Single polysaccharide assembly protein that integrates polymerization, termination, and chain-length quality control

Danielle M. Williams[a], Olga G. Ovchinnikova[a], Akihiko Koizumi[b,c], Iain L. Mainprize[a], Matthew S. Kimber[a], Todd L. Lowary[b,c], and Chris Whitfield[a,1]

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

Edited by Laura L. Kiessling, University of Wisconsin–Madison, Madison, WI, and approved December 27, 2016 (received for review August 16, 2016)

Lipopolysaccharides (LPS) are essential outer membrane glycolipids in most gram-negative bacteria. Biosynthesis of the O-antigenic polysaccharide (OPS) component of LPS follows one of three widely distributed strategies, and similar processes are used to assemble other bacterial surface glycoconjugates. This study focuses on the ATP-binding cassette (ABC) transporter-dependent pathway, where glycans are completed on undecaprenyl diphosphate carriers at the cytosol:membrane interface, before export by the ABC transporter. We describe *Raoultella terrigena* WbbB, a prototype for a family of proteins that, remarkably, integrates several key activities in polysaccharide biosynthesis into a single polypeptide. WbbB contains three glycosyltransferase (GT) modules. Each of the GT102 and GT103 modules characterized here represents a previously unrecognized GT family. They form a polymerase, generating a polysaccharide of [4)-α-Rha*p*-(1→3)-β-Glc*p*NAc-(1→] repeat units. The polymer chain is terminated by a β-linked Kdo (3-deoxy-D-*manno*-oct-2-ulosonic acid) residue added by a third GT module belonging to the recently discovered GT99 family. The polymerase GT modules are separated from the GT99 chain terminator by a coiled-coil structure that forms a molecular ruler to determine product length. Different GT modules in the polymerase domains of other family members produce diversified OPS structures. These findings offer insight into glycan assembly mechanisms and the generation of antigenic diversity as well as potential tools for glycoengineering.

microbial glycobiology | lipopolysaccharides | glycosyltransferases | glycan biosynthesis | molecular ruler

Bacterial surfaces possess an array of complex glycoconjugates (sugar-containing macromolecules) that play varied and vital roles in the biology of these organisms. In pathogens, glycoconjugates participate in adhesion, biofilm formation, and interaction with innate and adaptive immune responses. For example, in gram-negative bacteria such as *Escherichia coli*, long glycan chains in capsular polysaccharides and the O-antigen polysaccharide (OPS) components of lipopolysaccharide (LPS) molecules typically confer resistance to opsonophagocytosis and complement-mediated killing (1, 2). Effective protection depends on the amount and surface distribution of glycan, as well as glycan chain lengths suited for a given purpose. As an example, shorter OPS chains offer no resistance to complement-mediated killing, whereas chains longer than an optimal size potentially represent an unnecessary energy cost while offering no additional advantage (3, 4). Despite its fundamental importance in colonization and virulence of bacterial pathogens, the molecular mechanisms of glycan chain-length regulation are often poorly understood (5). However, OPS biosynthesis provides influential prototypes for understanding the guiding principles underpinning chain-length regulatory mechanisms in bacterial glycoconjugates in general.

The LPS glycolipid is a major component of the outer membrane of gram-negative bacteria. Classical LPS molecules consist of three structural regions: lipid A, core oligosaccharide, and OPS (reviewed in refs. 6 and 7). All OPS are long-chain polysaccharides comprised of repeating oligosaccharide subunits with hypervariable structures. A range of different repeat-unit structures may be produced by a given species, with each OPS structure defining an O-antigen serotype. The LPS molecule is synthesized by two pathways, one for the lipid A-core and the other for the OPS, and these converge in the periplasm with a ligation step that joins the two parts of the molecule, before its translocation to the cell surface (6–8). There are two well-distributed assembly strategies for OPS biosynthesis and both generate undecaprenyl diphosphate (Und-PP)–linked OPS intermediates and both end with O-antigen ligase (WaaL) enzyme transferring the nascent OPS from Un-PP–linked donor to lipid A-core acceptor. The two OPS-biosynthesis strategies differ in their polymerization mechanisms, components, and membrane topology of the reaction steps (reviewed in ref. 6). In nature, OPS chain lengths typically fall into a limited (modal) size range. The two OPS assembly systems use different approaches to establish the modal distribution before its ligation to lipid A-core (5).

One strategy for OPS biosynthesis is provided by the ATP-binding cassette (ABC) transporter-dependent assembly pathway and it provides the focus of this investigation. Here, the OPS is polymerized in the cytoplasm and then exported to the periplasm, by the pathway defining ABC transporter, for ligation (6, 9). This pathway has two variants distinguished by whether (or not) the ABC transporter is specific for the carbohydrate structure of a particular OPS substrate (reviewed in ref. 9). In systems where the ABC transporter is not specific, the Und-PP–linked glycan is polymerized by a membrane-associated complex of

Significance

Glycosyltransferase enzymes synthesize complex sugar-containing macromolecules that play pivotal roles in the biology of all cells. Bacteria produce a vast range of glycoconjugate structures that frequently dictate host–pathogen interactions. The composition and chain length of bacterial glycans are both important for biological function, yet the mechanisms for chain-length determination are generally poorly understood. Here, we describe a remarkable family of proteins in which each member contains multiple glycosyltransferase modules (often representing new glycosyltransferase families) for glycan polymerization and chain termination. The polymerase and terminator activities are separated by a molecular ruler that establishes product length. The modularity of these proteins provides an adaptable platform for generating antigenic diversity in nature and potential new strategies for glycoengineering.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: cwhitfie@uoguelph.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613609114/-/DCSupplemental.

glycosyltransferase (GT) enzymes. Polymerization and export are obligatorily coupled in this process and the resulting OPS products possess a broad range of chain lengths. *Klebsiella pneumoniae* O2a provides the prototype for this assembly strategy (10, 11). The polymannose OPSs shared by *E. coli* O8/O9/O9a and their *K. pneumoniae* O5/O3 counterparts offer the prototypes for the substrate-specific assembly pathway. A serotype-specific polymerase enzyme (designated WbdA), possessing two or more GDP-mannose-dependent GT catalytic sites, is responsible for building the repeat-unit structure of the Und-PP–linked polymannose OPS (12, 13). The polymerase activity is opposed by a chain terminator (WbdD) that adds a methyl or phosphomethyl residue (depending on the serotype) to the nonreducing ends, creating completed OPS chains with optimal lengths within a limited (modal) size range before the involvement of the ABC transporter (14, 15). Chain termination chemically defines OPS chain length by adding residues that prevent further chain extension. However, the system includes a mechanism for chain-length determination, which dictates when OPS chains reach the correct length for termination to take place. This is established by structural features of a membrane-associated complex composed of WbdA and WbdD (13, 16). Stoichiometry and geometry of the complex influences chain length control (5) but the distance between the polymerase and terminator catalytic sites provides the critical determinant of the OPS chain length. An extended coiled-coil structure in WbdD performs this function, acting as a molecular ruler (17). This cannot occur unless the polymerase operates with a distributive mechanism, where the growing nonreducing terminus of the glycan is released from the active site after each glycosyltransfer reaction. The in vitro properties of WbdA are consistent with this mechanism (12). Depending on the length of the glycan, it may then be available to the terminator (WbdD) active site, which is separated from the polymerase by the coiled-coil structure (17). To ensure that only optimal-sized (terminally capped) glycans are exported, the ABC transporter possesses a substrate-specific carbohydrate binding module (CBM) that recognizes the modified terminal structure (18, 19). This offers a quality-control step, so only mature OPS components are found in LPS molecules that reach the cell surface (18, 19). Although many of the details of this prototype system are now well established, it remains unclear how universally these key features apply to OPS synthetic pathways that produce dissimilar glycan structures.

Here, we report biosynthesis of an OPS produced by some isolates of *Raoultella terrigena* and *K. pneumoniae* (hereafter referred to as O12 after the *K. pneumoniae* serotype). This is an example of an OPS assembly pathway involving a substrate-specific ABC transporter. These bacteria share the same OPS repeat unit structure and gene cluster (Fig. 1), presumably reflecting lateral transfer (20). The α-Rha*p*-(1 →3)-β-Glc*p*NAc disaccharide repeat unit is capped with a nonreducing terminal residue of β-linked 3-deoxy-D-*manno*-octulosonic acid (β-Kdo*p*) (21, 22), which is recognized by a CBM in the ABC transporter (23). Here we describe the biosynthetic pathway and the central role of a remarkable enzyme that combines a chain terminator and polymerase, as well as a coiled-coil molecular ruler within a single polypeptide.

Results

The Biosynthesis Pathway for the O12 OPS. The gene cluster responsible for OPS biosynthesis in the O12 system is composed of eight ORFs (Fig. 1*A*). Several of the gene products have been investigated or can be confidently assigned functions from sequence similarities. This, together with the data provided in the experiments described below, leads to the functional assignments for the O12 biosynthesis proteins in Fig. 1*C*.

ABC transporters possess two transmembrane domains forming the membrane channel and two nucleotide binding domains that hydrolyze ATP to drive transport (24). In most OPS ABC exporters, these domains are encoded by separate genes

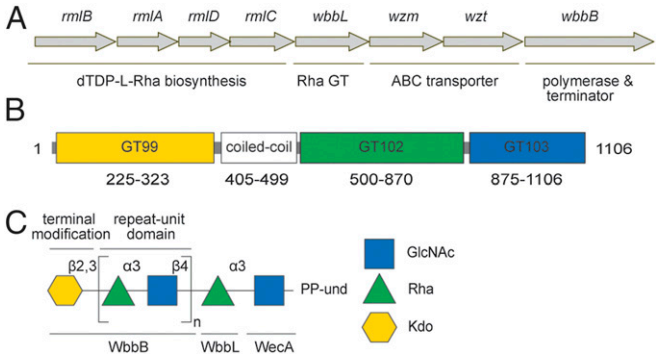

Fig. 1. Structure and biosynthesis of *K. pneumoniae* O12 OPS. (*A*) Organization of the *K. pneumoniae* O12 OPS biosynthesis gene cluster. (*B*) Predicted location of WbbB domains. The N-terminal GT99 (yellow) identifies the area covered in a validated functional construct (22). (*C*) Structure of the *K. pneumoniae* O12 Und-PP–linked OPS intermediate predicted from structures of the authentic OPS glycans (20, 21). The standard colored symbol nomenclature (www.functionalglycomics.org/static/consortium/Nomenclature.shtml) is used to describe oligosaccharide structures. The basis for the functional assignments of the gene products is described in the text.

(*wzm* and *wzt*, respectively) and the function of the O12 ABC transporter has been validated (23). The *rmlBADC* genes encode the conserved pathway converting glucopyranose-1-phosphate to dTDP-L-rhamnopyranose (dTDP-Rha*p*) (25), whereas the UDP-N-acetyl-D-glucosamine (UDP-Glc*p*NAc) and CMP-3-deoxy-D-*manno*-octulosonic acid (CMP-Kdo*p*) precursors are produced by "housekeeping enzymes," because they are essential for production of peptidoglycan (26) and the lipid A-core component of LPS (7). Biosynthesis of *K. pneumoniae* OPS is initiated by WecA (23), a member of the UDP-D-N-acetylhexosamine:polyprenol phosphate N-acetylhexosamine-1-phosphate transferases, which initiate a variety of glycan assembly pathways (27). In OPS assembly, the WecA product (Und-PP-Glc*p*NAc) is modified by one or more NDP-glycose-dependent GT activities, creating an appropriate acceptor for OPS polymerization. In the O12 OPS pathway, the candidate for this transition step is WbbL, an enzyme also found in *E. coli*, where it is a dTDP-Rha*p*-dependent GT forming α-Rha*p*-(1→3)-β-Glc*p*NAc disaccharide (28); the same catalytic activity has been demonstrated for the *K. pneumoniae* O12 homolog (29). O12 OPS polymerization requires only two GTs (WbbB and WbbL) in addition to the initiating WecA (23). However, it was unknown before the current study whether WbbL action is confined to a single early step or whether it also forms the same linkage during polymerization of repeat units. A previous study proposed WbbB was a UDP-Glc*p*NAc-dependent GT, with WbbL providing the rhamnosyltransferase for polymerization (20, 29).

Using domain/motif search tools, WbbB is predicted to possess three GT modules (Fig. 1*B*). The N-terminal domain contains the recently characterized chain-terminating β-Kdo*p* GT belonging to GT99 (22). BLAST conserved domain assignments (30) suggested two additional C-terminal putative GT domains annotated (in order from the N terminus) as "GT 1/RfaB-like" and "GT25." However, neither shares any significant sequence similarity with bona fide GT1 or GT25 proteins in the Carbohydrate-Active enzymes (CAZy) database (31) and the putative C-terminal GT domains in WbbB were not formally assigned to a CAZy family in their own right, because there were no data to support their function(s). Based on their novel sequence features and the biochemical assignment of their activities identified here, the C-terminal GTs define two new families in the CAZy database, GT102 and GT103, respectively. The data supporting their functions are presented below.

WbbB$^{540\text{-}1106}$ Is Sufficient for Polymerization of O12 OPS. In vitro reactions with purified proteins and synthetic acceptors defined

the minimal requirements for polymerization of the repeat-unit domain. The activity of WbbB$^{540\text{-}1106}$, possessing both the GT102 and GT103 domains (Fig. 1*B*), was examined using synthetic acceptors A and B (Fig. 2*A*).

Initial reactions were performed with a single donor substrate (UDP-Glc*p*NAc or dTDP-Rha*p*) to isolate specific activities. In these reactions, WbbB$^{540\text{-}1106}$ generated reaction products with slower migration than acceptors A and B on TLC (Fig. 2*B*) and their compositions were verified by MS (Fig. 2 *D* and *F*). Typically, TLC also revealed a series of fluorescein-labeled compounds reflecting photodegradation products, which were verified in subsequent MS analysis. Consistent with the authentic O12 OPS structure, modification of each acceptor occurred with only one of the donors. WbbB$^{540\text{-}1106}$ added a single Glc*p*NAc residue (Δ*m* 203) to acceptor A (MW 884), resulting in a major product with molecular weight (MW) 1,087. No activity was evident with dTDP-Rha*p* donor. Conversely, the MS spectrum of the reaction of WbbB$^{540\text{-}1106}$ with acceptor B and dTDP-L-Rha*p* revealed a major product with MW 1,233, consistent with the addition of a Rha*p* residue (Δ*m* 146) to acceptor B (MW 1,087). No addition to acceptor B was evident with UDP-Glc*p*NAc. These results indicated that WbbB$^{540\text{-}1106}$ possesses both Rha*p* and Glc*p*NAc GT activities, contrary to previous work (20, 29).

The ability of WbbB$^{540\text{-}1106}$ to polymerize O12 antigen was then examined using each acceptor in reactions containing both donor substrates. These reactions generated a range of slower-moving products on TLC (Fig. 2*B*). The MS spectrum of the reaction with acceptor A (Fig. 2*C*) contained a range of products, beginning with MW 1,087, which represents a tagged trisaccharide formed by the addition of a single Glc*p*NAc residue to acceptor A. The larger products are consistent with the extension of the trisaccharide by one or two additional disaccharide repeat units (Δ*m* 349), all with terminal Glc*p*NAc residues. The MS profile of reaction products using acceptor B revealed a series of products with additional repeat units (Δ*m* 349) (Fig. 2*E*). Interestingly, there was no evidence of any product terminating in Rha*p*.

To confirm the authentic O12 structure in the in vitro products, the products generated by WbbB$^{540\text{-}1106}$ using acceptor A and both donors were studied by NMR spectroscopy. The ^{1}H and ^{13}C NMR chemical shifts were assigned using a set of 2D experiments, including ^{1}H,^{1}H COSY, total correlation spectroscopy (TOCSY), rotating-frame nuclear Overhauser effect correlation spectroscopy (ROESY), ^{1}H,^{13}C heteronuclear single quantum coherence (HSQC), and heteronuclear multiple bond correlation (HMBC) (Table S1). The anomeric region of the ^{1}H NMR spectrum contained four signals of different intensities in the region δ 4.77–4.85, and a broad, low intensity signal at δ 4.46 (Fig. 3*A*). Analysis of the COSY and TOCSY spectra revealed spin systems for five sugar residues, including three β-Glc*p*NAc (labeled as **A**, **A′**, and **A′′**) and two α-Rha*p* residues (**B** and **B′**).

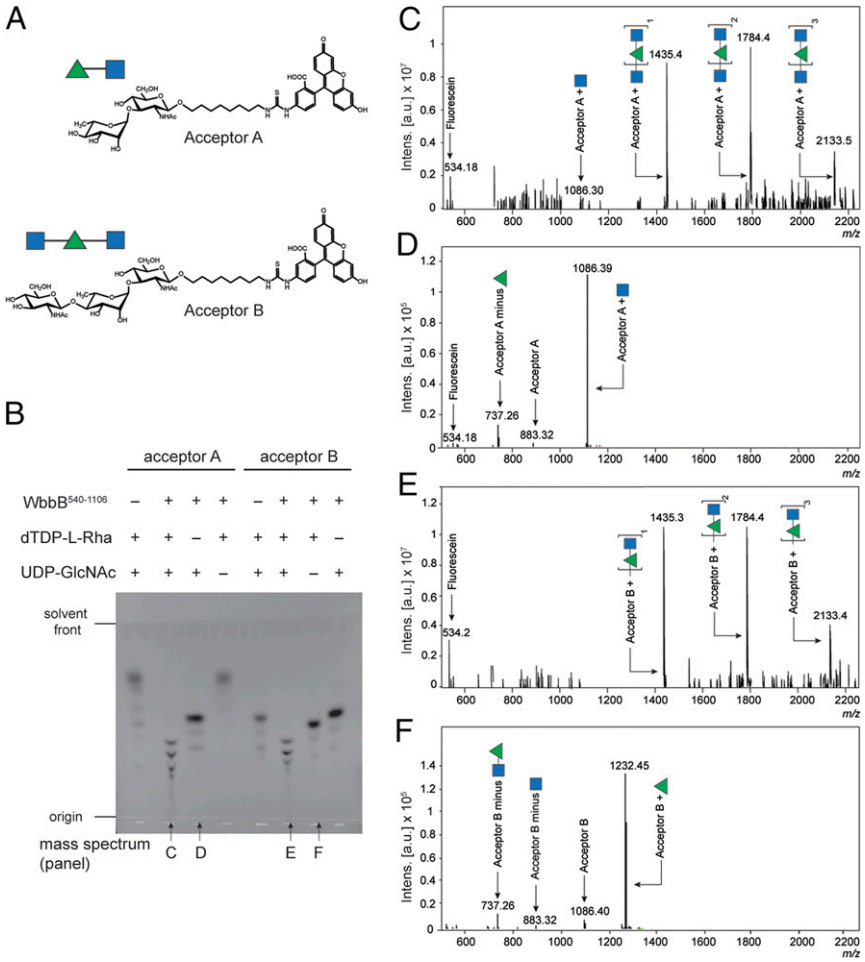

Fig. 2. In vitro activity of the putative polymerase, WbbB$^{540\text{-}1106}$, using single sugar donors. (*A*) Structures of the synthetic acceptors. (*B*) TLC analysis of the reaction mixtures obtained after incubating the enzyme with the identified acceptor (a synthetic di/trisaccharide attached via an aliphatic chain to fluorescein) and donor sugar(s). (*C*–*F*) Samples from the reactions indicated in *B* (identified under the relevant TLC lane) were analyzed by MS and the charge-deconvoluted electrospray ionization (ESI) mass spectra are presented.

The major series of signals belong to residues **A** and **B**, whereas the integral intensities of the signals belonging to residues **A′**, **A′′**, and **B′** are approximately four to five times lower. The positions of glycosylation were defined by the significant downfield displacement of the signals for residue **A** C-3 and residue **B** C-4 by 8.0 and 8.2 ppm, respectively, compared with their positions in the nonsubstituted monosaccharides (32). The sequence of sugar residues was determined using HMBC and ROESY experiments. The HMBC spectrum demonstrated interresidue correlations **A** H-1/**B** C-4 and **B** H-1/**A** C-3 at δ 4.81/81.4 and 4.85/82.8, respectively (Fig. 3*B*). The major series of signals in the NMR spectra corresponds to the internal disaccharide repeat units [4)-α-Rha*p*-(1→3)-β-Glc*p*NAc-(1→], which are identical to the native O12 repeat unit (21). Consistent with the MS data, no minor signals were observed that would correspond to terminal Rha*p* residues. Collectively, these data establish that the WbbB fragment containing GT102 and GT103 is sufficient for polymerization and rule out WbbL as an obligatory participant in chain extension.

In a two-site polymerase a distribution of glycan chains ending in either Glc*p*NAc or Rha*p* was anticipated, so the observation that all of the in vitro polymerization products terminated in Glc*p*NAc was surprising. The initial reactions included equimolar amounts of UDP-Glc*p*NAc and dTDP-Rha*p*. To determine whether this is a catalytic property of the polymerase, or a by-product of the reaction conditions, a series of reactions were performed using acceptor B, 5 mM dTDP-Rha*p*, and a titration series of UDP-Glc*p*NAc from 1 to 4 mM (Table S2). In these experiments, all of the higher mass peaks in the profile end in Glc*p*NAc, regardless of the dTDP-Rha*p*:UDP-Glc*p*NAc ratio. This observation could reflect different reaction rates in the two GT domains, where addition of Rha*p* is less efficient. The authentic O12 glycan is terminated by a β-(2→3)–linked Kdo*p* added to a terminal Rha*p* and this reaction is potentially complicated by competing with a more rapid Glc*p*NAc addition and further chain elongation. Interestingly, these reactions did not generate products extended by more than three O12 repeat units. Longer-chain products are not observed in reactions with varied NDP-sugar:acceptor ratios. Furthermore, the same product spectrum is obtained in reactions with added UDP or dTDP, indicating that product inhibition is not involved. It is unclear whether these unexpected catalytic properties translate to the full-length WbbB protein in its natural (membrane-associated) environment, or whether it is an artifact of the in vitro system. The results may reflect use of truncated proteins, where the absence of other domains may prevent optimal folding and/or acceptor binding. Unfortunately, native WbbB is prone to aggregation and degradation, precluding analysis of the GTs in the context of the full protein.

Assignment of Catalytic Activities to the GT102 and GT103 Domains. The data above establish WbbB$^{540\text{-}1106}$ as the O12 polymerase but additional experiments were required to clarify the identities of the Glc*p*NAc and Rha*p* GTs. A domain containing GT102 was expressed alone (WbbB$^{576\text{-}871}$) and shown to add a single Rha*p* residue to acceptor B, which was confirmed by TLC and MS (Fig. 4 *A* and *C*). This construct showed no activity with UDP-Glc*p*NAc and acceptor A, unequivocally identifying as a dTDP-L-Rha*p*-dependent GT. Attempts to generate active protein fragments containing only the GT103 domain were unsuccessful. To overcome this, site-directed mutagenesis was used to inactivate the GT102 domain to examine the activity of the GT103 domain in isolation. Three basic residues (R738, R739, and K740) were targeted for mutagenesis because they are strongly conserved among those (uncharacterized) proteins showing significant similarity to GT102 in BLAST searches. In addition, a role for basic residues in coordinating and stabilizing the phosphate leaving group is well-established in diverse GTs (e.g., refs. 33 and 34). In the absence of structural information to interpret possible functions, all three residues were mutated to alanine to maximize the strength of the phenotype; this variant eliminated the activity of the GT102 domain while retaining wild-type levels of N-acetylglucosaminyltransferase activity (Fig. 4 *B* and *D*). This result confirms the GT103 domain as the UDP-Glc*p*NAc-dependent Glc*p*NAc GT.

Chain-Length Regulation by WbbB. The chain-length-regulating molecular ruler provided by the coiled-coil structure in the *E. coli* O9a terminator component of the WbdD–WbdA complex has been established (17). Analysis of the WbbB sequence using the program COILS (35) predicted a coiled-coil region located

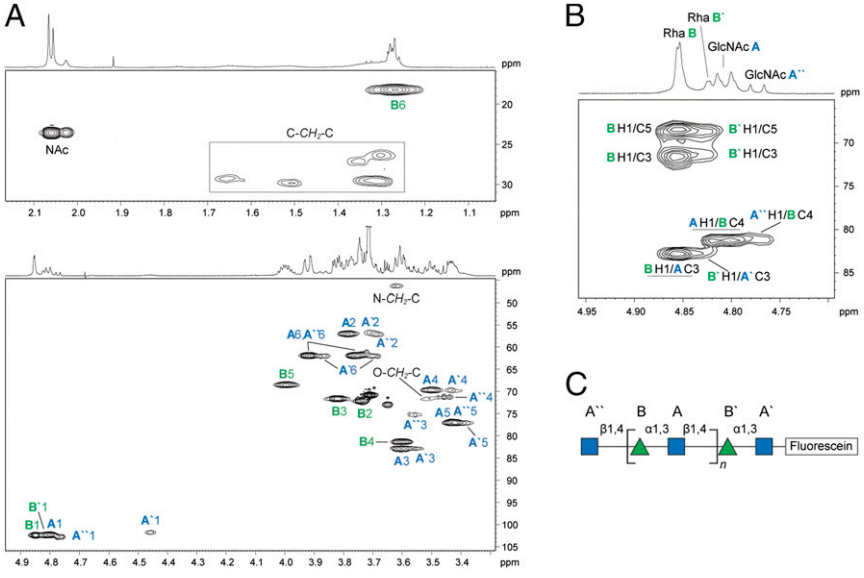

Fig. 3. NMR spectroscopy analysis of the polymeric product generated by WbbB$^{540\text{-}1106}$ using acceptor A. (*A*) Parts of the ^{1}H,^{13}C HSQC spectrum. The corresponding parts of the ^{1}H NMR spectrum are shown along the axis. The numbers refer to H/C pairs in sugar residues designated by letters as shown in *C*. The O-*CH*$_2$-C, C-*CH*$_2$-N and C-*CH*$_2$-C signals belong to an eight-carbon methylene linker. Stars indicate the signals from a polyethyleneglycol impurity whose source was not resolved. (*B*) Part of the ^{1}H,^{13}C HMBC spectrum. The underlined interresidue correlations between anomeric protons and the linkage carbon atoms demonstrate **A**(1→4)**B** and **B**(1→3)**A** linkages. (*C*) Structure of the enzymatic product.

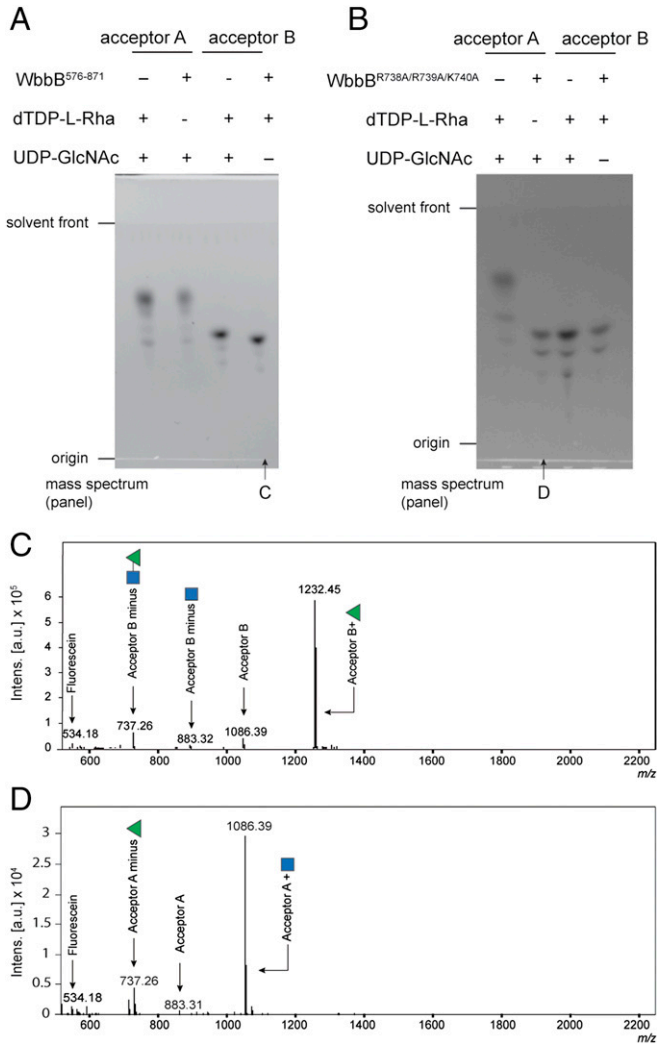

Fig. 4. In vitro activities of GT102 and GT103. (*A*) TLC analysis of the reaction mixtures obtained after incubating the WbbB$^{576\text{-}871}$ fragment with the identified acceptor and sugar donors. (*B*) The corresponding experiments with WbbB$^{R738A/R739A/K740A}$, which inactivated the GT102 site and allowed evaluation of GT103. As expected, each enzyme showed activity with a single combination of acceptor and NDP-sugar donor. (*C* and *D*) Charge-deconvoluted ESI mass spectra of the reaction mixtures highlighted in *A* and *B*.

between the GT99 and GT102 domains (Fig. 1*B*). Coiled-coils are widespread structural motifs and typically consist of two to five α-helices wrapped around one another forming a left-handed helical supercoil. The dimerization of the GT99 domain (22) suggests that two helices are involved in WbbB. Each helix is made of heptad repeats (36). COILS predicted that the WbbB coiled-coil consisted of two blocks of six heptad repeats flanking an 11-residue sequence that degenerates from the canonical heptad after six residues (Fig. 5*A*). The involvement of the WbbB coiled-coil domain in chain-length regulation was assessed by systematically varying the size of this region in WbbB. Silver-stained SDS/PAGE of whole-cell lysates identifies OPS species that are exported and linked to lipid A-core (23) (Fig. 5*B*). The OPS-carrying LPS species generated by wild-type WbbB (WbbB 6,6) formed a tight cluster of molecules, reflecting the native chain length. Reducing the number of heptads from either side of the nonheptad region resulted in an overall shift to shorter OPS chain lengths, compared with the wild-type construct. Increasing the number of heptad repeats resulted in an upward shift in the maximum chain length. Interestingly, the LPS species for the 2,2 construct were indistinguishable from those generated by WbbB with the entire coiled-coil domain removed. Most of these coiled-coil variants resulted in a broadening of the observed size distribution. Removal of the central region, which is not predicted to form a canonical heptad, led to a slight reduction in chain length, but, unlike the other variants, it did not result in any significant broadening of the range of chain lengths. To confirm that these results were not simply due to altered expression of the WbbB proteins their amounts were confirmed using the C-terminal FLAG tag (Fig. 5*D*). To focus only on the Und-PP–linked OPS the WbbB variants were examined in a background lacking the Wzt component of the ABC transporter, with detection performed by Western immunoblots with anti-O12 antibodies (Fig. 5*C*). The shifts in size distributions were consistent with those seen in completed LPS, but the distributions of Und-PP–OPS were generally broader with more molecules in the upper size ranges. The data are consistent with the WbbB coiled-coil domain acting as a molecular ruler, with a function analogous to the WbdD protein in *E. coli* O9a (17).

In *E. coli* O9a, the chain length of the OPS is affected by the stoichiometry of the interacting polymerase and terminator proteins, presumably within the context of the known complex (5). This can be manipulated experimentally (and potentially under natural conditions) because the two key activities are contributed by two separate polypeptides. To investigate whether the same principle exists in the O12 system, we investigated the size distribution of OPS assembled in vivo by cells overexpressing separate polymerase and catalytically active terminating β-Kdo*p* GT99 [WbbB$^{2\text{-}401}$ (22)] domains, in the presence of full-length WbbB. These experiments were performed in an export-deficient background. Two polymerase constructs, WbbB$^{401\text{-}1106}$ and WbbB$^{540\text{-}1106}$, were tested and these differed in the inclusion (or not) of the coiled-coil domain (Fig. 1*A*). Overexpression of the WbbB$^{2\text{-}401}$ terminating GT99 domain caused an overall downward shift in the size distribution and a broadening of the modal distribution in the Und-PP–OPS (Fig. 5*E*). In contrast, overexpression of the polymerase constructs had the opposite effect, with a significant increase in the maximum chain length and a depletion of shorter chain lengths. The presence or absence of the coiled-coil domain did not affect the outcome.

Discussion

The biosynthesis of the O12 OPS requires a remarkable trifunctional protein WbbB that is responsible for all of the steps in assembly: polymerization, termination (i.e., the act of adding a nonreducing terminal residue to prevent further extension), and chain-length regulation (i.e., dictating the length of OPS that are candidates for termination). The essential components (polymerase, terminator, and molecular ruler) are required in both the O12 system and the *E. coli* O9a prototype, suggesting a shared biosynthesis strategy. However, the latter requires two proteins (WbdA and WbdD) to provide these key assembly steps (Fig. 6). The N terminus of WbbB contains the GT99 β-Kdo*p* transferase (terminator) and its activity and structure have been reported (22). The C-terminal GTs show no significant similarity to any of the existing families in the CAZy database (31), and with the determination of their catalytic activities they now form founding members of the new GT102 and GT103 families. They operate in a two-site model of polymerization by inverting mechanisms (37), where the anomeric linkages in the activated donors (dTDP-β-L-Rha*p* and UDP-α-D-Glc*p*NAc) are inverted in the product [4)-α-Rha*p*-(1→3)-β-Glc*p*NAc-(1→]. Taken in aggregate, each of the three GT domains of WbbB represents a new GT family.

The role of a coiled-coil structure as a molecular ruler for OPS chain-length determination was first demonstrated in *E. coli* O9a (17). The specificity of the ABC transporter for terminated O9a OPS chains precludes the premature export of shorter chains (19).

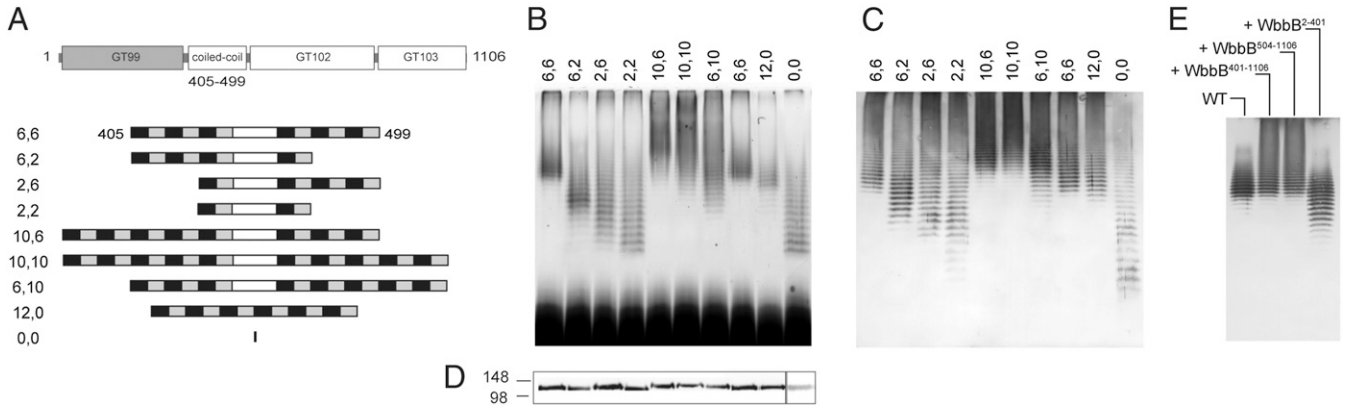

Fig. 5. The coiled-coil domain of WbbB is a molecular ruler to determine OPS chain length. (*A*) The architecture of the coiled-coil region identified by COILS (35). The wild-type WbbB coiled-coil domain (designated 6,6) consists of two identical blocks of six heptads (black/gray boxes) flanking an 11-residue region that does not correspond to a canonical heptad structure. The series of engineered coiled-coil regions are shown with their designations. Each variant is named for the number of heptad repeats on either side of the nonheptad region. In 12,0, the central non-coiled-coil region was deleted, whereas the entire coiled-coil region was removed in 0,0. (*B*) Silver-stained SDS/PAGE gel of the LPS produced by the WbbB variants transformed in transport-proficient *E. coli* CWG1219 (pWQ673; *wbbL-wzm-wzt*). (*C*) Western immunoblot with anti-O12 antibodies showing Und-PP–linked OPS in an export-deficient background provided by CWG1219 (pWQ676; *wbbL-wzm*). (*D*) Anti-FLAG Western immunoblots to assess amounts of WbbB-variants. The amount of the 0,0 protein was lower, so it was necessary to increase the contrast to detect it (hence the separate box). (*E*) Western immunoblot with anti-O12 antibodies showing the effect of overexpressing the GT99 terminator ($WbbB^{2\text{-}401}$) or polymerase ($WbbB^{401\text{-}1106}$; $WbbB^{504\text{-}1106}$) on the chain length distribution of Und-PP–OPS.

These general principles are conserved in *K. pneumoniae* O12. In *E. coli* O9a, a coiled-coil with an estimated size of ∼200 Å gives rise to a modal chain-length distribution with an average of four tetrasaccharide repeat units. Although accurate O12 repeat-unit counts cannot be made because of the absence of low-MW species in the profile, dictated by the regulatory system, the expanded ranges of chain lengths seen in some coiled-coil variants allow an estimated wild-type average of ∼20 repeat units (Fig. 5*B*). In WbbB, the presence of a noncanonical heptad sequence flanked by canonical heptads makes it difficult to assess the length of the molecular ruler with certainty. However, units of six heptads have an estimated size of ∼60 Å, based on the parameters determined for WbdD in *E. coli* O9a. This would give rise to an average chain length of seven disaccharide repeats, each contributing ∼9 Å in chain length (calculated using the GLYCAM server; glycam.org/). A ∼120-Å spacer seems too small for the observed LPS products in PAGE, if the coiled-coil domain alone dictates spacing. Given that the coiled-coil structure is fused to the polymerase domain, it seems likely that the physical bulk of the polymerase domain acts as a substantive (∼60 Å) extension to the membrane-proximal end of the ruler; this idea is consistent with our finding that a minimal degree of polymerization is retained even in the shortest coiled-coil variants. In, *E. coli* O9a the stoichiometry and architecture of the complex also influences chain length distribution (5), and although the O12 system can be experimentally manipulated to alter chain length the deployment of the key activities in a single WbbB protein limit this in vivo.

In *E. coli* O9a, the coiled-coil region plays additional roles in interacting with the active site of the terminating kinase and ordering two loops needed for proper substrate binding, as well as stabilizing the WbdD trimer (17). The true oligomeric state of the full-length WbbB protein is unknown but the GT99 β-Kdo*p* GT is a dimer in solution and in the solved crystal structures (22). For the coiled-coil domain of WbbB to be effective as a spacer, the C-terminal end of the molecule must presumably be anchored to the membrane. In addition, the polymerase domain requires proximity to the membrane to access its Und-PP–linked acceptor molecules. In the O9a system, the amphipathic helix of WbdD provides anchorage to the membrane with defined orientation. Analysis of the polymerase domain of WbbB indicates the presence of several candidate membrane-interacting helices that may play an analogous role.

An important open question is how distributed these types of modular enzymes are in bacterial glycan biosynthesis. Sequence data identify several other members from the genera *Klebsiella*, *Raoultella*, *Aeromonas*, and *Serratia* among proteins containing GT99 family modules in the CAZy database (Table S3). Their GT99 sequences share ∼60% or greater similarity and share the same N-terminal location, upstream of a high-confidence predicted coiled-coil region. They all possess one or more C-terminal GT candidates. Although the structures of the polysaccharide repeat units are available for only two other examples for correlation with potential GT modules (Fig. 7), some conserved functional properties clearly emerge. The carbohydrate backbones of the OPS from *K. pneumoniae* O12 and *Aeromonas hydrophila* AH-1 are identical, although AH-1 OPS also possesses variable acetylation at the 2 and 3 positions of the Rha*p* residue (38). The OPS-biosynthesis genetic locus of AH-1 contains a putative acetyltransferase gene. Terminal residues were not reported for AH-1 OPS but the presence of a GT99 domain is a strong indicator that a terminal β-Kdo*p* residue exists. Furthermore, the Wzt proteins from these bacteria possess conserved C-terminal sequences, which are predicted to encode a CBM similar to that recognizing the terminal β-Kdo*p* in *K. pneumoniae* O12 (23). Both bacteria possess similar WbbL homologs, indicating the biosynthesis pathways and principles are conserved. The polymerase domains from *A. hydrophila* and *R. terrigena* share <60% overall similarity but, as expected from the conserved OPS repeat-unit structure, sequence and secondary structure alignments reveal strong candidates for GT102 (59% similarity/41% identity) and GT103 (61% similarity/45% identity) (Fig. S1).

The *Serratia marcescens* O4 OPS repeat unit differs from *K. pneumoniae* O12 by possessing a [3)-α-Rha*p*-(1→4)-α-Glc*p*-(1→] disaccharide backbone (39). This organism contains genes encoding WbbL and the modular protein, WbbK, containing a GT99 β-Kdo*p* GT active site (Fig. 7). The Wzt protein from this OPS-assembly system also contains a putative β-Kdo*p*–specific CBM, suggesting a biosynthetic pathway and terminal modification similar to *K. pneumoniae* O12. However, previous work has indicated that the ABC transporters were not exchangeable between *S. marcescens* O4 and *K. pneumoniae* O12 (29), which presumably reflects linkage differences in the terminal structure recognized by the CBMs. BLAST predictions identify two C-terminal GT modules identified as "GT2-like" and "GT4 or RfaB-like." The former shares 57% similarity/29% identity with

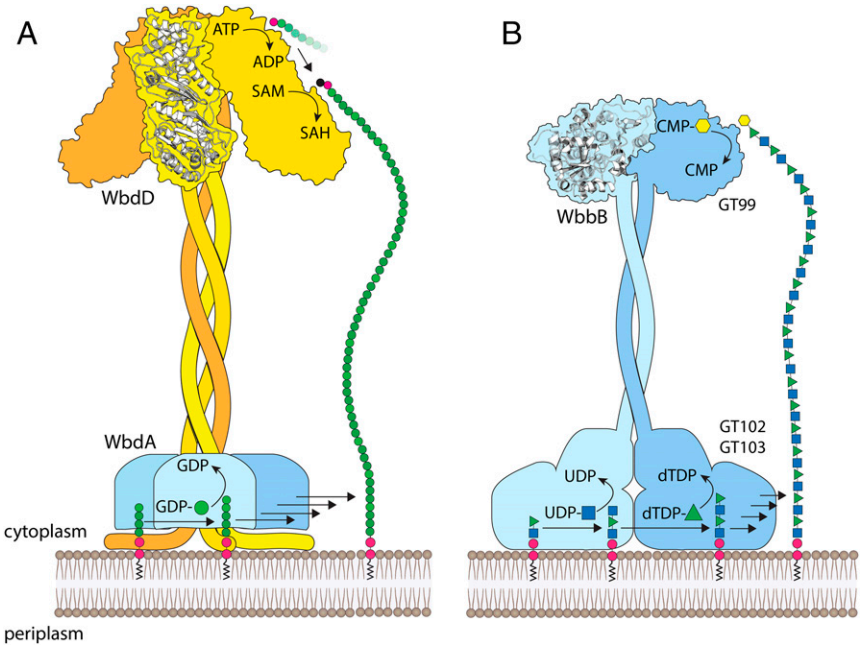

Fig. 6. Cartoon representation of the polymerization and chain termination mechanisms from *E. coli* O9a and *K. pneumoniae* O12. (*A*) In *E. coli* O9a, WbdA is a polymerase containing two GDP-mannose-dependent mannosyltransferase catalytic sites and generates a polymannose glycan with a tetrasaccharide repeat unit defined by linkage sequences. The terminator is WbdD and possesses a membrane-associating amphipathic helix in close proximity to a region that interacts with the polymerase. Chain termination requires a dual kinase, methyltransferase domain whose structure has been solved. (*B*) *K. pneumoniae* O12 WbbB provides both polymerase (GT102 and GT103) and terminator (GT99) functions. The GT99 domain of WbbB forms dimers and this is therefore predicted for the full-length protein. The positioning of the polymerase is also predicted based on the model for *E. coli* O9a and the need to elongate an Und-PP–linked glycan anchored in the membrane. The function of the coiled-coil domains as molecular rulers requires polymerase reactions that are distributive (i.e., the growing glycan must be released from the active site after each step in elongation), to afford the opportunity for the growing glycan to interact with the spatially separated terminator active site once the glycan length is sufficient. The in vitro properties of WbdA and the O12 polymerase are consistent with a distributive mechanism.

GT102, with the alignment being strongest in the predicted C-terminal part of the GT (Fig. S1). The "GT4 or RfaB-like" motif presumably identifies the retaining α-glucosyltransferase, which uses UDP-α-D-Glc*p* as its donor and this shares no significant similarity with GT103 but some sequence relatedness with GT102 (42% similarity/22% sequence identity) (Fig. S1).

In the other examples of these multidomain proteins there are different combinations of putative GT motifs that are identified in BLAST as "GT1," "GT2," "GT25," and "GT4/RfaB-like" (Table S3). Most possess two separate GT motifs downstream of a coiled-coil region and the resulting structures are therefore predicted to contain disaccharide repeat units. The proteins from *K. pneumoniae* (AAN06494.1) and *Serratia fonticola* (ALX94372) provide exceptions, where only one GT motif is identified from sequence alone. In terms of the evolution of glycan diversity, (re)combining different GT modules behind a more conserved GT99 module and coiled-coil region allows the production of new structures within the same conceptual model. This general concept has been seen in other assembly systems, such as the production of *Salmonella* O antigens via the Wzy-dependent pathway (40), but in those cases diversity requires acquisition of new independent GT genes, rather than modules within a single polypeptide.

Aspects of this modular protein approach may also be used for the O-glycosylation of some S-layer proteins. In *Geobacillus stearothermophilus*, one protein (WsaE) is partially responsible for polymerization, as well as catalyzing chain termination (41). Sequence analysis of the Wzt homolog from this system reveals a CBM. The WsaE protein is predicted to possess a coiled-coil between the methyltransferase and rhamnosyltransferase domains, which suggests chain-length regulation in this system as well. In contrast, the polymerases for some gram-negative capsular polysaccharides synthesized by ABC transporter-dependent pathways often contain more than one GT module (e.g., ref. 42) but these processes operate without chain-terminating residues or coiled-coil molecular rulers and the ABC transporters lack CBMs (43).

In summary, the findings reported here provide insight into the assembly of bacterial polysaccharides via the ABC transporter-dependent pathway and a molecular strategy for producing glycan diversity. The inclusion of polymerization, termination, and chain-length regulation in a single protein significantly extends our understanding of the range of organizational possibilities for mechanisms that operate via a defined chain terminator mechanism. In addition, this work confirms the principle of coiled-coil rulers for establishing glycan chain lengths in these systems. Parallels between WbbB and the proteins involved in the biosynthesis of polyketides are obvious. These polyketide synthases are also single polypeptides containing multiple domains with distinct functions in the assembly of natural products (44). The fascinating integration of activities in a single protein also points to a potential avenue for glycoengineering to produce therapeutic proteins and vaccines. The inclusion of selected GT modules in a recombinant protein would afford development of a glycosylation cassette that modifies appropriate acceptors, such as glycosylated proteins and lipids, with a glycan of defined structure and chain length. In this regard we note that there has been considerable success in engineering polyketide synthases to produce novel structures with a host of biological functions (45). Such applications will be dependent on understanding the architectural principles from structures of prototypes such as WbbB.

Experimental Procedures

Bacterial Strains, Plasmids, and Growth Conditions. The bacterial strains used in this study are *E. coli* BL21 [B F$^-$ *dcm ompT hsdS*(r_B- m_B-) *gal* [*malB*$^+$]$_{K-12}$ (λ^S)] (Novagen), Top10 [F$^-$, *mcr*A Δ(*mrr-hsdRMS-mcrBC*) φ80, *lacZΔM15, ΔlacX74, deoR, nupG, recA1, araD139*, Δ(*ara-leu*)*7697, galU, galK, rpsL*(Str^r), *endA1*] (Invitrogen), and CWG1219 (Top10 *Δwzx-wbbK, ΔgtrA*) (46). Bacteria were

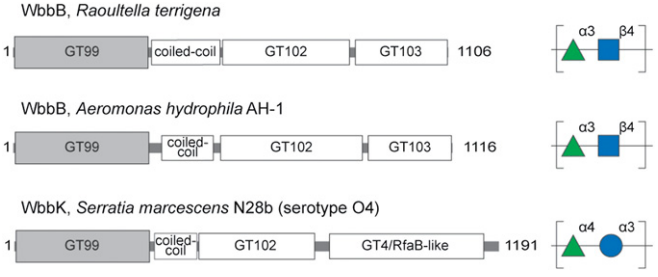

Fig. 7. Modular proteins related to WbbB in other bacterial genera. Several members of the GT99 family possess an organization resembling WbbB (Table S5) but only two are from organisms with known OPS structures. They are WbbB from *A. hydrophila* AH-1 (AKL88477) and WbbK [formerly WbbA (39)] from *S. marcescens* N28b (AAC00183).

grown in lysogeny broth (LB) at 37 °C. When required, media contained one or more of the following supplements: D-glucose (0.2%, wt/vol), L-arabinose (0.02%, wt/vol), isopropyl-β-D-1-thiogalactoside (IPTG; 1 mM), ampicillin (100 μg/mL), anhydrotetracycline (2.5 μg/mL), chloramphenicol (34 μg/mL), or kanamycin (50 μg/mL). The plasmids are listed in Table S4.

Purification of the WbbB GT Domains. Proteins containing both GT102 and GT103 GT modules (corresponding to WbbB$^{540\text{-}1106}$), or GT102 alone (WbbB$^{576\text{-}871}$), were purified from *E. coli* BL21 (DE3) transformed with either pWQ867 or pWQ878, respectively. Details of expression and purification conditions are provided in *SI Experimental Procedures*. Briefly, cells were resuspended in 25 mL Buffer A (0.1 M sodium phosphate buffer containing 250 mM NaCl, pH 7.4) with Complete Mini EDTA-free protease inhibitor tablets (Roche Applied Science) and lysed by sonication. Protein was purified from a cell and membrane-free supernatant obtained by sequential centrifugation culminating in a final step at 100,000 × *g* for 60 min. His_6-tagged proteins were purified using a Ni^{2+}-affinity column and the imidazole-containing eluant was exchanged with buffer A.

Synthesis of dTDP-L-Rhamnopyranose. Synthesis of dTDP-L-rhamnopyranose was performed essentially as previously described (47). Briefly, a 5-mL reaction contained 20 μmol of dTDP-D-glucopyranose, 1 μmol of NAD^+, 100 μmol of ammonium formate, 250 μg/mL each of purified RmlB, RmlC, and RmlD, and 3.5 units of formate dehydrogenase from *Candida boidinii* (Sigma) in buffer B (0.1 M Tris·HCl, pH 7.0). After reaction for 2 h at 37 °C, proteins were removed by ultrafiltration in a 3,000 molecular weight cutoff (MWCO) Vivaspin filtration unit (Sartorius Biolab Products). dTDP-L-Rha*p* product was confirmed using an Agilent LC-UHD Q-TOF instrument, operated in negative mode, in the Mass Spectrometry Facility at the University of Guelph Advanced Analysis Centre. The MS spectrum revealed a single major peak at *m/z* 547, which is the expected mass for dTDP-L-Rha*p*. Because the reaction resulted in essentially quantitative conversion of substrate to product, the protein-free reaction mixture was used without further purification in GT assays.

In Vitro GT Activities. Two synthetic oligosaccharide acceptors carrying a fluorescein moiety for detection (Fig. 2*A*) were used as acceptor substrates to determine GT functions. Synthesis of acceptor A has been reported (22) and the synthetic scheme for acceptor B is provided in Fig. S2, with the experimental details given in *SI Experimental Procedures*. Acceptor A [α-Rha*p*-(1→3)-β-Glc*p*NAc-octyl-fluorescein] represents the disaccharide repeat unit of *K. pneumoniae* O12 OPS and is the presumed acceptor for the *N*-acetylglucosaminyltransferase activity required for polymerization, whereas acceptor B [β-Glc*p*NAc-(1→4)-α-Rha*p*-(1→3)-β-Glc*p*NAc-octyl-fluorescein] is the presumed acceptor for the rhamnosyltransferase. Unless indicated otherwise, reactions were performed in 20-μL reaction volumes of buffer C (50 mM Hepes, pH 7.5), containing 10 μM enzyme (purified WbbB$^{576\text{-}871}$ or WbbB$^{540\text{-}1106}$), 0.2 mM acceptor, and 5 mM UDP-Glc*p*NAc and/or dTDP-L-Rha*p*. Reactions were incubated at 37 °C for 30 min. A 1-μL aliquot from each reaction mixture was then spotted on an aluminum foil silica gel 60 F_{254} TLC plate (EDM Millipore). TLC plates were developed with ethyl acetate:water:1-butanol:acetic acid (5:4:4:2.5) and fluorescent reaction products were detected with a hand-held UV lamp.

Structural Analysis of Reaction Products. Mass spectra of the reaction products were obtained in the University of Guelph Advanced Analysis Centre, using either an Agilent LC-UHD Q-TOF instrument operated in positive mode or, for polymerized product, a Bruker AmaZon SL LC-MS^n in positive mode, using UV detection at 437 nm. To generate sufficient product for NMR analysis, acceptor A was used and the reaction volume was scaled up to 500 μL. After 1 h, the reaction mixture was diluted in 10 mL of distilled H_2O and loaded into Sep-Pak Cartridge (Waters). The cartridge was washed with water, and the products were eluted in 6 mL of 60% (vol/vol) aqueous acetonitrile and then concentrated using a SpeedVac. The polymerized product was dried twice in a SpeedVac from a 99.9% D_2O solution and dissolved in 99.96% D_2O. NMR spectra were recorded at 35 °C on a Bruker Avancell 600-MHz spectrometer equipped with a cryoprobe in the NMR Facility at the University of Guelph Advanced Analysis Center. Internal sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0, δ_C −1.6) was used 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 of 100 and 200 ms were used in TOCSY and ROESY experiments, respectively. The HMBC spectroscopy experiment was optimized for the $J_{H,C}$ coupling constant 8 Hz.

Functional Analysis of WbbB Coiled-Coil Variants. Eight WbbB coiled-coil variants (Fig. 5) were constructed, generating plasmids pWQ869-pWQ876. Details of the engineering strategy are given in *SI Experimental Procedures*. Each plasmid carrying a mutant *wbbB* gene was cotransformed with either pWQ673 (*wbbL-wzm-wzt*) or pWQ676 (*wbbL-wzm*) into *E. coli* CWG1219 (*Δwzx-wbbK ΔgtrA*). To assess effects on OPS chain lengths, cell samples were collected for detection of LPS by silver-nitrate staining and Und-PP–linked OPS by Western immunoblotting using anti-O12 antiserum. C-terminal FLAG tags were exploited for Western immunoblotting to detect expression of WbbB derivatives.

ACKNOWLEDGMENTS. We thank J. D. King and M. Leeson for their contributions during the early stages of this project. This work was supported by individual Natural Sciences and Engineering Research Council Discovery Grants (to M.S.K., T.L.L., and C.W.). These authors are also supported by GlycoNet (the Canadian Glycomics Network, National Centres of Excellence Program). A.K. was the recipient of a postdoctoral fellowship from the Naito Foundation (Japan). T.L.L. and C.W. are Canada Research Chairs.

1. Cress BF, et al. (2014) Masquerading microbial pathogens: Capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol Rev* 38(4):660–697.
2. Lerouge I, Vanderleyden J (2002) O-antigen structural variation: Mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol Rev* 26(1):17–47.
3. Grossman N, et al. (1987) Lipopolysaccharide size and distribution determine serum resistance in *Salmonella montevideo*. *J Bacteriol* 169(2):856–863.
4. Burns SM, Hull SI (1998) Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic *Escherichia coli* O75:K5. *Infect Immun* 66(9):4244–4253.
5. King JD, Berry S, Clarke BR, Morris RJ, Whitfield C (2014) Lipopolysaccharide O antigen size distribution is determined by a chain extension complex of variable stoichiometry in *Escherichia coli* O9a. *Proc Natl Acad Sci USA* 111(17):6407–6412.
6. Raetz CRH, Whitfield C (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem* 71:635–700.
7. Whitfield C, Trent MS (2014) Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* 83:99–128.
8. Okuda S, Sherman DJ, Silhavy TJ, Ruiz N, Kahne D (2016) Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nat Rev Microbiol* 14(6):337–345.
9. Greenfield LK, Whitfield C (2012) Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* 356:12–24.
10. Kos V, Cuthbertson L, Whitfield C (2009) The *Klebsiella pneumoniae* O2a antigen defines a second mechanism for O antigen ATP-binding cassette transporters. *J Biol Chem* 284(5):2947–2956.
11. Kos V, Whitfield C (2010) A membrane-located glycosyltransferase complex required for biosynthesis of the D-galactan I lipopolysaccharide O antigen in *Klebsiella pneumoniae*. *J Biol Chem* 285(25):19668–19687.
12. Greenfield LK, et al. (2012) Biosynthesis of the polymannose lipopolysaccharide O-antigens from *Escherichia coli* serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. *J Biol Chem* 287(42):35078–35091.
13. Liston SD, et al. (2015) Domain interactions control complex formation and polymerase specificity in the biosynthesis of the *Escherichia coli* O9a antigen. *J Biol Chem* 290(2):1075–1085.
14. 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.
15. Hagelueken G, et al. (2012) Structure of WbdD: A bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in *Escherichia coli* O9a. *Mol Microbiol* 86(3):730–742.
16. Clarke BR, Greenfield LK, Bouwman C, Whitfield C (2009) Coordination of polymerization, chain termination, and export in assembly of the *Escherichia coli* lipopolysaccharide O9a antigen in an ATP-binding cassette transporter-dependent pathway. *J Biol Chem* 284(44):30662–30672.
17. 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.
18. Cuthbertson L, Powers J, Whitfield C (2005) The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in *Escherichia coli* serotypes O8 and O9a. *J Biol Chem* 280(34):30310–30319.
19. Cuthbertson L, Kimber MS, Whitfield C (2007) Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc Natl Acad Sci USA* 104(49):19529–19534.
20. 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.
21. 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.

22. Ovchinnikova OG, et al. (2016) Bacterial β-Kdo glycosyltransferases represent a new glycosyltransferase family (GT99). *Proc Natl Acad Sci USA* 113(22):E3120–E3129.
23. Mann E, Mallette E, Clarke BR, Kimber MS, Whitfield C (2016) The *Klebsiella pneumoniae* O12 ATP-binding cassette (ABC) transporter recognizes the terminal residue of its O-antigen polysaccharide substrate. *J Biol Chem* 291(18):9748–9761.
24. Locher KP (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* 23(6):487–493.
25. Dong C, et al. (2003) A structural perspective on the enzymes that convert dTDP-d-glucose into dTDP-l-rhamnose. *Biochem Soc Trans* 31(Pt 3):532–536.
26. Lovering AL, Safadi SS, Strynadka NCJ (2012) Structural perspective of peptidoglycan biosynthesis and assembly. *Annu Rev Biochem* 81:451–478.
27. Price NP, Momany FA (2005) Modeling bacterial UDP-HexNAc: Polyprenol-P HexNAc-1-P transferases. *Glycobiology* 15(9):29R–42R.
28. Stevenson G, et al. (1994) Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol* 176(13):4144–4156.
29. 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.
30. Marchler-Bauer A, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222–D226.
31. 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.
32. Jansson PE, Kenne L, Widmalm G (1991) CASPER: A computer program used for structural analysis of carbohydrates. *J Chem Inf Comput Sci* 31(4):508–516.
33. Pak JE, Satkunarajah M, Seetharaman J, Rini JM (2011) Structural and mechanistic characterization of leukocyte-type core 2 β1,6-N-acetylglucosaminyltransferase: A metal-ion-independent GT-A glycosyltransferase. *J Mol Biol* 414(5):798–811.
34. Martinez-Fleites C, et al. (2006) Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chem Biol* 13(11):1143–1152.
35. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252(5009):1162–1164.
36. Mason JM, Arndt KM (2004) Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem* 5(2):170–176.
37. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem* 77:521–555.
38. 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.
39. 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.
40. Liu B, et al. (2008) Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* 32(4):627–653.
41. Steiner K, et al. (2008) Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. *J Biol Chem* 283(30):21120–21133.
42. 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.
43. Cuthbertson L, Kos V, Whitfield C (2010) ABC transporters involved in export of cell surface glycoconjugates. *Microbiol Mol Biol Rev* 74(3):341–362.
44. Weissman KJ (2015) The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* 11(9):660–670.
45. Williams GJ (2013) Engineering polyketide synthases and nonribosomal peptide synthetases. *Curr Opin Struct Biol* 23(4):603–612.
46. Mann E, Ovchinnikova OG, King JD, Whitfield C (2015) Bacteriophage-mediated glucosylation can modify lipopolysaccharide O-antigens synthesized by an ATP-binding cassette (ABC) transporter-dependent assembly mechanism. *J Biol Chem* 290(42):25561–25570.
47. Graninger M, Nidetzky B, Heinrichs DE, Whitfield C, Messner P (1999) Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *J Biol Chem* 274(35):25069–25077.
48. Makarova O, Kamberov E, Margolis B (2000) Generation of deletion and point mutations with one primer in a single cloning step. *Biotechniques* 29(5):970–972.
49. Darveau RP, Hancock REW (1983) Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa and Salmonella typhimurium* strains. *J Bacteriol* 155(2):831–838.
50. Tsai CM, Frasch CE (1982) A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal Biochem* 119(1):115–119.
51. Schmidt RR, Hoffmann M (1983) Synthesis of C-α- and C-β-D-glucopyranosyl derivatives from O-(α-D-glucopyranosyl) trichloroacetimidate. *Angew Chemie Int Ed English* 22(5):406–406.
52. Greenfield LK, et al. (2012) Domain organization of the polymerizing mannosyltransferases involved in synthesis of the *Escherichia coli* O8 and O9a lipopolysaccharide O-antigens. *J Biol Chem* 287(45):38135–38149.