2014-03-24

Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in Streptomyces nogalater

Claesson, Magnus; Siitonen, Vilja; Dobritzsch, Doreen; Metsä-Ketelä, Mikko; Schneider, Gunter

FEBS J. 2012 Sep;279(17):3251-63
http://doi.org/10.1111/j.1742-4658.2012.08711.x
http://hdl.handle.net/10616/41974

Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater*

Magnus Claesson[1], Vilja Siitonen[2], Doreen Dobritzsch[1], Mikko Metsä-Ketelä[2] and Gunter Schneider[1]*

1. Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
2. Department of Biochemistry and Food Chemistry, University of Turku, Turku, Finland

**To whom correspondence should be addressed: Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden gunter.schneider@ki.se, phone:+46852487675, fax: +468327626*

Running title: Crystal structure of the glycosyltransferase SnogD

Keywords. Polyketide biosynthesis, glycosyltransfer, protein structure, nucleotide binding, enzyme mechanism

Abstract

The glycosyltransferase SnogD from *Streptomyces nogalater* transfers a nogalamine moiety to the metabolic intermediate 3´,4´-demethoxynogalose-1-hydroxynogalamycinone during the final biosynthetic steps of the aromatic polyketide nogalamycin. The crystal structure of recombinant SnogD, as apo-enzyme and with a bound nucleotide, 2-deoxyuridine-5´-diphosphate, was determined to 2.6 Å resolution. Reductive methylation of SnogD was crucial for the reproducible preparation of diffraction quality crystals due to the creation of an additional intermolecular salt bridge between the methylated lysine residue Lys384 and Glu374* from an adjacent molecule in the crystal lattice. SnogD is a dimer both in solution and the crystal and the enzyme subunit displays a fold characteristic of the GT-B family of glycosyltransferases. Binding of the nucleotide is associated with rearrangement of two active site loops. Site-directed mutagenesis shows that two active site histidine residues, His25 and His301, are critical for the glycosyltransferase activities of SnogD both *in vivo* and *in vitro*. The crystal structures and the functional data are consistent with a role of His301 in binding of the diphosphate group of the sugar donor substrate and a function of His25 as a catalytic base in the glycosyl transfer reaction.

Database

The atomic coordinates and structure factors have been deposited with the RCSB Protein Data Bank under accession numbers **4AMB**, **4AMG**, and **4AN4**. Streptomyces nogalater glycosyltransferase SnogD, GenBank accession no. **AAF01811.1**, **EC 2.4.1**, UniProt **Q9RN61**

Abbreviations

GT, glycosyltransferase; dUDP, 2´-deoxyuridine-5´-diphosphate; SnogD-wt, SnogD wild-type; SnogD-wtUDP, wild-type SnogD-dUDP binary complex; SnogD-m, methylated SnogD; SnogD-mdUDP, methylated SnogD-dUDP binary complex.

Introduction

Anthracylines, produced mainly by soil dwelling Gram-positive bacteria of the genus *Streptomyces*, are aromatic polyketides that often exhibit a high cytostatic potency. Several members of this group, for example epirubicin and doxorubicin, belong to the most used anti-cancer agents worldwide [1]. Anthracycline biosynthesis in *Streptomyces* starts with the production of the polyketide scaffold, which is formed by iterative additions of malonate extender units to acetate or propionate primers by type II polyketide synthase in concert with cyclases, aromatases and ketoreductases [2]. The aglycone backbone is further modified by tailoring enzymes including hydroxylases, methylases and aminotransferases, leading to the large diversity of these metabolites. An important class of tailoring enzymes are the glycosyltransferases, as glycosylation of the carbon skeleton is required in many cases to form the biologically active polyketides [3,4]. These glycosyltransferases use nucleotide diphosphate sugars as donors, typically TDP-conjugated sugars, with a significantly greater diversity amongst the sugar-acceptors [5].

Streptomyces nogalater produces the aromatic polyketide nogalamycin (Figure 1, **1**), which contains the carbohydrate moieties nogalose attached to carbon 7 and nogalamine attached to carbon 1 [6]. The attachment of nogalamine to the aglycone is highly remarkable as it involves the formation of a carbon-carbon bond linking atoms C5´´ of the sugar and C2 of the polyketide scaffold. A second link is formed via an O-glycosidic bond to carbon 1 of nogalamine. Attachment of sugar moieties to natural products via C-C bonds has been described in a few cases, for instance urdamycin [7], gilvocarcin [8], hedamycin [9] and granaticin [10], but the combination of both types of linkages in nogalamycin is rather unique and poses interesting questions about the order of the attachment and the chemistry of this glycosylation step.

The gene cluster responsible for the biosynthesis of nogalamycin in *S. nogalater* has been cloned and the functions of most of the encoded enzymes have been annotated by sequence comparisons and/or subsequent biochemical and structural studies [11–15]. The cluster contains three genes that code for glycosyltransferases, *snogD*, *snogE* and *snogZ*, related to each other by amino acid sequence identities of around 30 %. A recent study

showed that SnogE is responsible for the transfer of nogalose to the C-7 hydroxyl group of nogalamycinone (Figure 1, **2**), whereas SnogD is involved in the transfer of the nogalamine moiety to the glycosylated metabolic intermediate 3´,4´-demethoxynogalose-1-hydroxynogalamycinone (Figure 1, **3)** [15]. The glycosyltransferase SnogZ appears to be redundant as it is not necessary for the heterologous formation of double glycosylated anthracyclines in the expression host *S. albus*.

Sequence comparisons of SnogD (UniProt **Q9RN61**) with related enzymes show that it belongs to the glycosyltransferase type 1 family [16]. These enzymes display the GT-B fold [17] and are characterized by a S_N2-type mechanism of sugar transfer with accompanying inversion of the anomeric configuration at C1 of the donor-sugar. Here we present the X-ray crystal structure of SnogD, in the apo-form and in complex with the nucleotide 2´-deoxyuridine-5´-diphosphate (dUDP). We further report the results of *in vitro* and *in vivo* mutagenesis studies of the conserved active site residues His25 and His301.The structural and functional data are consistent with crucial roles of these residues in the mechanism of SnogD as a catalytic base and in nucleotide-sugar binding, respectively.

Results

Structure determination of SnogD

The crystal structure of recombinant SnogD, comprising residues 13-390 of the native enzyme, was determined in two different space groups to resolutions of 2.6 and 2.7 Å, respectively (Table 1). Crystals of wild-type SnogD (SnogD-wtdUDP) of sufficient diffraction quality (< 3 Å resolution) could not be reproduced despite numerous attempts. Reductive methylation of the protein sample proved critical for reproducibility of crystallisation and diffraction quality of the crystal. Chemical modification of SnogD resulted in a mass of 42023.5 Da determined by electrospray mass spectrometry, which compares well to the theoretical mass (42024.0 Da) of the SnogD construct including

eight methyl groups. This data thus suggest almost complete di-methylation of the N-terminal nitrogen atom and the four lysine residues.

The structure of the enzyme was determined in three different forms, wild-type SnogD in space group $P2_12_12$ and methylated SnogD in space groups $P2_12_12$ and P2, respectively (SnogD-mdUDP and SnogD-m). The orthorhombic space group contains two chains in the asymmetric unit (51 % solvent content). The refined models comprise all residues of the construct (13-390), with a few exceptions. Weak or no electron density is observed for two loop regions in most of the chains of all three reported structures, which were thus modelled to various extents. These peptide segments are in the vicinity of the acceptor substrate binding site of the enzyme and include residues 74-93 (loop FL1) and 323-327 (loop FL2). The asymmetric unit of the monoclinic crystal form contains four polypeptide chains, corresponding to a solvent content of 40 %. In addition to the above mentioned two loops, several other loop regions are disordered in at least one of the subunits of the corresponding model. These include residues 152-156, 196-199, 235-244, 275-276, and 339-350.

During refinement, residual difference electron density at the nucleotide binding site of chain B in SnogD-wtdUDP and chains B and D of SnogD-mdUDP indicated ligand binding. The shape of this electron density and the hydrogen bonding interactions with the enzyme were consistent with that of a diphosphorylated nucleoside lacking the 2´-hydroxyl group, and was subsequently modelled as dUDP (Figure 2). The ligand originates from the expression host as it was not added during crystallisation. Details of the refinement and the model quality are given in Table 1.

Overall structure of SnogD

Structure of the subunit: SnogD displays the twin-domain architecture of the GT-B fold, consisting of two Rossman-fold domains with the active site located at the domain interface (Figure 3). The N-terminal domain (residues 1-209) consists of a seven-stranded parallel β-sheet, eight α-helices and two 3_{10}-helices. The sheet itself is sandwiched between four of the helices on one side and three helices on the other side. The C-terminal domain (residues 228-390) is of similar topology with a six-stranded parallel β-

sheet flanked by six α- and four 3_{10}-helices. The last C-terminal helix, with a kink between residues Glu374-Pro377, a common feature of the GT-B fold, crosses over to complete the N-terminal domain through residues Pro378-Gly390.

Structural relationships to related proteins: A search for structural homologues of SnogD using the DALI-server [18] resulted in several GT-1 glycosyltransferases with significant structural similarity despite low sequence homology. The closest structural relatives are CalG3 [19], Z-score 42.5, r.m.s.d 2.3 Å, 38 % sequence identity; UrdGT2 [20], Z-score 37.4, r.m.s.d 2.6 Å, 28 % sequence identity; SpnG [21], Z-score 37.4, r.m.s.d 2.6 Å, 30 % sequence identity; CalG1 [22], Z-score 37.1, r.m.s.d 3.5 Å, 30 % sequence identity, EryCIII [23], Z-score 35.8, r.m.s.d 2.9 Å, 34 % sequence identity and OleI [24], Z-score 33.2, r.m.s.d 2.7 Å, 25 % sequence identity.

A pBLAST search returns the amino acid sequence of the hypothetical protein ML5_4439 (Uniprot id: **E8RWV0**) from *Micromonospora* sp. L5 as the closest relative, showing 63.1 % amino acid sequence identity.

Quaternary structure: Analysis of the crystal packing by manual inspection and with the PISA server [25] reveals that SnogD is a dimer in both crystal forms. This agrees well with the results from analytical gel filtration experiments that indicated a dimeric structure also in solution. The subunits in the SnogD dimer are oriented head-to-tail (Figure 3), and related by a twofold non-crystallographic symmetry axis. The dimer interface covers an area of 1430 $Å^2$, and includes three salt bridges and 16 hydrogen bonds between the subunits. Part of the interface is also the extension of the 7-stranded β-sheet of the N-terminal domain by an additional parallel β-strand comprising amino acids 215-217 from the long inter-domain peptide segment (210-227) linking the N-terminal to the C-terminal domain of the neighboring subunit. Superimposition of the two subunits forming the dimer results in an r.m.s.d. of 0.5-0.8 Å. The overall structures of the dUPD-free and dUDP bound subunits are all very similar, irrespective of crystal from, and superposition results in r.m.s.d values in the range of 0.6-0.8 Å. Structural deviations are restricted to the nucleotide binding site and described below.

Reductive methylation stabilizes crystal packing: Reductive methylation significantly improved the reproducibility in obtaining better diffracting crystals. Electron density corresponding to a methylated lysine side chain was however only observed for one (Lys384) of the four lysine residues in chain A. This residue was well defined in electron density in the two crystal forms, whereas the side chains of other lysine residues (including Lys384 from chains B, C and D in the asymmetric unit) were disordered to various extends. Methylated Lys384 forms an intramolecular stacking interaction with the side chain of Trp204, and most importantly is also involved in a salt bridge to Glu374* from the polypeptide chain related by crystallographic symmetry. This salt bridge, only observed for chain A, is not present in crystals of wild-type SnogD and the additional interaction in the crystal lattice most likely is involved in the improved crystallization behavior upon reductive methylation.

Nucleotide binding to SnogD: dUDP is bound in a cleft at the interface between the two domains of the SnogD subunit. Hydrogen bonds to the ligand involve residues from the C-terminal domain and one residue of the inter-domain linker (Figure 4). The uracil moiety packs against the side chain of Trp285. Atoms O2 and N3 of the pyrimidine ring are hydrogen-bonded to the backbone amide of Leu288 and carbonyl oxygen of Ile286, respectively, while the O4 atom is in close proximity (3.4 Å) of the backbone amide of the latter residue. The 3´-hydroxyl group and the O5´ atom of the deoxyribose moiety form hydrogen bonds to the side chains of Asn212 and Thr306, respectively.

The α-phosphate is located at the N-terminal end of the helix (303-315) and is linked via hydrogen bonds to the amide groups of Ser304 and Gly305 (Figure 3). Both residues are part of the conserved diphosphate-binding "PPi-motif" [26] represented in SnogD by GGSG (Figure S1). The highly conserved His301 is appropriately positioned to form hydrogen bonds to both α- and β-phosphate groups. A modest movement would bring the β-phosphate group into hydrogen bonding distance to main chain amide and side chain hydroxyl of Ser236. Such an interaction is observed in one of the nucleotide-binding subunits of SnogD-mdUDP.

In UDP-utilizing GTs the recognition of the ribose moiety is provided by hydrogen bonds of a glutamic acid residue located in the α-helix directly following the PPi-motif

[24,27,28]. In SnogD, this glutamic acid is replaced by Thr309, which does not interact with the deoxyribose group. Instead the 3'-hydroxyl group of the carbohydrate moiety forms a hydrogen bond to the side chain of Asn212. The different mode of hydrogen bonding interactions of the deoxyribose group in SnogD compared to the UDP-donor specific glycosyltransferases results in a difference in conformation of parts of the interdomain linker (residues 210-227). These conformational differences can be described as a shift of this peptide segment closer to the nucleotide binding site (Figure S2), a feature also recently observed in TDP-donor specific SpnG [21].

Another sequence motif, D/E-Q, frequently encountered in the nucleotide binding region of glycosyltransferases [29] and involved in hydrogen bonds to the C2´´-C4´´ hydroxyl groups of the donor-sugar substrate is not conserved in SnogD (Figure S1). However, within the FL2 loop such polar residues are present, which potentially could form hydrogen bonding interactions to the 3´´-amino and 4´´-hydroxyl groups of the donor sugar nogalamine. This loop is disordered in several of the subunits of SnogD, most likely due to flexibility in the absence of the activated sugar donor.

In SnogD-wtdUDP and SnogD-mdUDP, nucleotide binding is observed in only one of the two subunits of the dimer. Comparison of nucleotide-free with nucleotide-bound subunits reveals that ligand binding is accompanied by the rearrangements and/or stabilization of two nearby residue segments. One of these segments (n1) comprises part of the binding site for the α-phosphate, residues 235-237. In the nucleotide-free subunits these residues are either disordered or adopt different conformations and are placed more distant from the ligand binding site. Local structural differences extend to amino acids up to position 249, including the unwinding of the 3_{10}-helix (242-246) in nucleotide-free subunits A and C of SnogD-m (Figure 5). In the structure of the binary complex loop n2 (residues 266-268) is shifted away from the binding site to accommodate the pyrimidine ring of the bound nucleotide.

The binding of nucleotides to only one of the active sites in the dimer observed for SnogD-wtdUDP and SnogD-mdUDP may be directly linked of the participation of segments 235-249 (n1) and 266-277 (n2) in crystal packing interactions in both crystal forms. It is thus likely that "half-occupied" and "empty" dimers are preferably selected

for incorporation into the crystal lattice, as they interact more optimally with neighboring molecules. This may also explain the difficulties in obtaining crystals of SnogD with fully occupied binary complexes.

Model of SnogD with activated sugar donor and acceptor substrates: Based on the structures of the SnogD-dUDP binary complexes and of the complex of a plant 3-O-glucosyltransferase with the sugar donor analog uridine-5´-diphosphate-2-deoxy-2-fluoro-α-D-glucose and the acceptor kaempferol [29] we modeled binding of an activated sugar donor to the active site of SnogD (Figure 6). The position of the donor-sugar nogalamine is restricted by the covalent link to the dinucleotide, the small pocket for the carbohydrate moiety present in SnogD, and the requirement for the C1´-hydroxyl to be in an axial orientation so that glycosyl transfer can occur.

In the absence of experimental data on SnogD-acceptor substrate complexes the precise orientation of acceptor substrates in the binding pocket of SnogD is more difficult to define, due to the large size of the hydrophobic pocket and the lack of structural data for the FL1 loop. Acceptor substrate binding to SnogD most likely involves hydrophobic residues lining the cleft between the domains and closure of the FL1 loop over the acceptor substrate as seen in CalG3 [22]. The acceptor substrate **3** (Figure 1) can be modeled in the predominantly hydrophobic substrate cleft in such a way that the C1 hydroxyl group of the acceptor is located suitably to act as nucleophile and attack the C1 atom of the activated nogalamine sugar (Figure 6). In this model His25 is close to the C1 hydroxyl group of the aglycon, which is the attacking nucleophile in the glycosyl-transfer reaction.

In vitro characterization of active site mutants of SnogD

Based on the modeled ternary complex of SnogD with the nucleotide-sugar donor and acceptor substrates, we selected the highly conserved residues His25 and His301 (Figure S1) as target for site directed mutagenesis. The mutations His25Ala, His25Asn and His301Ala were introduced and the mutant proteins purified according to the protocol used for wild-type SnogD. The CD spectra of mutants and wild-type enzyme were very similar, indicating that the mutants were correctly folded. The natural sugar donor TDP-

nogalamine was not available and the enzymatic reaction was therefore studied in the reverse direction using the assay described previously [15] (Figure 7A). The *in vitro* assays with the mutants His25Ala, His25Asn and His301Ala showed a significant reduction of UDP-dependent deglycosylation of nogalamycin F **4** to **3** to 1.5 – 2.4 % of the wild type enzyme activity (Figure 7B and Table 2).

In vivo studies of active site mutants

In order to probe the glycosyltransfer reaction in the forward direction, the mutated genes were expressed *in vivo* in *S. albus* using a two-plasmid system. The polyketide acceptor **3** and the nucleotide diphosphate deoxysugar donor substrates were produced using cosmid pSnoΔgD, which contains the majority of the nogalamycin gene cluster and from which *snogD* has been deleted [15]. The native and mutated *snogD* genes were expressed in the heterologous host using the high copy number vector pIJ486 from their own promoter sequences as described in the experimental section. The production profiles of the strains indicated that formation of the aminoglycosylated SnogD reaction product, nogalamycin R **5** (Figure 8A), was impaired in the mutant strains (Figure 8B), consistent with the *in vitro* activity measurements.

Discussion

Previous studies on the glycosylation reactions in nogalamycin biosynthesis have shed some light on the order of these steps and the enzymes involved [15]. Gene inactivation experiments have shown that glycosyltransfer of nogalose to the C7-hydroxyl group of nogalamycinone by SnogE precedes transfer of nogalamine to the C1-hydroxyl by SnogD. The studies further suggested that formation of the C1-O-glycosidic linkage precedes formation of the C-C bond between the C5´´ carbon of nogalamine and the C2 carbon of the polyketide scaffold. The nogalamine moiety has been demonstrated to be involved in the DNA-nogalamycin interactions [30], which are important for poisoning of human topoisomerases [31], one of the major biological activities of nogalamycin. Therefore detailed understanding of these biosynthetic steps and the enzymes may facilitate engineering of novel anthracyclines with improved activities.

The structure analysis of SnogD provides a framework for the last glycosylation step, the transfer of nogalamine to **3**. The enzyme belongs to the type 1 branch of the GT-B fold family of glycosyltransferases. Fortuitously the enzyme crystallized with bound dUDP derived from the expression host that allowed structural characterization of nucleotide binding to SnogD. Sugar donors for the glycosylation reactions in anthracycline biosynthesis in *Streptomyces* species are often formulated as activated TDP-sugars, although the biochemical evidence in the case of nogalamycin biosynthesis is lacking. SnogD can utilize both TDP and UDP in the *in vitro* assay, the deglycosylation of **4** [15]. This is consistent with structural data, as the binding pocket could accommodate the uracil and thymidin rings. SnogD however lacks the glutamic acid residue proposed to be characteristic for UDP-donor specific glycosyltransferases [27] and thus might not be able to discriminate efficiently between UDP- and TDP-donors.

Since attempts to obtain crystals of ternary complexes of SnogD with acceptor substrates were not successful, the structure of SnogD with bound dUDP was used as templates to model the Michaelis complex of SnogD with UDP-nogalamine and **3**. This model suggests one active site residue, His25 close to the C1´´ atom of the activated sugar that might be able to facilitate the enzymatic reaction. This residue could be involved in proton abstraction, thus enhancing nucleophilicity of the acceptor substrate and facilitating the attack onto the C1´´ of the donor-sugar and subsequent glycosyl transfer. Replacements of His25 by alanine and asparagine result in mutants that are severely impaired in catalytic activity *in vitro.* Analysis of the production profiles *in vivo* using these SnogD mutants show the same trend with a significant reduction in the amounts of **5** produced. These data demonstrate a critical role of this residue in the enzyme reaction and are consistent with a function as catalytic base.

Another histidine residue, His301 close to the diphosphate group of the donor sugar, is also essential for catalysis as substitution of this residue by alanine results in significant loss of enzyme activity *in vitro*. This residue is involved in binding of the diphosphate group and could further stabilize the negative charge developing at the β-phosphate of the leaving group [32]. These two residues are conserved in many glycosyltransferases and

their counterparts have been implicated in the catalytic mechanisms of these enzymes [19,24,27–29,33].

The evidence for the function of SnogD as a glycosyltransferase involved in the formation of the C1-O-glycosidic bond to nogalamine is now rather compelling. It is however much less clear whether or not SnogD is directly involved in the formation of the carbon-carbon linkage between C5´´ of nogalamine and the C2 atom of the polyketide. This reaction requires formally abstraction of hydrogen, and the structure of SnogD does not contain any catalytic groups and/or cofactors that might support such chemistry. This step in nogalamycin biosynthesis remains enigmatic and most likely requires another enzyme, either working in concert with SnogD or acting after the glycosyltransfer reaction.

Materials and methods

Cloning, expression and purification of recombinant SnogD

The cloning, expression and purification of recombinant SnogD was performed as described previously [15]. In short, eight different constructs were designed, all with an N-terminal His-tag, amplified from genomic *Streptomyces nogalater* DNA by PCR and cloned using ligation-independent (LIC) cloning [34]. The production of one construct, comprising residues 13-390, using *E. coli* as expression host followed a protocol described elsewhere [15]. For the production of SnogD used for subsequent methylation experiments, cells were grown with supplementation of the culture medium with a mixture of metal salts with final concentrations of 40µM $CaCl_2$, 20µM $MnCl_2$, 20µM $ZnSO_4$, 4µM $CoCl_2$, 4µM $CuSO_4$, 4µM Na_2MoO_4, 4µM H_3BO_3, 4µM $NiSO_4$, 4µM Na_2SeO_4 and 4µM $FeCl_3$. Recombinant SnogD was purified using affinity, anion-exchange and size exclusion chromatography. The purity of the protein was monitored by SDS-PAGE and the identity of the recombinant protein was confirmed by electrospray mass spectrometry (ES-MS).

Mutagenesis

Residues for mutagenesis were selected based on the structure of SnogD and protein sequence alignment to structural homologues (Figure S1). The mutations were introduced

by PCR using the full-length expression plasmid, encoding residues 1-390 of SnogD cloned as described previously [15], as template using primers containing the mutations (His25Alaf 5´-GCCATCCTGCCGACGGTGCCGCTGGC-3´, His25Alar 5´-GGCGCTGAGC CCGGGTGAAGTGATGAACAACGCA-3´, His25Asn_f 5´-AACATCCTGCCGACGGTGCCGCT GGC-3´, His25Asn_r 5´-GTTGCTGAGCCCGGGTGAAGTGATGAACAACGCA-3´, His301Alaf 5´-GCCGGGGGCAGCGGCACACTGCTGACG-3´, His301Alar 5´-GGCATGGATGATCGCGTCGC ACGTCTCCAGCAG-3´), which were combined with vector specific primers (f5´-GAATAATT TTGTTTAACTTTAAGAAGGAGATATACATATGCACCATCA-3´, r5´-TGCGGCCGCAAG CTTGTCGACGG-3´) to amplify the mutated DNA fragments using Phusion DNA polymerase (Finnzymes, Espoo, Finland)). These fragments were subsequently used to generate the full dsDNA sequence of SnogD harbouring the mutation as well as vector compatible restriction sites. After restriction enzyme digestion (*Nde*I and *Hind*III) ligation using T4 ligase (all enzymes from New England Biolabs, MA, USA or Fermentas International Inc., Canada) was performed to insert the dsDNA into the vector (pET28-pNic BasI). The sequences of the respective mutants were verified by DNA sequencing.

Reductive methylation of SnogD

Purified SnogD was diluted to a final concentration of 1 mg/ml in a solution containing 45 mM Na_3PO_4, 0.3 M NaCl, and 10 % (v/v) glycerol pH 8.0. Reductive methylation was performed based on a protocol described previously [35], with two exceptions; the formaldehyde solution was prepared freshly by dissolving paraformaldehyde in 50 mM Na_3PO_4 pH 8.0 and heating to 333 K during 4 hours, and the methylation reaction was quenched by addition of 1 M glycine to a final concentration of 0.1 M. After filtration methylated SnogD was purified using anion exchange and size exclusion chromatography. The extent of methylation of SnogD was determined by ES-MS.

Crystallization of SnogD

Wild-type SnogD (SnogD-wtdUDP) was crystallized by hanging drop vapour diffusion against a reservoir containing 26 % (w/v) PEG 3350 and 0.1 M citrate pH 5.2. Droplets of the protein solution (10 mg/ml SnogD-wtdUDP in 50 mM Tris pH 7.4, 0.2 M NaCl) were mixed with equal volumes of the reservoir solution. Crystals belonging to space group $P2_12_12$ appeared after three days equilibration at 293 K.

Since diffraction quality crystals were difficult to reproduce using wild type enzyme, screening for crystallization conditions was repeated with methylated SnogD, which in combination with streak seeding reproducibly yielded diffraction quality crystals at two conditions. The SnogD-mdUDP and SnogD-m data were collected from crystals obtained by vapour diffusion against a reservoir of 15 % (w/v) PEG 8000, 0.16 M calcium acetate, 0.08 M sodium cacodylate pH 6.8, and 16 % (w/v) PEG 3350, 0.2 M $MgCl_2$, 0.1 M BisTris pH 5.7, respectively.

Data collection and structure determination

SnogD-wtdUDP crystals were briefly transferred to a solution of 30 % (v/v) glycerol, 28 % (w/v) PEG 3350, 0.1 M BisTris pH 5.3 prior to flash-freezing in liquid nitrogen. Crystallographic data were collected at 100 K and at a wavelength of 0.934 Å at beamline ID14-EH1 of the European Synchrotron Radiation Facility (Grenoble, France), to a resolution of 2.6 Å. Crystals of methylated SnogD were cryo-protected using mineral oil. Data were collected from two crystals, to 2.7 Å (SnogD-mdUDP) and 2.6 Å (SnogD-m) resolution, respectively, at a wavelength of 1.072 Å at beamline ID23-EH1 at ESRF (Grenoble, France). Data processing was performed using the programs MOSFLM and SCALA from the CCP4 package [36]. Statistics of the diffraction data sets are given in Table 1.

The SnogD-m crystal belonged to space group $P2_12_12$ with similar unit cell dimension as the SnogD-wt crystal, containing two enzyme subunits per asymmetric unit. In contrast, the SnogD-mdUDP crystal belongs to space group P2 with 4 molecules (2 dimers) per asymmetric unit.

The crystal structure of SnogD-wtdUDP was determined by molecular replacement using PHASER [37] and an ensemble of monomer structures of five homologues enzymes (PDB accession codes **1F0K**, **1NLM**, **2IYA**, **2P6P**, and **3D0Q**) as search model, of which CalG3 from *Micromonospora echinospora* shows the highest sequence identity to SnogD (36 %). The structures of methylated SnogD were determined by molecular replacement using PHASER (SnogD-mdUDP) or MOLREP [38] (SnogD-m) and the refined ligand-free subunit of SnogD-wt as search model.

Model building and refinement was performed using WinCOOT [39] and REFMAC5 [40], employing tight main chain NCS restraints in the first and automatically determined local NCS restraints in the final cycles of refinement. 5 % randomly selected reflections were used for monitoring R_{free}. Water molecules were added in WinCOOT. The presence of residual electron density indicated ligand binding to the active sites of one subunit per dimer for SnogD-wtdUDP and SnogD-mUDP. Based on shape of the density and the hydrogen-bonding pattern the ligand was identified and modeled as dUDP, an analogue of one of the products of the SnogD-catalyzed reaction.

The final model of SnogD-wtdUDP contains residues 13-389 of SnogD, one residue from the affinity tag for chain A and residues 13-390 for chain B, two residues from the affinity tag, a total of 115 water molecules and one dUDP molecule. Two loop regions, residues 76-84 and 325-327, are disordered to various extents in the two subunits. The refined model of SnogD-mdUDP contains four subunits. Also in this space group several loop regions were disordered to various extents in the four subunits and not included in the model. The structural model contains 48 water and 2 dUDP molecules. The final model of SnogD-m contains residues 13-389 (except for the following loop regions, which are disordered in the electron density maps, A79-A90, A239-A241, B78-82, and B324-327) and 67 water molecules. Refinement statistics are given in Table 1.

The atomic coordinates and crystallographic data have been deposited with the Protein Data Bank under accession numbers **4AMB** (SnogD-wtdUDP), **4AMG** (SnogD-m), and **4AN4** (SnogD-mdUDP).

Quaternary structures, crystal packing and interfaces were analyzed using the PISA-server at the European Bioinformatics Institute [25]. Database searches for structural homologues were carried out using Dali [18]. Sequence alignments were performed using ClustalW [41]. Figures were prepared with PyMol [42].

In vitro enzyme assays

Full-length SnogD and point mutants were used for the activity assays. Since activated TDP-nogalamine or UDP-nogalamine were not available as sugar donors the reaction was monitored in the reverse direction, using UDP as acceptor and **4** as donor using the protocol described previously [15]. In short 8.1 μg of SnogD was added to 40 μl of a mixture of 20 mM UDP as acceptor- and 0.5 mM **4** as donor-substrate in 50 mM Tris, 200 mM NaCl, pH 7.4 and incubated in the dark at 295 K for 13h. The reactions were quenched by addition of chloroform and the anthracyclines were extracted from the resulting mix of chloroform and water. The chloroform phases were analyzed by HPLC to detect glycosyl transferase activity, which was defined as the ratio between formed product **3** over remaining substrate **4**, after correction for **3** present in minor amounts in the substrate solution.

In vivo activity assays

The mutated *snogD* genes were moved as *Nde*I-*EcoR*I fragments to digested and CIAP-treated pBluescript SK(-) (Stratagene; La Jolla, CA, USA) harbouring the PCR amplified and sequenced *snogD* promoter region. The PCR was carried out using pSnogaori [15] as the template and the following primers with vector compatible restriction sites (*EcoR*I, *Hind*III) and added *Nde*I-site downstream of the promoter region: FWgDp AATAAGCTTAGGTCTCCGGGCGGGTCA REVgDp: TTAGAATTCTTACATATGGACGGCGCCTTCTGTTGC. The promoter and gene region were moved as an *EcoR*I-*Hind*III fragment (1.5 kb) to a digested and CIAP-treated pIJ486-vector [43]. The plasmids were amplified in *S. lividans* TK24, verified by

digestion analysis and moved to *S. albus*/pSnoΔgD by PEG-induced protoplast transformation [44]. The wild type SnogD plasmid was constructed as described earlier [15].

For analysis of the activity of SnogD and mutated SnogD the strains were cultivated in 250 ml Erlenmeyer-flasks with 30 ml of NoS-soyE1 [15] supplemented with thiostrepton (Calbiochem; San Diego, CA, USA 50 μg ml^{-1}) and apramycin (Sigma-Aldrich, St. Louis, MO, USA 50 μg ml^{-1}) for 5 days at 303 K in vigorous shaking. The produced metabolites were collected by binding to Amberlite® XAD-7 (2 days, Rohm and Haas, Philadelphia, PA, USA, 20 g l^{-1}) after which the medium was decanted and the XAD-7 was washed with water. The metabolites were extracted with MeOH and filtered (0.2 μm PTFE, VWR, Radnor, PA, USA) prior to analysis by HPLC (Shimadzu SCL-10Avp; Kyoto, Japan) with a reversed-phase column (SunFireTM C18 3.5 μm 2.1 x 150 mm, Waters, Milford, MA, USA) using a gradient from 15% acetonitrile in 0.1% formic acid to 100% acetonitrile.

Acknowledgements

We thank Dr. Jarmo Niemi, University of Turku, for DNA coding for SnogD. We acknowledge access to synchrotron radiation at the ESRF, Grenoble, France and at the EMBL Outstation, DESY, Hamburg, Germany and thank the beamline staff for helpful support. This work was supported by a grant from the Swedish Research Council and the Academy of Finland (grant no 136060).

References

1 Minotti G, Menna P, Salvatorelli E, Cairo G & Gianni L (2004) Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity. *Pharmacol Rev* **56**, 185-229.

2 Metsä-Ketelä M, Niemi J, Mäntsälä P & Schneider G (2008) Anthracycline Chemistry and Biology I: Biological Occurence and Biosynthesis, Synthesis and Chemistry. *Top Curr Chem* **282**, 101-140.

3 Weymouth-Wilson AC (1997) The role of carbohydrates in biologically active natural products. *Nat Prod Rep* **14**, 99-110.

4 La Ferla B, Airoldi C, Zona C, Orsato A, Cardona F, Merlo S, Sironi E, D'Orazio G & Nicotra F (2011) Natural glycoconjugates with antitumor activity. *Nat Prod Rep***28**, 630-648.

5 Thibodeaux CJ, Melançon CE & Liu H-wen (2008) Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew Chem* **47**, 9814-59.

6 Arora SK (1983) Molecular structure, absolute stereochemistry, and interactions of nogalamycin, a DNA-binding anthracycline antitumor antibiotic. *J Am Chem Soc* **105**, 1328-1332.

7 Trefzer A, Hoffmeister D, Künzel E, Stockert S, Weitnauer G, Westrich L, Rix U, Fuchser J, Bindseil KU, Rohr J & Bechthold A (2000) Function of glycosyltransferase genes involved in urdamycin A biosynthesis. *Chem Biol* **7**, 133-42.

8 Nakano H, Matsuda Y, Ito K, Ohkubo S, Morimoto M & Tomita F (1981) Gilvocarcins, new antitumor antibiotics. 1. Taxonomy, fermentation, isolation and biological activities. *J Antibiot* **34**, 266-70.

9 Schmitz H, Crook KE & Bush JA (1966) Hedamycin, a new antitumor antibiotic. I. Production, isolation, and characterization. *Antimicrob Agents Chemother* **6**, 606-12.

10 Ichinose K, Bedford DJ, Tornus D, Bechthold A, Bibb MJ, Revill WP, Floss HG & Hopwood DA (1998) The granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tü22: sequence analysis and expression in a heterologous host. *Chem Biol* **5**, 647-59.

11 Torkkell S, Kunnari T, Palmu K, Mäntsälä P, Hakala J & Ylihonko K (2001) The entire nogalamycin biosynthetic gene cluster of Streptomyces nogalater: characterization of a 20-kb DNA region and generation of hybrid structures. *Mol Genet Genomics* **266**, 276-288.

12 Sultana A, Kallio P, Jansson A, Wang J-S, Niemi J, Mäntsälä P & Schneider G (2004) Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J* **23**, 1911-21.

13 Beinker P, Lohkamp B, Peltonen T, Niemi J, Mäntsälä P & Schneider G (2006) Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *J Mol Biol* **359**, 728-40.

14 Grocholski T, Koskiniemi H, Lindqvist Y, Mäntsälä P, Niemi J & Schneider G (2010) Crystal structure of the cofactor-independent monooxygenase SnoaB from Streptomyces nogalater: implications for the reaction mechanism. *Biochemistry* **49**, 934-44.

15 Siitonen V, Claesson M, Patrikainen P, Aromaa M, Mäntsälä P, Schneider G & Metsä-Ketelä M (2012) Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin. *Chembiochem* **13**, 120-8.

16 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, D233-8.

17 Lairson LL, Henrissat B, Davies GJ & Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-55.

18 Holm L & Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-9.

19 Zhang C, Bitto E, Goff RD, Singh S, Bingman C A, Griffith BR, Albermann C, Phillips GN & Thorson JS (2008) Biochemical and structural insights of the early glycosylation steps in calicheamicin biosynthesis. *Chem Biol* **15**, 842-53.

20 Mittler M, Bechthold A & Schulz GE (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.

21 Isiorho EA, Liu H-W & Keatinge-Clay AT (2012) Structural Studies of the Spinosyn Rhamnosyltransferase, SpnG. *Biochemistry* **51**, 1213-22.

22 Chang A, Singh S, Helmich KE, Goff RD, Bingman CA, Thorson JS & Phillips GN (2011) Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity. *Proc Natl Acad Sci USA* **108**, 17649-54.

23 Moncrieffe MC, Fernandez M-J, Spiteller D, Matsumura H, Gay NJ, Luisi BF & Leadlay PF (2012) Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *J Mol Biol* **415**, 92-101.

24 Bolam DN, Roberts S, Proctor MR, Turkenburg JP, Dodson EJ, Martinez-Fleites C, Yang M, Davis BG, Davies GJ & Gilbert HJ (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc Natl Acad Sci USA* **104**, 5336-41.

25 Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97.

26 Mulichak A M, Losey HC, Walsh CT & Garavito RM (2001) Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* **9**, 547-57.

27 Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M & Walker S (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.

28 Mulichak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT & Garavito RM (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc Natl Acad Sci USA* **100**, 9238-43.

29 Offen W, Martinez-fleites C, Yang M, Kiat-lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ & Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J* **25**, 1396-1405.

30 Smith CK, Davies GJ, Dodson EJ & Moore MH (1995) DNA-nogalamycin interactions: the crystal structure of d(TGATCA) complexed with nogalamycin. *Biochemistry* **34**, 415-25.

31 Sim SP, Gatto B, Yu C, Liu A A, Li TK, Pilch DS, LaVoie EJ & Liu LF (1997) Differential poisoning of topoisomerases by menogaril and nogalamycin dictated by the minor groove-binding nogalose sugar. *Biochemistry* **36**, 13285-91.

32 Ha S, Gross B & Walker S (2001) E. Coli MurG: A Paradigm for a Superfamily of Glycosyltransferases. *Curr Drug Targets: Infect Disord* **1**, 201-213.

33 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 **18***, 5170-5180.

34 Aslanidis C & de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**, 6069-74.

35 Walter TS, Meier C, Assenberg R, Au K-F, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI & Grimes JM (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-22.

36 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta cryst D, Biolo Crystallogr* **67**, 235-42.

37 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

38 Vagin AA & Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Cryst D, Biol Crystallogr* **57**, 1451-6.

39 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Cryst D, Biol Crystallogr* **66**, 486-501.

40 Winn MD, Murshudov GN & Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374**, 300-21.

41 Larkin M a, Blackshields G, Brown NP, Chenna R, McGettigan P A, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8.

42 DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

43 Ward JM, Janssen GR, Kieser T, Bibb MJ, Buttner MJ & Bibb MJ (1986) Construction and characterisation of a series of multi-copy promoter-probe plasmid vectors for Streptomyces using the aminoglycoside phosphotransferase gene from Tn5 as indicator. *Molecular & Gen Genet* **203**, 468-478.

44 Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton J. C, Kieser HM, Lydiate DJ, Smith CP, Ward JM & Schrempf H (1986) *Genetic Manipulation of Streptomyces: A Laboratory Manual*

45 Gouet P, Courcelle E, Stuart DI & Métoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

46 Modolo LV, Li L, Pan H, Blount JW, Dixon R a & Wang X (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol* **392**, 1292-302.

Supporting information

The following supplementary information is available:

Fig. S1. Structure based sequence alignment of SnogD and homologues.

Fig. S2. Inter-domain linker reorganization upon hydrogen bonding to Asn212 of SnogD to deoxy-ribose C3´-hydroxyl of dUDP.

Figure legends

Figure 1. Glycosylation steps in nogalamycin biosynthesis. **1**, nogalamycin; **2**, nogalamycinone: **3**, 3´,4´-demethoxynogalose-1-hydroxynogalamycinone; SnogY, nogalose C2´ methyltransferase; SnogM and SnogL, C3´ and C4´ O-methyltransferases; SnoaL2, C1 hydroxylase; SnogA and SnogX, nogalamine C3´´ aminomethyltransferases; SnoN and SnoT putative nogalamine C2´´ hydroxylases.

Figure 2: Composite 2Fo-Fc simulated annealed omit map at the position of bound dUDP in chain B of SnogD-wtdUDP, shown in green. The omit electron density map was calculated with Phenix [37] and is contoured at 1.3σ. The refined 2Fo-Fc map is overlaid in blue contoured at 1.5σ.

Figure 3. Stereo view of the three-dimensional structure of SnogD. Chain A is shown coloured after secondary structure elements (red/yellow/green), chain B in blue with bound dUDP as a stick model. The flexible loops FL1 and FL2, likely to be involved in acceptor substrate binding are shown as dashed lines. The putative binding location of the acceptor-substrate is indicated by an asterisk.

Figure 4. Binding of dUDP to SnogD. Interactions of dUDP, shown as yellow sticks, with amino acids in the nucleotide binding pocket of SnogD. Putative hydrogen bonds (< 3.2 Å distance between donor and acceptor atoms) are indicated as dotted lines. For clarity the N-terminal domain of SnogD is not shown.

Figure 5. Structural changes upon nucleotide binding to SnogD illustrated by superimposition of apo-enzyme and binary complexes of SnogD with dUDP. Structures of SnogD with bound nucleotide, SnogD-wtdUDP (chain B) and SnogD-mdUDP (chain B), are shown in blue and green, respectively. Nucleotide-free chains of SnogD-mdUDP (chain C) and SnogD-m (chain B) in red and yellow. The loops n1 and n2, undergoing conformational changes are indicated. For clarity the N-terminal domain of SnogD is not shown.

Figure 6. A. Structure of ThDP-nogalamine B. Model of the SnogD Michaelis complex with the acceptor substrate **3** and the donor substrate TDP-nogalamine at the active site. The calculated

electrostatic surface potential is coloured blue to red, after increasing negative charge. The mutated histidine residues are highlighted.

Figure 7. Enzymatic assays of wild-type and active-site mutants of SnogD. A. Scheme of the deglycosylation of nogalamycin F **4** to 3´, 4´-demethoxynogalose-1-hydroxynogalamycinone **3**, used in the SnogD *in vitro* assay. B. HPLC chromatogram traces of the UDP dependent deglycosylation of **4** *in vitro*. The traces were recorded at 460 nm. The numbering of the compounds is as in Figure 7A. The absorbance (*A)* is shown on the Y-axis scale, in arbitrary units.

Figure 8: *In vivo* experiments with wild type SnogD and active site mutants. A. Structure of the bi-glycosylated SnogD reaction product nogalamycin R **5** produced during *in vivo* cultivations in *S. albus*. B. HPLC chromatogram traces of crude extracts of the anthracycline compounds produced during *in vivo* cultivations, recorded at 460 nm. Compound numbering as in Figure 7A. The absorbance (*A)* is shown on the Y-axis scale, in arbitrary units.

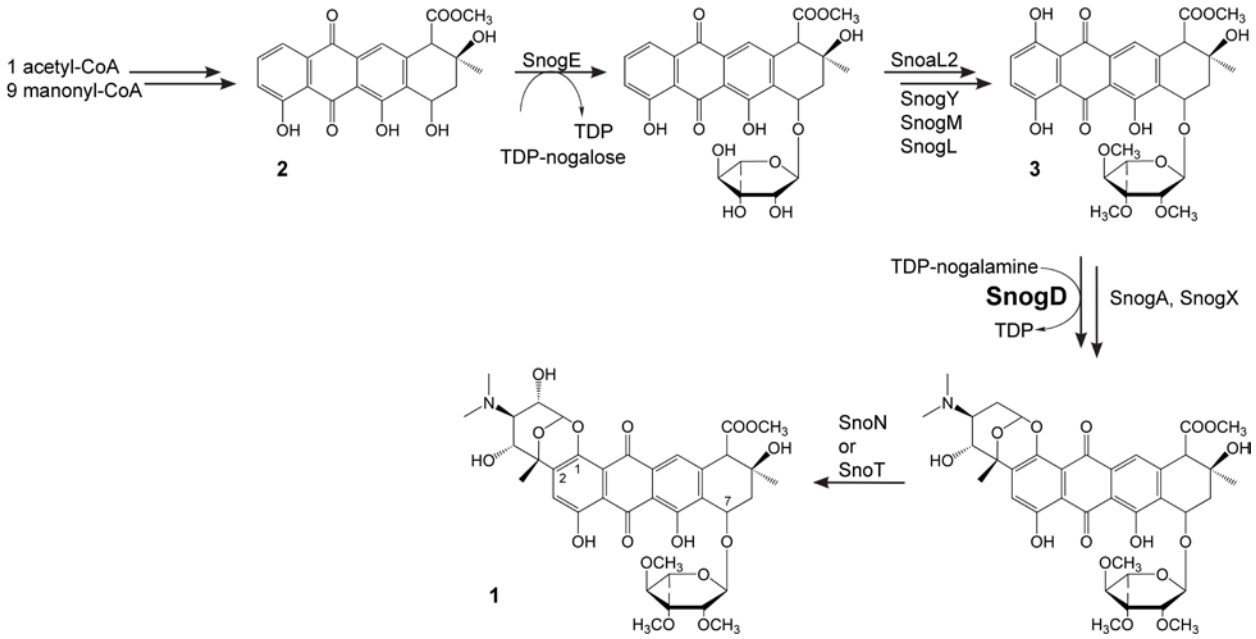
Table 1 – Statistics of data collection and refinement.

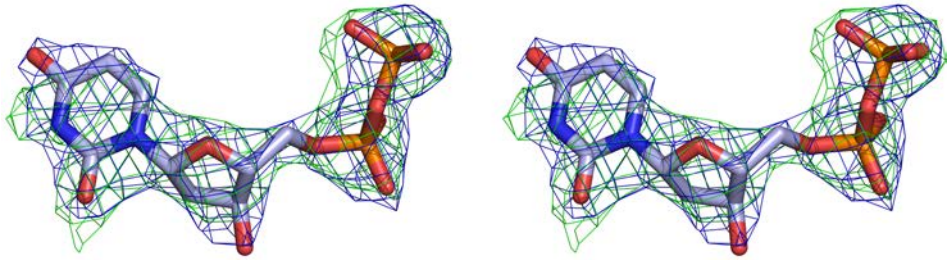
	SnogD-wtdUDP	**SnogD-mdUDP**	**SnogD-m**
Data collection [#]			
Beam line	ID14-EH1 (ESRF)	ID23-EH1 (ESRF)	ID23-EH1 (ESRF)
Wavelength (Å)	0.934	1.072	1.072
Space group	$P2_12_12$	P2	$P2_12_12$
Cell axes (Å)	64.9, 171.0, 69.4	64.6, 70.1, 176.0	66.7, 179.8, 70.2
Cell angles (deg.)	90.0, 90.0, 90.0	90.0, 91.7, 90.0	90.0, 90.0, 90.0
Resolution (Å)	60-2.6 (2.76-2.62)	58.6-2.7 (2.85-2.70)	48.3-2.6 (2.73-2.59)
R_{sym} (%)	13.7 (45.9)	7.0 (50.2)	5.6 (35.1)
I/Iσ	7.9 (2.0)	17.7 (2.2)	15.2 (2.0)
Completeness (%)	97.6 (86.7)	98.3 (92.7)	96.2 (82.4)
Multiplicity	3.0 (2.3)	4.5 (3.7)	3.0 (2.1)
No. of reflections	69886 (6703)	192976 (21392)	77517 (6780)
No. of unique reflections	23310 (2915)	42845 (5800)	26051 (3158)
Wilson B-factor ($Å^2$)	43.0	81.3	66.9
Refinement			
Resolution	50.0-2.6	60.0-2.7	60.0-2.6
R/R_{free}	22.3 / 29.7	22.8 / 26.6	22.1 / 24.3
No. of non H-atoms / average B-factor ($Å^2$)			
Protein	5450 / 33.1	10233 / 82.0	5370 / 54.5
Water	115 / 24.5	48 / 43.4	67 / 37.3
Ligands (dUDP)	24 / 37.6	48 / 84.2	- / -
R.m.s.d. bond lengths (Å)	0.011	0.009	0.010
R.m.s.d. bond angles (°)	1.5	1.4	1.3
Ramachandran plot			
Residues in favoured regions (%)	96.2	96.3	97.2
Residues in allowed regions (%)	3.7	3.4	2.7
Residues in disallowed regions (%)	0.1	0.3	0.1

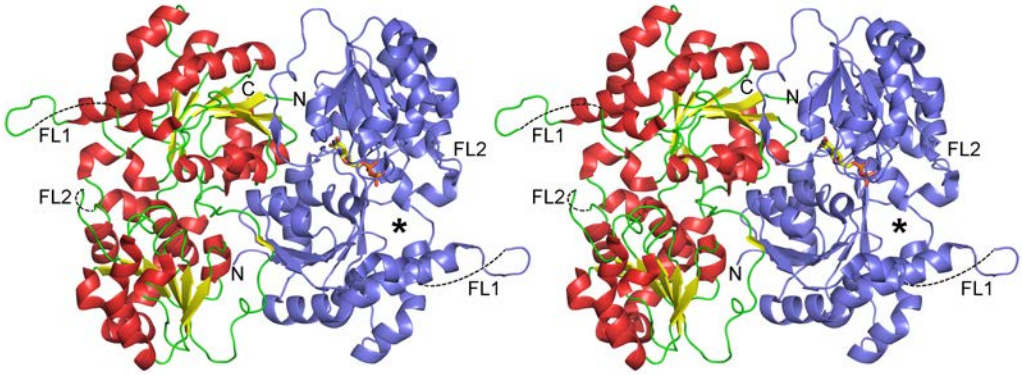
[#]Numbers in parentheses are for the highest resolution shell

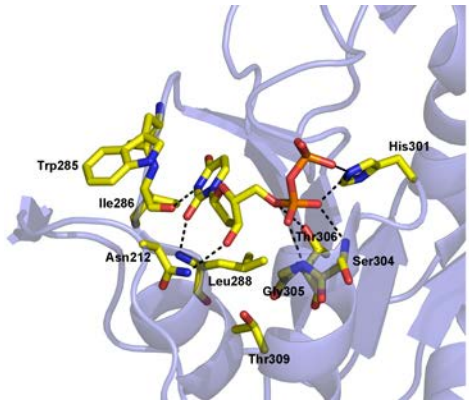
Table 2 – Relative *in vitro* activities of SnogD mutants

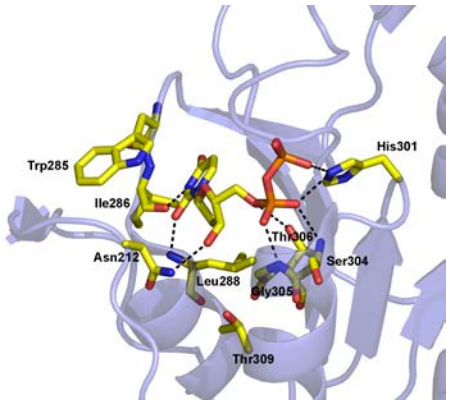
	Relative activity of triplicates	+/- SD
No enzyme	1.1%	1.7%
No UDP	0.2%	0.1%
His25Ala	1.5%	0.6%
His25Asn	1.9%	1.3%
His301Ala	2.4%	1.6%
WT	100%	5.8%

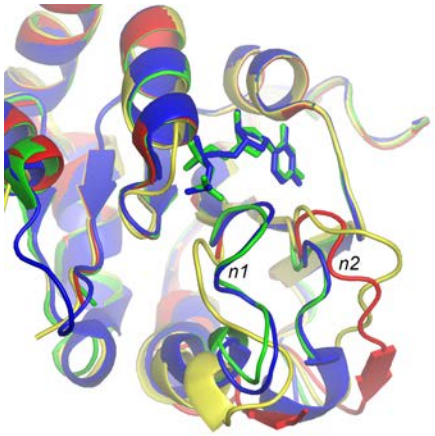

1 acetyl-CoA
9 manonyl-CoA
2
SnogE
TDP
TDP-nogalose
SnoaL2
SnogY
SnogM
SnogL
3
TDP-nogalamine
SnogD
TDP
SnogA, SnogX
SnoN
or
SnoT
1

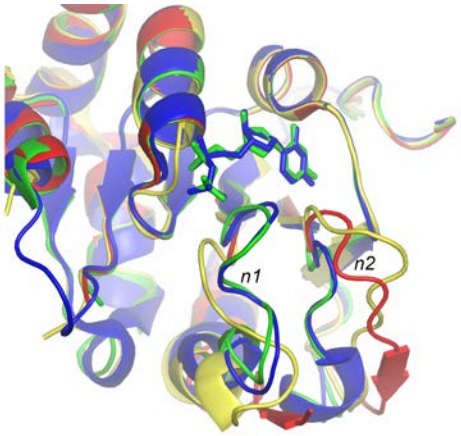



C
N
FL1
FL2
FL2
N
FL1
C
N
FL1
FL2
FL2
N
FL1

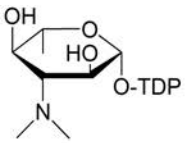

Trp285
His301
Ile286
Thr306
Asn212
Ser304
Leu288
Gly305
Thr309


Trp285
His301
Ile286
Thr306
Asn212
Ser304
Leu288
Gly305
Thr309

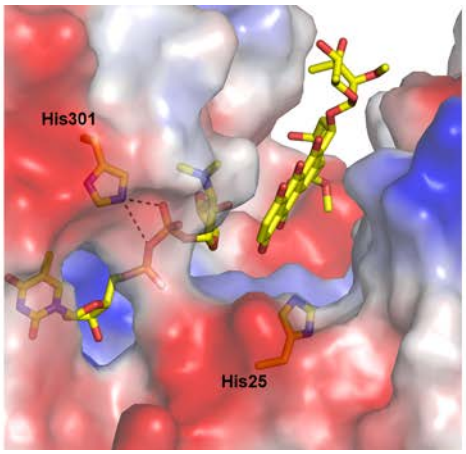

n1
n2

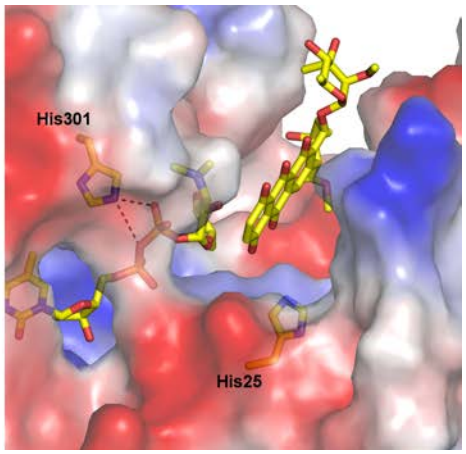

n1
n2

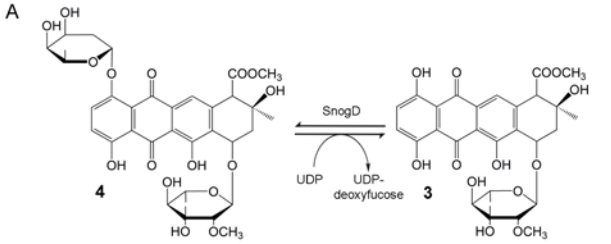
A

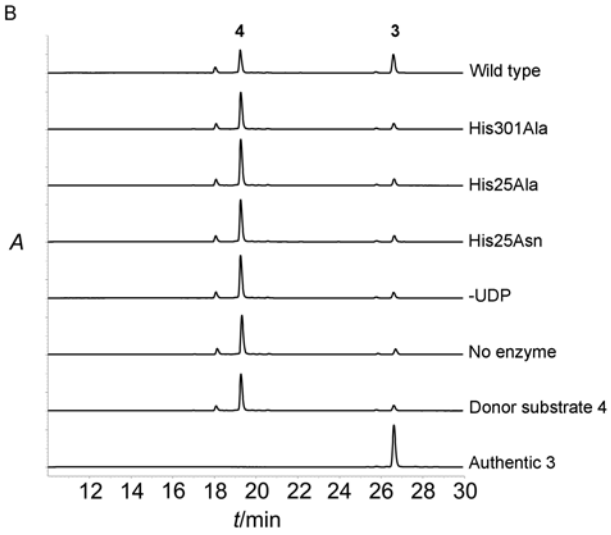

OH
O
HO
O-TDP
N

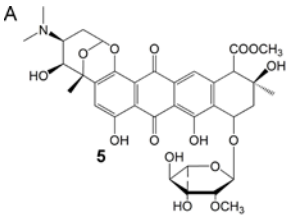
B

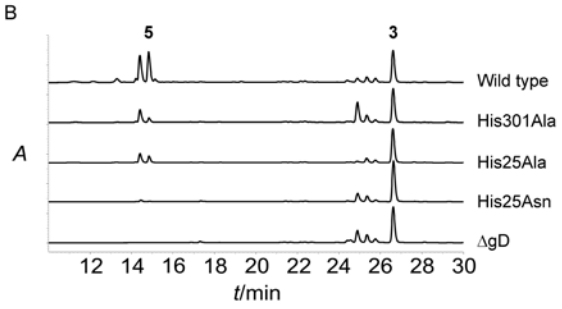

His301
His25


His301
His25


A
SnogD
UDP
UDP-deoxyfucose
4
3


B
4
3
Wild type
His301Ala
His25Ala
His25Asn
-UDP
No enzyme
Donor substrate 4
Authentic 3
A
12 14 16 18 20 22 24 26 28 30
t/min


A
5


B
5
3
Wild type
His301Ala
His25Ala
His25Asn
ΔgD
A
12 14 16 18 20 22 24 26 28 30
t/min

Supplementary information

Figure S1. Structure based sequence alignment of SnogD and homologues. The active site histidine residues mutated in this study are highlighted by black boxes and the PPi motif in a green box. Invariant residues are shown in red boxes, conserved residues are shown in red. The figure was prepared with ESPript [45].

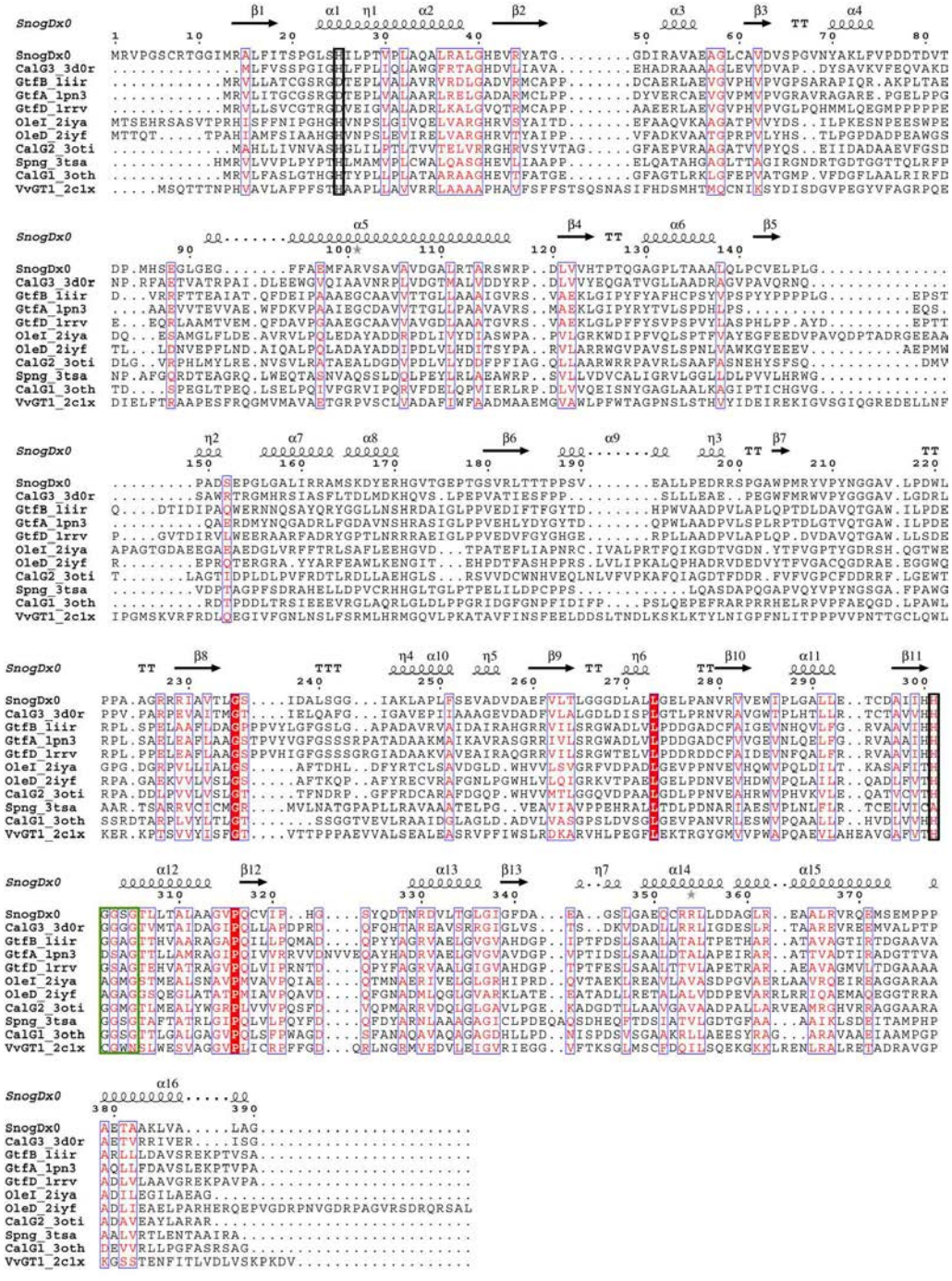
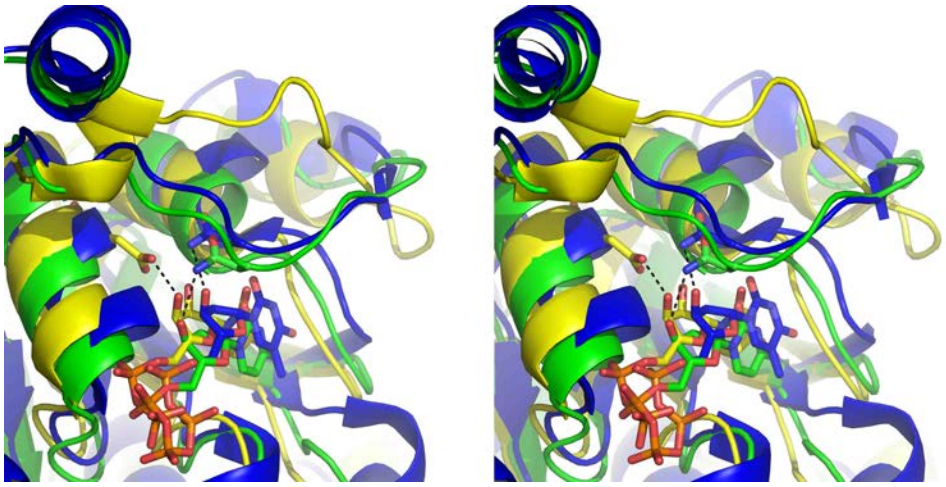


Figure S2. Inter-domain linker reorganization upon hydrogen bonding of Asn212 of SnogD to deoxy-ribose C3´-hydroxyl of dUDP. Protein chains are shown as cartoons with SnogD (blue), SpnG (green), and UGT78G1 (yellow) [46]. The bound NDPs are shown as sticks with colours matching the protein chains. Hydrogen bonds to the C3´-hydroxyl are shown as dashed lines. For clarity the respective N-terminal domains are not shown.



References

45 Gouet P, Courcelle E, Stuart DI & Métoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

46 Modolo LV, Li L, Pan H, Blount JW, Dixon R a & Wang X (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol* **392**, 1292-302.