Bacterial β-Kdo glycosyltransferases represent a new glycosyltransferase family (GT99)

Olga G. Ovchinnikova[a], Evan Mallette[a], Akihiko Koizumi[b], Todd L. Lowary[b], Matthew S. Kimber[a,1], and Chris Whitfield[a,1]

[a]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; and [b]Department of Chemistry and Alberta Glycomics Centre, University of Alberta, Edmonton, AB, Canada T6G 2R3

Edited by Barbara Imperiali, Massachusetts Institute of Technology, Cambridge, MA, and approved April 22, 2016 (received for review February 24, 2016)

Kdo (3-deoxy-D-*manno*-oct-2-ulosonic acid) is an eight-carbon sugar mostly confined to Gram-negative bacteria. It is often involved in attaching surface polysaccharides to their lipid anchors. α-Kdo provides a bridge between lipid A and the core oligosaccharide in all bacterial LPSs, whereas an oligosaccharide of β-Kdo residues links "group 2" capsular polysaccharides to (lyso)phosphatidylglycerol. β-Kdo is also found in a small number of other bacterial polysaccharides. The structure and function of the prototypical cytidine monophosphate-Kdo–dependent α-Kdo glycosyltransferase from LPS assembly is well characterized. In contrast, the β-Kdo counterparts were not identified as glycosyltransferase enzymes by bioinformatics tools and were not represented among the 98 currently recognized glycosyltransferase families in the Carbohydrate-Active Enzymes database. We report the crystallographic structure and function of a prototype β-Kdo GT from WbbB, a modular protein participating in LPS O-antigen synthesis in *Raoultella terrigena*. The β-Kdo GT has dual Rossmann-fold motifs typical of GT-B enzymes, but extensive deletions, insertions, and rearrangements result in a unique architecture that makes it a prototype for a new GT family (GT99). The cytidine monophosphate-binding site in the C-terminal α/β domain closely resembles the corresponding site in bacterial sialyltransferases, suggesting an evolutionary connection that is not immediately evident from the overall fold or sequence similarities.

microbial glycobiology | 3-deoxy-D-*manno*-oct-2-ulosonic acid | Kdo | glycosyltransferase | polysaccharide

Carbohydrates coat the surface of cells in the form of complex macromolecules, collectively known as glycoconjugates. These molecules are implicated in many crucial biological processes and the diversity of carbohydrate structures is a result of the activities of a large group of enzymes, called glycosyltransferases (GTs). GTs catalyze formation of glycosidic linkages by transferring a sugar moiety from a glycosyl donor to an acceptor substrate. The glycosyl donors are various sugar-1-phosphates and their derivatives, with sugar mono- or diphospho-nucleotides being most common. GTs are classified as either retaining or inverting, depending on whether the anomeric configuration of the glycosyl donor is retained or inverted in the final product (1). Based on amino acid sequence similarities, GTs are currently classified to 98 families in the Carbohydrate-Active enZYmes (CAZy) database (2), reflecting the breadth of biological systems and product chemistry. Each family typically has members with different donor or acceptor specificities. However, the 3D structures of GTs are more conserved; for nucleotide sugar-dependent enzymes, only two structural folds, GT-A and GT-B, are currently known. The fold does not dictate activity (i.e., inverting or retaining). Currently less than 1% of the overall ∼225,000 GT sequences in the CAZy database have been biochemically characterized, and less than 0.1% have solved 3D structures. Understanding the relationship between sequence, 3D structure, and substrate specificity of GTs represents a fundamental challenge in glycobiology. Moreover, increasing interest in GTs is emerging from the perspective of developing new therapeutics, as well as application to enzymatic and chemo-enzymatic synthesis of medically and industrially important glycoconjugates, including glycoconjugate vaccines (3, 4).

Bacteria possess a remarkable range of GTs, which transfer sugars found nowhere else in nature: they generate glycoconjugates involved in the interface between pathogens and the host immune system. We recently discovered a novel family of 3-deoxy-β-D-*manno*-oct-2-ulosonic acid (β-Kdo) transferases (5). Kdo is eight-carbon keto-sugar structurally related to sialic acids. Kdo is a ubiquitous component of bacterial LPSs. In eukaryotes, Kdo has been identified as a component of rhamnogalacturonan II in the primary cell walls of higher plants (6, 7), and in cell wall polysaccharides of some green algae (8, 9), but it has never been found in the animal kingdom. LPS is a glycolipid that forms the outer leaflet of the characteristic outer membrane of the Gram-negative bacterial cell; it is essential for the viability of almost all of these bacteria. LPS molecules typically consist of three domains: a hydrophobic moiety (lipid A) forming the outer leaflet of the membrane, a polysaccharide chain (O-antigen, OPS), and a bridging oligosaccharide called the core. Kdo is the most conserved sugar residue in all studied LPS, where it possesses the α-configuration and provides the linkage between lipid A and carbohydrate moieties. In contrast, β-Kdo is found in a

Significance

Glycosyltransferase enzymes synthesize complex sugar-containing macromolecules that play pivotal roles in the biology of all cells. Bacteria produce a remarkable range of these glycoconjugate structures, often containing unusual sugars. For example, Gram-negative bacteria exploit an unusual eight-carbon sugar (Kdo) as a linkage point between diverse glycan structures and conserved lipid termini in LPS and (some) capsules. Here, we describe the distribution and phylogenetic relationships of a new family of β-Kdo glycosyltransferases. Although these enzymes resemble some other glycosyltransferases, including those forming α-Kdo linkages, they are not readily identified as glycosyltransferases by bioinformatics approaches. The structure of a prototypical enzyme reveals extensive insertions, deletions, and rearrangements in the normally highly conserved GT-B–fold, highlighting the unusual structure of this glycosyltransferase family.

Author contributions: O.G.O., T.L.L., M.S.K., and C.W. designed research; O.G.O., E.M., and A.K. performed research; O.G.O., E.M., A.K., T.L.L., M.S.K., and C.W. analyzed data; and O.G.O., T.L.L., M.S.K., and C.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession numbers for region 2 of CPS gene clusters sequences of *Escherichia coli* K12 and K14 are KU314824 and KU314825, respectively). The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (coordinates for the *Raoultella terrigena* $WbbB_{2-401}$ and for $WbbB_{2-401}$ in complex with CMP are PBD ID codes 5FA0 and 5FA1, respectively).

[1]To whom correspondence may be addressed. Email: mkimber@uoguelph.ca or cwhitfie@uoguelph.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603146113/-/DCSupplemental.

small number of bacterial glycoconjugates. Its most conserved application lies in a recently discovered (10) β-Kdo oligosaccharide that bridges a lipid anchor (lyso-phosphatidylglycerol) to a long-chain capsular polysaccharide (CPS) in organisms, including *Escherichia coli*, *Neisseria meningitidis*, *Haemophilus influenzae*, and others (11). These capsules are collectively known as "group 2" based on their shared biosynthetic pathway. Identification of the β-Kdo oligosaccharide led to definition of the β-Kdo GTs activities of a group of enzymes (pfam05159), previously described only as "capsule polysaccharide biosynthesis proteins" with no assigned function (5). These retaining β-Kdo transferases show no significant sequence similarity to their well-studied inverting α-Kdo counterparts of the LPS core biosynthetic pathway (Fig. 1), or to any other known GT family.

Here, we report the structure and function of a prototype β-Kdo transferase. The enzyme acts as the chain-terminating component of the multidomain WbbB protein, which is involved in biosynthesis of the OPS in *Raoultella terrigena*. The protein possesses a GT-B fold that shows Rossmann-folds of reduced numbers of α/β repeats, as well as additional novel structural features that place it in a new family of GTs, assigned as family GT99 in the CAZy database (2).

Results

β-Kdo-Transferases in the Synthesis of Bacterial Glycans. In the biosynthesis of the conserved glycolipid anchor for many Gram-negative capsules, the KpsS enzyme adds the first Kdo residue to the lipid acceptor and this product is then extended by KpsC, which contains duplicated domains with predicted β-Kdo GT activity. Definitive biochemical assignment of KpsS and KpsC as β-Kdo GTs in the CPS assembly (5) facilitated bioinformatic assignments of other candidates from O-antigen and CPS biosynthesis confined to Gram-negative bacteria. A search of the Bacterial Carbohydrate Structure Database for β-Kdo–containing glycans resulted in <40 unique glycan structures, and some could be correlated with existing sequence data in GenBank. Combining the structural data on bacterial polysaccharides with sequenced gene clusters for their synthesis allowed prediction of functionally related β-Kdo GTs (Fig. 2, Fig. S1, and Table S1), although overall shared sequence similarity was low (Table S2). Details of the basis for the functional assignments are given in *Materials and Methods*.

In general, the glycans containing β-Kdo fell into three functional categories: (*i*) capsule linkers, (*ii*) repeat units of CPSs and OPSs, and (*iii*) nonreducing terminal residues in a group of OPSs synthesized in a particular chain-terminating strategy. Multiple sequence alignments were prepared using the identified β-Kdo GTs (domains) along with homologs of KpsS and the individual domains of KpsC from a number of pathogens (Fig. S2 and Table S1). Because of variations in amino acid sequences and protein length (ranging from 295 to 326 aa for N-terminal KpsC domains, to 537 aa for *E. coli* K12 Orf3) the alignments introduced multiple gaps. Amino acid sequence variations reflect not only evolutionary relationships between proteins but also the differences in acceptor substrates (e.g., lysophosphatidylglycerol, another Kdo residue, hexoses, 6-deoxyhexoses, and HexNAc) and linkage specificities. Elimination of poorly aligned and divergent regions using GBlocks resulted in 68 of the original 906 positions that fulfill the GBlocks relaxed parameters (12), and these positions were used to calculate a maximum-likelihood phylogenetic tree (Fig. 3).

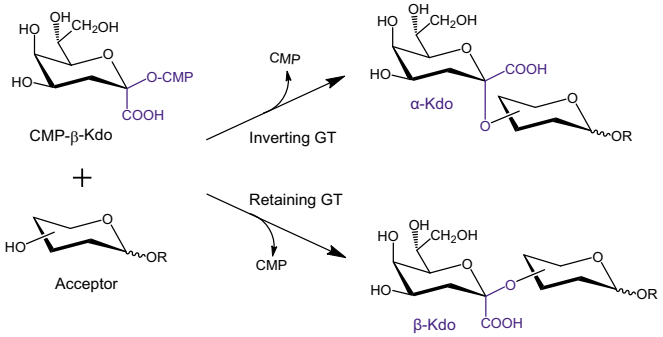

Fig. 1. Glycosylation reactions catalyzed by Kdo transferases.

Known and predicted polymerizing β-Kdo GTs (KpsC and RkpZ homologs) form a well-define clade, with each of KpsC domain grouping in separate subclades with counterparts from different bacteria. All analyzed KpsS proteins also group together, with higher similarity occurring between sequences from closely related species. Despite the difference in the chemical nature of acceptor substrates, KpsS enzymes are more closely related to single-addition β-Kdo GTs participating in CPS and O-antigen synthesis. The high sequence divergence between these GTs (i.e., long terminal branches) precludes deeper phylogenetic analysis. Four chain-terminating enzymes group together, as might be expected from the similar roles of product structures. *R. terrigena* WbbB ($WbbB_{Rt}$) and *Klebsiella pneumoniae* WbbB ($WbbB_{KpO12}$) are the most closely related, reflecting taxonomic closeness of recently delineated *Raoultella* and *Klebsiella* genera (13). Each of these enzymes possess multiple GT modules with different predicted specificities. For example, $WbbB_{Rt}$ is a large (1,106-aa residue) enzyme that we predict to be solely responsible for the synthesis of the capped OPS repeat unit domain. At the C terminus of $WbbB_{Rt/Kp}$, there are two GT domains sharing homology with GT1 and GT25 families, presumably one each for transfer of Rha and GlcNAc. The N terminus contains the candidate β-Kdo GT domain and this activity is verified below.

In Vitro Analysis of the Activity of the β-Kdo GT Domain of $WbbB_{Rt}$. The β-Kdo GT domain from the *R. terrigena* WbbB protein was selected as a model for biochemical and structural characterization. To generate a construct expressing only the β-Kdo GT domain, $WbbB_{Rt}$ was examined using Phyre2 for regions of predicted disordered sequence that might mark the separation of its three GT domains. The β-Kdo GT and the central GT1 domains are separated by a region with strong predicted propensity to form a parallel coiled-coil domain according to COILS (14) prediction. The coiled-coil segment might function as a molecular ruler in regulation of OPS chain length, analogous to a similar role in WbdD, an OPS chain-terminating dual methyltransferase/kinase in *E. coli* serotype O9a (15, 16). The break point after Ser401 was selected, which is within a small disordered region and is right before coiled-coil segment. A truncated polypeptide comprising amino acids 2–401 was cloned and overexpressed in *E. coli* BL21 (DE3). $WbbB_{2–401}$ was soluble and the added C-terminal His_6-tag facilitated purification to apparent homogeneity using nickel affinity chromatography. Purified $WbbB_{2–401}$ migrated on SDS/PAGE in accordance with the theoretical molecular mass of a monomer of 46,047 Da (Fig. S3*A*). The molecular mass of 120 kDa estimated by size-exclusion chromatography is consistent with a dimer or a trimer (Fig. S3*B*).

The activity of $WbbB_{2–401}$ was assayed in vitro using a synthetic acceptor (**1**) comprising a *R. terrigena* OPS disaccharide repeat unit [α-L-Rha-(1→3)-β-D-GlcNAc] linked to a fluorescein tag (Fig. 4*A*). This structure represents the terminal disaccharide that is capped by a β-Kdo residue in the native OPS (17). The nucleotide-sugar donor of Kdo residues, CMP-β-Kdo, is unstable (18), and was generated in situ using CTP, Kdo and purified CMP-Kdo synthethase (KdsB) from *E. coli* (19). The reaction products were analyzed by thin-layer chromatography (TLC) and liquid chromatography–mass spectrometry (LC-MS) on a reverse-phase C_{18} column (Fig. S4 *A* and *B*). In the absence of enzyme, two minor (photo)degradation products of the substrate were observed on TLC. Accordingly, in the LC-MS profile the signal corresponding to substrate ion $[M+H]^+$ at *m/z* 884.31 was accompanied by two

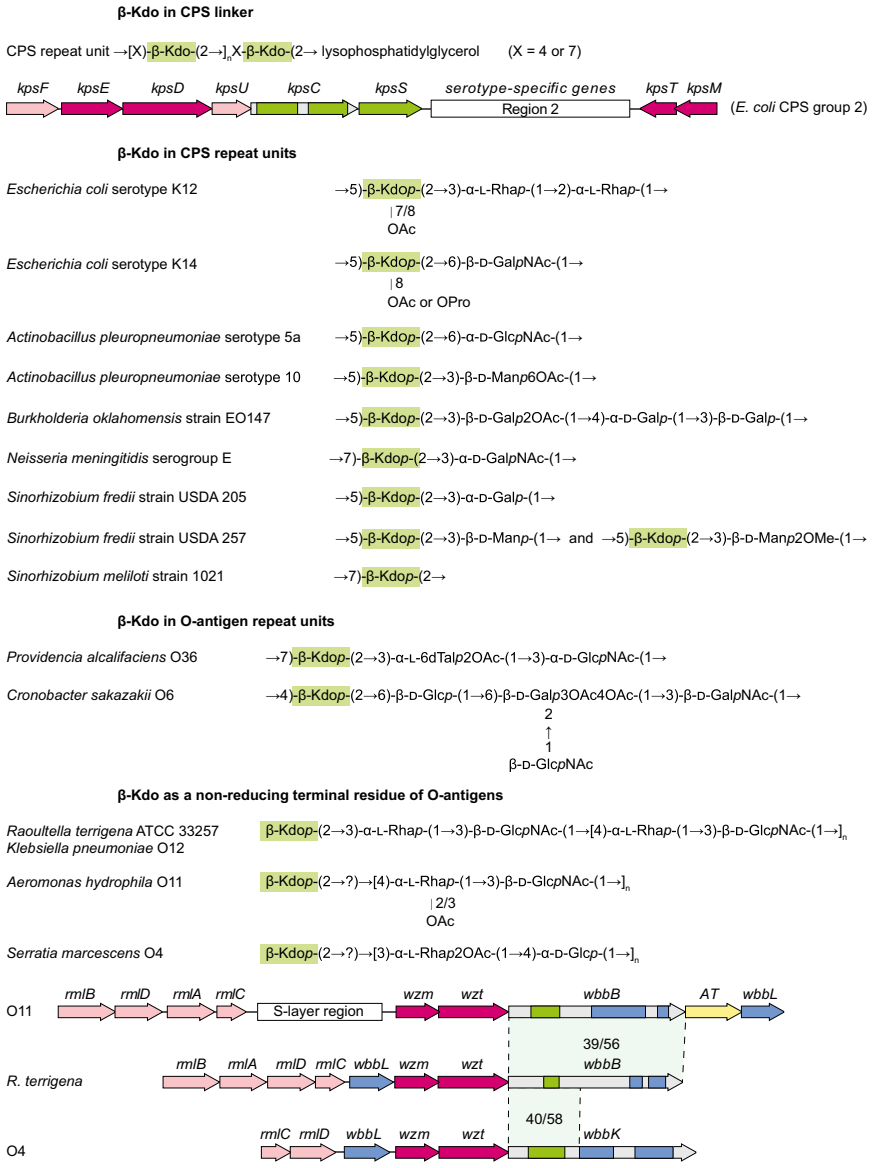

Fig. 2. β-Kdo–containing bacterial glycans and the gene clusters associated with their biosynthesis. Only those gene clusters encoding enzymes central to the experimental approach are shown; others are given in Fig. S1. The analysis is confined to examples with known gene clusters and unambiguous polysaccharide structures. For all structures except CPS of *Burkholderia oklahomensis*, O-acetylation and O-propionylation are nonstoichiometric. In multidomain proteins, the conserved domains identified using the NCBI Conserved Domain Search service are indicated. β-Kdo GTs are shown in green, other GTs are represented in blue, acyltransferases (*AT*) in yellow, nucleotide sugar precursor synthesis genes in pink, and PS export machinery genes in magenta. For the WbbB protein from *R. terrigena*, the percentage of amino acid identity/similarity is indicated. The organization of O-antigen biosynthesis gene cluster of *K. pneumoniae* O12 is identical to that of *R. terrigena*, with WbbB sharing 72% identity and 83% similarity.

minor peaks at *m/z* 868.33 and 1197.39. The observed in-source fragmentation of these ions, as well as MS/MS data, indicated that the structural heterogeneity of acceptor arises from the fluorescein moiety (Fig. S4 *C* and *D*), not the carbohydrate. $WbbB_{2–401}$ catalyzed apparent complete conversion of all acceptor species into products that migrated slower on TLC. The major peak in the LC-MS profile eluted at 7.3 min and showed a $[M+H]^+$ ion peak at *m/z* 1104.36. This result represents a gain of 220.05 u compared with **1**, which corresponds to one added Kdo residue. The presence of ions at *m/z* 1088.38 and 1417.44 indicated that $WbbB_{2–401}$ also transferred a Kdo residue to both degraded acceptors. As expected, no product was observed when the substrate was incubated in the absence of either KdsB or $WbbB_{2–401}$.

$WbbB_{2–401}$ Adds a Single β-Kdo Residue to Position O-3 of Rhamnose. To determine the position and anomeric configuration of ketosidic linkage of Kdo, the products from a scaled up reaction were purified using a Sep-Pak C_{18} cartridge. The structure **2** was established by a combination of ^{1}H NMR, proton-coupled ^{13}C NMR, 2D ^{1}H,^{1}H COSY, total correlation spectroscopy (TOCSY), rotating-frame nuclear Overhauser effect spectroscopy (ROESY), ^{1}H,^{13}C heteronuclear single-quantum coherence (HSQC), and heteronuclear multiple band correlation (HMBC) experiments (NMR chemical shifts are given in Table S3). In addition to the signals for fluorescein tag, GlcNAc, and Rha residues, the HSQC spectrum of **2** (Fig. 4*B*) contained two methylene signals (δ_C 36.0, δ_H 1.86, 2.45), two HOCH_2-C signals (δ_C 65.6, δ_H 3.78, 3.87), and four other signals in the region δ_C/δ_H 66.6–74.7/3.63–3.97, consistent with the presence of a single Kdo residue. The signal at δ_C 36.0 is characteristic of Kdo*p* C-3 and indicated pyranosidic form (20) [C-3 of Kdo*f*, the furanose form, is at δ ∼45 ppm (21)]. Correlations from H-3 to H-5 were traced in the COSY spectrum, and the signals for the corresponding carbon atoms from C-3 to C-5 were assigned based on HSQC data. The coherence transfer was interrupted after H-5, because of a small $^3J_{5,6}$ coupling constant typical of Kdo. Signals for H-6 and C-6 were identified based on H-5/H-6 correlation observed in the ROESY spectrum, as well as C-5/H-6 and C-6/H-4 correlations in the HMBC spectrum (Fig. 4*C*). The remaining assignments were completed by tracing the coherence transfer of H-6 to the exocyclic protons H-7 and H-8 and confirmed by HMBC data. The signals for Kdo C-1 and C-2 were found from correlations H-3ax/C-1 at δ 1.86/175.5 and H-3ax/C-2 at δ 1.86/103.3 in the HMBC spectrum. The β-anomeric configuration was inferred based on $^3J_{C\text{-}1,H\text{-}3ax}$

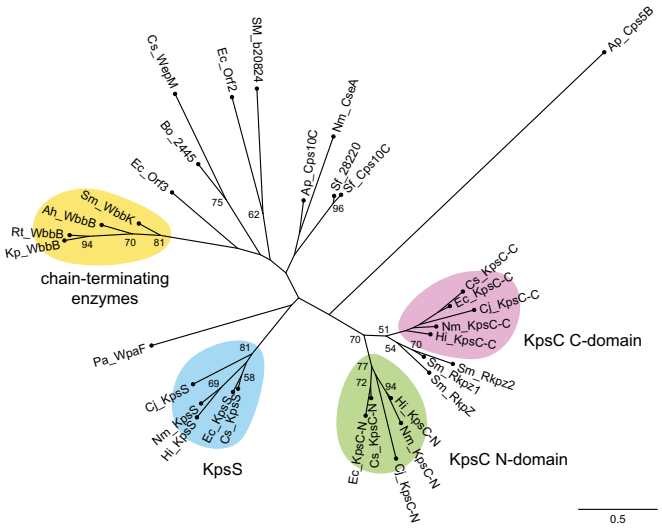

Fig. 3. Unrooted phylogenetic tree of β-Kdo GTs. Kdo-transferase domains in multidomain proteins were separated based on the conserved domains identified using the NCBI Conserved Domain Search, disordered regions in Phyre2 secondary structure models, and biochemical data (see *SI Materials and Methods* for details). Multiple sequence alignment was performed with MAFFT followed by curation in GBlocks (Fig. S2). The maximum-likelihood tree was calculated using PhyML. Bootstrap values above 50% are shown adjacent to nodes and are based on 1,000 replications. The accession numbers of the proteins are given in Table S1.

coupling constant of 4.6 Hz, characteristic for *trans*-diaxial orientation of C-1 and H-3ax (22), as well as the relatively large difference between chemical shifts of H-3ax and H-3eq of 0.59 ppm (20, 23). The ^{13}C NMR chemical shifts of Kdo*p* were indicative of its terminal position (23), whereas the Rha C-3 signal was shifted downfield evidently because of a mild α-glycosylation effect. The Kdo-(2→3)-Rha linkage was unambiguously confirmed by the correlation between Kdo C-2 and the Rha H-3 at δ 103.3/4.07 in the HMBC spectrum (Fig. 4*C*). Thus, the identity of the reaction product **2** confirmed the predicted terminal Kdo at position O-3 of Rha in a structure identical to the native glycan.

The $WbbB_{2–401}$ Structure Reveals a Novel Variant of the Canonical GT-B Fold. We determined the structure of $WbbB_{2–401}$ in two crystal forms: as a ligand-free protein in $P2_12_12_1$ (2.3 Å resolution) and as a cytidine monophosphate (CMP) complex in $C222_1$ (2.1 Å resolution). $WbbB_{2–401}$ is built around two α/β domains with recognizable Rossmann-fold topology, and therefore broadly meets the criteria of conforming to a GT-B–fold. However, comparison of the $WbbB_{2–401}$ structure with a canonical GT-B enzyme, WaaA (an α-Kdo GT; PDB ID code 2XCU) (24) shows that the organization of WbbB is profoundly different (Fig. 5 *A*–*C*). Rather than the usual six or seven parallel β-strands for each domain, the α/β repeats are reduced in number, with five parallel β-strands (plus one antiparallel) for the N-terminal, and four β-strands for the C-terminal Rossmann-fold domain. The N-terminal Rossmann-fold domain is supplemented by the insertion of extended loops and additional α-helices, principally into the Nβ2–Nβ3 and Nβ3–Nβ4 loops. WbbB also includes an additional α-helical domain made from five α-helices contributed by the C-terminal portion of the chain, and two helices formed by the linker between the N- and C-terminal α/β domains. Instead of forming the two-lobed structure divided by a clear cleft typical of GT-B enzymes, the Rossmann-fold domains in WbbB occupy opposite corners of a roughly trapezoidal protein, with the remaining two corners filled by internal extensions of the Rossmann-fold domains and the α-helical domain. The Rossmann-fold domains of $WbbB_{2–401}$ are therefore considerably rearranged relative to WaaA, with the N-terminal α/β domain being rotated and displaced some 25 Å from the position it occupies in WaaA. The interactions between these elements are extensive, suggesting minimal capacity for the large-scale interdomain movements usually characteristic of the functional cycle of GT-B enzymes.

Searching the PDB with DALI finds no proteins with global resemblance to $WbbB_{2–401}$, whereas the conserved core of the individual α/β-fold domains show weak resemblance (rmsd values around 3.5 Å) to many proteins, including various GTs. The α-helical domain

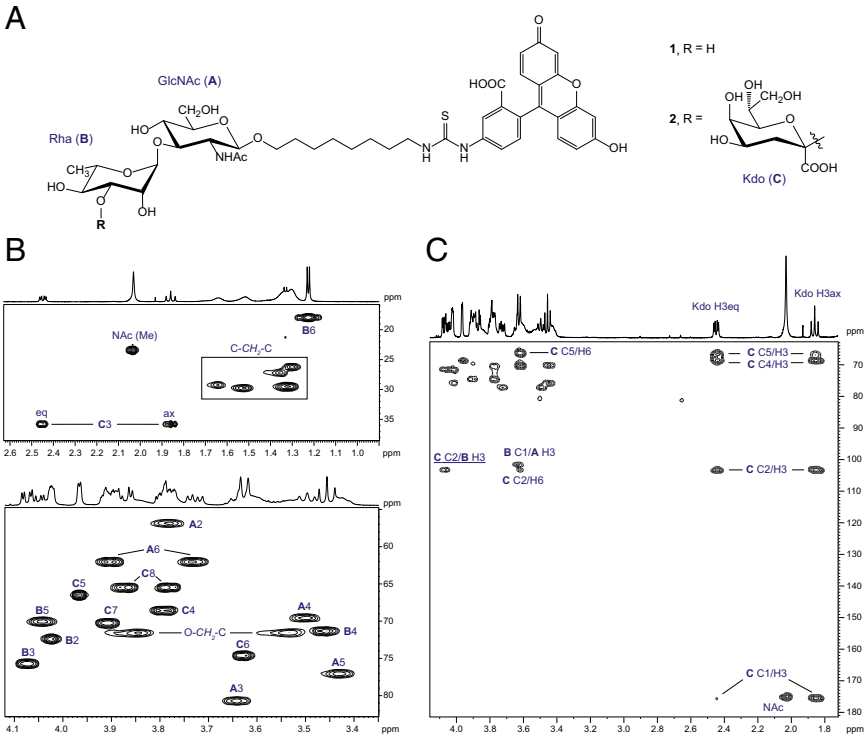

Fig. 4. NMR spectroscopic analysis of product **2**. (*A*) Structures of synthetic acceptor **1** and the product **2**, obtained in vitro using $WbbB_{2–401}$. (*B*) Parts of ^{1}H,^{13}C HSQC spectrum. Numbers refer to H/C pairs in sugar residues designated as **A** (GlcNAc), **B** (Rha), and **C** (Kdo). O-*CH*$_2$-C and C-*CH*$_2$-C signals belong to eight-carbon methylene linker. (*C*) Part of ^{1}H,^{13}C HMBC spectrum. The inter-residue correlation between Kdo C-2 and Rha H-3 (underlined) demonstrates Kdo-(2→3)-Rha linkage.

shows no significant resemblance to anything else in the PDB. Although its ancestry among GT-B enzymes remains apparent, WbbB represents an unprecedentedly radical departure from the typical GT-B organization. On the basis of its sequence and structural novelty, $WbbB_{2–401}$ was assigned to a new family of GTs, family GT99.

Oligomeric Organization of $WbbB_{2–401}$. Analysis of both crystal structures with PISA (25) indicates that the $WbbB_{2–401}$ forms a dimer. In the ligand-free structure, this dimer occupies the asymmetric unit, whereas in the CMP costructure, the two chains each represent half of a distinct dimer straddling a crystallographic symmetry axis. The interface is made up from Cα1a (C-terminal α/β domain) packing on the Nβ1–Nα1 and Nα3a–Nα3b loops and the N-terminal end of Nα3b (N-terminal α/β domain); in addition, residues 270–272 in the long Cβ2–Cα3 loop interact as a short antiparallel strand, Cβ2a (Fig. 5 *B* and *D*). The interface buries 1,440 $Å^2$, or 9% of the total surface buried per subunit, and with a predicted dissociation energy of 9.6 kcal/mol. Gel-filtration analysis is consistent with this dimer being present in solution (Fig. S3*B*).

$WbbB_{2–401}$ is one domain of a larger protein that includes two other candidate GT domains in the C-terminal part of the protein. The β-Kdo GT domain is separated from the other GT domains by a 70-aa region with low sequence identity to other proteins, followed by the 90-aa predicted parallel coiled-coil

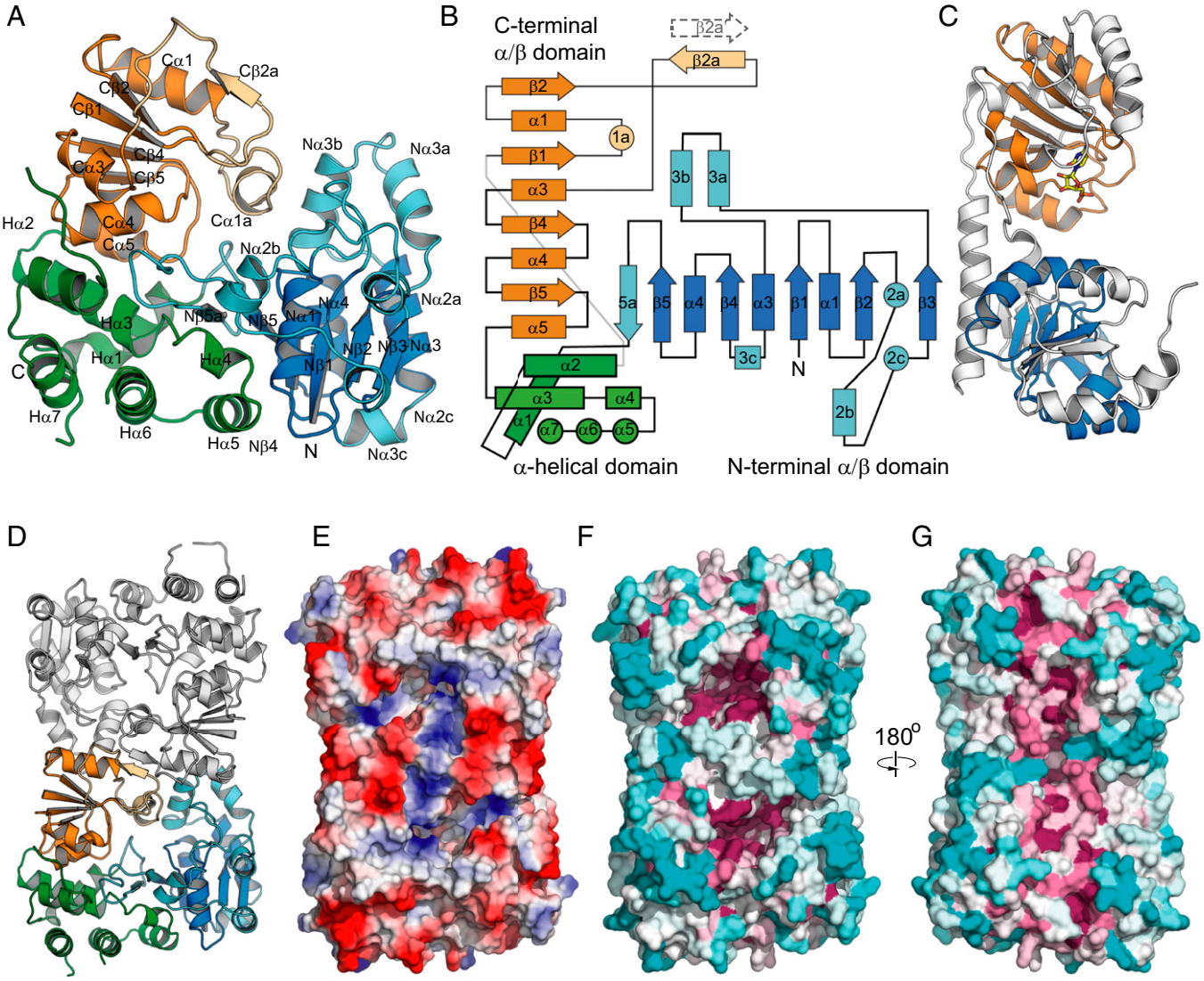

Fig. 5. Structure of $WbbB_{2–401}$. (*A*) Organization of the protomer. The protomer has three distinct domains: an N-terminal α/β domain (blue), a C-terminal α/β domain (orange), and an α-helical domain (green). Elements that do not correspond to the Rossmann-fold architecture are shown in a lighter shade. Secondary structure elements are numbered in analogy to the six-stranded α/β domain of canonical GT-B proteins. Helices are numbered by the preceding β-strand; elements with letter suffixes denote elements inserted into the standard Rossmann-fold. (*B*) Topology of $WbbB_{2–401}$. Secondary structural elements are colored the same as in *A*. (*C*) Structure of WaaA (PDB ID code 2XCU) for comparison. WaaA is an α-Kdo transferase with a canonical GT-B organization. Elements of the N-terminal and C-terminal α/β domains that have corresponding elements in WbbB are shown in blue and orange, respectively. Image orientation places the C-terminal α/β domain in a similar orientation to *A*. Note that in $WbbB_{2–401}$ the α-helical domain occupies the approximate space occupied in WaaA by the N-terminal α/β domain, which is rotated and displaced. (*D*) Structure of the $WbbB_{2–401}$ dimer, viewed down the twofold axis. The second protomer is colored white. (*E*) Electrostatic surface of the $WbbB_{2–401}$ CMP complex, showing the substrate binding face. (*F*) Sequence conservation mapped onto the surface. A multiple sequence alignment of WbbB homologs was mapped onto the $WbbB_{2–401}$ CMP structure using Consurf. Magenta denotes the most conserved residues, cyan denotes the most variable. The two clusters of absolutely conserved residues represent the substrate binding and catalytic sites of each chain. (*G*) Sequence conservation on the $WbbB_{2–401}$ face distal to the CMP binding site. The conserved central vertical band of residues likely mediate interactions with the other domains of WbbB in the complete protein.

domain. The organization of the dimer places the C termini at opposite corners of the molecule on the same face as the active site, suggesting that the linker may be required to bridge the C termini of $WbbB_{2–401}$ to the N-terminal end of the coiled-coil. The presence of a wide medial strip of surface-exposed conserved residues on the face opposite the CMP binding site possibly reflects the position of this linker joining $WbbB_{2–401}$ to the coiled-coil domain (Fig. 5*G*).

Active Site Organization and CMP Binding. Cocrystallization trials with 1 mM CMP yielded a new crystal form not seen with the ligand-free protein. Examination of the electron density showed that CMP binding occurs with good occupancy in both active sites in the asymmetric unit. In common with other GT-B enzymes, the nucleotide, CMP, binds to the C-terminal α/β domain. Superposition of the ligand-free and CMP-bound structures shows that CMP binding induces fairly small movements (<2 Å) that are localized primarily within the C-terminal α/β domain; this region partakes in the dimeric interface, and CMP binding also somewhat alters the conformation of the dimer (Fig. 6*A*). This finding is in contrast with other GT-B enzymes, where nucleotide binding typically results in a large rotation (on the order of 20°) of the two domains, closing the catalytic cleft (Fig. S5*A*), but is consistent with the extensive interactions between all three domains, which seems to produce a more rigid platform than is typically seen for GT-B enzymes.

Although GT-B–fold proteins appear to have no absolutely conserved residues, there are several features identified as being conserved across multiple families. These features include a glutamate residue that forms hydrogen bonds with the O2 and O3 groups of ribose, a glycine-rich loop that coordinates the phosphate ion, and a (D/E)(D/E)G motif between Nβ5 and Nα5 that may contribute the general base (Fig. S5*B*) (26, 27). $WbbB_{2–401}$ lacks all of these substrate-interacting motifs. Instead, the cytosine ring of CMP is sandwiched between Val302 and Pro266, whereas the Watson and Crick face makes hydrogen bonds with the carbonyl oxygen and guanidinyl group of Arg263 (Fig. 6*B*). The amine also forms a hydrogen bond with the carbonyl oxygen of Leu227. The phosphate group forms hydrogen bonds with His265, as well as Ser300 (Oγ) and Ser301 (both N and Oγ) at the N terminus of helix Cα4. His265 and Pro266 (the HP motif) are invariant in all analyzed β-Kdo GTs (Fig. S2) and aligned structurally with highly conserved HP motifs of CMP-sialic acid-dependent sialyltransferases. H265A has a more severe catalytic defect than P266A; however, neither of the mutations abolishes enzyme activity completely (Fig. 6 *D* and *E*). The ribose hydroxyl O2 and O3 do not form hydrogen bonds

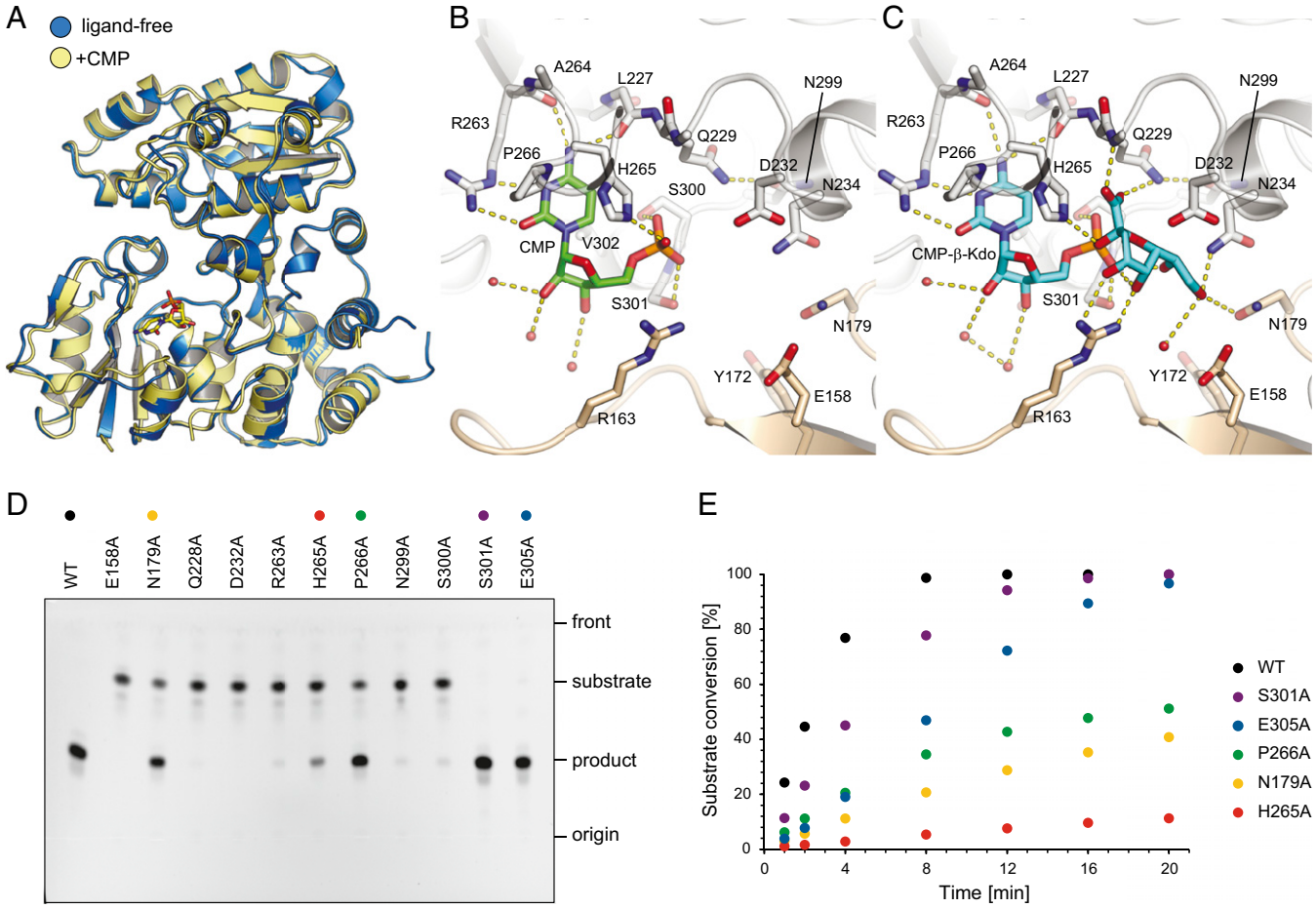

Fig. 6. CMP binding to $WbbB_{2–401}$ and in vitro activity of $WbbB_{2–401}$ and site-directed mutants. (*A*) Superposition of ligand-free (blue) and CMP bound (yellow) of $WbbB_{2–401}$. Local movement in the C-terminal α/β domain accompanies CMP binding; note that this movement occurs at the dimeric interface. (*B*) Details of CMP binding site. (*C*) Model of CMP-β-Kdo in the binding site. The ligand was placed by superimposing the CMP portion of the molecule, then positioning the sugar ring by torsions around the P–O and O–C2 bonds. Potential hydrogen bonds are shown in dashed lines. (*D*) The reaction mixtures were incubated for 20 min and analyzed by TLC. Of six mutants showing little or no activity on TLC, E158A and D232 were completely inactive, whereas others retained residual activity (<3% conversion after 20 min). (*E*) Time course of Kdo-transferase reaction. The percentage of conversion was determined using LC-MS by integrating the peak areas of the substrate and the product in the merged extracted ion chromatograms.

directly with the protein, but instead interact with structured water molecules. This CMP binding mode is similar to what is seen in GT-B sialyltransferases of CAZy families GT52 and GT80. For example, *Pasteurella multocida* multifunctional sialyltransferase (GT80; PDB ID code 2EX1) coordinates CMP using homologous (K/R)XPH and SS motifs to mediate identical interactions to bind both the base and phosphate group (27) (Fig. S5). Note that this resemblance seems localized to the CMP binding pocket; although GT80 sialyltransferases also show insertion of extra α-helices, these occur at topologically distinct sites. Overall, these similarities are unlikely to be convergent, leading us to conclude that sialyltransferases are likely the most closely related previously characterized relatives of β-Kdo transferases.

The CMP binding site is part of a larger conserved pocket that is formed at the conjunction of the three domains. Immediately adjacent to the CMP binding pocket is an extended, conserved, basic pocket that is a good candidate for Kdo binding (Fig. 5 *E* and *F*). This pocket is formed mostly by elements of the C-terminal α/β domain, as well as the β-hairpin motif linking the N-terminal α/β domain and the α-helical domain. Arg163 repositions itself in the CMP costructure so the guanidine group stacks on the ribose C4 and C5. This residue is then well positioned to donate one or two hydrogen bonds to β-Kdo, as well as stabilize the negative charges associated with the carboxylate and phosphate groups. The rest of the pocket is lined by hydrogen bonding groups, including Glu158, Tyr172, Asn179, Gln228, Asp232, and Asn234. E158A was found be completely inactive, whereas N179A shows significant catalytic defects. Gln228 and Asp232 are of special note as together they form a QXXXD motif that is conserved across all identified β-Kdo transferases (Fig. S2). Asp232 is positioned 4.4 Å from the nearest phosphate oxygen. The D232A mutant proved completely inactive, whereas the Q228A mutant was severely compromised in activity (Fig. 6*D*).

Modeling potential CMP-β-Kdo binding modes suggests that interactions with critical residues such as Gln228 and Asn179 can only be achieved if the Kdo ring is inserted edgewise into the cleft between Asp232 and Arg163; binding here positions the carboxylate group between His265 and Asp232, allowing it to form hydrogen bonds with Gln228Nε and Val229N. Placing the Kdo ring in a boat conformation (shown in Fig. 6*C*) seems to maximize interactions with the pocket. In CMP-β-Kdo, this conformation is stabilized by internal hydrogen bonds (between O5, O7, and a phosphate oxygen) and closely resembles the conformation argued to explain the unusually high CMP-β-Kdo autohydrolytic reaction rate (18). In this boat conformation, O4 and O5 can form hydrogen bonds with Arg163, whereas O8 forms hydrogen bonds with Asn179 and Asn234. Interestingly, in this model the anomeric carbon atom is completely buried from solvent, whereas the Asp232 carboxylate is 3.2 Å away, and in line with the C2–O2 bond that is broken during the reaction.

The active site is extended from the CMP binding site into a 20 Å long groove (Fig. 5 *E* and *F*), which may accommodate the long-chain acceptor molecule during substrate binding. The location of the acceptor-binding site is uncertain, although the presence of an extended exposed hydrophobic patch (Trp20, Trp54, Leu64, Ile159, and Phe348) contributed mainly by the N-terminal Rossmann-fold domain presents a reasonable candidate. These residues are, however, too distant to interact with the terminal residue of the acceptor. Of note here is Glu158, which appears absolutely required for activity, but is located too far from CMP to do more that form a single hydrogen bond with the donor. Possibly, this residue plays an important role in acceptor positioning or activation.

Discussion

The recent discovery of β-Kdo transferases (KpsC and KpsS) involved in the assembly of a conserved β-Kdo oligosaccharide linker in a large family of CPSs provided a foundation for a survey of homologs in bacterial strains known to contain β-Kdo as a component of their surface glycans. This approach allowed identification of new potential β-Kdo transferases, all being members of the pfam05159 family. The phylogenetic analysis reflects highly divergent amino acid sequences of identified GTs (Fig. 3). When the search was performed without limitations of the glycan structural data (i.e., PSI-BLAST search against the nonredundant protein sequences database) the KpsC and KpsS homologs were confined to Proteobacteria and one species of green sulfur bacteria. The predicted function for some β-Kdo GTs are supported by functional complementation (28) and in vitro biochemical studies (5); however, the structures of enzymatic products were not completely characterized (5). Although the overall sequence similarities between the β-Kdo transferases are low (Table S2), as reflected in the pylogenetic tree (Fig. 3), there is substantially more conservation in the putative catalytic sites (Fig. S2). Divergence in the flanking sequence presumably reflects the differences in acceptor molecules. We consider these proteins to be GT99 members in the absence of detailed structural or mechanistic data that would indicate otherwise.

We successfully delineated the N-terminal β-Kdo GT domain from the modular *R. terrigena* GT and demonstrated that its coiled-coil region and two other GT domains are not required for either folding or in vitro activity of $WbbB_{2–401}$. In vivo, the β-Kdo GT domain of WbbB transfers a Kdo residue to the nonreducing end of a long polysaccharide chain. In the *E. coli* O9a prototype for this type of OPS-assembly system, a nonreducing terminal methylphosphate serves as chain-terminator, creating a modal distribution of long O-antigen chains (15, 16) as well as a signal to export OPS to the periplasmic side of the inner membrane. However, in *E. coli* O9a, different proteins dictate polymerization and chain termination, whereas the GT99, GT1-like, and GT25-like modules in WbbB are consistent with it performing both functions in *R. terrigena*. The involvement of β-Kdo transferases (KpsC and KpsS) in the assembly of a conserved capsule linker in a large family of CPSs found in important Gram-negative pathogens (5, 10) has stimulated interest in whether these enzymes offer a new potential therapeutic target for small-molecule inhibitors. A solved structure for a β-Kdo GT prototype provides an important asset in such strategies.

To our knowledge, the crystal structure of $WbbB_{2–401}$ provides the first insight into this new GT99 family. The complex of $WbbB_{2–401}$ with CMP reveals a characteristic donor recognition HP motif that is involved in coordinating the cytosine ring; this interaction is also conserved among different bacterial sialyltransferase families (29). A detailed evaluation of the donor substrate binding site is challenging because of the instability of CMP-Kdo, which undergoes rapid hydrolysis to CMP and Kdo (18). However, relative activity assays showed that the H265A and P266A mutants maintain reduced catalytic activity. Another bacterial "sialyl motif" S (S/T) (30, 31) is reduced to a single invariant Ser residue (S300 in $WbbB_{Rt}$) in most β-Kdo GTs, although the four closest WbbB homologs also possess an adjacent Ser residue. Mutation S300A dramatically decreases enzyme activity, whereas S301A only has a slight catalytic defect. β-Kdo transferases diverge, however, from the sialyltransferases in possessing an invariant QXXXD motif. Mutagenesis of these residues severely impedes activity, suggesting an essential role for this motif in $WbbB_{2–401}$'s mechanism.

The mechanisms used by retaining GTs have proved difficult to pin down. Initially, in an analogy with retaining glycosyl hydrolases, it was thought that these enzymes would use a double-displacement mechanism, wherein a protein carboxylate group attacks the anomeric carbon as a nucleophile, displacing the nucleotide leaving group. The resulting covalent intermediate is resolved by the incoming acceptor displacing the carboxylate in a second S_N2 reaction, restoring the original anomeric configuration. Although this mechanism remains at least partially supported by available data for family GT6 (32, 33), most candidate acidic groups in retaining GTs have alternative roles in substrate binding, are weakly conserved, or are too distant from the anomeric carbon to allow them to readily act as nucleophiles (1). As a result,

Table 1. Data collection, phasing and refinement statistics

Data collection and statistics	Sel Met (apo)	CMP complex
Data collection		
Space group	$P2_12_12_1$	$C222_1$
Cell dimensions		
a, *b*, *c* (Å)	82.93, 82.93, 120.71	92.75, 159.39, 120.04
Resolution (Å)	50–2.3 (2.36–2.3)*	50–2.1 (2.15–2.10)
R_{sym}	0.073 (0.93)	0.094 (0.64)
$I/\sigma I$	19.1 (2.2)	13.3 (2.85)
Completeness (%)	99.8 (100)	100 (100)
Redundancy	5.8 (5.8)	7.2 (7.5)
Refinement		
Resolution (Å)	41.5–2.3	50–2.1
No. reflections	37,670	
R_{work}/R_{free}	0.186/0.229	0.169/0.211
No. atoms		
Protein	6,211	6,187
Ligand/ion	1	42
Water	562	836
B-factors		
Protein	33.2	24.1
Ligand/ion	42.9	11.6
Water	34.3	31.7
Rms deviations		
Bond lengths (Å)	0.003	0.005
Bond angles (°)	0.636	0.852

A single crystal was used in each case to determine the structure.
*Values in parentheses are for highest-resolution shell.

recent attention has focused on the alternative S_Ni "internal return" mechanisms. Here, the incoming acceptor is positioned adjacent to the phosphate group; dissociation of the phosphate-donor bond forms an oxocarbenium ion, and transfer of a proton from the acceptor hydroxyl to the phosphate-leaving group allows the deprotonated acceptor to recombine with the oxocarbenium ion. Anomeric configuration is retained because the acceptor and nucleotide are positioned on the same face of the donor (1).

$WbbB_{2–401}$ is unique among known retaining GTs in that the donor sugar is a ketose. Although the C2-carboxylate group should contribute substantially to stabilizing an oxocarbenium ion transition state (34), Kdo's bulky carboxylate group may impede the acceptor sugar from closely approaching the anomeric carbon while the nucleotide is present. Modeling potential donor binding modes suggests that the anomeric carbon is likely to be wholly buried in the active site, leaving little room for the acceptor to approach with a geometry suitable for S_Ni mechanisms. The same model predicts that Asp232—a residue indispensable for $WbbB_{2–401}$ activity and one of six residues absolutely conserved among all β-Kdo transferases—may be ideally positioned (about 3.2 Å from the anomeric carbon, and in line with the nucleotide leaving group) to attack C2 as a nucleophile (Fig. 6*C*). It is also worth noting that $WbbB_{2–401}$ likely evolved from an inverting sialyltransferase family that uses a conserved histidine to accelerate the reaction, most likely by protonating the leaving group (35, 36). $WbbB_{2–401}$ similarly requires His265 for maximal activity. Because S_Ni mechanisms require no such general acid (as the phosphate group acquires a proton directly from the acceptor), this suggests that the $WbbB_{2–401}$ reaction mechanism retains considerable S_N2-like characteristic. Together, these observations imply that retaining β-Kdo transferases are at least strong candidates for using a glycosyl hydrolase-like double-displacement mechanism.

Materials and Methods

Bacterial Strains and Growth Conditions. Bacterial cultures were grown with aeration at 37 °C in Luria Broth Base (LB) (Invitrogen) supplemented with kanamycin (50 μg/mL^{-1}) when necessary. *R. terrigena* ATCC 33257 K$^-$ was a gift from U. Mamat (Research Centre Borstel, Leibnitz Centre for Medicine and Biosciences, Borstel, Germany). *E. coli* O4:K12:H1 was obtained from Wyeth-Lederle Vaccines. *E. coli* O15:K14:H4 (strain SSI 85370) was purchased from Cedarlane. *E. coli* strain CWG1217 [F$^-$, *mcrA*, Δ(*mrr-hsdRMS-mcrBC*), ϕ80, *lacZΔM15*, Δ*lacX74*, *deoR*, *nupG*, *recA1*, *araD139*, Δ(*ara-leu*)7697, *galU*, *galK*, *rpsL*(Str^r), *endA1*, Δ(*wzx-glf-wbbHIJK*)] used for cloning is a derivative of strain TOP10 from Invitrogen that lacks part of the K-12 *wb** gene cluster. Protein expression was performed in *E. coli* BL21 (DE3) [F$^-$ *ompT gal dcm* $hsdS_B(r_B^- m_B^-)$ λ(DE3)] or the methionine auxotroph *E. coli* B834 (DE3) [F$^-$ *ompT gal dcm* $hsdS_B(r_B^- m_B^-)$ *met* λ(DE3)].

Sequencing of CPS Gene Clusters. Primers OL1052 and OL1053, based on the *kpsS* and *kpsT* genes, respectively, were used to amplify the region 2 of CPS gene cluster in *E. coli* K14 with the Phusion DNA polymerase. This process took advantage of the observation that the serotype-specific CPS biosynthesis gene cassettes (named region 2) usually lie between two conserved gene blocks with *kpsS* and *kpsT* as the flanking genes (Fig. 2). The PCR product was sequenced by primer walking on both strands. However, with *E. coli* serotype K12, only a short 2,810-bp PCR product was obtained, which contained two insertion elements, *insA* and *insB*, indicating that serotype-specific genes were relocated in this isolate.

E. coli K12 genomic DNA was sequenced using Illumina MiSeq technology but the CPS gene cluster region was represented as several contigs. The *E. coli* K12 CPS biosynthesis genes showed 100% identity to those from *E. coli* KTE195 (GenBank accession no. ASUJ01000047.1), the serotype of which is not indicated. The region 2 genes of *E. coli* KTE195 are located in the correct position on the chromosome but are flanked by two pairs of *insA* and *insB* sequences. We used an intergenic region sequence between flanking IS elements and region 2 to designed new primers OL1078 and OL1079 and successfully amplified the relocated region 2 in the *E. coli* K12 serotype reference strain. The sequence was confirmed by primer walking. Custom oligonucleotide primers (Sigma) are listed in Table S4.

Bioinformatic Assignment of β-Kdo GTs. The assignments for the candidate β-Kdo GTs (Table S1) involved integrated assessment of the GT requirements imposed by the glycan structure, organization of genetic loci responsible for biosynthesis (Fig. 2 and Fig. S1), and multiple sequence alignments of the proteins (Fig. S2 and Table S2) (see *SI Materials and Methods* for details). Glycan structures containing β-Kdo were retrieved from the Bacterial Carbohydrate Structure Database (37).

Multiple Sequence Alignment and Phylogenetic Analysis. The proteins and domains used in alignment are listed in Table S1. The protein secondary structure was modeled using the Phyre2 server (38). The coiled-coil segment was predicted using COILS (14). $WbbB_{Rt}$ was aligned with its homologs $WbbB_{KpO12}$, $WbbB_{AhO11}$, and $WbbK_{SmO4}$ to identify the break points in their sequences corresponding to Ser401 in $WbbB_{Rt}$. *E. coli* Orf2 and *Sinorhizobium meliloti* SM_b20824 domains were separated based on disordered regions in Phyre2 model and the conserved domains identified using the National Center for Biotechnology Information (NCBI) Conserved Domain Search service. *N. meningitidis* $KpsC_{1–325}$ and $KpsC_{327–704}$ were shown to be active as separate polypeptides in vivo (5). Multiple-sequence alignment of KpsC homologs was performed to identify the break points corresponding to Pro326 in *N. meningitidis* KpsC. A total of 33 proteins and separated domain sequences were aligned using MAFFT (39) with the iterative refinement method E-INS-i and BLOSUM62 scoring matrix. Poorly aligned and divergent regions were eliminated using GBlocks running under relaxed parameters (12). Phylogenetic analysis was performed using the PhyML 3.0 server (40) (LG matrix; SPR tree improvement) with 1,000 bootstrap replicates. The multiple-sequence alignment and the tree were visualized with ESPript (41) and FigTree, respectively.

Overexpression and Purification of $WbbB_{2–401}$ and Its Mutants. $WbbB_{2–401}$-His_6 and its mutants were expressed in *E. coli* BL21 (DE3). Overnight cultures were used to inoculate 250 mL LB media at 1/100 dilution. Cultures were grown at 37 °C until OD_{600} ~0.2 was reached and then placed at 18 °C. Protein expression was induced at OD_{600} ~0.6 by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and cultures were grown at 18 °C overnight. Cells were harvested by centrifugation at 5,000 × *g* for 10 min and either used for immediate protein purification or kept at −80 °C until needed. The cell pellet was resuspended in 25 mL buffer A (50 mM Tris·HCl, 250 mM NaCl, pH 7.4) containing 10 mM imidazole and supplemented with 1/2 Complete mini EDTA-free protease inhibitor tablets (Roche). The cells were lysed by ultrasonication for a total of 3 min 30 s, with 30% amplitude in pulses (10 s on/15 s off). The suspension was cleared by successive centrifugation steps at 4,000 × *g* for 10 min and 100,000 × *g* for 1 h at 4 °C. The $WbbB_{2–401}$-His_6

protein was purified from the supernatant using 2 mL Ni-NTA Agarose resin columns (Qiagen). The column was washed sequentially with 10 mL of binding buffer, 20 mL of buffer A containing 25 mM imidazole, and eluted with 10 mL of buffer A containing 250 mM imidazole; 1-mL fractions were collected. Protein was exchanged into imidazole-free buffer A using a PD-10 desalting column (GE Lifesciences). Protein purity was assessed by SDS/PAGE analysis using 10% (wt/vol) acrylamide resolving gels, and the concentration was estimated from A280 values and a theoretical extinction coefficient of 40,800 $M^{-1}{\cdot}cm^{-1}$ predicted by ProtParam (42). Folding of the mutant proteins was confirmed by a differential-scanning fluorescence method (43) using the Protein Thermal Shift Dye kit (ThermoFisher). Proteins were stored in buffer A containing 10% glycerol at −80 °C.

Overexpression and Purification of Selenomethionine-Labeled $WbbB_{2–401}$. $WbbB_{2–401}$-His_6 was expressed in the methionine auxotrophic strain *E. coli* B834 (DE3). A starter culture was grown at 37 °C overnight in LB media. Cells from the overnight culture were collected by centrifugation (4,000 × *g* for 10 min) and resuspended in cold sterile water. The cell suspension was used to inoculate 750 mL selenomethionine minimal media (44) supplemented with L-SeMet (50 μg/mL) to achieve a final OD_{600} 0.3. The culture was then grown at 18 °C until OD_{600} ~0.6 and protein expression was induced with IPTG to a final concentration of 0.5 mM. The culture was then incubated at 18 °C for 24 h and protein purification followed the procedure described above.

Overexpression and Purification of KdsB. Plasmid pJKB72 encoding His_6-tagged CMP-Kdo synthethase (KdsB) from *E. coli* K-12 was a gift from H. Brade (Research Centre Borstel, Leibnitz Centre for Medicine and Biosciences, Borstel, Germany). KdsB was expressed in *E. coli* LE392 and purified essentially as described previously (45).

SDS/PAGE. SDS/PAGE was performed using 10% resolving gels in Tris-glycine buffer, and the gels were stained with Simply Blue SafeStain (ThermoFisher).

FPLC. Purified $WbbB_{2–401}$ (0.2 mg at 2 mg/mL) was eluted through a Superose 6 HR 10/30 size-exclusion column (GE Healthcare) with buffer A (50 mM Tris·HCl, 250 mM NaCl, pH 7.4) at 0.5 mL/min. Elution profiles of freshly purified $WbbB_{2–401}$ (immediately after elution from the Ni-NTA Agarose column in buffer A containing 250 mM imidazole) and after buffer exchange and storage for 5 d at 4 °C were essentially identical. Molecular mass was estimated from a standard curve derived from the standards β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), BSA (66 kDa), ovalbumin (44 kDa), and carbonic anhydrase (29 kDa).

Synthesis of Acceptor 1. The route used for the synthesis of acceptor disaccharide **1** is provided in *SI Materials and Methods* (Fig. S6). The *N*-phthaloyl protected glucosamine derivative **3** (46) was converted to the corresponding glycosyl bromide and coupled to 8-azido-1-octanol in the presence of *N*-iodosuccinimide, silver triflate, and 2,4,6-collidine giving the desired product **4**, in 92% yield. Deacetylation of **4** by treatment with sodium methoxide and then protection of the C4 and C6 hydroxyl groups as a benzylidene acetal afforded acceptor **5** in 79% yield over the two steps. Glycosylation of **5** with rhamnose donor **6** (47) gave disaccharide **7** in 72% yield. Full deprotection of sugar moiety was smoothly carried out by first ethylenediamine-mediated cleavage of the phthalimide protecting group and acetyltion of the resulting amine to give disaccharide **8** in 91% over the two steps. Next, acid hydrolysis of the acetal-protecting groups in **8** and then acetylation of liberated hydroxyl groups produced, over the two steps, a 91% yield of compound **9**. The acetylation step was done to facilitate purification of the compound. Removal of the *O*-acetate protecting groups in **9** afforded a 92% yield of compound **10**. Hydrogenation of the azido group led to primary amine **11**, which was obtained in 84% yield. Finally, **11** was treated with FITC to generate **1** in 46% yield.

In Vitro Determination of $WbbB_{2–401}$ Activity. In vitro reactions were performed in PCR tubes at 37 °C. The acceptor concentration was estimated by measuring absorbance at 490 nm in 0.01 M NaOH, using extinction coefficient 88,000 $M^{-1}{\cdot}cm^{-1}$ (48). To optimize the amount of enzymes in a coupled assay, the reactions (20 μL) contained 50 mM Hepes (pH 8.0), 10 mM $MgCl_2$, 2 mM Kdo, 5 mM CTP, 0.17 mM synthetic acceptor, and variable amounts of KdsB (0.5–4 μg) and WT $WbbB_{2–401}$ (1-4 μg). The reaction was initiated by addition of KdsB. The reaction was stopped by spotting 1 μL on TLC plate (Silicagel 60 F_{254}) after 1, 2, 4, 8, 12, 16, and 20 min and the TLC plate was developed in freshly prepared chloroform-methanol-water-acetic acid mixture (25:15:4:2, vol/vol). The products were visualized using a hand-held UV lamp.

The activities of site-directed mutants were examined under conditions where CMP-Kdo is expected to be in excess. The reaction was scaled up to 80 μL and contained all of the components at the same concentrations as above, with 2 μg KdsB and 4 μg $WbbB_{2–401}$. Ten-microliter aliquots were withdrawn after 1, 2, 4, 8, 12, 16, and 20 min. For mutants showing little or no activity after 20 min in TLC analysis, the reaction mixture was incubated for longer time and aliquots were withdrawn after 20, 60, and 120 min. The reactions were quenched by incubating at 98 °C for 5 min. Aliquots were kept on ice until the end of experiment and centrifuged at 13,000 × *g* for 5 min. An equal amount of water was added to the samples and 1-μL aliquots were analyzed by LC-MS (see *SI Materials and Methods* for details). The assay was performed in triplicate for WT and each of $WbbB_{2–401}$ mutants shown in Fig. 6*E*. The percentage of conversion was determined by integrating the peak areas of the substrate and the product in the merged extracted ion chromatograms using MassHunter Workstation Software (Agilent).

Purification of Reaction Product for Structural Analysis. Reaction mixtures (9 × 100 μL) containing 50 mM Hepes (pH 8.0), 10 mM $MgCl_2$, 2 mM Kdo, 5 mM CTP, 0.34 mM synthetic acceptor, 4 μg KdsB, and 8 μg $WbbB_{2–401}$ were incubated at 37 °C for 30 min. Completion of reaction was confirmed by TLC as described above. The reaction mixtures were pooled, diluted with water up to 10 mL, and loaded to a Sep-Pak C_{18} column, previously washed with 10 mL acetonitrile and 20 mL water. The column was washed with 20 mL water and the product was eluted in 50% (vol/vol) acetonitrile-water mixture and dried using a Speed-Vac.

NMR Spectroscopy. NMR studies were performed at the NMR Centre in the University of Guelph Advanced Analysis Center. The product **2** was deuterium-exchanged by dissolving in 99.9% D_2O and drying using a Speed-Vac. NMR spectra were obtained at 30 °C in 99.96% D_2O using a Bruker Avance II 600 MHz spectrometer equipped with a cryoprobe. Internal sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 ($\delta_H = 0$ ppm, $\delta_C = -1.6$ ppm) served as a reference. Two-dimensional experiments were performed using standard Bruker software, and the Bruker TopSpin 2.1 program was used to acquire and process the NMR data. Mixing times in TOCSY and ROESY experiments were set up to 100 and 200 ms, respectively. The HMBC experiment was optimized for the $J_{H,C}$ coupling constant 8 Hz.

$WbbB_{2–401}$ Structure Determination. All crystallization experiments were conducted in a sitting-drop configuration at room temperature, by mixing protein in a 1:1 ratio with well solution, and then conducting vapor diffusion against ~100 μL of the same well solution. Selenomethionine-labeled $WbbB_{2–401}$ protein was crystallized by using 1 μL of 10 mg/mL protein and a well solution of 0.2 M $MgCl_2$, 0.1 M Tris·HCl pH 8.5, 25% polyethylene glycol 4000, 10 mM Tris(2-carboxyethyl)phosphine (TCEP). Crystals grew as prisms (200 mm) after three days. For the CMP complex, protein was concentrated to 10 mg/mL in a buffer that included 20 mM Tris·HCl pH 8.0, 150 mM NaCl, 1 mM TCEP, 1 mM EDTA, and 1 mM CMP. Crystals were grown using a well solution of 0.2 M NaCl, 0.1 M Bis-Tris pH 5.5, and 25% PEG 3350. Prismatic crystals up to 150 μm in length were observed within 14 d. Both crystal forms were cryoprotected with paratone-N oil before freezing in liquid nitrogen for data collection at 100 K. Data were collected for the selenomethionyl protein crystals at the Canadian Light Source beamline CLS08B1 at a wavelength of 0.97836 Å; crystals diffracted to 2.3 Å. The structure was determined using single anomalous scattering. Anomalous substructure searching with Phenix autosol found 12 anomalous scattering centers, with an overall figure of merit 0.231; autotracing in Phenix was able to then correctly trace most the structure. Manual rebuilding in Coot (49) and refinement in Phenix (50) was used to complete the structure. Ramachandran statistics for the ligand-free structure were 96.8% favored, 2.8% allowed, and 0.4% outliers (as determined by Molprobity). Data for CMP-complex crystals were collected at the Canadian Light Source, beamline CLS08ID at wavelength of 0.97949 Å, with diffraction to 2.1 Å. The structure of the CMP complex crystals was determined using molecular replacement in Phaser (51) in Phenix, with the complete ligand-free protomer as a search model. Ramachandran statistics for the CMP costructure were 98.1% favored, 1.8% allowed, and 0.1% outliers. In both structures, several residues in the loops between Cα5 and Hα3, and between Hα6 and Hα7 are disordered. Data collection and refinement statistics are shown in Table 1. Structural figures were prepared in Pymol.

ACKNOWLEDGMENTS. This work was supported by operating funding from the Canadian Institutes of Health Research (C.W.); Discovery Grants from the National Science and Engineering Research Council of Canada (to M.S.K., T.L.L., and C.W.); and GlycoNet (the Canadian Glycomics Network, National Centres of Excellence Program) (to M.S.K, T.L.L., and C.W.). T.L.L. and C.W. are recipients of Canada Research Chairs.

1. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem* 77:521–555.
2. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(Database issue):D490–D495.
3. Baker JL, Çelik E, DeLisa MP (2013) Expanding the glycoengineering toolbox: The rise of bacterial N-linked protein glycosylation. *Trends Biotechnol* 31(5):313–323.
4. Koeller KM, Wong CH (2000) Synthesis of complex carbohydrates and glycoconjugates: Enzyme-based and programmable one-pot strategies. *Chem Rev* 100(12):4465–4494.
5. Willis LM, Whitfield C (2013) KpsC and KpsS are retaining 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. *Proc Natl Acad Sci USA* 110(51):20753–20758.
6. Smyth KM, Marchant A (2013) Conservation of the 2-keto-3-deoxymanno-octulosonic acid (Kdo) biosynthesis pathway between plants and bacteria. *Carbohydr Res* 380:70–75.
7. York WS, Darvill AG, McNeil M, Albersheim P (1985) 3-Deoxy-D-*manno*-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants. *Carbohydr Res* 138(1):109–126.
8. Becker B, Hård K, Melkonian M, Kamerling JP, Vliegenthart JF (1989) Identification of 3-deoxy-*manno*-2-octulosonic acid, 3-deoxy-5-O-methyl-*manno*-2-octulosonic acid and 3-deoxy-*lyxo*-2-heptulosaric acid in the cell wall (theca) of the green alga *Tetraselmis striata* Butcher (Prasinophyceae). *Eur J Biochem* 182(1):153–160.
9. Becker B, Lommerse JPM, Melkonian M, Kamerling JP, Vliegenthart JFG (1995) The structure of an acidic trisaccharide component from a cell wall polysaccharide preparation of the green alga *Tetraselmis striata* Butcher. *Carbohydr Res* 267(2):313–321.
10. Willis LM, et al. (2013) Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *Proc Natl Acad Sci USA* 110(19):7868–7873.
11. Willis LM, Whitfield C (2013) Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr Res* 378:35–44.
12. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56(4): 564–577.
13. Drancourt M, Bollet C, Carta A, Rousselier P (2001) Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. *Int J Syst Evol Microbiol* 51(Pt 3):925–932.
14. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252(5009):1162–1164.
15. Hagelueken G, et al. (2015) A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nat Struct Mol Biol* 22(1):50–56.
16. Clarke BR, et al. (2011) In vitro reconstruction of the chain termination reaction in biosynthesis of the *Escherichia coli* O9a O-polysaccharide: The chain-length regulator, WbdD, catalyzes the addition of methyl phosphate to the non-reducing terminus of the growing glycan. *J Biol Chem* 286(48):41391–41401.
17. Mertens K, et al. (2010) Antiserum against *Raoultella terrigena* ATCC 33257 identifies a large number of *Raoultella* and *Klebsiella* clinical isolates as serotype O12. *Innate Immun* 16(6):366–380.
18. Lin CH, Murray BW, Ollmann IR, Wong CH (1997) Why is CMP-ketodeoxyoctonate highly unstable? *Biochemistry* 36(4):780–785.
19. Gronow S, Brabetz W, Brade H (2000) Comparative functional characterization in vitro of heptosyltransferase I (WaaC) and II (WaaF) from *Escherichia coli*. *Eur J Biochem* 267(22):6602–6611.
20. Lenter M, Jann B, Jann K (1990) Structure of the K16 antigen from *Escherichia coli* O7: K16:H-, a Kdo-containing capsular polysaccharide. *Carbohydr Res* 197:197–204.
21. Ahrens R, Jann B, Jann K, Brade H (1988) Structure of the K74 antigen from *Escherichia coli* O44:K74:H18, a capsular polysaccharide containing furanosidic beta-KDO residues. *Carbohydr Res* 179:223–231.
22. Unger FM, Stix D, Schulz G (1980) The anomeric configurations of the two ammonium (methyl 3-deoxy-D-*manno*-2-octulopyranosid)onate salts (methyl α- and β-ketopyranosides of KDO). *Carbohydr Res* 80(1):191–195.
23. Vinogradov E, et al. (2002) Structures of lipopolysaccharides from *Klebsiella pneumoniae*. Eluicidation of the structure of the linkage region between core and polysaccharide O chain and identification of the residues at the non-reducing termini of the O chains. *J Biol Chem* 277(28):25070–25081.
24. Schmidt H, et al. (2012) Structural and mechanistic analysis of the membrane-embedded glycosyltransferase WaaA required for lipopolysaccharide synthesis. *Proc Natl Acad Sci USA* 109(16):6253–6258.
25. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
26. Wrabl JO, Grishin NV (2001) Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J Mol Biol* 314(3):365–374.
27. Ni L, et al. (2006) Cytidine 5′-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. *Biochemistry* 45(7):2139–2148.
28. Sharypova LA, Chataigné G, Fraysse N, Becker A, Poinsot V (2006) Overproduction and increased molecular weight account for the symbiotic activity of the *rkpZ*-modified K polysaccharide from *Sinorhizobium meliloti* Rm1021. *Glycobiology* 16(12):1181–1193.
29. Freiberger F, et al. (2007) Biochemical characterization of a *Neisseria meningitidis* polysialyltransferase reveals novel functional motifs in bacterial sialyltransferases. *Mol Microbiol* 65(5):1258–1275.
30. Yamamoto T, Ichikawa M, Takakura Y (2008) Conserved amino acid sequences in the bacterial sialyltransferases belonging to Glycosyltransferase family 80. *Biochem Biophys Res Commun* 365(2):340–343.
31. Romanow A, et al. (2014) Dissection of hexosyl- and sialyltransferase domains in the bifunctional capsule polymerases from *Neisseria meningitidis* W and Y defines a new sialyltransferase family. *J Biol Chem* 289(49):33945–33957.
32. Monegal A, Planas A (2006) Chemical rescue of alpha3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases. *J Am Chem Soc* 128(50):16030–16031.
33. Soya N, Fang Y, Palcic MM, Klassen JS (2011) Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* 21(5):547–552.
34. Horenstein BA (1997) Quantum mechanical analysis of an α-carboxylate-substituted oxocarbenium ion. Isotope effects for formation of the sialyl cation and the origin of an unusually large secondary ^{14}C isotope effect. *J Am Chem Soc* 119(5):1101–1107.
35. Ni L, et al. (2007) Crystal structures of *Pasteurella multocida* sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* 46(21):6288–6298.
36. Lin LY-C, et al. (2011) Structure and mechanism of the lipooligosaccharide sialyltransferase from *Neisseria meningitidis*. *J Biol Chem* 286(43):37237–37248.
37. Toukach PV (2011) Bacterial carbohydrate structure database 3: Principles and realization. *J Chem Inf Model* 51(1):159–170.
38. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858.
39. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
40. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
41. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320–W324.
42. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *Proteomics Protocols Handbook*, ed Walker J (Humana Press, Totowa, NJ), pp 571–607.
43. Phillips K, de la Peña AH (2011) The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr Protoc Mol Biol* Chapter 10:Unit10.28.
44. Guerrero SA, Hecht HJ, Hofmann B, Biebl H, Singh M (2001) Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl Microbiol Biotechnol* 56(5-6):718–723.
45. Frirdich E, Vinogradov E, Whitfield C (2004) Biosynthesis of a novel 3-deoxy-D-*manno*-oct-2-ulosonic acid-containing outer core oligosaccharide in the lipopolysaccharide of *Klebsiella pneumoniae*. *J Biol Chem* 279(27):27928–27940.
46. Akiya S, Osawa T (1960) Nitrogen-containing sugars. VI. On the N, N-phthaloyl derivatives of D-glucosamine. *Chem Pharm Bull (Tokyo)* 8(7):583–587.
47. Song G, Liu H, Zhang W, Geng M, Li Y (2010) Synthesis and biological evaluation of cytotoxic activity of novel anthracene L-rhamnopyranosides. *Bioorg Med Chem* 18(14):5183–5193.
48. Klonis N, Sawyer WH (1996) Spectral properties of the prototropic forms of fluorescein in aqueous solution. *J Fluoresc* 6(3):147–157.
49. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
50. Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58(Pt 11):1948–1954.
51. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
52. Cuthbertson L, Kos V, Whitfield C (2010) ABC transporters involved in export of cell surface glycoconjugates. *Microbiol Mol Biol Rev* 74(3):341–362.
53. Jann B, Ahrens R, Dengler T, Jann K (1988) Structure of the capsular polysaccharide (K19 antigen) from uropathogenic *Escherichia coli* O25:K19:H12. *Carbohydr Res* 177:273–277.
54. Ward CK, Lawrence ML, Veit HP, Inzana TJ (1998) Cloning and mutagenesis of a serotype-specific DNA region involved in encapsulation and virulence of *Actinobacillus pleuropneumoniae* serotype 5a: Concomitant expression of serotype 5a and 1 capsular polysaccharides in recombinant *A. pleuropneumoniae* serotype 1. *Infect Immun* 66(7):3326–3336.
55. Fraysse N, et al. (2005) *Sinorhizobium meliloti* strain 1021 produces a low-molecular-mass capsular polysaccharide that is a homopolymer of 3-deoxy-D-*manno*-oct-2-ulosonic acid harboring a phospholipid anchor. *Glycobiology* 15(1):101–108.
56. Whitfield C (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* 75:39–68.
57. Raetz CRH, Whitfield C (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem* 71: 635–700.
58. Ovchinnikova OG, et al. (2012) Localization and molecular characterization of putative O antigen gene clusters of *Providencia* species. *Microbiology* 158(Pt 4):1024–1036.
59. Sun Y, et al. (2012) Genetic analysis of the *Cronobacter sakazakii* O4 to O7 O-antigen gene clusters and development of a PCR assay for identification of all *C. sakazakii* O serotypes. *Appl Environ Microbiol* 78(11):3966–3974.
60. Greenfield LK, Whitfield C (2012) Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* 356:12–24.
61. Merino S, Canals R, Knirel YA, Tomás JM (2015) Molecular and chemical analysis of the lipopolysaccharide from *Aeromonas hydrophila* strain AH-1 (Serotype O11). *Mar Drugs* 13(4):2233–2249.
62. Oxley D, Wilkinson SG (1988) Structural studies of glucorhamnans isolated from the lipopolysaccharides of reference strains for *Serratia marcescens* serogroups O4 and O7, and of an O14 strain. *Carbohydr Res* 175(1):111–117.
63. Izquierdo L, Merino S, Regué M, Rodriguez F, Tomás JM (2003) Synthesis of a *Klebsiella pneumoniae* O-antigen heteropolysaccharide (O12) requires an ABC 2 transporter. *J Bacteriol* 185(5):1634–1641.