enediyne core. Like all enediynes, CLM-induced oxidative DNA strand scission is enabled by rapid enediyne cycloaromatization to form a highly reactive diradical species (Zein et al., 1989, 1988). This reactive intermediate is exquisitely positioned by the CLM aryltetrasaccharide (Figure 1), the critical DNA docking element of CLM (Kumar et al., 1997; Walker et al., 1993). The incredible potency of CLM has also been harnessed for clinical use by way of conjugation to tumor-targeting antibodies, as exemplified by the γ-CD33 antibody conjugate (Mylotarg) approved by FDA in 2000 to treat acute myelogenous leukemia (AML) (Sievers and Linenberger, 2001). Similarly appended CLM-antibody conjugates to treat other cancers are steadily progressing through clinical trials (Boghaert et al., 2004; DiJoseph et al., 2005; Hamann et al., 2005).

Consistent with their many novel structural and pharmacological features, enediyne biosynthetic pathways are also rich with unique enzyme-catalyzed biotransformations. Early metabolic labeling studies suggested the 9- and 10-membered enediynes derive from distinct biosynthetic pathways (Hensens et al., 1989; Lam et al., 1993; Tokiwa et al., 1992). In contrast, the recent cloning and characterization of gene clusters encoding 9-membered enediynes, including C-1027 (Figure 1, **4**; Liu et al., 2002), neocarzinostatin (Figure 1, **5**; Liu et al., 2005), and maduropeptin (Figure 1, **6**; Van Lanen et al., 2007), and 10-membered enediynes, including CLM (Ahlert et al., 2002), esperamicin (Figure 1, **2**; and J.S.T., unpublished data, AY267372), and dynemicin (Figure 1, **3**; Gao and Thorson, 2008; Zazopoulos et al., 2003), revealed a unified, divergent polyketide paradigm for enediyne core biosynthesis (Ahlert et al., 2002; Liu et al., 2002; Zhang et al., 2008). Some enediyne-producing organisms also rely on a unique "self-sacrifice" resistance protein (as exemplified by the CLM protein CalC) for enediyne self-resistance (Biggins et al., 2003; Singh et al., 2006). Shen and coworkers were the first to demonstrate the elegant application of pathway engineering to produce chromoprotein enediyne analogs with drastically differing activities (Kennedy et al., 2007a, 2007b; Liu et al., 2002), whereas sugar exchange and aglycon exchange reactions catalyzed by the CLM glycosyltransferases (GTs) CalG1 and CalG4 recently enabled the production of more than 70 differentially glycosylated CLM variants (Zhang et al., 2006b). Though this latter study also provided in vitro biochemical characterization of CalG1 and CalG4 as the CLM 3-*O*-methyl-rhamnosyltransferase and aminopentosyltransferase, respectively (Zhang et al., 2006b), the function of the remaining CLM glycosyltransferases CalG2 and CalG3 remain unresolved. In addition, though the structures for various natural product-associated glycosyltransferase have emerged in recent years (Bolam et al., 2007; Mittler et al., 2007; Mulichak et al., 2001, 2003, 2004), enediyne GTs remain structurally uncharacterized (Van Lanen and Shen, 2008).

Herein we report the further study of the internal stages of CLM glycosylation. Specifically, using a combination of GT reaction reversibility and sugar nucleotide surrogates, CalG3 was verified as the requisite calicheamicinone 4,6-dideoxy-4-hydroxylamino-α-D-glucosyltransferase and demonstrated to accept a set of 10 alternative sugar nucleotide donors. The structural studies

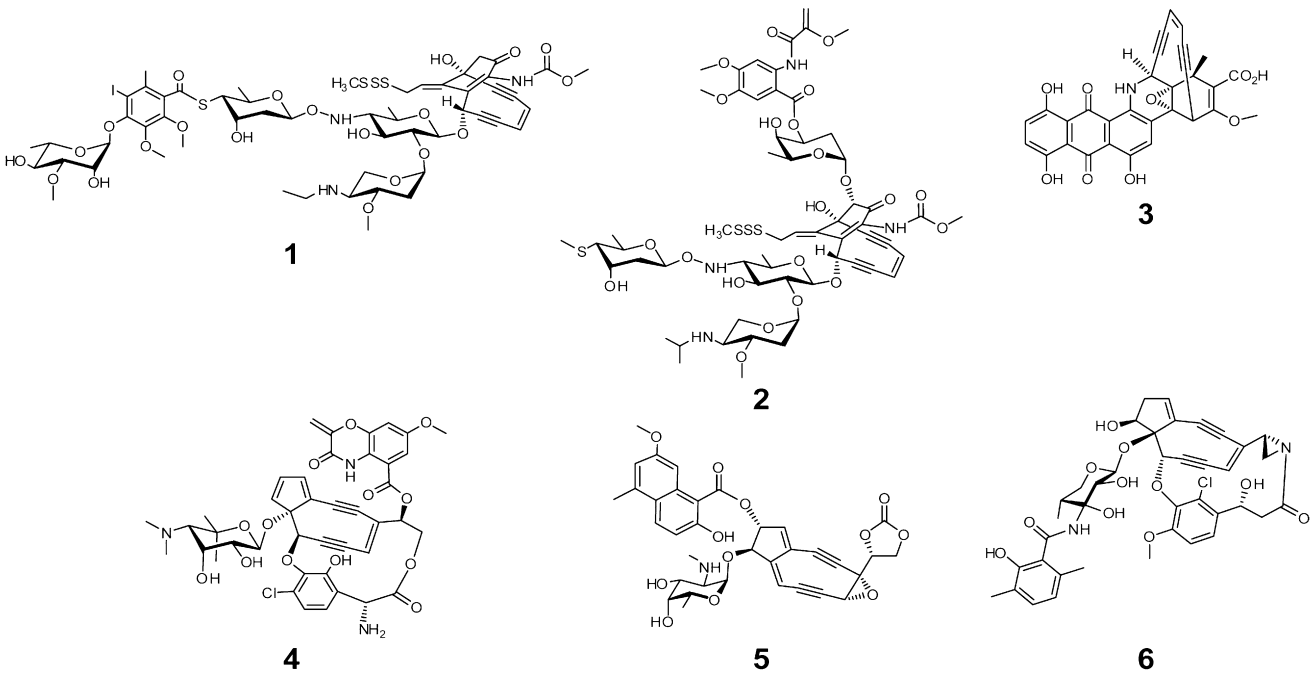



**Figure 1. Representative Naturally Occurring Enediynes**

Ten-membered enediynes calicheamcin $\gamma_1^I$ (**1**), esperamicin (**2**), and dynemicin (**3**); and nine-membered chromoprotein enediynes C-1027 (**4**), neocarzinostatin (**5**), and maduropeptin (**6**).

highlighted herein also revealed that enediyne GTs such as CalG3 adopt a GT-B fold common to natural product GTs. This structural study illuminated key catalytic residues and snapshots of a dynamic loop anticipated to participate in NDP binding. The application of a surrogate sugar nucleotide also enabled the confirmation of CalG2 as the remaining internal GT—the 4-deoxy-thio-α-D-digitoxosyltransferase—as the first characterized hydroxylamino glycosidic bond-forming GT. In conjunction with our previous report (Zhang et al., 2006b), this study completes the functional assignment of all four CLM GTs, extends the concept of reversibility of enediyne GT-catalyzed reactions, and highlights a crystal structure of an enediyne GT.

## RESULTS

### Overexpression and Purification of CalG3 and CalG2

Analysis of the CLM biosynthetic gene cluster from *M. echinospora* revealed four putative GT-encoding genes, *calG1*, *calG2*, *calG3*, and *calG4* (Ahlert et al., 2002), implicating a distinct GT for each sugar of the CLM aryltetrasaccharide. Consistent with this, in vitro biochemical characterization confirmed CalG1 and CalG4 as the CLM 3-*O*-methyl-rhamnosyltransferase and aminopentosyltransferase, respectively (Zhang et al., 2006b). Analysis of CalG3 revealed highest homology to characterized GTs, which operate on aromatic acceptors such as the nogalamycin SnogD (37% identity; Torkkell et al., 1997) and elloramycin ElmGT (36% identity; Blanco et al., 2001), whereas CalG2 more closely resembled GTs, which act on carbohydrate acceptors such as the CLM CalG4 (50% identity; Zhang et al., 2006b) and avermectin AveBI (42% identity; Wohlert et al., 2001; Zhang et al., 2006a). Based on this simple analysis, CalG3 was proposed as the putative calicheamicinone 4,6-dideoxy-4-hydroxylamino-α-D-glucosyltransferase and CalG2 postulated as the subsequent 4-deoxy-thio-α-D-digitoxosyltransferase (anticipated to form the signature CLM hydroxylamino glycosidic bond). To complete the CLM GT annotation, recombinant *N*-$His_{10}$-CalG2 and *N*-$His_{10}$-CalG3 fusion proteins were overproduced in *E. coli* and subsequently purified by nickel-affinity chromatography to >95% homogeneity (see Figure S1 available online) with overall yields of 10–15 mg/l culture.

### Preparation of the Putative Acceptor CLM $T_0$

The availability of an appropriate acceptor was critical to the in vitro characterization of CalG2 and CalG3. To address this issue, the truncated CLM analog CLM $T_0$ (Figure 2A, **9**) was prepared with a slight modification of a literature procedure (Walker et al., 1992). Specifically, CLM $\alpha_3^I$ (Figure 2A, **7**, and Figure 2B, i) was refluxed in wet acetone with pyridine *p*-toluenesulfonate and the reaction monitored via HPLC. After refluxing for 19 h, a product clearly distinct from starting material, but with the characteristic CLM core UV signature, was generated (Figure 2B, ii). The product was purified (Figure 2B, iii) and determined to have a mass of 625.1 [M + H] by APCI-MS analysis—40 daltons greater than expected product **9**. Subsequent $^1$H NMR analysis (Figure S2) revealed the new product to be **8** (Figure 2A), an isopropylidene adduct formed by the dehydration of acetone in the presence of tosylate. Reasoning that the adduct could be easily removed under mild acidic conditions, **8** was incubated with 0.2% TFA to provide **9** in 4 hr (Figure 2B, iv and v). The identity of **9** was verified by high-resolution ESI-MS ($[M + Na]^+$, *m/z* $C_{24}H_{28}N_2NaO_9S_3$, calculated 607.0855, found 607.0866) and $^1$H NMR (Figure S3).

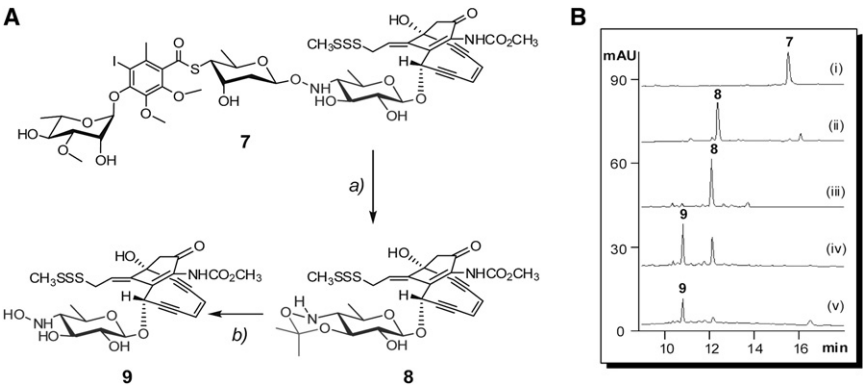



**Figure 2. Preparation of CLM $T_0$ (9) from CLM $\alpha_3^I$ (7)**

(A) Schematic of the strategy: (a) refluxing in acetone, 65°C, 19 hr; (b) incubated in 0.2% TFA, RT, 4 hr.

(B) HPLC analyses of the preparation: (i) starting material **7**; (ii) refluxed for 19 hr; (iii) purified **8** from the reaction mixture of (ii); (iv) **8** incubated with 0.2% TFA at RT for 2 hr; and (iv) **8** incubated with 0.2% TFA at RT for 4 hr.

## Reversibility of CalG3-Catalyzed Reaction

Following recently established protocols for GT reversibility (Minami et al., 2005; Zhang et al., 2006a, 2006b, 2007), CalG3 and CalG2 were first investigated for their ability to catalyze the excision of 4,6-dideoxy-4-hydroxylamino-α-D-glucosyl moiety from CLM $T_0$ (**9**) (Figure 3A). No reaction was observed with **9** (50 μM) and CalG2 (7.5 μM) in the presence of 2 mM NDP (ADP, CDP, GDP, UDP, or TDP) in Tris-HCl (10 mM [pH 7.6]). In contrast, the incubation of **9** (50 μM) with CalG3 (7.5 μM) in the presence of various NDPs (2 mM) led to new product after 2 hr at 30°C (Figure 3B, i–v). This transformation was determined to be both CalG3 and NDP dependent (Figure 3B, vi), and the new product was subsequently identified as the deglycosylated calicheamicinone (Figure 3A, **10**) by LC-MS (calculated 423, found 446.0 $[M + Na]^+$ and 422 $[M - H]^-$) (Table S1). Unlike prior reports on the reversibility of GT-catalyzed reactions that indicated such transformations to be NDP specific, CalG3 reversibility was observed with ADP (conversion rate of 40%; Figure 3B, i), GDP (23%; Figure 3B, ii), UDP (24%; Figure 3B, iii), TDP (trace; Figure 3B, iv), and even CDP (trace; Figure 3B, v). The apparent reversibility of this reaction was also enhanced at lower pH (Figure S4A). Cumulatively, these studies highlight the clear reversibility of the CalG3-catalyzed reaction and are consistent with CalG3 as the requisite 4,6-dideoxy-4-hydroxylamino-α-D-glucosyltransferase involved in **1** biosynthesis.

## CalG3-Catalyzed Sugar Exchange

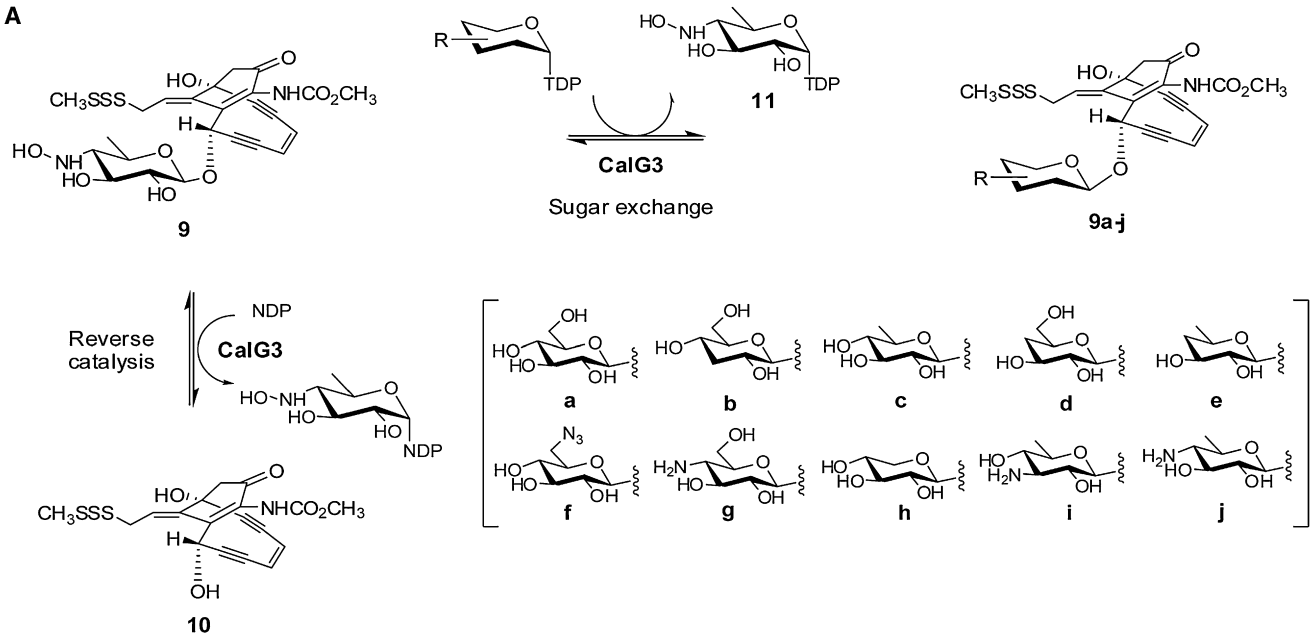
Given the lack of calicheamicinone availability (Figure 3A, **10**), the CalG3 sugar nucleotide specificity was alternatively probed with **9** as an acceptor for putative GT-catalyzed "sugar exchange" reactions. In a GT-catalyzed sugar exchange reaction, first observed in the context of CalG1 catalysis (Zhang et al., 2006b), the native sugar of a natural glycoside can be substituted in situ by unnatural sugars supplied as NDP sugar donors. Five commercially available NDP-glucoses, including ADP-, CDP-, GDP-, UDP-, and TDP-Glc were examined in the CalG3 sugar exchange reaction with **9**. Remarkably, all five NDP-glucoses were established as CalG3 donor substrates, albeit with varying sugar exchange efficiencies in the end point assay (50 μM **9**, 2 mM NDP-Glc, 7.5 μM CalG3, 30°C overnight). TDP glucose exhibited the highest conversion rate of **9**–**9a** (49.3%), followed by UDP-Glc (36.0%), CDP-Glc (25.3%), GDP-Glc (10.6%), and ADP-Glc (9.8%). In contrast to the influence of pH on reaction reversibility, the CalG3-catalyzed sugar exchange reaction was enhanced at higher pH (Figures S4B and S4C).

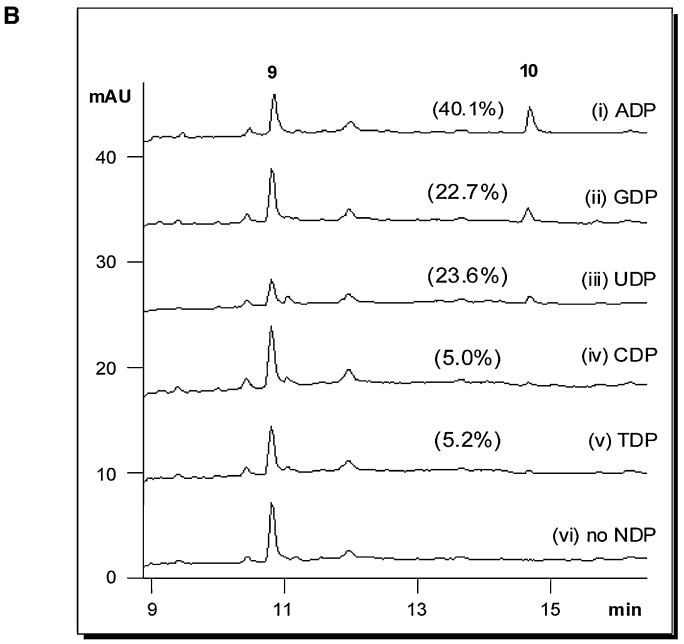
The sugar exchange promiscuity of CalG3 was subsequently probed directly with **9** and a small library of 22 TDP sugars (Figure S5) comprised of 20 TDP-D sugars (including commercially available TDP-α-D-glucose and unnatural sugar nucleotides generated via chemoenzymatic synthesis with functionality variations such as deoxy, amino, and azido at the sugar C2, C3, C4, C5, or C6 positions) and two TDP-L sugars (TDP-β-L-rhamnose and TDP-α-L-rhamnose). From this substrate specificity analysis, 10 sugar nucleotides were identified as CalG3 substrates (Figures 3A and 3C, i–x) to ultimately provide 10 unique CLM variants **9a–j** (Figure 3A and Table S1) with sugar exchange conversions ranging from 15% to 70% (Figure 3C). Cumulatively, these studies revealed CalG3 to be a relatively promiscuous GT and further highlighted GT-catalyzed sugar exchange as an expeditious method for natural product diversification.

## CalG3 Structure

The structure of CalG3 was determined by single wavelength anomalous diffraction using a 2.8 Å data set collected from Se-Met-labeled crystals (from I222 crystal form) and refined against this data set as well as a native data set to a resolution of 1.9 Å (from $P2_1$ crystal form). Data collection, phasing, and refinement statistics are summarized in Table 1. Consistent with gel filtration, asymmetric units of both crystal forms revealed two molecules of CalG3 arranged into dimers with $C_2$ symmetry (Figure 4A). Analysis of the binding interface of these dimers by the PISA server (Krissinel and Henrick, 2007) revealed the total buried surface area for the complex of 3300–3600 $Å^2$. The dimers of both crystal forms align with an all-atom rmsd of 0.76 Å, whereas all observed monomer conformers from both crystal forms align with rmsd ~0.48 Å. These results confirm that CalG3 forms dimers with a closely similar quaternary arrangement in both crystal forms, and this form of protein thus may be relevant in vivo. A detailed backbone comparison of the CalG3 molecules found in the two crystal forms revealed that several secondary structure elements undergo slight shifts—these include surface-exposed segments containing residues 54–76, 205–212, and 219–226. In addition, a tetraglycine loop spanning residues 285–288 (Figure 4B, pink) undergoes a conformational change.

The structure of the CalG3 monomer revealed that this protein consists of two closely opposed globular domains, the



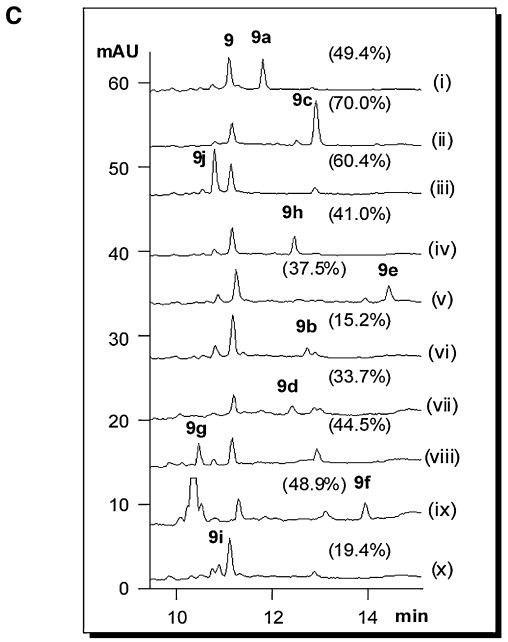




**Figure 3. CalG3-Catalyzed Reverse Reaction and Sugar Exchange Reaction**

(A) Schematic of the CalG3-catalyzed formation of **10** from **9** via reverse catalysis and the production of 10 unique CLM variants **9a–j** via sugar exchange.

(B) HPLC analyses of CalG3-catalyzed reverse reactions. In these reactions, 50 μM **9** was incubated with 7.5 μM CalG3 for 2 hr at 30°C in the presence of (i) 2 mM ADP; (ii) 2 mM GDP; (iii) 2 mM UDP; (iv) 2 mM CDP; (v) 2 mM TDP; and (vi) no NDP. Percent conversions were indicated in the parentheses.

(C) The production of **9a–j** via CalG3-catalyzed sugar exchange. A total of 50 μM **9** was incubated with 7.5 μM CalG3 at 30°C overnight in the presence of various TDP sugars (300 μM; Figure S4). Percent conversions were indicated in the parentheses.

N-terminal domain (residues 1–193; Figure 4B, cyan) and the C–terminal domain (residues 209–360; Figure 4B, khaki) connected with a linker (residues 194–208; Figure 4B, yellow) and stabilized by interaction of the C-terminal helix (residues 362–375; Figure 4B, green) and the N-terminal domain. Both the N- and C-terminal domain adopt a fold with Rossmann topology and a three-layer α-β-α sandwich architecture that shows homology to glycogen phoshorylase B. The structurally homologous cores of these domains can be aligned with an rmsd of 3.1 Å for 103 residues. However, the domains show dramatic variations in size and conformation of several loop regions as well as topology of their C-termini. Full-length CalG3 belongs to the GT-B clan of GTs and adopts a UDP-glycosyltransferase/glycogen phosphorylase fold found in many GTs. Based

**Table 1. Crystal Parameters, X-Ray Data Collection, Phasing, and Refinement Statistics**

| | SeMet | Native |
|---|---|---|
| Crystal parameters | | |
| Space group | I222 | $P2_1$ |
| Unit-cell parameters (Å, °) | a = 106.7, b = 119.3, c = 156.0 | a = 57.4, b = 97.7, c = 63.0 β = 90.6 |
| Data collection statistics | | |
| Wavelength (Å) | 0.97918 | 0.97918 |
| Energy (eV) | 12,662 | 12,662 |
| Resolution range (Å) | 26.75–2.80 (2.90–2.80) | 38.76–1.68 (1.74–1.68) |
| No. of reflections (measured/unique)[a] | 180,890/24,883 | 482,455/68,513 |
| Completeness (%) | 99.2 (94.9) | 86.3 (43.6) |
| $R_{merge}$[b] | 0.121 (0.426) | 0.051 (0.496) |
| Redundancy | 7.3 (6.3) | 7.0 (3.9) |
| Mean I/sigma(I) | 9.8 (4.0) | 25.1 (2.4) |
| Phasing statistics[c] | | |
| Mean FOM (centric/acentric) | 0.344/0.090 | |
| Phasing power (isomorphous/anomalous) | 0.0/1.17 | |
| Cullis R-factor (isomorphous/anomalous) | 0.0/0.79 | |
| Refinement and model statistics | | |
| Resolution range | 26.75–2.79 (2.86–2.79) | 38.76–1.90 (1.95–1.90)[d] |
| No. of reflections (work/test) | 23,613/1,270 | 50,954/2,745 |
| $R_{cryst}$[e] | 0.187 (0.311) | 0.160 (0.186) |
| $R_{free}$[f] | 0.243 (0.378) | 0.212 (0.275) |
| Rmsd bonds (Å) | 0.011 | 0.014 |
| Rmsd angles (°) | 1.336 | 1.417 |
| ESU from $R_{free}$ (Å)[g] | 0.342 | 0.146 |
| B factor, Wilson plot (Å$^2$) | 50.3 | 24.3 |
| B factor, monomer A/B/waters (Å$^2$)[h] | 31.4/33.0/23.3 | 31.4/30.2/36.4 |
| No. of protein molecules/all atoms | 2/5,783 | 2/6,378 |
| No. of waters | 45 | 579 |
| No. of auxiliary molecules | 2 MOPS | 1 PEG |
| Ramachandran plot by MolProbity (%) | | |
| Favored regions | 97.5 | 98.3 |
| Additional allowed regions | 2.5 | 1.6 |
| Outliers | 0.0 | 0.1 |
| PBD code | 3DOQ | 3DOR |

[a] Values in parentheses are for the highest-resolution shell.
[b] $R_{merge} = \sum_h \sum_i |I_i(h) - < I(h) > | / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $< I(h) >$ is the mean intensity of the reflection.
[c] Resolution range for phasing in SHARP was (26.66–3.2) Å.
[d] Resolution range for refinement was cut (38.76–1.90) Å due to low completeness and signal in the remaining resolution shells.
[e] $R_{cryst} = \sum_h ||F_{obs}| - |F_{calc}|| / \sum_h |F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure-factor amplitudes, respectively.
[f] $R_{free}$ was calculated as $R_{cryst}$ using ~5.0% of the randomly selected unique reflections that were omitted from structure refinement.
[g] Estimated standard uncertainty based on $R_{free}$.
[h] B-factors from the model refined without TLS.

on the VAST server (Madej et al., 1995), the closest overall structural homologs of CalG3 include *Streptomyces fradiae* TDP-D-olivose-transferase UrdGT2 (which structurally aligns with CalG3 for 369 residues with 24% identity and rmsd of 3.1 Å, PDB ID 2p6p) (Mittler et al., 2007) and *Amycolatopsis orientalis* TDP-*epi*-vancosaminyltransferase GtfA (363 residues with 20% identity and rmsd of 4.0 Å, PDB ID 1pnv, 1pn3, 1rrv; Mulichak et al., 2003).

## In Vitro Characterization of CalG2

Given the lack of availability of the natural substrate NDP-4-deoxy-4-thio-α-D-digitoxose, the biochemical function of CalG2 was examined with a small library of 22 surrogate TDP sugar substrates (Figure S5). Of this set, only one, TDP-4,6-dideoxy-α-D-glucose (Figure 5, **12**), led to the CalG2-catalyzed transformation of **9** to a new product with a mass (calculated 714, found 737 $[M + Na]^+$) (Table S1) consistent with **13** (Figure 5, i). Notably,

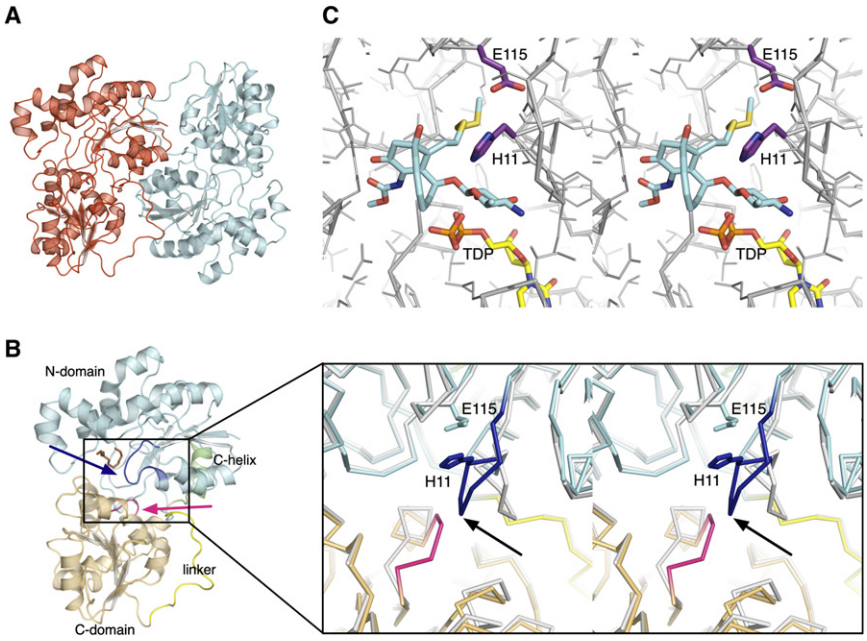



**Figure 4. Structure of CalG3**

(A) A ribbon diagram of the CalG3 dimer with monomers color-coded in red and cyan.

(B) The CalG3 monomer is formed by closely opposed N-terminal (cyan) and the C-terminal domains (khaki). These distinct domains are connected by a linker (yellow) and their interaction is stabilized by the C-terminal helix (green). The blue arrow indicates the putative catalytic loop; the magenta arrow points to a pyrophosphate-binding tetraglycine loop spanning residues 285–288. An ordered portion of a polyethylene glycol molecule (brown) has been found in the cavity formed by the N-domain. The inset highlights Cα-trace of CalG3 (cyan, yellow, khaki) in the active site with putative catalytic diad residues H11 and E115 highlighted. The gray Cα-trace is that of a docked model, which incorporates experimentally observed conformational changes in the pyrophosphate-binding loop (magenta) and modeled changes of the catalytic loop (blue, black arrow).

(C) Manually docked model of the CalG3 with CLM $T_0$ (carbon, cyan; oxygen, red; sulfur, yellow; nitrogen, blue) and a dinucleotide TDP (carbon, yellow; oxygen, red; nitrogen, blue; phosphorus, orange) in the active site.

this new product was clearly distinct from the CalG3-catalyzed sugar exchange product **9e** (Figure 5, iii), derived from the same set of starting materials **9** and **12** (Figure 5), and its production was enzyme dependent (Figure 5, ii). The incubation of **8** with **12** and CalG2 under identical conditions led to no change, consistent with the **8** isopropylidene as masking the CalG2 acceptor nucleophile. Cumulatively, these in vitro studies confirmed that CalG2 was capable of adding a sugar to **9**, consistent with CalG2 as the requisite CLM 4-deoxy-4-thio-α-D-digitoxosyltransferase.

## DISCUSSION

### CLM GTs and CLM Biosynthesis

Calicheamicin (**1**) has two distinct structural regions: the enediyne aglycon, or warhead, which consists of a highly functionalized bicyclo[7.3.1.]tridecadiyne core structure with an allylic trisulfide serving as the initial trigger for warhead cycloaromatization, and the aryltetrasaccharide, which is composed of a set of unusual carbohydrate and aromatic units and docks the metabolite specifically into the minor groove of DNA (Kumar et al., 1997; Walker et al., 1994). Two possible scenarios for the final stages of CLM biosynthesis have been put forth: sequential GT-catalyzed glycosylation of the calicheamicinone core (as highlighted in Figure 6) or coupling an intact aryltetrasaccharide unit to the calicheamicinone core (Rothstein and Love, 1991; Thorson et al., 1999). The first of these was based on the conventional routes to secondary metabolite glycosylation; the second was derived from the structural similarities between the CLM aryltetrasaccharide and the orthosomycins avilamycin or evernimicin (Hosted et al., 2001; Weitnauer et al., 2001). Consistent with either putative NDP-sugar or aryltetrasaccharide intermediates, random chemical mutagenesis of the CLM-producing strain *M. echinospora* led to undefined water-soluble intermediates that could be transformed to CLM by other blocked mutant strains (Rothstein and Love, 1991). Characterization of the CLM biosynthetic gene cluster (Ahlert et al., 2002) revealed four putative GT-encoding genes (*calG1*, *calG2*, *calG3*, and *calG4*) that enabled the initial biochemical characterization of the CLM GTs CalG1 and CalG4 (Zhang et al., 2006b), providing support for the pathway highlighted in Figure 6. Specifically, this prior work revealed CalG1 and CalG4 as the 3-*O*-rhamonsyltransferase and aminopentosyltransferase, respectively, the order of which appeared indiscriminate. The current study extends this work by confirming CalG3 and CalG2 as the internal sequential 4,6-dideoxy-4-hydroxylamino-α-D-glucosyltransferase and 4-deoxy-thio-α-D-digitoxosyltransferase, respectively. Thus, this study completes the functional assignment of the four CLM GTs and provides further support for a sequential glycosylation pathway in CLM biosynthesis.

### CalG3 Reaction Reversibility and Sugar Exchange

The reversibility of GT-catalyzed reactions has enabled GT biochemical characterization, the syntheses of exotic sugar nucleotides, and the differential glycosylation of various complex natural products (Zhang et al., 2006a, 2006b, 2007). As described in the present study, a similar strategy facilitated the functional assignment of CalG3 as the requisite CLM 4,6-dideoxy-4-hydroxylamino-α-D-glucosyltransferase and allowed for the generation of a small set of novel CLM analogs via CalG3-catalyzed sugar exchange. Though this study revealed CalG3 to be among the growing list of inherently promiscuous GTs (Salas and Mendez, 2007; Thibodeaux et al., 2007), a remarkable distinction of the CalG3 reaction from previously studied examples (Minami et al., 2005; Zhang et al., 2006a, 2006b, 2007) is the apparent reversibility with nonnative NDPs. Specifically, biochemical characterization of the CLM sugar nucleotide pathways reveals all four aryltetrasaccharide sugars to derive

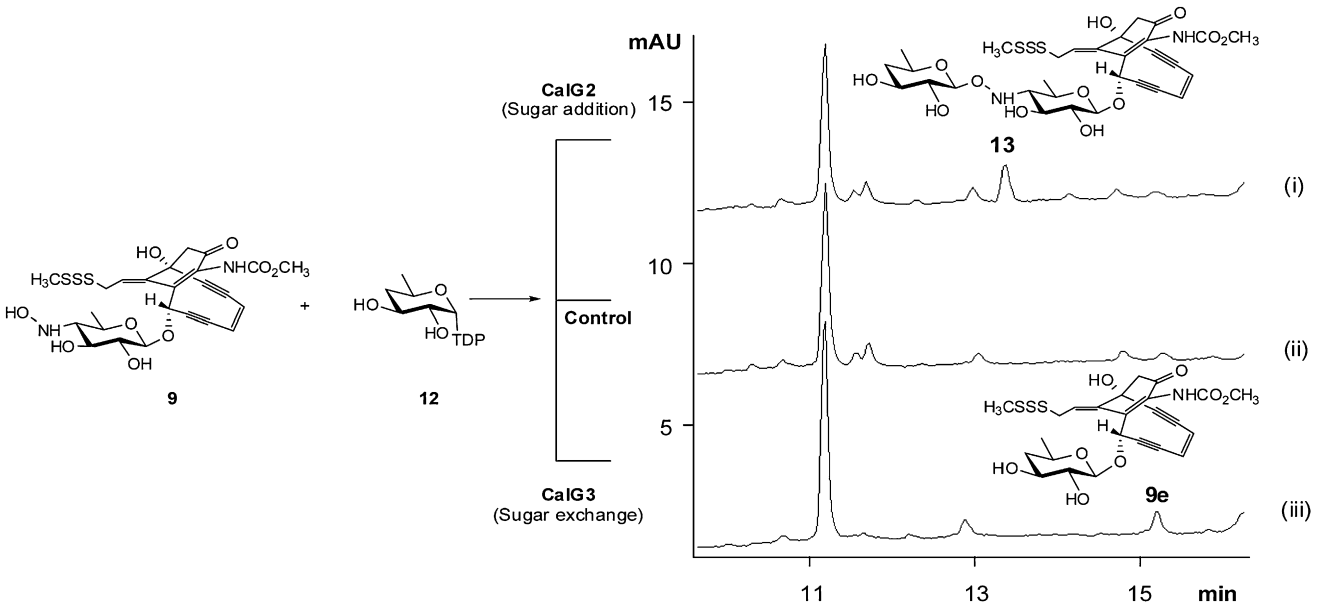



**Figure 5. Differential Reactions with 9 and 12 in the Presence of CalG2 and CalG3**

(i) 50 μM **9**, 300 μM **12** in the presence of 7.5 μM CalG2 at 30°C overnight; (ii) 50 μM **9**, 300 μM **12** in the absence of enzymes at 30°C overnight; (iii) 50 μM **9**, 300 μM **12** in the presence of 7.5 μM CalG3 at 30°C overnight.

from TDP/UDP-sugar precursors (Bililign et al., 2002; Johnson and Thorson, unpublished data), whereas the present study reveals purine-based nucleotides to be optimal substrates in the reverse direction. Although this highlights the use of caution when employing reversibility as a means to determine GT substrate specificity, the nucleotide specificity of CalG3 sugar exchange reactions with the NDP-Glc series is consistent with a TDP/UDP sugar-dependent process. Simulation of GT catalysis has revealed the equilibrium constant ($K_{eq}$) to be the single most critical factor governing reaction efficiency (Melancon et al., 2006). The observed modulation of the CalG3 equilibrium via nucleotides (Figure 3 and, specifically, Figure 3B) and/or pH (Figure S4) is also consistent with a thermodynamically controlled process.

## CalG3 Structure

Though the sequence homology of GalG3 and its closest structural homologs is low (<25% identity), the modular design and the structural similarity between GT-B glycosyltransferases implicate the C-terminal domain as involved in sugar nucleotide binding. Specifically, the tetraglycine loop 285–288 (Figure 4B, inset), which adopted different conformations within the distinct CalG3 crystal forms, most likely interacts with the sugar nucleotide pyrophosphate while an adjacent large cavity within the C-terminal domain of CalG3 is anticipated to accommodate the corresponding nucleoside. The N-terminal domain also contains a large cavity, which in our high-resolution model binds an extended chemical entity (likely an ordered fragment of polyethylene glycol used during crystallization; Figure 4B, brown). Based on structural homology of CalG3 to other natural product GTs, this cavity likely binds calicheamicinone.

Using this information as a guide, the products of the reaction, CLM $T_0$ and TDP, were manually docked into the CalG3 model (Figure 4C). Specifically, the dinucleotide positioning was initially guided by the high structural conservation of C-terminal domains of CalG3 and oleandomycin glycosytransferase OleI in complex with UDP (PDB ID 2iya; Bolam et al., 2007). Subsequent CLM $T_0$ orientation in the corresponding N-terminal cavity was defined by the surface of the cavity and the geometric constraints of a putative $S_N2$ reaction between the acceptor hydroxyl group and the sugar nucleotide anomeric carbon. Based on this model, the acceptor hydroxyl group is located near a putative catalytic diad (His11 and Glu115). The first residue of the diad, His11, is highly conserved as His or Asp among GT homologs (based on a multiple sequence alignment of 1070 sequences from NR85S database by FFA03) (Jaroszewski et al., 2005) and a structure-based multiple sequence alignment (Figure S6). Consistent with this, mutation of the His11 equivalent in the oleadromycin GT OleI (His25 to Ala) led to complete loss of catalytic activity (Bolam et al., 2007). Moreover, the CalG3 His11-containing loop (N1 loop, residues 5–12; Figure 4B, inset, black arrow) has to undergo conformational changes during the substrate binding and catalysis. In the CalG3 model, this dynamic loop resides between the calicheamicinone-binding cavity and an internal cavity filled with a cluster of well-ordered water molecules, the removal of which allows for an expansion of the calicheamicinone-binding cavity to accommodate the requisite hexose.

The second member of the CalG3 putative catalytic diad (Glu115), though less conserved in general, is most often found to be Asp or Glu in GT sequence homologs (Figure S6). The CalG3-substrate complex model also predicts the access to the calicheamicinone hydroxyl nucleophile to be sterically hindered, and thus, activation of this hydroxyl may require a water-mediated process. Similar water-containing catalytic triads have been described in variety of enzymes, including phospholipases (Scott et al., 1990). In some flavonoid GTs, a neighboring

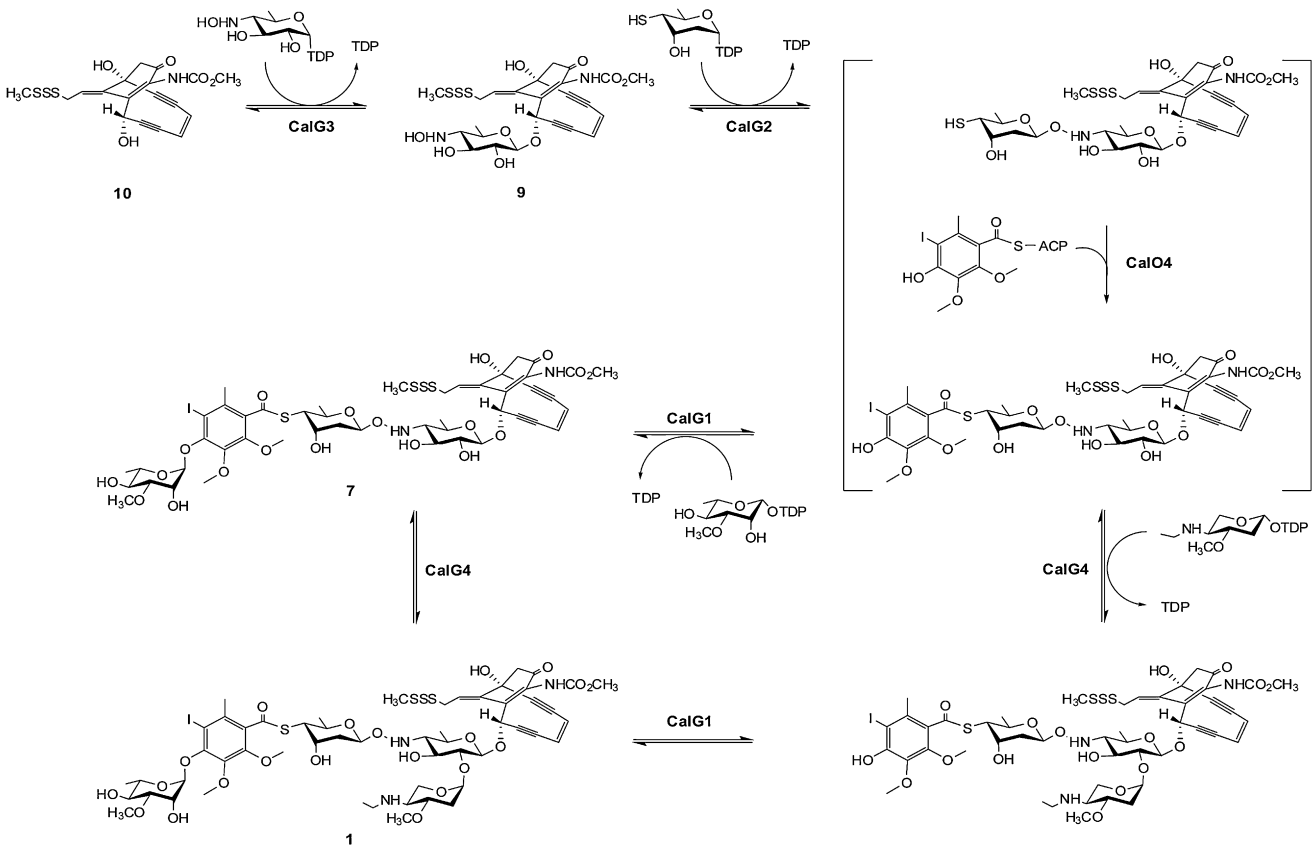



**Figure 6. Proposed CLM Glycosylation Pathway**

serine within the same loop has been implicated (along with the His/Asp-Asp/Glu diad) as part of a catalytic triad (Offen et al., 2006). CalG3 also contains a Ser within the N1 loop but in a different position, the function of which remains to be elucidated. In addition, consistent with other GTs, the CalG3-substrate model is consistent with the conserved Gln311-sugar, His284-pyrophosphate, and Trp268-base stacking components of the sugar nucleotide (Figure S6). In comparison with other UDP/TDP sugar nucleotide-utilizing GTs, CalG3 Leu271 is topologically equivalent to the TDP sugar-utilizing GT GtfA Leu280, an amino acid noted as a potential selectivity filter for 2′-deoxy nucleotides. In contrast, the TDP sugar-utilizing GT OleI has a Gln315 at the same position that hydrogen bonds the nucleotide 2′-OH. Finally, this model also implicates the cavity formed by the residues Leu14, Pro15, Gln137, and Arg135 may accommodate the CLM trisulfide-SSSMe trigger.

## SIGNIFICANCE

**In conjunction with our previous report (Zhang et al., 2006b), this study completes the functional assignment of all four CLM GTs, extends the concept of reversibility of enediyne GT-catalyzed reactions, and highlights a unique crystal structure of an enediyne GT. From a structural biology perspective, this work reveals the common GT-B structural fold can now be extended to include enediyne GTs. Given the notable architectural distinctions of enediynes from other natural products for which GT structures have been elucidated (including glycopeptides, aromatic polyketides, and macrolides), this work adds to the structural blueprints for engineering and/or evolving novel glycosylation catalysts (Williams et al., 2007, 2008; Williams and Thorson, 2008). From a biosynthetic perspective, this work also completes the functional annotation of the four calicheamicin GTs, highlights the first characterization of a hydroxylamino glycosidic bond-forming GT, and presents conclusive evidence of a sequential glycosylation pathway in CLM biosynthesis. Furthermore, this work highlights the utility of the reversibility of GT-catalyzed reactions and sugar nucleotide surrogate substrates for both the elucidation of enzyme function and the diversification of therapeutically important natural products.**

## EXPERIMENTAL PROCEDURES

### Materials

*E. coli* DH5α and BL21(DE3) competent cells were purchased from Invitrogen (Carlsbad, CA). The pET-16b *E. coli* expression vector was purchased from Novagen (Madison, WI). Primers were purchased from Intergrated DNA Technology (Coralville, IA). *Pfu* DNA polymerase was purchased from Stratagene (La Jolla, CA). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). All other chemicals were reagent grade or better and purchased from Sigma (St. Louis, MO). Calicheamicin $\alpha_3^I$

(Figure 2, **7**) was provided by Wyeth Research (Pearl River, NY). Analytical HPLC was run on a Varian Prostar 210/216 system connected to a Prostar 330 photodiode array detector (Varian; Walnut Creek, CA). Mass spectra (MS) were obtained by using electrospray ionization on an Agilent 1100 HPLC-MSD SL quadrupole mass spectrometer (Agilent Technologies; Palo Alto, CA) connected with a UV/Vis diode array detector.

### Chemoenzymatic Synthesis of TDP Sugars

The set of TDP sugars employed were generated chemoenzymatically as previously described (Barton et al., 2002; Jiang et al., 2000, 2001). Specifically, the RmlA ($E_p$, glucose-1-phosphate thymidylyltransferase) reaction was carried out in Tris-HCl buffer (50 mM [pH 8.0]) containing 5 mM $MgCl_2$, 1 U inorganic pyrophosphatase, 10 μM of purified $E_p$, 8 mM sugar-1-phosphate and 6 mM TTP, and incubated at 37°C for 2 hr. The formation of sugar nucleotides (Figure 2, **5**–**24**) was analyzed by HPLC using a reverse-phase column Luna C18, 5 μm, 250 × 4.6 mm with UV detection at 254 nm. A phosphate buffer A (30 mM potassium phosphate [pH 6.0], 5 mM tetrabutylammonium hydrogensulfate, 2% acetonitrile) was used as mobile phase, and the HPLC was run with a 0%–50% gradient of buffer B (acetonitrile) over 30 min.

### Preparation of CLM $T_0$ (9) from Calicheamicin $\alpha_3^I$ (7)

CLM $\alpha_3^I$ (**7**, 20 mg, 0.017 mM) was dissolved in wet acetone (20 ml) followed by the addition of pyridine *p*-toluenesulfonate (0.4 mg, 0.002 mM). The reaction was heated to reflux (65°C) and monitored by TLC (9:1 $CHCl_3$:MeOH) and reverse-phase HPLC (Phenomenex [Torrance, CA] Luna C18, 4.6 × 250 mm column; Figure 2). After 19 hr, full conversion of **7** to a product with a UV spectrum characteristic to the CLM core structure had occurred. Purification was performed by silica gel column chromatography (9:1 $CHCl_3$:MeOH) to yield the isopropylidenylated analog **8** as a light brown oil (10 mg, 97%, TLC $R_f$ = 0.35, 9:1 $CHCl_3$:MeOH), which was characterized by $^1$H NMR on a Varian $^{UNITY}$Inova 400 MHz instrument (Figure S3) and by APCI-MS using an Agilent 1100 Series LC/MSD. Compound **8** (10 mg, 0.016 mM) was hydrolyzed by dissolving in 1 ml of a 1:1 MeOH:$H_2O$ solution in the presence of TFA (12 μl) and agitating for 4 hr using TLC and reverse-phase HPLC to observe the reaction (Figure 2). The solvent was removed in vacuo and the crude oil purified by preparatory reverse-phase HPLC (Supelco Discovery BIO [St. Louis, MO], 10 × 250 mm, 5 μm column) using a gradient of 10%–100% acetonitrile in water over 20 min at a rate of 10 ml/min with UV detection at 280 nm. Lyophilization yielded **9** as a white powder (5.8 mg, 62%, TLC $R_f$ = 0.32, 9:1 $CHCl_3$:MeOH). Characterization was performed on a Varian $^{UNITY}$Inova 500 MHz NMR with a capillary probe (Figure S2) and by high-resolution MS using a Waters LCT time-of-flight mass spectrometer. Compound **8**: $^1$H NMR ($d_6$-acetone, 400 MHz) δ 7.86 (s, 1 H), 6.46 (dd, *J* = 10.3, 5.0 Hz, 1 H), 6.15 (d, *J* = 1.5 Hz, 1 H), 6.04 (d, *J* = 9.4 Hz, 1 H), 5.98 (dd, *J* = 9.4, 1.5 Hz, 1 H), 4.75 (d, *J* = 7.9 Hz, 1 H), 4.23 (t, *J* = 9.1 Hz, 1 H), 4.19–4.13 (m, 2 H), 4.03 (t, *J* = 9.2 Hz, 1 H), 3.94 (dd, *J* = 14.9, 5.0 Hz, 1 H), 3.65 (s, 3 H), 3.48 (t, *J* = 8.4 Hz, 1 H), 3.05 (d, *J* = 17.2 Hz, 1 H), 2.73 (d, *J* = 17.2 Hz, 1 H), 2.54 (s, 3 H), 2.28 (s, 3 H), 2.13 (s, 3 H), 1.18 (d, *J* = 6.2 Hz, 3 H); MS (APCI) *m/z* $C_{27}H_{33}N_2O_9S_3$ ([M + H]$^+$) 625.1, calculated 625.1. Compound **9**: $^1$H NMR ($d_6$-acetone, 500 MHz) δ 6.46 (dd, *J* = 10.3, 4.9 Hz, 1 H), 6.14 (s, 1 H), 6.03 (d, *J* = 9.3 Hz, 1 H), 5.97 (d, *J* = 9.3 Hz, 1 H), 4.72 (d, *J* = 7.7 Hz, 1 H), 4.23 (t, *J* = 9.2 Hz, 1 H), 4.18–4.11 (m, 2 H), 4.00 (t, *J* = 9.3 Hz, 1 H), 3.93 (dd, *J* = 14.9, 4.9 Hz, 1 H), 3.65 (s, 3 H), 3.49 (dd, *J* = 9.2, 7.7 Hz, 1 H), 3.05 (d, *J* = 17.0 Hz, 1 H), 2.73 (d, *J* = 17.0 Hz, 1 H), 2.53 (s, 3 H), 1.17 (d, *J* = 6.0 Hz, 3 H); HRMS (ESI) *m/z* $C_{24}H_{28}N_2NaO_9S_3$ ([M + Na]$^+$) 607.0866, calculated 607.0855.

### Cloning, Expression, and Purification of GTs

The *calG2* and *calG3* genes from the calicheamicin producer *Micromonospora echinospora* LL6600 were amplified from genomic DNA by using primer pairs: 5′-cacggacggagtcgcatatggcccacctc-3′ (forward, NdeI) and 5′-gccggtggatccgcggggcg-3′ (reverse, BamHI) for *calG2*; 5′-gaagggctccca tatgcgcgtgctgttc-3′ (forward, NdeI) and 5′-gggcgacgagatctgctcaacccga gatg-3′ (reverse, BglII) for *calG3,* using *Pfu* DNA polymerase. PCR products were digested with NdeI/BamHI (*calG2*) or NdeI/BglII (*calG3*) and ligated into the pET16b expression vector (NdeI/BamHI, to generate the *N*-terminal MGHHHHHHHHHH fusion) to give plasmids pCAM3.2 (CalG2) and pCAM11.2 (CalG3), respectively.

For CalG3 expression, a single transformant of *E. coli* BL21(DE3)/pCAM11.2 was inoculated into 4 ml LB medium supplemented with 100 μg/ml of ampicillin and grown at 37°C overnight. The starter cultures were inoculated into 1 liter of LB medium with 100 μg/ml of ampicillin and initially grown at 28°C to an $OD_{600}$ value of 0.5–0.7. Expression was induced with the addition of 0.4 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) followed by continued growth with shaking for 16 hr. The cells obtained from 1 liter of culture were washed twice with buffer A (20 mM $NaH_2PO_4$ [pH 7.5], 500 mM NaCl, 10 mM imidazole) and resuspended in 30 ml of buffer A supplemented with 1 mg/ml of lysozyme. After 10 min incubation on ice, the cells were lysed by three rounds of French press (1200 psi, Thermo IEC, Milford, MA) and the insoluble material was removed by centrifugation at 30,000 g for 1 hr (4°C). The supernatant was loaded onto the HisTrap HT column (1 ml, GE Healthcare, Piscataway, NJ) and the N-$(His)_{10}$-tagged CalG3 was eluted with a linear gradient of imidazole (10–500 mM) in buffer A using a FPLC-AKTA system (GE Healthcare). The purified protein was desalted through PD-10 column (GE Healthcare) and stored in the buffer containing 10 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 10% glycerol until use. Protein concentration was determined by Bradford assay. *N*-$(His)_{10}$-tagged CalG2 was expressed and purified following the same protocol from *E. coli* overexpression strain BL21(DE3)/pCAM3.2.

### CalG3 Assays

Generally, CalG3 assays were performed in a total volume of 100 μl containing 50 μM of CLM $T_0$ (**9**) and 2 mM NDP (Figure 3B) or 300 μM of TDP sugars (Figure S4) with incubation at 30°C for 2 hr (reverse reaction) or overnight (sugar exchange) in the presence of 7.5 μM of CalG3, in Tris-HCl buffer (10 mM [pH 7.6]) containing 1 mM of $MgCl_2$. The assay mixtures lacking CalG3 served as controls. The reactions were subsequently quenched by the addition of 100 μl methanol and centrifuged to remove proteins. The reactions were monitored by HPLC (Phenomenex Luna C18, 5 μm, 250 × 4.6 mm; 10%–100% $CH_3CN$ over 20 min, 1 ml/min, 280 nm). The conversion rate was calculated by dividing the integrated area of glycosylated product with the sum of integrated area of product and the remaining substrate. All newly formed products were also analyzed by LC-MS (ESI) with both positive (+) and negative (−) modes.

To assess the pH range for CalG3 catalysis, potassium phosphate buffers (50 mM [pH 6.0, pH 7.0, pH 7.6, and pH 8.0]) were used (**9** was instable > pH 8.0). The reaction mixtures contained 50 μM **9**, 2 mM NDP (or 2 mM NDP glucose), and 7.5 μM CalG3 and were incubated at 30°C for 2 hr (reverse reaction) or overnight (sugar exchange).

### CalG2 Assays

Generally, CalG2 assays were performed in a total volume of 100 μl containing 50 μM of CLM $T_0$ (**9**) and 2 mM of NDP glucoses (Figure 3B) or 300 μM of TDP sugars (Figure S4) with incubation at 30°C overnight in the presence of 7.5 μM CalG2, in Tris-HCl buffer (10 mM [pH 7.6]) containing 1 mM of $MgCl_2$. The assay mixtures without addition of CalG2 served as controls. The reactions were quenched by the addition of 100 μl methanol and centrifuged to remove proteins. The formation of new CLM products was monitored by HPLC analysis as described above for CalG3 assays.

### CalG3 Crystallization

Crystals of selenomethionine-labeled (SeMet) CalG3 were grown at 277 K by the hanging drop method from a 10 mg/ml protein solution in a protein buffer (50 mM NaCl, 10 mM TRIS [pH 7.5]) mixed with an equal amount of the well solution (16% w/v polyethylene glycol 4000, 200 mM triammonium citrate, 100 mM MOPS [pH 7.0]). Rod-shaped crystals with dimensions up to 400 × 20 × 20 μm grew fused in parallel clusters, only occasionally as usable single needles. Crystals were cryoprotected at 277 K by soaking in the well solution containing 0%, 10%, and 20% (v/v) ethylene glycol and were flash frozen in a stream of cryogenic nitrogen gas at 100 K. The native crystals of CalG3 that lead to a high-resolution data set were obtained by sitting drop method from a 10 mg/ml protein solution in the protein buffer mixed with an equal amount of the well solution (25% w/v polyethylene glycol 1500; condition D1 of Hampton IndexHT screen). Crystals were discovered 8 months after the initial setup and were never again successfully reproduced. Crystals were cryoprotected at 277 K in Fomblin MW2500 (Aldrich) and were flash frozen in a stream of cryogenic nitrogen gas at 100 K. The diffraction quality of crystals

was evaluated using a laboratory X-ray diffraction instrument equipped with a Bruker AXS Proteum R CCD detector and a Microstar rotating anode generator using copper K$\alpha$ radiation (Bruker; Madison, WI). All attempts to prepare crystals of complexes of CalG3 with its substrates, via cocrystallization with ligands or soaking crystals with ligands, were unsuccessful.

### CalG3 Structure Determination

X-ray diffraction data for both SeMet-labeled and native CalG3 were collected at the General Medicine and Cancer Institute Collaborative Access Team (GM/CA-CAT) 23-ID-D beamline at the Advanced Photon Source at Argonne National Laboratory. Each of the 290 diffraction images for the native crystal was collected at the crystal-to-detector distance of 220 mm and exposed for 2 s with 100-fold attenuation of the incident beam. The data was collected in a single pass with 1.25° oscillation per frame. Each of the 180 diffraction images for the selenomethionine-labeled crystal was collected at wavelength of 0.97918 Å at the crystal-to-detector distance of 300 mm and exposed for 6 s with 100-fold attenuation of the incident beam. The data was collected in a single pass with 1° oscillation per frame. The diffraction images were integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). The native crystals belong to the space group $P2_1$ with unit cell parameters a = 57.4 Å, b = 97.7 Å, c = 63.0 Å, $\beta$ = 90.6°. The selenomethionine-labeled crystals belong to the space group I222 with unit cell parameters a = 106.7 Å, b = 119.3 Å, c = 155.9 Å.

The selenomethionine substructure of the SeMet-labeled crystals of CalG3 was determined using HySS (Adams et al., 2002; Uson and Sheldrick, 1999). These programs identified 13 consensus anomalous sites. The structure was automatically phased using autoSHARP (de La Fortelle and Bricogne, 1997) with the help of auxiliary programs from the CCP4 (Collaborative Computational Project Number 4, 1994) suite. The initial phase information was significantly improved by twofold averaging during the density modification as implemented in CNS (Brunger et al., 1998). Resulting map at 2.8 Å resolution was high quality and allowed for partial building of the model in ARP/wARP (Perrakis et al., 1999). This model was improved by manual building in COOT (Emsley and Cowtan, 2004) and used as a search model in molecular replacement trials in MOLREP (Vagin and Teplyakov, 1997) against the 1.9 Å native diffraction data. The high-resolution model of CalG3 was next built in ARP/wARP using the phase information derived from the successfully placed low-resolution model. The automatically built model contained 614 residues, of which 591 had side chains assigned and R = 25.3% (Rfree = 31.0%). The structure was completed in multiple cycles of manual building in COOT and refinement in REFMAC5 (Murshudov et al., 1997). Final refinement protocol included TLS refinement with five TLS groups per monomer based on the TLSMD server (Painter and Merritt, 2006) analysis. The final refined model has R = 16.0% (Rfree = 21.2%). In addition to residues −2 to 375 of monomer A and residues −6 to 374 of monomer B, the final model contains 579 waters and a molecule of polyethylene glycol. Finally, we used the final high-resolution model of CalG3 (in $P2_1$) as a starting point for refinement of the lower-resolution model of the SeMet-labeled CalG3 (in I222).

## ACCESSION NUMBERS

The CalG3 1.9 Å native and 2.8 Å Se-Met structures have been deposited in the Protein Data Bank under accession numbers 3DOR and 3DOQ, respectively.

## SUPPLEMENTAL DATA

Supplemental Data include six figures, one table, and Supplemental References, and can be found with this article online at http://www.chembiol.com/cgi/content/full/15/8/842/DC1/.

## ACKNOWLEDGMENTS

We thank the University of Wisconsin–Madison School of Pharmacy Analytical Facility for analytical support, Philip R. Hamann (Wyeth Research) for providing CLM $\alpha_3^I$, and Gavin J. Williams for helpful discussion. B.R.G. is a postdoctoral fellow of the American Cancer Society (PF-05-016-01-CDD) and J.S.T is a University of Wisconsin H.I. Romnes Fellow. This research was supported in part by National Institutes of Health Grant CA84374 and National Cooperative Drug Discovery Group Grant U19 CA113297 from the National Cancer Institute. 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). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-ENG-38. The authors report competing interests. J.S.T. is a cofounder of Centrose (Madison, WI).

Received: May 6, 2008
Revised: June 12, 2008
Accepted: June 20, 2008
Published: August 22, 2008

## REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B.O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R.E., et al. (2002). The calicheamicin gene cluster and its iterative type I enediyne PKS. Science *297*, 1173–1176.

Barton, W.A., Biggins, J.B., Jiang, J., Thorson, J.S., and Nikolov, D.B. (2002). Expanding pyrimidine diphosphosugar libraries via structure-based nucleotidylyltransferase engineering. Proc. Natl. Acad. Sci. U.S.A. *99*, 13397–13402.

Biggins, J.B., Onwueme, K.C., and Thorson, J.S. (2003). Resistance to enediyne antitumor antibiotics by CalC self-sacrifice. Science *301*, 1537–1541.

Bililign, T., Shepard, E.M., Ahlert, J., and Thorson, J.S. (2002). On the origin of deoxypentoses: evidence to support a glucose progenitor in the biosynthesis of calicheamicin. ChemBioChem *3*, 1143–1146.

Blanco, G., Patallo, E.P., Brana, A.F., Trefzer, A., Bechthold, A., Rohr, J., Mendez, C., and Salas, J.A. (2001). Identification of a sugar flexible glycosyltransferase from *Streptomyces olivaceus*, the producer of the antitumor polyketide elloramycin. Chem. Biol. *8*, 253–263.

Boghaert, E.R., Sridharan, L., Armellino, D.C., Khandke, K.M., DiJoseph, J.F., Kunz, A., Dougher, M.M., Jiang, F., Kalyandrug, L.B., Hamann, P.R., et al. (2004). Antibody-targeted chemotherapy with the calicheamicin conjugate hu3S193-N-Acetyl gamma calicheamicin dimethyl hydrazide targets lewisy and eliminates Lewis(y)-positive human carcinoma cells and xenografts. Clin. Cancer Res. *10*, 4538–4549.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J., and Gilbert, H.J. (2007). The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. Proc. Natl. Acad. Sci. U.S.A. *104*, 5336–5341.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

de La Fortelle, E., and Bricogne, G. (1997). SHARP: a maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

DiJoseph, J.F., Popplewell, A., Tickle, S., Ladyman, H., Lawson, A., Kunz, A., Khandke, K., Armellino, D.C., Boghaert, E.R., Hamann, P.R., et al. (2005). Antibody-targeted chemotherapy of B-cell lymphoma using calicheamicin conjugated to murine or humanized antibody against CD22. Cancer Immunol. Immunother. *54*, 11–24.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Galm, U., Hager, M.H., Van Lanen, S.G., Ju, J., Thorson, J.S., and Shen, B. (2005). Antitumor antibiotics: bleomycin, enediynes, and mitomycin. Chem. Rev. *105*, 739–758.

Gao, Q., and Thorson, J.S. (2008). The biosynthetic genes encoding for the production of the dynemicin enediyne core in *Micromonospora chersina* ATCC53710. FEMS Microbiol. Lett. *282*, 105–114.

Hamann, P.R., Hinman, L.M., Beyer, C.F., Lindh, D., Upeslacis, J., Shochat, D., and Mountain, A. (2005). A calicheamicin conjugate with a fully humanized anti-MUC1 antibody shows potent antitumor effects in breast and ovarian tumor xenografts. Bioconjug. Chem. *16*, 354–360.

Hensens, O.D., Giner, J.L., and Goldberg, I.H. (1989). Biosynthesis of NCS chrom-A, the chromophore of the antitumor antibiotic neocarzinostatin. J. Am. Chem. Soc. *111*, 3295–3299.

Hosted, T.J., Wang, T.X., Alexander, D.C., and Horan, A.C. (2001). Characterization of the biosynthetic gene cluster for the oligosaccharide antibiotic, evernimicin, in *Micromonospora carbonacea* var. *africana* ATCC 39149. J. Ind. Microbiol. Biotechnol. *27*, 386–392.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005). FFAS03: a server for profile-profile sequence alignments. Nucleic Acids Res. *33*, W284–W288.

Jiang, J., Biggins, J.B., and Thorson, J.S. (2000). A general enzymatic method for the synthesis of natural and "unnatural" UDP- and TDP-nucleotide sugars. J. Am. Chem. Soc. *122*, 6803–6804.

Jiang, J., Biggins, J.B., and Thorson, J.S. (2001). Expanding the pyrimidine diphosphosugar repertoire: yhe chemoenzymatic synthesis of amino- and acetamidoglucopyranosyl derivatives. Angew. Chem. Int. Ed. Engl. *40*, 1502–1505.

Kennedy, D.R., Gawron, L.S., Ju, J., Liu, W., Shen, B., and Beerman, T.A. (2007a). Single chemical modifications of the C-1027 enediyne core, a radiomimetic antitumor drug, affect both drug potency and the role of ataxia-telangiectasia mutated in cellular responses to DNA double-strand breaks. Cancer Res. *67*, 773–781.

Kennedy, D.R., Ju, J., Shen, B., and Beerman, T.A. (2007b). Single modifications of an enediyne chromophore can confer an ability to induce DNA double strand breaks, interstrand crosslinks or both with concominant changes in PIKK regulation of DNA damage response. Proc. Natl. Acad. Sci. U.S.A. *104*, 17632–17637.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Kumar, R.A., Ikemoto, N., and Patel, D.J. (1997). Solution structure of the calicheamicin gamma$_1^I$-DNA complex. J. Mol. Biol. *265*, 187–201.

Lam, K.S., Veitch, J.A., Golik, J., Krishnan, B., Klohr, S.E., Volk, K.J., Forenza, S., and Doyle, T.W. (1993). Biosynthesis of esperamicin-a(1), an enediyne antitumor antibiotic. J. Am. Chem. Soc. *115*, 12340–12345.

Liu, W., Christenson, S.D., Standage, S., and Shen, B. (2002). Biosynthesis of the enediyne antitumor antibiotic C-1027. Science *297*, 1170–1173.

Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S.D., Bae, J., Van Lanen, S.G., Zazopoulos, E., Farnet, C.M., Yang, C.F., and Shen, B. (2005). The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. Chem. Biol. *12*, 293–302.

Madej, T., Gibrat, J.F., and Bryant, S.H. (1995). Threading a database of protein cores. Proteins *23*, 356–369.

Melancon, C.E., III, Thibodeaux, C.J., and Liu, H.W. (2006). Glyco-stripping and glyco-swapping. ACS Chem. Biol. *1*, 499–504.

Minami, A., Kakinuma, K., and Eguchi, T. (2005). Aglycon switch approach toward unnatural glycosides from natural glycoside with glycosyltransferase VinC. Tetrahedron Lett. *46*, 6187–6190.

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.

Mulichak, A.M., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2001). Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. Structure *9*, 547–557.

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

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2004). Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. Biochemistry *43*, 5170–5180.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. *25*, 1396–1405.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. *62*, 439–450.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Rothstein, D.M., and Love, S.F. (1991). Isolation of mutants blocked in Calicheamicin biosynthesis. J. Bacteriol. *173*, 7716–7718.

Salas, J.A., and Mendez, C. (2007). Engineering the glycosylation of natural products in actinomycetes. Trends Microbiol. *15*, 219–232.

Sievers, E.L., and Linenberger, M. (2001). Mylotarg: antibody-targeted chemotherapy comes of age. Curr. Opin. Oncol. *13*, 522–527.

Scott, D.L., White, S.P., Otwinowski, Z., Yuan, W., Gelb, M.H., and Sigler, P.B. (1990). Interfacial catalysis: the mechanism of phospholipase A2. Science *250*, 1541–1546.

Singh, S., Hager, M.H., Zhang, C.S., Griffith, B.R., Lee, M.S., Hallenga, K., Markley, J.L., and Thorson, J.S. (2006). Structural insight into the self-sacrifice mechanism of enediyne resistance. ACS Chem. Biol. *1*, 451–460.

Thibodeaux, C.J., Melancon, C.E., and Liu, H.W. (2007). Unusual sugar biosynthesis and natural product glycodiversification. Nature *446*, 1008–1016.

Thorson, J.S., Shen, B., Whitwam, R.E., Liu, W., Li, Y., and Ahlert, J. (1999). Enediyne biosynthesis and self-resistance: a progress report. Bioorg. Chem. *27*, 172–188.

Thorson, J.S., Sievers, E.L., Ahlert, J., Shepard, E., Whitwam, R.E., Onwueme, K.C., and Ruppen, M. (2000). Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research. Curr. Pharm. Des. *6*, 1841–1879.

Tokiwa, Y., Miyoshisaitoh, M., Kobayashi, H., Sunaga, R., Konishi, M., Oki, T., and Iwasaki, S. (1992). Biosynthesis of dynemicin-A, a 3-ene-1,5-diyne antitumor antibiotic. J. Am. Chem. Soc. *114*, 4107–4110.

Torkkell, S., Ylihonko, K., Hakala, J., Skurnik, M., and Mantsala, P. (1997). Characterization of *Streptomyces nogalater* genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis. Mol. Gen. Genet. *256*, 203–209.

Van Lanen, S.G., Oh, T.-J., Liu, W., Wendt-Pienkowski, E., and Shen, B. (2007). Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. J. Am. Chem. Soc. *129*, 13082–13094.

Uson, I., and Sheldrick, G.M. (1999). Advances in direct methods for protein crystallography. Curr. Opin. Struct. Biol. *9*, 643–648.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J. Appl. Cryst. *30*, 1022–1025.

Van Lanen, S.G., and Shen, B. (2008). Biosynthesis of enediyne antitumor antibiotics. Curr. Top. Med. Chem. *8*, 448–459.

Walker, S., Landovitz, R., Ding, W.D., Ellestad, G.A., and Kahne, D. (1992). Cleavage behavior of calicheamicin gamma$_1^I$ and calicheamicin T. Proc. Natl. Acad. Sci. U.S.A. *89*, 4608–4612.

Walker, S., Murnick, J., and Kahne, D. (1993). Structural characterization of a calichemicin-DNA complex by NMR. J. Am. Chem. Soc. *115*, 7954–7961.

Walker, S.L., Andreotti, A.H., and Kahne, D.E. (1994). NMR characterization of calicheamicin $\gamma_1^I$ bound to DNA. Tetrahedron *50*, 1351–1360.

Williams, G.J., and Thorson, J.S. (2008). A high-throughput fluorescence-based glycosyltransferase screen and its application in glycosyltransferase directed-evolution. Nat. Protocols *3*, 357–362.

Williams, G.J., Zhang, C., and Thorson, J.S. (2007). Directed evolution of a natural product glycosyltransferase. Nat. Chem. Biol. *3*, 657–662.

Williams, G.J., Goff, R.D., Zhang, C., and Thorson, J.S. (2008). Optimizing glycosyltransferase specificity via 'hot spot' saturation mutagenesis presents a new catalyst for novobiocin glycorandomization. Chem. Biol. *15*, 393–401.

Weitnauer, G., Muhlenweg, A., Trefzer, A., Hoffmeister, D., Sussmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U., and Bechthold, A. (2001). Biosynthesis of the orthosomycin antibiotic avilamycin A: Deductions from the molecular analysis of the *avi* biosynthetic gene cluster of *Streptomyces viridochromogenes* Tü57 and production of new antibiotics. Chem. Biol. *8*, 569–581.

Wohlert, S., Lomovskaya, N., Kulowski, K., Fonstein, L., Occi, J.L., Gewain, K.M., MacNeil, D.J., and Hutchinson, C.R. (2001). Insights about the biosynthesis of the avermectin deoxysugar L-oleandrose through heterologous expression of *Streptomyces avermitilis* deoxysugar genes in *Streptomyces lividans*. Chem. Biol. *8*, 681–700.

Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B., and Farnet, C.M. (2003). A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. *21*, 187–190.

Zein, N., Sinha, A.M., McGahren, W.J., and Ellestad, G.A. (1988). Calicheamicin gamma$_1^I$: An antitumor antibiotic that cleaves double-stranded DNA site specifically. Science *240*, 1198–1201.

Zein, N., Poncin, M., Nilakantan, R., and Ellestad, G.A. (1989). Calicheamicin gamma$_1^I$ and DNA: Molecular recognition process responsible for site-specificity. Science *244*, 697–699.

Zhang, C., Albermann, C., Fu, X., and Thorson, J.S. (2006a). The in vitro characterization of the iterative avermectin glycosyltransferase AveBI reveals reaction reversibility and sugar nucleotide flexibility. J. Am. Chem. Soc. *128*, 16420–16421.

Zhang, C., Griffith, B.R., Fu, Q., Albermann, C., Fu, X., Lee, I.K., Li, L., and Thorson, J.S. (2006b). Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. Science *313*, 1291–1294.

Zhang, C., Fu, Q., Albermann, C., Li, L., and Thorson, J.S. (2007). The in vitro characterization of the erythronolide mycarosyltransferase EryBV and its utility in macrolide diversification. ChemBioChem *8*, 385–390.

Zhang, J., Van Lanen, S.G., Ju, J., Liu, W., Dorrestein, P.C., Li, W., Kelleher, N.L., and Shen, B. (2008). A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. Proc. Natl. Acad. Sci. U.S.A. *105*, 1460–1465.