Chemistry & Biology 13, 1143–1152, November 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.chembiol.2006.09.005

Insights into the Synthesis of Lipopolysaccharide and Antibiotics through the Structures of Two Retaining Glycosyltransferases from Family GT4

Carlos Martinez-Fleites,[1] Mark Proctor,[2] Shirley Roberts,[1] David N. Bolam,[2] Harry J. Gilbert,[2] and Gideon J. Davies[1,*]
[1]York Structural Biology Laboratory
Department of Chemistry
University of York
York, YO10 5YW
United Kingdom
[2]Institute for Cell and Molecular Biosciences
Medical School, University of Newcastle upon Tyne
Framlington Place
Newcastle upon Tyne, NE2 4HH
United Kingdom

Summary

Glycosyltransferases (GTs) catalyze the synthesis of the myriad glycoconjugates that are central to life. One of the largest families is GT4, which contains several enzymes of therapeutic significance, exemplified by WaaG and AviGT4. WaaG catalyses a key step in lipopolysaccharide synthesis, while AviGT4, produced by *Streptomyces viridochromogenes*, contributes to the synthesis of the antibiotic avilamycin A. Here we present the crystal structure of both WaaG and AviGT4. The two enzymes contain two "Rossmann-like" (β/α/β) domains characteristic of the GT-B fold. Both recognition of the donor substrate and the catalytic machinery is similar to other retaining GTs that display the GT-B fold. Structural information is discussed with respect to the evolution of GTs and the therapeutic significance of the two enzymes.

Introduction

Glycosyltransferases (GTs) are central to life. The synthesis of the glycosidic bond impacts not only on the formation of structural and food storage polysaccharides, but also on the metabolic orchestration of small molecules and the formation of numerous glycolipids, glycopeptides, and lipopolysaccharides (LPSs). GTs harness activated sugar donors, typically nucleotide or lipid-phospho sugars, to drive glycosyl transfer to the appropriate acceptor. GT genes represent approximately 1%–2% of all coding regions of genomes, and there are around 20,000 open-reading frames predicted as GTs (by virtue of similarity to characterized enzymes) known to date [1, 2]. These sequences are currently classified into 87 distinct sequence families (CAZy; carbohydrate-active enzymes database at http://www.cazy.org/). Of these families, just 21 (GT1, 2, 5–9, 13, 15, 20, 27, 28, 35, 42–44, 63, 64, 72, 78, and 80) currently have a known three-dimensional structure. In contrast to glycoside hydrolases, which display a myriad of topologies, GT structures have, thus far, revealed just two folds (and variants thereof), termed GT-A and GT-B [2–4].

*Correspondence: davies@ysbl.york.ac.uk

GTs specifically catalyze a reaction of defined stereochemistry; in other words, the configuration of the activated donor may be inverted or retained, leading to the terms "inverting" and "retaining" enzymes. Currently, there is no correlation between the overall fold of the GT and its mechanism of action: both inverting and retaining enzymes of both fold types are known. Furthermore, most eukaryotic and bacterial genomes display a wide diversity of GTs from many of the sequence-derived families. *Homo sapiens*, for example, have enzymes that sparsely populate 42 of these families. Whereas most organisms marshal a similarly diverse array of GTs, the sequences of Archael genomes, over 20 of which are now known, reveal a different picture. Primitive Archae often display GTs from just two sequence-based families: GT2 and GT4. GT2 enzymes form a family of inverting enzymes, while all (currently characterized) GT4 enzymes are retaining. Thus, it seems likely that GT2 and GT4 are the ancestral inverting and retaining families from which enzymes of both stereochemistries have evolved. While GT2 enzymes display the GT-A fold [5], it was unconfirmed whether GT4 GTs adopt the GT-B fold.

GT4 is one of the largest of all the GT families, with over 3700 members. This family contains examples of enzymes that not only utilize nucleotide-sugar donors, but also simple phospho-sugars and lipid-phospho sugar donors, perhaps hinting at the ancient origin of the family. Family GT-4 functions appear as diverse as their sequences, reflecting differences in both donor and acceptor specificity. Several enzymes in this family are of potential therapeutic significance, exemplified here by WaaG and AviGT4. WaaG is an α-1,3 glucosyltransferase, which transfers a glucose from UDP-Glc onto L-*glycero*-D-*manno*-heptose II, contributing to the core structure of LPSs [6, 7], while Bechtold's outstanding work has shown that AviGT4 catalyses the attachment of an as-yet undefined precursor of the eurekanate sugar moiety to L-lyxose [8] during the synthesis of the orthosomycin antibiotic avilamycin A [9], respectively (Figure 1). Although both located in family GT4, WaaG and AviGT4 display just 16% sequence identity.

Here, we present the 3D structures, determined by X-ray crystallography, of the first two family GT4 members, WaaG and AviGT4. Structural approaches provide insight into these two systems, where lack of substrate availability, or even identification, precludes kinetic approaches. We show that both enzymes exhibit a GT-B fold, consistent with the hypothesis that GT2 and GT4 are progenitors of the numerous GT families, while structure-based sequence alignments of the C-terminal domains of diverse GT-B enzymes suggest a structural basis for the evolution of enzyme mechanism.

Results and Discussion

Initial attempts to produce AviGT4 and WaaG generated insoluble recombinant protein. The inclusion of Triton X-100 in the cell lysis buffer yielded soluble forms of both GTs, and the proteins remained soluble despite

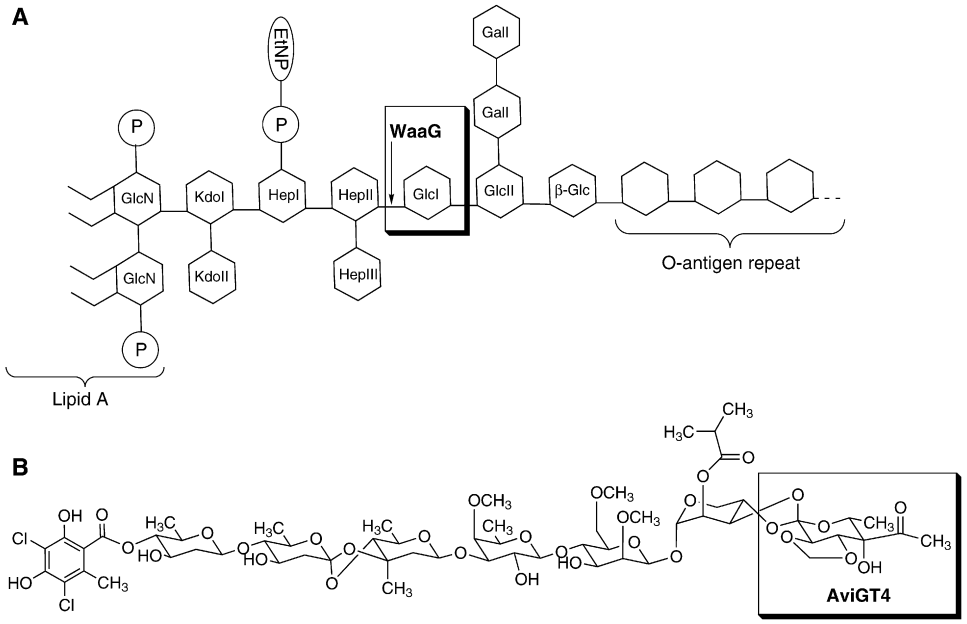

Figure 1. The Catalytic Action of Two Family GT4 GTs

(A) Schematic diagram of *E. coli* LPS. GlcN, D-glucosamine; Hep, L-*glycero*-D-*manno*-heptose; P, phosphate; EtNP, 2-aminoethyl phosphate; Glc, D-glucose; Gal, D-galactose. WaaG is responsible for the addition of the first glucose moiety via an α-1,3 glycosidic linkage to Hep II. (B) Structure of the antibiotic avilamycin A. Disruption of the *avigt4* gene results in a product lacking the eurekanate moiety (boxed) normally bonded to the L-lyxose residue [8].

the exclusion of the detergent from the subsequent buffers used in the purification schemes. Thus, in the absence of detergent, the two GTs appear to bind to hydrophobic particulate matter present in *Escherichia coli* cells; however, once this material had been removed when the recombinant enzymes were bound to the column matrix, the inclusion of Triton X-100 was no longer required to maintain the two proteins in soluble form. It is generally recognized that the production of soluble recombinant GTs is problematic, and the capacity of these proteins to bind to hydrophobic structures in *E. coli* may contribute to the general insolubility of recombinant GTs expressed in *E. coli*.

GT4 Glycosytransferases Display the GT-B Fold

The structure of WaaG was solved by single-wavelength anomalous dispersion (SAD) with a selenomethionine (SeMet) version of the protein, at a resolution of 1.6 Å (Table 1), in the presence of UDP, while the crystal structure of AviGT4 was determined by multiple-wavelength anomalous dispersion (MAD) methods, also employing SeMet protein to a resolution of 2.1 Å (Table 1). The electron density allows for the modeling of the WaaG sequence as a continuous chain from residue Ile2 to Gly371 (Figure 2). The WaaG sequence folds into two "Rossmann-like" (β/α/β) domains, with a similar topology of other previously described GTs displaying the GT-B fold. The N-terminal domain comprises residues Ile2–Gly168 and displays seven parallel β strands flanked by five α helices on both sides of a characteristically twisted β sheet. The C-terminal domain (residues Ile169–Gly371) displays six parallel β strands flanked by six α helices, and terminates with the recurrent pattern of a kinked α helix (Ser338–Gly371) that crosses over to the N-terminal domain (residues Ser359–Gly371), adding an extra α helix to the N-terminal Rossmann-like fold.

Crystals of AviGT4 contain two copies of the protein in the asymmetric unit, interacting over an area of 3750 Å^2 per chain, which have both been modeled from residues Pro13–Lys352. AviGT4 also displays two Rossmann-like (β/α/β) domains, and thus also adopts the GT-B fold. The N-terminal domain comprises residues Pro13–Pro156, and its core is composed of six parallel β strands alternating with three connecting α helices (Figure 2B). The C-terminal domain (Val157–Lys352) is composed of a central parallel β sheet of six strands flanked by seven α helices. As with WaaG and all known GT-B enzymes (reviewed in [10]), the last C-terminal helix is kinked at residue Trp329, located in the interface between the two domains, and the rest of the helix crosses to the N-terminal domain (residues His331–Lys352).

Structural homology searches (DALI [11]) reveal that the closest, currently deposited, structural neighbor to WaaG structure is the retaining GT-20, OtsA (the trehalose-6-phosphate synthase whose *E. coli*-derived structure is known; DALI Z-score, 25.6), with an rmsd of 3.4 Å and with which WaaG shares 12% sequence identity. The *Agrobacterium tumefaciens* glycogen synthase, GagA, also retaining, from family GT5 appears second (DALI Z-score, 21.4; 13% sequence identity). Interestingly, the next closest structural match is not with a GT but with a UDP-sugar epimerase (*E. coli* UDP-N-acetylglucosamine 2-epimerase), with a DALI Z-score of 20.4 and 12% sequence identity. Homology searches for AviGT4 reveal similar results; the closest structural homologs represented by the *A. tumefaciens* glycogen synthase (DALI Z-score, 22.5) and GT20 OtsA (DALI Z-score, 22.4). A family GT-35 retaining glycogen phosphorylase from *Oryctolagus cuniculus* (rabbit) appeared

Table 1. Data Collection and Refinement Statistics for AviGT4 and the GT4 Enzyme WaaG

	AviGT4 (SeMet)	AviGT4 + UDP-Glc	WaaG (SeMet) + UDP	WaaG + UDP-2F-Glc
Data Collection				
Space group	$P2_1$	P1	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	47.5, 73.8, 92.1	51.5, 74.4, 90.6	50.7, 88.7, 89.6	50.7, 88.3, 89.3
α, β, γ (°)	90.0, 90.6, 90.0	89.9, 92.7,100.7	90.0, 90.0, 90.0	90, 90, 90
	Peak[a]		Peak	
Wavelength (Å)	0.97770	0.98000	0.97950	0.93300
Resolution (Å)	40.00–2.10	90.53–2.30	33.6–1.60	40.67–1.50
R_{merge}	0.064 (0.183)	0.089 (0.324)	0.058 (0.196)	0.126 (0.283)
$I/\sigma I$	24.7 (8.1)	9.8 (2.6)	22.1 (8.6)	13.0 (2.2)
Completeness (%)	99.7 (98.1)	96.9 (92.3)	100 (100)	99.0 (93.6)
Redundancy	7.2 (6.4)	2.5 (2.4)	7.2 (7.3)	4.3 (2.6)
Refinement				
Resolution (Å)	20.00–2.10	20.00–2.30	20.00–1.60	20.00–1.50
No. reflections	35,293	53,958	51,242	61,003
R_{work}/R_{free}	0.17/0.23	0.20/0.26	0.17/0.20	0.18/0.22
B factors				
Protein	17	35	12	10
/UDP/UDP-2FGlc	No ligand	44	7	7
Water	23	36	29	24
Rms deviations				
Bond lengths (Å)	0.017	0.019	0.009	0.011
Bond angles (°)	1.597	1.863	1.192	1.945

[a] Although the structure was solved by MAD, only the peak data statistics used for refinement are given.

immediately below, with a DALI Z-score of 15, and, more distantly, GtfB, an antibiotic modifying, but inverting, GT from family GT1 (DALI Z-score, 12.8).

Donor Substrate Recognition by AviGT4 and WaaG

The structure of AviGT4 was solved in complex with UDP-Glc at a resolution of 2.3 Å, and that of WaaG solved

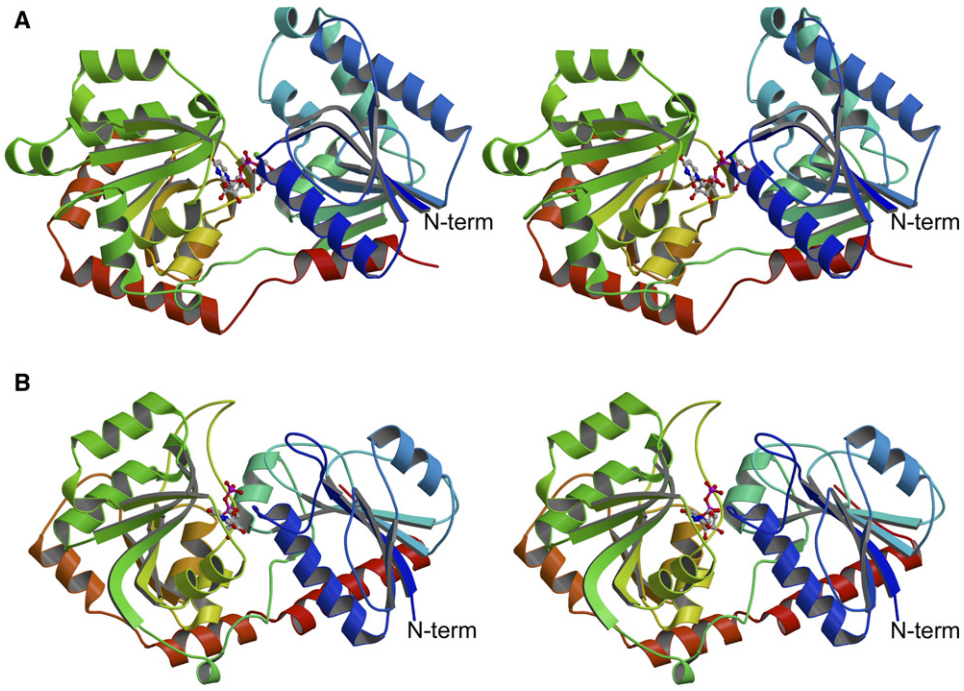

Figure 2. Three-Dimensional Structures of Two Family GT4 GTs

(A) Divergent (wall-eyed) stereo cartoon of the UDP-2FGlc complex of the GT4 enzyme WaaG from *E. coli*. The structure is "color-ramped" from N terminus (blue) to C terminus (red).

(B) Divergent (wall-eyed) stereo cartoon of the UDP complex of the GT4 enzyme AviGT4 from *Streptomyces viridochromogenes* Tü57.

This figure was drawn with MOLSCRIPT [42].

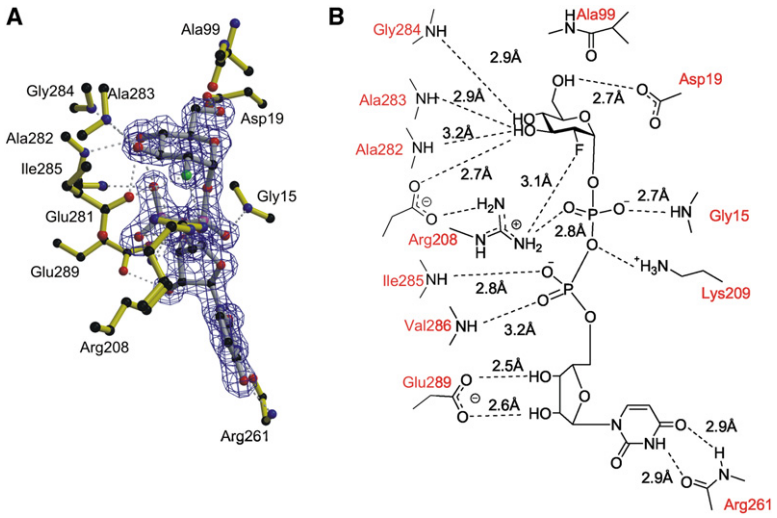

Figure 3. Donor Site Interactions of *E. coli* GT4 WaaG

(A) Observed electron density (calculated with the coefficients $2F_{obs} - F_{calc}$ and α_{calc}) for the UDP-2F-glucose complex of *E. coli* WaaG (drawn with BOBSCRIPT [43]).

(B) Schematic diagram of the interactions of WaaG with UDP-2F-glucose.

in complex with UDP-2-deoxy-2-fluoro glucose (UDP-2FGlc; synthesis described in [12]) at a resolution of 1.5 Å, providing insight into nucleotide-sugar donor binding (Figures 3 and 4). The interactions of the uracil and ribose moieties were as observed previously for GT-B clan retaining GTs, exemplified by GT20 trehalose-6-phosphate synthase OtsA [12, 13]. Uracil recognition is conferred by hydrogen bonds to the main chain carbonyl and amide of Arg261, by O7 and N3, respectively. Ribose binding is conferred by Glu289, which is a signature of the GT-B fold UDP-dependent enzymes (e.g., [10, 13–18]).

In the donor sugar site itself, while there are many similarities with the recognition of glucose in diverse GT-B enzymes, such as glycogen and maltodextrin phosphorylases [19, 20], OtsA [12, 13], and glycogen synthase [21], there are also significant differences that had not previously been predicted. Recognition of Glc O3 and O4 by two main-chain amide hydrogens and a carboxylate is again seen with WaaG; but here the carboxylate moiety is donated by a glutamate (Glu281) in place of the aspartate observed in OtsA. In 2001, Wrabl and Grisin proposed a sequence-derived superfamily of GT-B fold GTs, which they termed the "glycogen phosphorylase GT" (GPGTF) superfamily [18]. In the GPGTF superfamily, a carboxylate at this position, described above, is indeed considered the second signature motif of the superfamily. Most intriguingly, the 3D structure reveals that WaaG provides a different environment around O6 to all enzymes described previously. In glycogen and maltodextrin phosphorylases (enzymes that are mechanistically and structurally related to GT-B fold GTs), OtsA and glycogen synthase O6 interacts with a conserved histidine side chain, the main-chain carbonyl of which

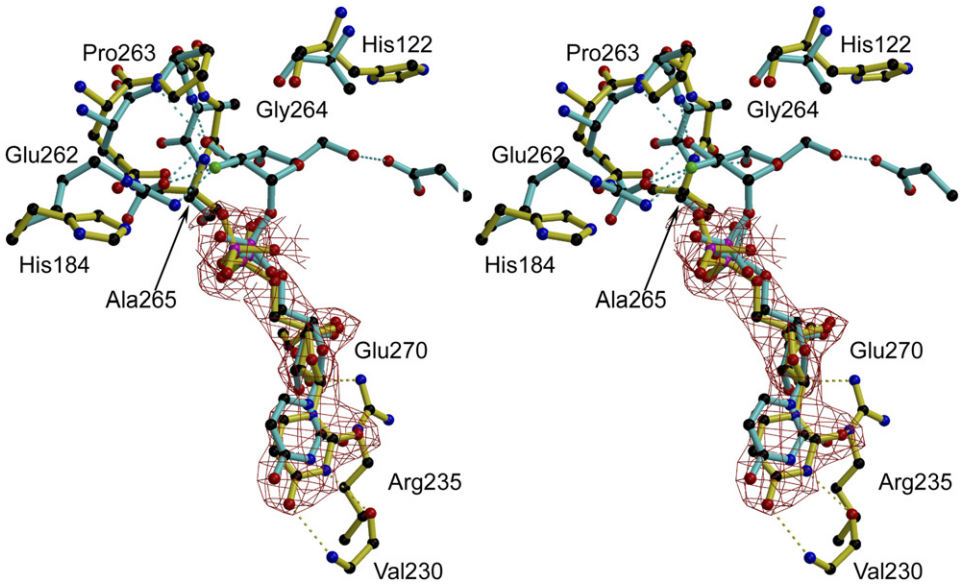

Figure 4. Overlay of the Active Centers of GT4 Enzymes AviGT4 and WaaG

The UDP-2FGlc complex of WaaG is shown in cyan with AviGT4 in yellow. Electron density for the ordered UDP portion of the UDP-Glc in complex with AviGT4 is shown in red and is an electron density map calculated with the coefficients $2F_{obs} - F_{calc}$ and α_{calc} and contoured at 1.2σ. AviGT4 residues discussed in the text are shown labeled. This figure was drawn with BOBSCRIPT [43].

sits "above" C1, and is implicated in catalysis, perhaps stabilizing the incipient oxocarbenium-ion-like transition state [22] (and for review, see [23]). In WaaG, an alanine is present in this position and, while its main-chain carbonyl is again poised above C1, the absence of the side-chain imidazole allows Asp19 to bind O6 in its place. WaaG also provides a different environment around O2 (here present as a fluorine substituent of the 2-deoxy-2-fluoro sugar) and around the β-phosphate of UDP. Unique to WaaG is an arginine residue, Arg208, which interacts with the superfamily invariant carboxylate of Glu281 and with both F2 and O2 of the β-phosphate.

The orthosomycin class of antibiotics, which are effective against many gram-positive bacteria, including emerging pathogens such as glycopeptide-resistant enterococci, methicillin-resistant staphylococci, and penicillin-resistant pneumococci [24], consist of a dichloroisoeverninic acid moiety and a heptasaccharide side chain (Figure 1) [25]. AviGT4 catalyzes the attachment of an as-yet undefined (both in terms of nucleotide and sugar) precursor of the eurekanate sugar moiety to L-lyxose (Figure 1B) [8]. Furthermore, as the eurekanate moiety displays a double glycosidic linkage to the lyxose, there also must be chemical events subsequent to glycosyl transfer; indeed in the absence of structural information, it is unclear as to which bond is formed by the AviGT4 catalyzed step. The crystal structure of the enzyme implicates UDP as a potential dinucleotide component of the donor substrate. Indeed, the interactions between AviGT4 and the uracil base and ribose sugar of the dinucleotide are conserved in retaining GTs that display the GT-B fold (Figure 4). Thus, uracil recognition is primarily conferred by hydrogen bonds between O7 and N3, respectively, and the main-chain carbonyl and amide of Val230, while O2 and O3 of the ribose interact with the carboxylate side chain of Glu270. The guanidino side chain of Arg235 also appears to make weak hydrogen bonds with O8 of the uracil and O2 of the ribose. The NH2 moiety of Arg235 of AviGT4 makes a 2.9 Å H bond to O2 of ribose. Chemically, this mirrors a similar interaction in WaaG with Arg173, but the latter is derived from a completely different secondary structural element. The α-phosphate of UDP makes weak interactions with the backbone amides of Thr266 and Val267, while the β-phosphate hydrogen bonds to the side chain of Tyr39. Finally, the two faces of the uracil ring make hydrophobic interactions with the side chain of Tyr39 and Met178.

Although the true nature of the donor sugar is unknown, AviGT4 is demonstrably the first retaining GT-B fold structure reported whose donor sugar is not glucose. Based upon the overlap of the WaaG UDP-2FGlc complex with AviGT4, it is clear that, at the 6 position, where the AviGT4 donor is a 6-deoxy D-sugar, the environment is entirely consistent with a deoxy species in which Val45 and Ile155 flank the C6 position (Figure 4). Conversely, the histidine residue implicated in binding O6 in all GT-B enzymes, apart from WaaG, is also present in AviGT4. Although it is likely that AviGT4 harnesses a 6-deoxy species, the presence of the highly conserved histidine leaves this open to speculation. It is possible that AviGT4 may only accommodate C6-deoxy sugars, and the retention of the conserved histidine may simply be a remnant of its role in stabilizing the oxocarbenium transition state formed during the course of the reaction.

AviGT4 contains a number of residues whose interactions are likely invariant with respect to WaaG. These include Glu262, which interacts with O3 and the main-chain amides of Pro263, Gly264, and Ala265, which correspond to Ala282, Ala283, and Gly284 of WaaG and which are thus implicated in the recognition of O2 and O3. Together, these observations imply that the donor sugar for AviGT4 may be in *gluco* configuration, consistent with the equatorial orientations of these groups in the eurekanate moiety. The mechanism by which AviGT4 interacts with the acetyl group appended to C4 and of the *galacto*-configured O4 is unclear. Indeed, as the nature of the sugar is unknown, it is possible that the donor sugars lack some of the eurekanate-specific signatures seen in the final product. It is interesting to note, however, that in a complex of AviGT4 with UDP-Glc, the glucosyl moiety of UDP-Glc is located in the acceptor ("+1" nomenclature adapted from [26]) subsite rather than the donor sugar binding site of the *Streptomyces* enzyme, indicating that the glycon transferred by AviGT4 is likely quite distinct from glucose.

The Acceptor Sites of WaaG and AviGT4 Show Remarkable Diversity

Inspection of WaaG and AviGT4 shows small changes in the donor site, consistent with the nature of the donor sugar, but much more substantial changes in the acceptor site, reflecting the very different nature of the acceptor species. This is reflected in the electrostatic surfaces (Figure 5), which show a very positively charged surface for WaaG, reflecting its interaction with the negatively charged LPS acceptor, yet a very negatively charged surface for AviGT4; the reason for which is not immediately obvious. The acceptor binding cavity of WaaG is a deep and open basin in which the glucosyl moiety of the donor is found covering the deepest point of its floor. This pocket encloses an area of about 940 $Å^2$, and its shape and volume suggest that it can accommodate the first two heptose moieties of the acceptor LPS, including, if necessary, the heptosyl ramification on the first of these residues. The walls of the cavity are lined by four tyrosine residues and with various lysine and arginine residues pointing directly to the cavity, in line with the generic signature of carbohydrate binding sites. The acceptor binding cavity of AviGT4 is also a depression with a wide aperture, but it is shallower than the WaaG binding surface. This cavity encloses an accessible surface area of approximately 550 $Å^2$. The elongated shape of its opening is reminiscent of the nonbranched nature of its acceptor, and its dimensions suggest that perhaps only the first two residues of the avilamycin acceptor are recognized. These distinct acceptor binding architectures, in stark contrast with, for example, the narrow cavity that accommodates a single sugar of its closest structural neighbor, OtsA from family GT-20, expand the current view of the amazing plasticity that the GT-B fold can support. These structural variations are found mainly expressed as alterations in the lengths and orientation of the connecting α helices of the Rossmann fold in the N-terminal domain.

Catalytic Mechanism of WaaG and AviGT4

The catalytic mechanism of retaining GTs remains elusive (reviewed in [4, 23, 27]). In the absence of a more

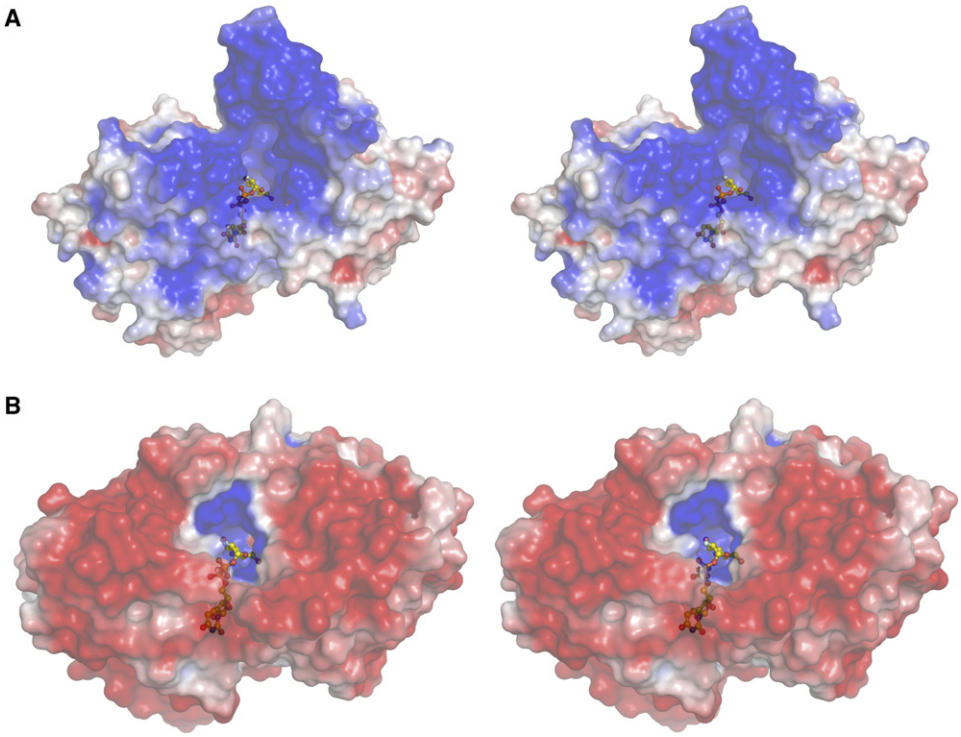

Figure 5. Electrostatic Surface Figures of WaaG and AviGT4

Surface representation of (A) WaaG and (B) AviGT4 colored by electrostatic potential (red, −3kT; blue +3kT, where k is the Boltzmann constant and T is temperature), calculated by the Adaptive Poisson-Boltzmann Solver (APBS) program [44] and visualized with Pymol (DeLano Scientific LLC, http://pymol.sourceforge.net/); ligands are shown in licorice representation.

convincing mechanism, most authors have converged on a mechanism involving "front-face" chemistry; an S_Ni-like reaction in which leaving-group departure and nucleophilic attack occur in a concerted, but necessarily stepwise, manner on the same face of the glycoside (see [13, 28], for example). In practice, this mechanism is favored primarily because of the lack of evidence (such as the presence of a conserved potential nucleophile that could attack C1 from the "β-face" of the sugar in a classic double displacement retaining reaction) supporting any alternative mechanism. The structure of WaaG is consistent with these proposals, for it displays the signature features of the GT-B fold retaining enzymes, notably a carbonyl moiety "above" C1, believed to stabilize the developing positive charge on the transition state [22] and the environment around the β-phosphate, which favors its interaction with the incoming nucleophile.

The mechanism by which AviGT4 catalyzes glycosyl transfer is less clear, as the eurekanate displays both α and β glycosidic bonds with L-lyxose, and, thus, could mediate the first glycosyl transfer via an inverting or retaining mechanism. It is generally believed that, as with glycoside hydrolase families, the mechanism of glycosyl transfer is conserved in the members of the same GT family, suggesting that AviGT4 is a retaining GT. This view is strongly supported by the topology and nature of its active site, which displays the signature features of a retaining enzyme, such as the conserved carbonyl moiety above C1 and the carboxylate interaction around O3. The mechanism by which the second glycosidic linkage is made is unknown; however, one can speculate that hydride abstraction from C1 would be required to activate this center for attack, but the enzyme(s) responsible have not been identified. Similarly, the double substituent at C4 might favor nucleophilic attack to C4, on a species already containing the ketone moiety, but, again, the synthetic pathway remains speculative.

Comparison with Other GT Families

The CAZY classification of protein sequences into families is organized on the basis of sequence similarity to one or more experimentally characterized founding family members. With the improvement of sequence detection methods, and with solved 3D structures, more distant relationships can be detected among different families. These interfamily relationships allow for the implementation of a hierarchical organization of families into clans and folds [2]. For GTs, only two different folds (and minor variations thereof), the GT-A and GT-B fold, are currently described for the 87 sequence families, but there is currently no correlation of fold with mechanism; nor is there a robust way of ascertaining the mechanism of catalysis from sequence alone, especially since the sequence similarities between families are extremely low. Three-dimensional structure comparison, through experimentally determined 3D structure, is arguably the most powerful method of inferring protein structure-function relatedness, especially when sequence identities are very low and sequence-based comparison methods are at the border of significance. Through the structural data for two members of the GT4 family,

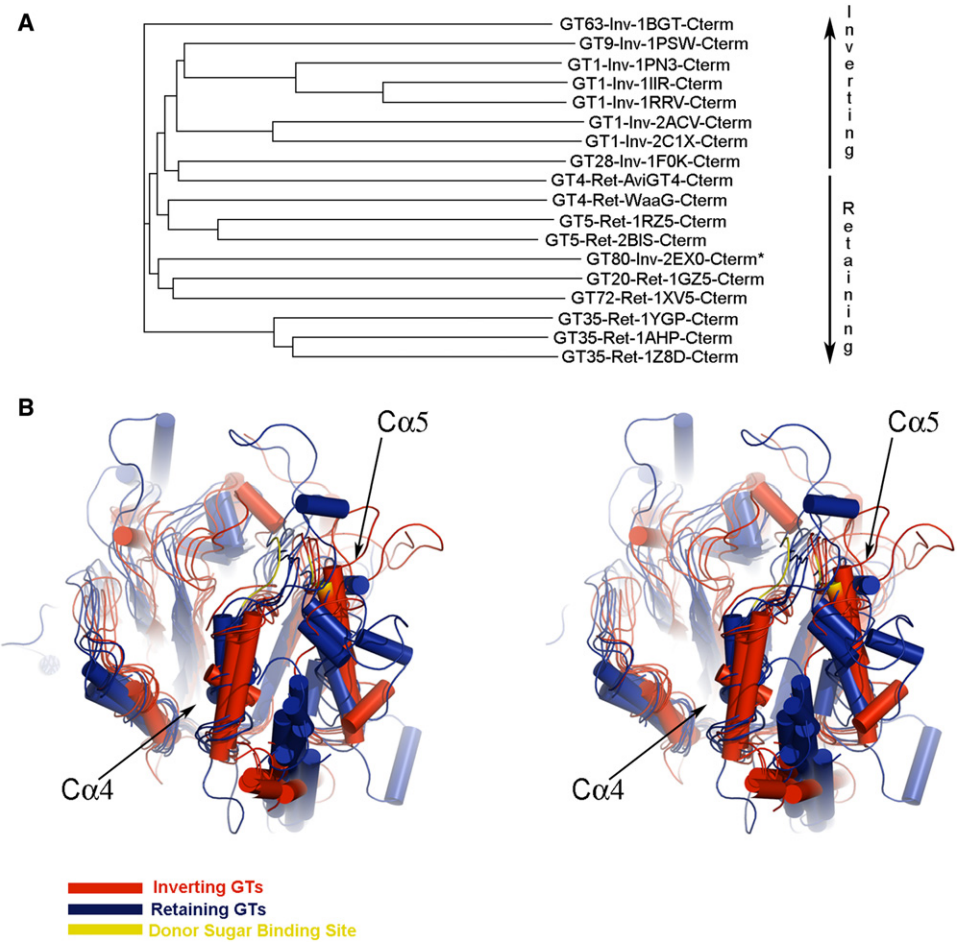

Figure 6. Overlay of GT-B Fold GTs

(A) Phylogenetic tree displaying the distances among the 18 C-terminal domains of GTs and glycogen phosphorylases obtained from a multiple sequence alignment derived, in turn, from the structural superposition of the 3D coordinates. The phylogram was drawn with TreeView [41], and the structural superposition and alignment was carried out with SSM [39]. An asterisk indicates the unusual inverting sialyltransferase, which is found in the "retaining" side of the division (discussed in text).

(B) Cartoon representation of the 3D alignment. Inverting GTs are colored red, while retaining GTs are colored blue. Two glucosyl recognition sites are highlighted in yellow: one derived from the retaining GT4 WaaG (this work), and the other in the inverting GT1 *Vv*GT1 [14].

reported here, it is possible to investigate the evolutionary relationship of this family with the rest of the structurally characterized members of the GT-B fold. Phylogenetic trees, based upon 3D structure-derived overlaps of the divergent sequences, treating the flexible N- and C-terminal domains independently, present intriguing results.

When sequence alignments, derived from the structural superposition of the separated 18 C-terminal domains, are visualized in phylogenetic trees, a clear separation of the inverting and retaining enzymes is apparent (Figure 6A). The sole "outlier" is the inverting GT80 cytidine 5′-monophosphate sialyltransferase (marked with an asterisk in the phylogenetic tree), which appears on the "retaining" side of the tree. The GT80 enzyme is also unusual in many other ways in that it uses a nucleoside monophosphate donor, has also been reported to catalyze transglycosylation with retention (in addition to glycosyl transfer with inversion), and lacks the typical signature of the GT-B fold of C-terminal α helix crossing over to the N-terminal domain [29]. The tree, the construction of which was only possible with the information provided by the 3D structures, not only allows the discrimination of catalytic activity, but also places both GT4 sequences clearly on the "cusp" of retaining and inverting enzymes, again providing support for their role as ancestral GTs first suggested by analysis of genome sequences (discussed previously).

The structural overlap of these C-terminal domains (Figure 6B) is equally provocative. While there is a remarkable structural conservation of the nucleotide binding region, there is a clear division between those elements used to bind the donor sugar moiety in inverting and retaining enzymes. In retaining GTs, this recognition site is generally located in the loop that connects the Cβ4 strand with the Cα4 helix, as observed in the complex of retaining GT4s, AviGT4 and WaaG (reported here), and GT20 OtsA [12] with their intact nucleotide-sugar donors. In retaining enzymes, and in AviGT4 especially, this loop is longer and more flexible than its counterparts in the inverting architectures. Inverting enzymes display a short and structural conserved loop that leaves the region

unblocked in such a way that the sugar recognition is implemented in the loop connecting the Cβ5 strand with the Cα5 helix instead, as observed in the complexes of the two plant GT-1 GTs [14, 30] and GT-28 MurG [17] with their intact nucleotide-sugar donors. Exceptions to this generality are the retaining GT-72 [16] and inverting GT-63 DNA glucosyltransferase [31], where the glucosyl moiety is principally recognized by residues in the N-terminal domain in both cases. Although the enzymatic mechanisms for retaining GTs are ill-defined, the different structural preferences of the inverting and retaining GTs in relation to the relative spatial disposition of the sugar moiety of the sugar-nucleotide to the acceptor, in otherwise very similar molecular architectures, again presents an easy evolutionary switch between the S_N2-like inversion mechanism and the "front face only" mechanism for retention—a feature that we had discussed previously for GT-A fold enzymes [28].

Significance

This report describes the crystal structures of two glycosyltransferase (GT) family 4 enzymes that, although displaying low sequence identity, exhibit a very similar protein structure, termed the GT-B fold, and active site topology. Differences in the donor sugar sites reflect the nature of the sugar donors of the two enzymes. UDP-glucose is the known donor for the lipopolysaccharide synthesizing α-1,3 glucosyltransferase WaaG, whereas AviGT4 likely uses a deoxy sugar donor during the synthesis of the orthosomycin antibiotic avilamycin A. The GT-B fold displays two domains, which, again with GT4, reflects a diverse acceptor binding N-terminal domain, but a more conserved nucleotide-sugar binding C-terminal domain. Such a GT-B fold has now been observed in around 18 different GT structures, both those performing catalysis with inversion and those with retention of anomeric configuration. The structures of GT4 members of this fold family are especially important, since GT4 is the largest of the GT-B fold families and the largest family of retaining GTs of any fold. The structures add to an emerging picture that GT4 may be an "ancestral" retaining family. It is the only family of retaining GTs seen in small, primitive genomes, it was only through GT4 sequences that Wrabl and Grisin [18] were able to detect the links between many of the diverse GTs in their superfamily classification (the GT4 family displays the highest level of connectivity with other families), and, here, we additionally show that phylogenetic trees based upon the structure-derived sequence alignments place both GT4 sequences at the cusp of a clear inverting versus retaining divide. Furthermore, the structural overlap of all 18 GT-B enzymes, described here, also shows a clear structural basis for the inverting versus retaining enzymes based upon the topology of the loops interacting with the donor sugar itself.

Experimental Procedures

AviGT4 Protein Production

The AviGT4 gene was amplified by the polymerase chain reaction (PCR) [32] from *Streptomyces viridochromogenes* (DMS 40,721) with genomic DNA and the primers 5′-**CACCACCACCAC** ATGAGG CCGTTGAAGGTGGCGCTGGTC-3′ and 5′-**GAGGAGAAGGCGCG TTA**TCACTTCCAGGTCGCTCC-3′ to incorporate sequences (in bold) designed to generate compatible overhangs with a pET28a-based ligation-independent cloning (LIC) vector [32], incorporating an N-terminal hexahistidine tag. Competent *E. coli* BL21(DE3)-RP strain (Stratagene) was transformed with the recombinant plasmid pETYSBLIC-*aviGT4* (pET28a containing *avigt4*) and grown at 37°C in Luria-Bertani broth supplemented with 35 μg ml^{-1} kanamycin and 50 μg ml^{-1} chloramphenicol to an OD_{600} of 0.4. Cultures were transferred to a 16°C environment and grown until an OD_{600} of 0.7 was achieved. Expression of *avigt4* was induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside and incubation at 16°C for 20 hr.

Harvested cells were resuspended in 25 mM HEPES-NaOH (pH 7.3), 400 mM NaCl, 10% glycerol, 5 mM DTT, and 0.5% Triton X-100 and incubated for 30 min on ice after the addition of 0.1 mg/ml lysozyme. Cell disruption was carried out by ultrasonication, and the lysate was centrifuged at 15,000 × g for 40 min. Clear, cell-free extract was loaded onto a 5 ml Ni^{2+}-charged His-Trap column (Amersham). Unbound proteins and detergent were removed by washing extensively with 20 mM HEPES-NaOH (pH 7.3), 400 mM NaCl, and 10% glycerol. AviGT4 was eluted with a 0–500 mM imidazole gradient. Fractions containing AviGT4 were pooled, concentrated, and buffer-exchanged to 20 mM HEPES-NaOH (pH 7.3), 150 mM NaCl, 10% glycerol with a 10 kDa cut-off ultrafiltration device (Vivaspin). AviGT4 purity and homogeneity was assessed by SDS-PAGE and gel filtration chromatography. SeMet-labeled AviGT4 was produced in *E. coli* B834 (DE3) cells containing pETYSBLIC-*aviGT4* after first culturing in minimal medium containing seleno-L-methionine. SeMet-labeled AviGT4 expression and purification was carried out following the same protocols used with the native protein. SeMet incorporation was checked by mass spectrometry (data not shown).

Crystallization and Structure Determination of AviGT4

Crystals of AviGT4 and Se-Met AviGT4 were grown by vapor phase diffusion by the hanging-drop method. Crystallization of AviGT4 (13 mg/ml) and SeMet AviGT4 (10 mg/ml) was conducted in 18% PEG 3350, 0.2 M Li_2SO_4, and 0.1 M MES (pH 6.5). Crystals of the complex of AviGT4 with UDP-glucose (Sigma) were prepared by addition of 10 mM UDP-glucose to the protein solution prior to crystallization. For cryoprotection, 20% (v/v) glycerol was added to the crystallization solution, and crystals were transferred directly into liquid nitrogen prior to data collection. A three-wavelength MAD data set was collected at the European Synchrotron Radiation Facility (ESRF) on beamline ID 14.3 from a single SeMet AviGT4 crystal. Diffraction data for the AviGT4 UDP-glucose complex were collected at the Synchrotron Radiation Source (SRS, Daresbury, UK) on beamline PX14.1. Both data sets were processed by MOSFLM and scaled and reduced by SCALA, both from the CCP4 suite [33]. The three merged Se-MAD datasets were input to SOLVE (version 2.08) [34], and a total of six Se sites were found corresponding to two AviGT4 molecules in the asymmetric unit. RESOLVE [34] was used to further improve the phases incorporating two-fold noncrystallographic symmetry averaging, with the operators automatically derived from the Se positions. The program ARP-wARP [33] was used to trace and refine the model. PHASER [35] was used to place the four AviGT4 molecules in the P1 cell of the AviGT4 UDP-glucose complex. Both models, AviGT4 with and without ligand, were manually corrected by COOT [36]. Final refinement cycles were performed by REFMAC [37], and TLS [38] corrections were incorporated in the AviGT4 UDP-glucose complex model refinement. Water molecules were added with the solvent-building options of ARP-wARP [33] and were manually inspected at the graphic display.

E. coli WaaG Protein Production

The *waaG* gene was amplified by the PCR from *E. coli* (strain W3110) with genomic DNA and the primers 5′-ATA**GGATCC**ATCGTGGCGT TTTGTT TATATAAATATTTTCCATTTGGTGG GCTTCAACG -3′ and 5′-TA**GAATTC**AACCATCTAAACCACCTGTAATGATATCCGCGGCT TTTTCTGG-3′ (enzyme restriction sites are in bold). The amplified DNA was digested with BamHI and EcoRI and ligated into suitably restricted pHISNOTHR (a gift from Prof. A.R. Fersht at the Medical

Research Council, Cambridge, UK; precut with BamHI and EcoRI) downstream of the sequence encoding a hexahistidine tag.

The *waaG* gene was expressed in *E. coli* cell strain C41(DE3) (Avidis) by induction with isopropyl-β-D-thiogalactopyranoside (100 μM) at 16°C for 16 hr in LB medium supplemented with ampicillin at 50 μg ml^{-1}. SeMet-labeled *waaG* expression in *E. coli* cell strain B834 (Novagen), cultured in SeMet growth medium (Molecular Dimensions) supplemented with ampicillin at 50 μg ml^{-1}, was induced as described above. Bacterial cells were harvested and resuspended in 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 0.5% Triton X-100, and 1 mM TCEP (buffer A), disrupted by ultrasonication, and the cell-free extract was applied to a 5 ml TALON affinity resin (Clontech). After washing with buffer B (buffer A lacking Triton X-100), WaaG was eluted from the resin with Buffer B, containing an imidazole gradient (5–100 mM over 100 ml), and the imidazole removed by size exclusion chromatography. WaaG was further purified with a G75 26/60 column (Amersham Pharmacia), and the eluted enzyme was concentrated to 10 mg/ml (determined with a molar extinction coefficient value of 41,390 at 280 nm wavelength).

Crystallization and Structure Determination of *E. coli* WaaG

Crystals of WaaG and of its SeMet derivative were grown in 0.1 M MES (pH 6.75), 0.2 M NaBr, and 15% PEG 3350 at a protein concentration of 10 mg/ml. Microseeding was necessary to achieve reproducibility of the crystals. Crystals of the complex of WaaG with UDP-2F-glucose were prepared by addition of 10 mM UDP-2F-glucose to the protein solution prior to crystallization. For cryoprotection, 22% (v/v) glycerol was added to the crystallization solution, and crystals were transferred directly into liquid nitrogen prior to data collection. A SAD data set was collected at the ESRF on beamline ID 14.4 from a single SeMet WaaG crystal. Diffraction data for the WaaG UDP-2F-glucose complex were collected in house. All data were processed by MOSFLM from the CCP4 suite [33]. SOLVE/RESOLVE (version 2.09) [34] were used to find the six Se sites and phase the diffraction data, resulting in a readily interpretable map in which ARP-wARP [33] could trace and refine the model. Subsequently, UDP and UPD-2F-Glc were added manually and the model corrected by COOT [36]. Final refinement cycles were performed with REFMAC [37].

Structure-Based Sequence Alignment

The structural superposition of the 18 C-terminal domains was carried out with secondary-structure matching (SSM; [39]). This program produces a sequence alignment derived from the structural superposition that was read into CLUSTAL W program [40] in order to calculate intersequence distances and produce a neighbor-joining tree. This tree was visualized as a phylogram with TreeView [41].

Acknowledgments

The authors thank the Biotechnology and Biological Sciences Research Council for funding this work and the European Synchrotron Radiation Facility (Grenoble) for provision of data collection facilities.

Received: May 15, 2006
Revised: August 21, 2006
Accepted: September 11, 2006
Published: November 27, 2006

References

1. Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino-acid similarities. Biochem. J. *326*, 929–942.
2. Coutinho, P., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An evolving heirarchical family classification for glycosyltransferases. J. Mol. Biol. *328*, 307–317.
3. Bourne, Y., and Henrissat, B. (2001). Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr. Opin. Struct. Biol. *11*, 593–600.
4. Davies, G.J., Gloster, T.M., and Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr. Opin. Struct. Biol. *15*, 637–645.
5. Charnock, S.J., and Davies, G.J. (1999). Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry *38*, 6380–6385.
6. Yethon, J.A., Vinogradov, E., Perry, M.B., and Whitfield, C. (2000). Mutation of the lipopolysaccharide core glycosyltransferase encoded by waaG destabilizes the outer membrane of *Escherichia coli* by interfering with core phosphorylation. J. Bacteriol. *182*, 5620–5623.
7. Whitfield, C., Heinrichs, D.E., Yethon, J.A., Amor, K.L., Monteiro, M.A., and Perry, M.B. (1999). Assembly of the R1-type core oligosaccharide of *Escherichia coli* lipopolysaccharide. J. Endotoxin Res. *5*, 151–156.
8. Hofmann, C., Boll, R., Heitmann, B., Hauser, G., Durr, C., Frerich, A., Weitnaer, G., Glaser, S.J., and Bechthold, A. (2005). Identification of genes encoding enzymes responsible for the biosynthesis of L-lyxose and the attachment of the eurekanate during avilamycin biosynthesis. Chem. Biol. *12*, 1137–1143.
9. 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* Tu57 and production of new antibiotics. Chem. Biol. *8*, 569–581.
10. Hu, Y.N., and Walker, S. (2002). Remarkable structural similarities between diverse glycosyltransferases. Chem. Biol. *9*, 1287–1296.
11. Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.
12. Gibson, R., Tarling, C.A., Roberts, S., Withers, S.G., and Davies, G.J. (2004). The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. J. Biol. Chem. *279*, 1950–1955.
13. Gibson, R.P., Turkenburg, J.P., Lloyd, R.M., Charnock, S.J., and Davies, G.J. (2002). Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem. Biol. *9*, 1337–1346.
14. 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 and activity of a flavonoid 3-O glucosyltransferase reveals the basis for plant natural product modification. EMBO J. *25*, 1396–1405.
15. Vrielink, A., Rüger, W., Driessen, H.P.C., and Freemont, P.S. (1994). Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J. *13*, 3413–3422.
16. Larivière, L., Sommer, N., and Moréra, S. (2005). Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. J. Mol. Biol. *352*, 139–150.
17. Ha, S., Walker, D., Shi, Y., and Walker, S. (2000). The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan synthesis. Protein Sci. *9*, 1045–1052.
18. Wrabl, J.O., and Grishin, N.V. (2001). Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. J. Mol. Biol. *314*, 365–374.
19. Watson, K.A., McCleverty, C., Geremia, S., Cottaz, S., Driguez, H., and Johnson, L.N. (1999). Phosphorylase recognition and phosphorolysis of its oligosaccharide substrate: answers to a long outstanding question. EMBO J. *18*, 4619–4632.
20. Oreilly, M., Watson, K.A., Schinzel, R., Palm, D., and Johnson, L.N. (1997). Oligosaccharide substrate binding in *Escherichia coli* maltodextrin phosphorylase. Nat. Struct. Biol. *4*, 405–412.
21. Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A., and Alzari, P.M. (2004). Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. EMBO J. *23*, 3196–3205.
22. Mitchell, E.P., Withers, S.G., Ermert, P., Vasella, A.T., Garman, E.F., Oikonomakos, N.G., and Johnson, L.N. (1996). Ternary complex crystal structures of glycogen phosphorylase with

the transition state analogue nojirimycin tetrazole and phosphate in the T and R states. Biochemistry *35*, 7341–7355.

23. Withers, S.G., Wakarchuk, W.W., and Strynadka, N.C.J. (2002). One step closer to a sweet conclusion. Chem. Biol. *9*, 1270–1273.
24. Weinstein, M.J., Luedemann, G.M., Oden, E.M., and Wagman, G.H. (1964). Everninomicin, a new antibiotic complex from *Micromonospora carbonacea*. Antimicrob. Agents Chemother. *10*, 24–32.
25. Belova, L., Tenson, T., Xiong, L., McNicholas, P.M., and Mankin, A.S. (2001). A novel site of antibiotic action in the ribosome: interaction of evernimicin with the large ribosomal subunit. Proc. Natl. Acad. Sci. USA *98*, 3726–3731.
26. Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. *321*, 557–559.
27. Davies, G.J., Sinnott, M.L., and Withers, S.G. (1997). Glycosyl transfer. In Comprehensive Biological Catalysis, Volume 1, M.L. Sinnott, ed. (London: Academic Press), pp. 119–209.
28. Flint, J., Taylor, E., Yang, M., Bolam, D.N., Tailford, L.E., Martinez-Fleites, C., Dodson, E.J., Davis, B.G., Gilbert, H.J., and Davies, G.J. (2005). Structural dissection and high-throughput screening of mannosylglycerate synthase. Nat. Struct. Mol. Biol. *12*, 608–614.
29. Ni, L., Sun, M., Yu, H., Chokhawala, H., Chen, X., and Fisher, A.J. (2006). Cytidine 5′-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. Biochemistry *45*, 2139–2148.
30. Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A., and Wang, X. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. Plant Cell *17*, 3141–3154.
31. Larivière, L., Gueguen-Chaignon, V., and Morera, S. (2003). Crystal structures of the T4 phage beta-glucosyltransferase and the D100A mutant in complex with UDP-glucose: glucose binding and identification of the catalytic base for a direct displacement mechanism. J. Mol. Biol. *330*, 1077–1086.
32. Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. *18*, 6069–6074.
33. CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.
34. Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.
35. Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr. D Biol. Crystallogr. *60*, 432–438.
36. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.
37. 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.
38. Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. *57*, 122–133.
39. Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.
40. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
41. Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. *12*, 357–358.
42. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
43. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. J. Mol. Graph. *15*, 133–138.
44. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank with accession codes 2IUY and 2IV3 (AviGT4) and 2IV7 and 2IW1 (WaaG).