Modification and periplasmic translocation of the biofilm exopolysaccharide poly-β-1,6-*N*-acetyl-D-glucosamine

Dustin J. Little[a,b], Grace Li[a,b], Christopher Ing[a,b], Benjamin R. DiFrancesco[c], Natalie C. Bamford[a,b], Howard Robinson[d], Mark Nitz[c,1], Régis Pomès[a,b,1], and P. Lynne Howell[a,b,1]

[a]Program in Molecular Structure and Function, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada M5G 1X8; [b]Department of Biochemistry, University of Toronto, Toronto, ON, Canada M5S 1A8; [c]Department of Chemistry, University of Toronto, Toronto, ON, Canada M5S 3H6; and [d]Photon Sciences Division, Brookhaven National Laboratory, Upton, NY 11973-5000

Edited by Gregory A. Petsko, Weill Cornell Medical College, New York, NY, and approved June 3, 2014 (received for review April 8, 2014)

Poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) is an exopolysaccharide produced by a wide variety of medically important bacteria. Polyglucosamine subunit B (PgaB) is responsible for the de–*N*-acetylation of PNAG, a process required for polymer export and biofilm formation. PgaB is located in the periplasm and likely bridges the inner membrane synthesis and outer membrane export machinery. Here, we present structural, functional, and molecular simulation data that suggest PgaB associates with PNAG continuously during periplasmic transport. We show that the association of PgaB's N- and C-terminal domains forms a cleft required for the binding and de–*N*-acetylation of PNAG. Molecular dynamics (MD) simulations of PgaB show a binding preference for *N*-acetylglucosamine (GlcNAc) to the N-terminal domain and glucosammonium to the C-terminal domain. Continuous ligand binding density is observed that extends around PgaB from the N-terminal domain active site to an electronegative groove on the C-terminal domain that would allow for a processive mechanism. PgaB's C-terminal domain ($PgaB_{310–672}$) directly binds PNAG oligomers with dissociation constants of ∼1–3 mM, and the structures of $PgaB_{310–672}$ in complex with β-1,6-$(GlcNAc)_6$, GlcNAc, and glucosamine reveal a unique binding mode suitable for interaction with de–*N*-acetylated PNAG (dPNAG). Furthermore, $PgaB_{310–672}$ contains a β-hairpin loop (βHL) important for binding PNAG that was disordered in previous $PgaB_{42–655}$ structures and is highly dynamic in the MD simulations. We propose that conformational changes in $PgaB_{310–672}$ mediated by the βHL on binding of PNAG/dPNAG play an important role in the targeting of the polymer for export and its release.

exopolysaccharide biosynthesis | glycobiology | carbohydrate binding | deacetylase

Planktonic bacteria often switch to a surface-associated biofilm mode of growth under stress. Within the biofilm, aggregated clusters of bacteria are localized with a self-produced extra cellular matrix composed of proteinaceous adhesins, nucleic acids, and exopolysaccharides (1–3). The biofilms of several Gram-positive and numerous Gram-negative bacteria have been shown to contain partially de–*N*-acetylated poly-β-1,6-*N*-acetyl-D-glucosamine (dPNAG) exopolysaccharides. dPNAG mediates cell-to-cell and cell-to-surface adhesion, contributes to the structural integrity of the biofilm, and reduces the susceptibility of the bacteria to antimicrobials and the innate immune system. In *Escherichia coli*, the production, modification, and export of poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) require the polyglucosamine (Pga) machinery encoded by the *pgaABCD* operon (4). The inner membrane proteins PgaC and PgaD interact in the presence of the second messenger bis-(3′-5′)-cyclic dimeric GMP to form the active biosynthetic complex (5). PgaC contains a cytosolic glycosyltransferase domain that uses UDP *N*-acetylglucosamine (GlcNAc) to synthesize the polymer, whereas the transmembrane regions of PgaC and PgaD have been proposed to facilitate its translocation across the inner membrane (5). PgaA has a predicted C-terminal β-barrel that likely facilitates dPNAG export across the outer membrane and a periplasmic domain predicted to contain tetratricopeptide repeat consensus sequence motifs, which may be involved in PgaA–PgaB or PgaA–dPNAG interactions (6–10). PgaB is a two-domain outer membrane lipoprotein that is required for the partial de–*N*-acetylation and export of PNAG (6, 8). The N-terminal domain of PgaB belongs to the family four carbohydrate esterases (CE4s) but has a unique circularly permuted arrangement of the canonical CE4 motifs (8, 11). This sequence permutation results in the absence of an aspartic acid residue typically involved in catalysis (8), and has been proposed to attenuate PgaB activity to maintain the low levels of PNAG de–*N*-acetylation (∼3–5%) observed in vivo (4, 6). The C-terminal domain of PgaB ($PgaB_{310–672}$) has been proposed to bind PNAG and assist in de–*N*-acetylation and export (6); however, the role it plays in these processes remains to be elucidated.

Research over the past decade on synthase-dependent polysaccharide systems like alginate, the pel polysaccharide, PNAG, and cellulose has started to provide structural and functional details on how these polysaccharides are synthesized, modified, and exported (9). The structure and characterization of the

Significance

Extracellular polysaccharides are important for bacterial aggregation and surface attachment during the formation of a biofilm. Bacteria living within a biofilm are more resistant to antibiotics and host defenses than those living in a free planktonic state. Poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) is produced by a number of pathogenic bacteria but is an insoluble polymer, making it difficult to study in vitro. Polyglucosamine subunit B (PgaB) is an outer membrane lipoprotein responsible for the deacetylation of PNAG, a key modification required for biofilm formation. Herein, we address a number of key questions related to the modification and translocation of PNAG/de–*N*-acetylated PNAG through the periplasmic space. The study provides valuable insight for synthase-dependent exopolysaccharide systems and a brute-force molecular dynamics approach for studying insoluble polymers using monosaccharides.

Author contributions: D.J.L., G.L., M.N., R.P., and P.L.H. designed research; D.J.L., G.L., C.I., N.C.B., and H.R. performed research; B.R.D. contributed new reagents/analytic tools; D.J.L., G.L., C.I., M.N., R.P., and P.L.H. analyzed data; and D.J.L., G.L., M.N., R.P., and P.L.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The structure factors and coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4P7L, 4P7O, 4P7Q, 4P7N, and 4P7R).

[1]To whom correspondence may be addressed. Email: howell@sickkids.ca, mnitz@chem.utoronto.ca, or pomes@sickkids.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406388111/-/DCSupplemental.

cellulose biosynthetic machinery, BcsA and BcsB, provide atomistic details into polymer synthesis and translocation into the periplasm (12, 13). What remains poorly understood is whether these long polymers are free or protected throughout periplasmic transport. Because the length of the polysaccharides and their insolubility make them challenging to study in vitro, we have characterized the structure and function of PgaB with short PNAG oligomers and used molecular dynamics (MD) simulations to study binding of *N*-acetylglucosamine (GlcNAc) and glucosammonium ($GlcNH_3^+$). We demonstrate herein that $PgaB_{310–672}$ is crucial for de–*N*-acetylation, because the N-terminal domain alone is enzymatically inactive. Modification of PNAG requires a cleft formed by the association of the N- and C-terminal domains, wherein both domains contribute residues required for polymer binding. Extensive MD simulations define a continuous and almost mutually exclusive binding surface for GlcNAc and $GlcNH_3^+$ that extends from the de–*N*-acetylation active site around PgaB to the C-terminal domain. The C-terminal domain shows a preference for $GlcNH_3^+$ binding, suggesting that de–*N*-acetylation occurs first and PNAG/dPNAG associates continuously with PgaB in a processive manner. The crystallographic structures and MD simulations identify a dynamic β-hairpin loop (βHL) involved in saccharide binding that we propose propagates the signal for polymer export. The brute-force MD approach using monosaccharides to define a global binding landscape for an insoluble polymer is generally applicable for other polysaccharide processing enzymes, and provides insight into the periplasmic modification and transport processes that occur during polymer biosynthesis.

Results

PNAG de–*N*-Acetylation Requires the Association of PgaB's N- and C-Terminal Domains. A distinct difference in the PNAG biosynthetic machinery between Gram-positive and Gram-negative bacteria is the PNAG de–*N*-acetylases IcaB and PgaB, respectively. IcaB is a single-domain protein, whereas PgaB is a two-domain protein. Both contain a CE4 domain and can de–*N*-acetylate PNAG oligomers in vitro (8, 14), but the function of $PgaB_{310–672}$ in this process is unknown. To characterize the role of $PgaB_{310–672}$ in PNAG de–*N*-acetylation, fluorescamine assays were performed and the activities of $PgaB_{22–672}$, $PgaB_{22–309}$, and $PgaB_{310–672}$ were compared. $PgaB_{22–672}$ de–*N*-acetylated β-1,6-$(GlcNAc)_5$ at a similar level to that shown previously (8), whereas neither $PgaB_{22–309}$ nor $PgaB_{310–672}$ shows appreciable levels of activity (Fig. 1*A*). This suggests that both domains of PgaB are required for de–*N*-acetylation of PNAG. To determine if $PgaB_{310–672}$ could rescue the de–*N*-acetylation activity of $PgaB_{22–309}$, the N- and C-terminal domains were purified separately, mixed at equal concentrations, and assayed. The $PgaB_{22–309}$/$PgaB_{310–672}$ solution did not show any significant levels of de–*N*-acetylation of β-1,6-$(GlcNAc)_5$ (Fig. 1*A*). Analytical size exclusion chromatography showed that when $PgaB_{22–309}$ and $PgaB_{310–672}$ are mixed, they elute at their expected monomeric molecular weight, suggesting the domains do not have appreciable affinity (K_d < 100 μM, change in Gibbs free energy (ΔG) of −5.5 kcal·mol^{-1} or stronger) for each other in the absence of the interdomain linker (IDL; residues 309–313) (Fig. S1*A*). The buried interaction surface between the N- and C-terminal domains in $PgaB_{42–655}$ is ∼800 Å^2, with a predicted ΔG of −1.2 kcal·mol^{-1} as calculated by the Protein Interfaces, Surfaces, and Assemblies (PISA) server (15). The buried interface area is comparable to that of a transient protein–protein interaction (16), but the low ΔG suggests the weak interaction is not sufficient for the two domains to associate without additional influences (e.g., the IDL). The domain association creates a cleft between the N- and C-terminal domains that we propose is the binding site for PNAG during de–*N*-acetylation. Docking studies with a β-1,6-$(GlcNAc)_5$, where the central GlcNAc was modeled as a tetrahedral intermediate to mimic the preferential de–*N*-acetylation pattern of PgaB, were carried out (8, 11). The top docking result shows that the tetrahedral oxyanion coordinates the metal in a bidentate fashion and hydroxyl at carbon 3 (OH-3) of the central GlcNAc binds the metal ion (Fig. 1*B*), similar to the binding mode proposed for other CE4 members (17). The modeled β-1,6-$(GlcNAc)_5$ reaction intermediate makes contacts with the N- and C-terminal domains with predicted interaction surfaces of ∼480 Å^2 and ∼300 Å^2, respectively. The interaction surface on the C-terminal domain consists of nine residues. Three of these residues, W387, T391, and R392, which reside on helix-α2 of $PgaB_{310–672}$ (Fig. 1*B* and Fig. S1*B*), are 100% conserved among PgaB homologs, suggesting this region is important for binding PNAG during de–*N*-acetylation.

Structure of $PgaB_{310–672}$ Reveals a Previously Disordered βHL. The structure of $PgaB_{42–655}$ was determined previously to 1.9 Å. However, to facilitate crystallization, the last 17 residues of PgaB's C-terminal domain were truncated (8, 18). To characterize the role of $PgaB_{310–672}$ further, we determined its structure to 1.8 Å (Table S1). The structure of $PgaB_{310–672}$ reveals a complete $(\beta/\alpha)_8$ triosephosphate isomerase (TIM)-barrel fold, which superimposes with the C-terminal domain of $PgaB_{42–655}$ with an rmsd of 1.2 Å over 324 equivalent Cα atoms (Fig. 2 *A* and *B*). There are two main structural differences between $PgaB_{310–672}$ and the C-terminal domain of $PgaB_{42–655}$. First, the presence of the eighth helix of the $(\beta/\alpha)_8$ barrel, helix α8, causes displacements of 8.0 Å and 2.9 Å to the N- and C-termini of helix α7, respectively (Fig. 2*B*). Second, residues 610–623, which are disordered in the $PgaB_{42–655}$ structure, form a large βHL between strand β6 and helix α7 (Fig. 2*B*). This βHL extends over the top of the $(\beta/\alpha)_8$ barrel, narrowing the pronounced electronegative groove to ∼7 Å (Fig. 2*C*), as opposed to ∼14 Å in the $PgaB_{42–655}$ structure.

$PgaB_{310–672}$ Binds PNAG Oligomers. A structural comparison search of $PgaB_{310–672}$ using the Dali server (19) revealed similarities to

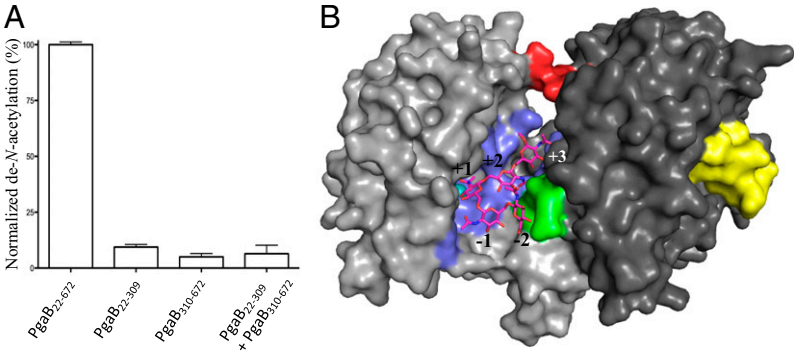


Fig. 1. $PgaB_{310–672}$ is required for de–*N*-acetylation and PNAG binding to the active site. (*A*) Fluorescamine activity assay for $PgaB_{22–672}$, $PgaB_{22–309}$, $PgaB_{310–672}$, and $PgaB_{22–309}$ mixed with $PgaB_{310–672}$ and incubated with β-1,6-$(GlcNAc)_5$ at 37 °C for 24 h. Data points are mean values, with error bars representing the SD between triplicate experiments. (*B*) Top docked β-1,6-$(GlcNAc)_5$ tetrahedral intermediate. The N-terminal and C-terminal domains of PgaB are colored light and dark gray, respectively. Residues involved in ligand binding are colored blue, except for the conserved patch on α2 of $PgaB_{310–672}$, which is colored green. The nickel ion, IDL, and βHL are colored teal, red, and yellow, respectively.

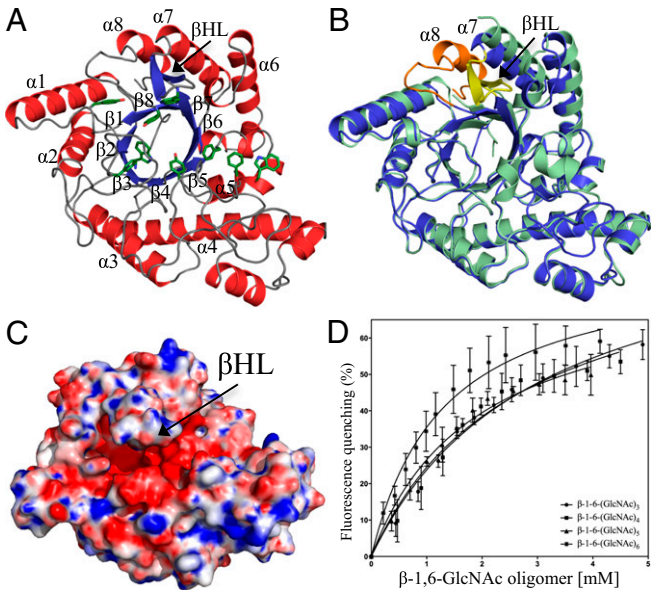


Fig. 2. Structure of $PgaB_{310–672}$ and binding of PNAG oligomers. (*A*) Structure of $PgaB_{310–672}$ shown in cartoon representation with α-helices and β-strands colored red and blue, respectively, with the canonical $(\beta/\alpha)_8$ barrel labeled β1–β8 and α1–α8. Aromatic residues that line the groove are shown in stick representation and are colored green. (*B*) Superposition of $PgaB_{310–672}$ (blue) and the C-terminal domain of $PgaB_{42–655}$ (pale green) reveals the final eighth helix of the $(\beta/\alpha)_8$ barrel fold (orange) and a long βHL (yellow). (*C*) Electrostatic surface potential of $PgaB_{310–672}$ shows the βHL extends over an electronegative groove pinching off the binding pocket to ~7 Å. Quantitative electrostatics are colored from red (−10 kT/e) to blue (+10 kT/e). (*D*) PgaB intrinsic fluorescence quenching binding curves for titrations with β-1,6-$(GlcNAc)_3$, β-1,6-$(GlcNAc)_4$, β-1,6-$(GlcNAc)_5$, and β-1,6-$(GlcNAc)_6$. Data points are mean values, with error bars representing the SD between triplicate experiments.

members of glycoside hydrolase family 18 (GH18) and GH20 as defined by the Carbohydrate-Active Enzymes (CAZy) database (20) (Fig. S2). These GH families bind and/or hydrolyze GlcNAc substrates, such as chitin, gangliosides, and PNAG, which suggests $PgaB_{310–672}$ could be a GH. However, previous and ongoing efforts to show $PgaB_{310–672}$ hydrolase activity with PNAG oligomers and artificial *para*-nitrophenyl glycoside substrates have proven unsuccessful (8). To test whether $PgaB_{310–672}$ had the ability to bind PNAG oligomers, intrinsic fluorescence quenching assays were conducted because the electronegative groove of $PgaB_{310–672}$ is lined with numerous aromatic residues (Fig. 2*A*). $PgaB_{310–672}$ fluorescence showed a maximum wavelength peak λ_{max} at 338 nm, suggesting the presence of solvent-exposed tryptophans, because apolar tryptophan environments have a λ_{max} ranging from 308 to 330 nm (21). The addition of β-1,6-$(GlcNAc)_3$, β-1,6-$(GlcNAc)_4$, β-1,6-$(GlcNAc)_5$, and β-1,6-$(GlcNAc)_6$ to $PgaB_{310–672}$ resulted in a concentration-dependent decrease (quenching) in fluorescence intensity, with no shift in the λ_{max} peak (Fig. 2*D*). The fluorescence data fitted to a one-site model with dissociation constants of 3.0 ± 0.4 mM, 2.7 ± 0.6 mM, 1.9 ± 0.3 mM, and 1.3 ± 0.2 mM for β-1,6-$(GlcNAc)_3$, β-1,6-$(GlcNAc)_4$, β-1,6-$(GlcNAc)_5$, and β-1,6-$(GlcNAc)_6$, respectively (Fig. 2*D*). Equivalent titrations using chitin oligomers showed minimal fluorescence quenching (≤5%) of $PgaB_{310–672}$, suggesting it does not bind chitin and is specific for β-1,6-GlcNAc oligomers.

Comparison of $PgaB_{310–672}$ with the GH18 and GH20 Dali server hits acidic mammalian chitinase (AMCase) (22) and DispersinB (DspB) reveals the absence of respective catalytic consensus sequences D*XX*D*X*D*X*E and GGDE (Fig. S2*A*). Structural alignment shows D466 in $PgaB_{310–672}$ is in an equivalent position to D138 in AMCase and D183 in DspB (Fig. S2 *B*, *C*, and *E*), residues that are responsible for stabilizing the oxazolinium intermediate during catalysis (22, 23). Despite this, $PgaB_{310–672}$ lacks a second residue equivalent to either E140 in AMCase or E184 in DspB that could act as the catalytic acid (22, 23) (Fig. S2*E*). Even though D467 could fulfill the role of the catalytic acid, the residue is buried in the structure and is not solvent-accessible (Fig. S2*E*).

Structure of $PgaB_{310–672}$ in Complex with β-1,6-$(GlcNAc)_6$. Limited structural data are available for PNAG, but NMR and modeling studies have shown it to be flexible in solution (24), with many low-energy conformations. To understand how the polymer binds to $PgaB_{310–672}$, we crystallized the protein in the presence of β-1,6-$(GlcNAc)_6$ and determined the structure to a resolution of 1.8 Å. Continuous density (2.2σ) was observed in the unbiased $|F_o - F_c|$ difference map along the end of the electronegative groove (Fig. 3*A*). Based on conserved acidic and aromatic residues, we predict 12 possible binding sites and modeled β-1,6-$(GlcNAc)_4$ into the density from the nonreducing to reducing direction in sites 9–12 (Fig. 3). GlcNAc molecules in sites 9 and 11 were well defined and make the most contacts with $PgaB_{310–672}$ (Fig. 3*B*). GlcNAc at site 9 makes bidentate hydrogen bonds with E576 via its OH-3 and OH-4 hydroxyls, a hydrogen bond with W552 via the *N*-acetyl carbonyl, and stacking interactions with F540 and F553 (Fig. 3*B*). GlcNAc at site 11 makes bidentate hydrogen bonds with D472 through the OH-3 and OH-4 hydroxyls and stacking interactions with W552 (Fig. 3*B*). No hydrogen bond contacts with $PgaB_{310–672}$ are observed for GlcNAc at site 10 or 12; these moieties are fixed in place by the linkages with GlcNAc sites 9 and 11 (Fig. 3*B*). Further inspection of the $|F_o - F_c|$ difference map shows discontinuous density in close proximity to W613 and Y432 (site 7 in Fig. 3*A*), that could not be accounted for by molecules in the crystallization solution. It is likely that weakly bound or multiple conformations of β-1,6-$(GlcNAc)_6$ are present at sites 7 and 8. The poor quality of the electron density does not allow confident modeling of the remaining two GlcNAc moieties.

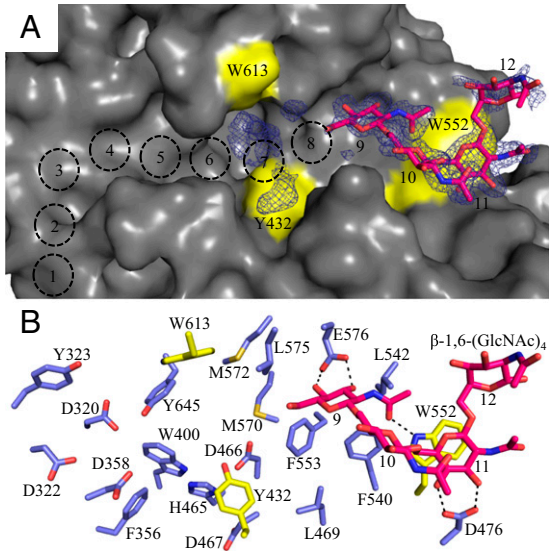


Fig. 3. Structure of $PgaB_{310–672}$ in complex with β-1,6-$(GlcNAc)_6$. Surface (*A*) and stick (*B*) representations of $PgaB_{310–672}$ with residues 614–619 omitted for clarity and Y432, W552, and W613 highlighted in yellow. The modeled β-1,6-$(GlcNAc)_4$ (magenta, stick representation) is shown with the corresponding unbiased $|F_o - F_c|$ density omit map displayed as blue mesh contoured at 2.20σ. Predicted binding sites 1–8 from the nonreducing terminus are depicted as dashed black circles. Hydrogen bonds between $PgaB_{310–672}$ and β-1,6-$(GlcNAc)_4$ are shown as dashed lines.

MD Simulations Suggest a Continuous Interdomain PNAG Binding Surface. PNAG oligomers longer than a hexasaccharide have limited solubility and are inherently flexible. This makes it very difficult to study PNAG binding to PgaB in vitro. To overcome these issues, we conducted MD simulations with no spatial restraints on a composite structure of PgaB ($PgaB_{43–667}$) successively in the apo-form and in the presence of the monosaccharides β-D-GlcNAc and β-D-$GlcNH_3^+$. PgaB did not show any significant rearrangements or large domain perturbations during the aggregate 4.83 μs of sampling time in the apo-form, β-D-GlcNAc, or β-D-$GlcNH_3^+$ simulations (Fig. S3*A*). The average rmsd of the backbone atoms of PgaB (excluding the βHL) from the X-ray crystal structure was ∼2 Å (Fig. S3*B*). Reversible binding and unbinding of monosaccharides occurred spontaneously at multiple locations on the protein surface. The binding probability density of β-D-GlcNAc shows significant occupation of three different regions [Fig. 4 (purple density) and Movie S1]. Binding region 1 is located along the cleft formed between the N- and C-terminal domains (Fig. 4*A*, Fig. S4*A*, and Movie S1). The density for β-D-GlcNAc in region 1 starts at the conserved residues W387, T391, and R392, which were implicated in the docking studies to bind β-1,6-$(GlcNAc)_5$, and extends across the de–*N*-acetylation active site to the IDL (Fig. 4*A*, Fig. S4*A*, and Movie S1). Region 2 extends from underneath the IDL toward the electronegative groove of the C-terminal domain to binding site 1 in $PgaB_{310–672}$ (Fig. 4 *B* and *C* and Movie S1). Region 3 is in the electronegative groove of the C-terminal domain extending from W613 to W552, binding sites 6–11 (Fig. 4*D* and Movie S1). The binding density of β-D-$GlcNH_3^+$ is predominantly along the electronegative groove of $PgaB_{310–672}$ [Fig. 4*D* (taupe density) and Movie S1], and binds predominantly through stacking interactions with aromatic residues and hydrogen bonding with acidic and polar side chains (Fig. S4*B*). Moreover, β-D-$GlcNH_3^+$ was observed to bind in the C-terminal groove by forming linear hydrogen-bonded clusters (Fig. S4*B*). This property is not solely determined by the net charge of β-D-$GlcNH_3^+$, because Na^+ and Cl^- salt ions, which do not possess the pyranose ring, show minimal binding to $PgaB_{43–667}$ (Fig. S4*C*). Together, these results suggest that the electronegative groove of $PgaB_{310–672}$ is a preferential binding site for dPNAG over PNAG. Superimposing the binding densities for β-D-GlcNAc and β-D-$GlcNH_3^+$ reveals contiguous density from the de–*N*-acetylation active site to the C-terminal domain. This finding suggests that after de–*N*-acetylation, PNAG/dPNAG may continually associate with PgaB and the polymer moves from the N-terminal domain to the C-terminal domain in a processive manner.

MD Simulations Predict Crystallographic Binding Sites. To validate the results of the MD simulations, structures of $PgaB_{310–672}$ in complex with GlcNAc and glucosamine (GlcN) were determined. For the $PgaB_{310–672}$-GlcNAc structure, four GlcNAc molecules could be modeled. The first GlcNAc molecule binds at site 11, making bidentate hydrogen bonds with D472 and stacking interactions with W552 (Fig. S5*A*). The second GlcNAc is slightly below site 7, making stacking interactions with Y432 and a hydrogen-bonding network mediated strictly through backbone carbonyl oxygens of G361, D362, P429, E430, and Q431 (Fig. S5*A*). The third GlcNAc is located at site 5, making hydrogen bonds with D320, D322, Y323, D358, Q611, W613, and Y645. The fourth GlcNAc is located away from the electronegative groove at I427, making hydrogen-bonding contacts with a symmetry-related molecule, and no density was observed in the simulations. In the $PgaB_{310–672}$-GlcN structure, the first GlcN binds at subsite 11 in the same manner as GlcNAc (Fig. S5*B*). The second GlcN binds at subsite 9 with E576 and makes stacking interactions with F540 and F553. The third GlcN binds at subsite 7, stacking with W613 and making hydrogen bonds with D466 and Y432 (Fig. S5*B*). The binding densities of β-D-GlcNAc and β-D-$GlcNH_3^+$ from the simulations overlap with the subsites occupied by GlcNAc, GlcN, and β-1,6-$(GlcNAc)_6$ in the crystallographic data (Fig. 5). This overall agreement suggests that brute-force computational sampling is a viable tool in predicting molecular binding surfaces for monosaccharides and more complex oligosaccharides to carbohydrate binding proteins.

βHL Is Conformationally Flexible. Structural data from our current study suggest that the βHL (residues 610–623) is flexible and plays a role in binding PNAG. During the MD simulations, this loop was found to be one of the most dynamic regions of the protein (Fig. S6*A*). Comparing βHL fluctuation analysis for the β-D-GlcNAc and β-D-$GlcNH_3^+$ simulations showed a difference, with rms fluctuations of 4.5 Å and 6 Å, respectively (Fig. S6*A*). The loop fluctuations result in variations in the width of the electronegative groove from ∼7–21 Å for the β-D-GlcNAc simulation

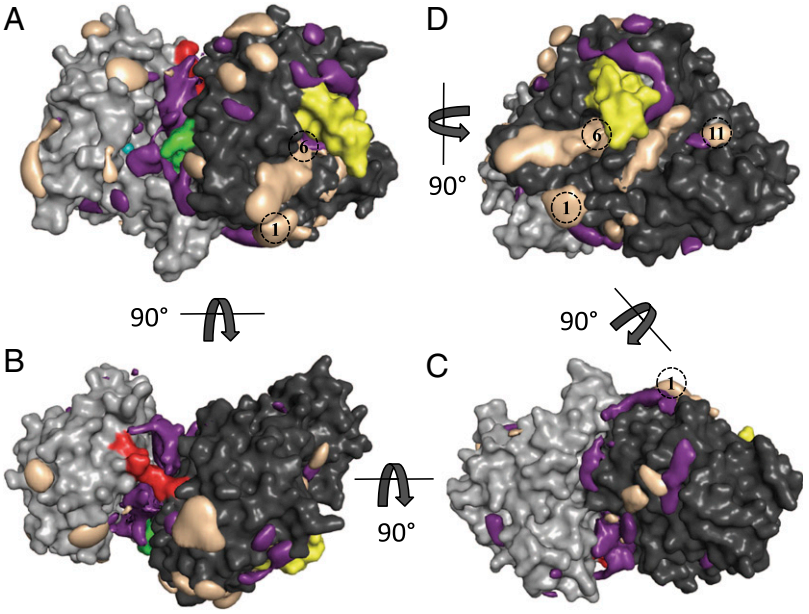


Fig. 4. MD simulations show a continuous monosaccharide binding surface. β-D-GlcNAc (purple) and β-D-$GlcNH_3^+$ (taupe) densities are overlapped with a cartoon representation of the $PgaB_{43–667}$ structure (N- and C-terminal domains are colored light gray and dark gray, respectively). In each view, binding densities are depicted at occupancies of 0.15; nickel ion is colored teal; IDL is colored red; W387, T391, and R392 are colored green; βHL is colored yellow; and predicted C-terminal domain binding sites are shown as dashed black circles. (*A*) β-D-GlcNAc density binds along the cleft formed between the N- and C-terminal domains from the de–*N*-acetylation active site to the IDL. (*B*) Continuous stretch of density for β-D-GlcNAc is seen underneath the IDL. (*C*) Slightly discontinuous density for β-D-GlcNAc and β-D-$GlcNH_3^+$ extends from the IDL to binding site 1 of $PgaB_{310–672}$. (*D*) β-D-$GlcNH_3^+$ density extends across the entire length of $PgaB_{310–672}$ and overlaps with β-D-GlcNAc density from binding sites 6–11.

and from 9–25 Å for the β-D-$GlcNH_3^+$ and apo-simulations (Fig. S6*B*). Binding of β-D-GlcNAc to the C-terminal domain reduces the flexibility of the electronegative groove, whereas binding of β-D-$GlcNH_3^+$ results in the same or a slightly more open conformation as found in the apo-simulation. In an attempt to locate additional GlcN binding sites, we determined the structure of $PgaB_{310–672}$ in the presence of 0.5 M GlcN. Crystallization under these conditions resulted in a different crystal form that diffracted to higher resolution. Although we were unable to locate additional density that could be modeled as GlcN, this P1 crystal form supports the findings that the electronegative groove is inherently flexible, because little to no density was observed for the βHL or the loop below sites 7–11 connecting E422 to Y432 (Fig. S6 *C–E*).

Discussion

Applying a combined approach of brute-force MD simulations with structural and functional studies has revealed mechanistic details for the role of PgaB in the de–*N*-acetylation and translocation of PNAG/dPNAG. The C-terminal domain of PgaB has previously been shown to be required for the de–*N*-acetylation of PNAG, because truncation resulted in export impairment and abolishment of biofilm formation in *E. coli* (6). Our activity assays and simulation data suggest that PNAG binds to the cleft between the N- and C-terminal domains, because $PgaB_{22–309}$ is unable to de–*N*-acetylate PNAG oligomers (Fig. 1*A*) and β-D-GlcNAc binds the cleft between the N- and C-terminal domains (Fig. 4 and Movie S1). Docking studies with a β-1,6-$(GlcNAc)_5$ reaction intermediate also predict binding to the cleft and identified residues W387, T391, and R392 on $PgaB_{310–672}$ as important for coordinating GlcNAc in the −2 and +2 de–*N*-acetylation subsites. The conserved patch of residues also showed the strongest density of occupancy in the simulations. It is quite possible that this conserved patch of residues initially binds PNAG and the force of biosynthesis moves the polymer in an extended conformation along the de–*N*-acetylation site to the IDL. Structural studies on β-1,6-GlcN oligomers show they have conformational freedom with the tendency to adopt stacked or helical structures (25). GlcNAc residues in −2 to +1 de–*N*-acetylation subsites show a stacked conformation (Fig. 1*B*), whereas GlcNAc residues at the +1 to +3 subsites are in an extended conformation (Fig. 1*B* and Fig. S4*A*). This suggests the interdomain cleft binds PNAG in an extended conformation, which may be required for catalysis. The presence of multiple low-energy conformations of PNAG oligomers in solution may contribute to the low rates of catalysis previously reported (8), because an extended conformation may be required for binding. The simulation data also support these findings, showing almost contiguous binding density for β-D-GlcNAc after de–*N*-acetylation (Fig. 4*A*, Fig. S4*A*, and Movie S1) along the interdomain cleft that extends beneath the IDL toward the C-terminal domain (Fig. 4 *B* and *C* and Movie S1). Although the sequence conservation is low in the region connecting the IDL and $PgaB_{310–672}$ (Fig. S7 *B* and *C*), it is quite possible PNAG interactions are mediated by backbone carbonyls, as seen at binding site 7 of the $PgaB_{310–672}$-GlcNAc structure (Fig. S5*A*).

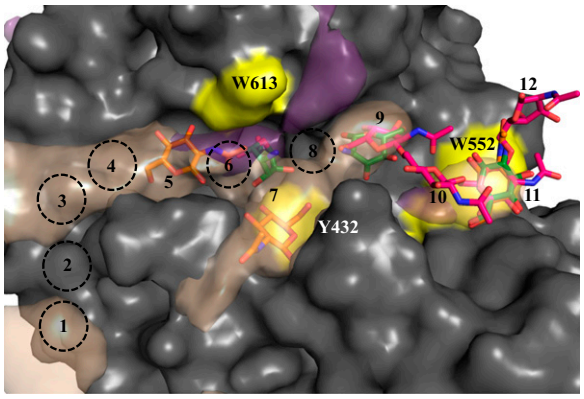


Fig. 5. Comparison of the crystallographic and MD simulation binding data. Binding densities of β-D-GlcNAc (purple) and β-D-$GlcNH_3^+$ (taupe) from MD simulations depicted at occupancies of 0.15 are overlapped with the GlcNAc (orange), GlcN (green), and β-1,6-$(GlcNAc)_4$ (magenta) molecules from the $PgaB_{310–672}$ crystal structures. Residues 614–619 were omitted for clarity, and Y432, W552, and W613 are highlighted in yellow. Binding sites are labeled 1–12, with predicted sites shown as dashed black circles.

Structural and functional data reported here show that $PgaB_{310–672}$ is similar to but lacks the catalytic motifs of GH18 and GH20 members and can bind short PNAG oligomers. This suggests the domain may be inactive. A number of GH18 members (chi-lectins) with mutations in the catalytic consensus motif (Fig. S2) retain the ability to bind chitin (26–32). Ala substitutions of conserved residues in the C-terminal domain of HmsF, the *Yersinia pestis* homolog of PgaB, had no effect on biofilm formation (33), further supporting the idea that $PgaB_{310–672}$ may be acting as a carbohydrate binding domain. Whereas the dissociation constant of $PgaB_{310–672}$ for β-1,6-$(GlcNAc)_6$ is weaker than those observed for chi-lectins, which have K_d values for β-1,4-$(GlcNAc)_6$ (chitohexaose) in the low (<20 μM) range (26–30), stronger interactions may be observed with longer or partially de–*N*-acetylated PNAG oligomers. The simulation data support this notion because β-D-$GlcNH_3^+$ density is only located along the electronegative groove of $PgaB_{310–672}$, suggesting the domain may preferentially bind dPNAG (Fig. 4*D* and Movie S1). Alternatively, the low affinity for PNAG oligomers may, in fact, be a characteristic of the domain. The β-1,6-$(GlcNAc)_6$–$PgaB_{310–672}$ structure shows an alternating pattern of binding with GlcNAc residues at sites 9 and 11 contacting the protein (Fig. 3). This alternating binding pattern was also present in the GlcNAc and GlcN structures with carbohydrate bound at sites 5, 7, 9, and 11 (Fig. 5). The simulation data further support this pattern because little to no binding density was seen at sites 2, 6, 8, 10, and 12 (Fig. 5). This ligand binding mechanism would also accommodate both GlcNAc and GlcN residues and would allow for continual movement of dPNAG across the electronegative groove in a processive manner. Sliding of dPNAG in a screw-like mechanism across the domain would prevent the net loss of binding energy to $PgaB_{310–672}$ as it is continuously synthesized and transported through the periplasm. This may be essential for efficient biosynthesis and transport because the only known source of energy in the system is from polymerization.

Binding of β-D-GlcNAc to $PgaB_{310–672}$ during MD simulations results in decreased flexibility of the βHL and stabilization of the electronegative groove (Fig. S6 *A* and *B*). In contrast, increased GlcN concentrations used during the crystallization of the P1 crystal form structure increased the flexibility of the electronegative groove (Fig. S6 *C–E*). Modulation of the groove may be an important step for binding and releasing dPNAG for export and/or initiating an interaction with PgaA. A transient PgaB–PgaA interaction during PNAG synthesis is supported by the interaction between the *Y. pestis* homologs HmsF and HmsH, as observed in vivo (34). Similar observations are well documented for chi-lectins when binding to long chitin oligomers, because conformational changes upon ligand binding have been proposed to propagate a molecular signal (28–30).

Molecular simulations are well suited for characterizing carbohydrate–protein interactions (35). The weak binding affinity and high dissociation rates of monosaccharides to the protein led to spontaneous, reversible binding on the nanosecond time scale. This property allowed us to probe for binding sites on the protein surface using independent, unconstrained MD simulations (with brute-force sampling) in the presence of a high concentration of

monosaccharides. This approach has been previously used successfully to examine the binding mechanism of inositol, a carbohydrate-like amyloid inhibitor, with amyloidogenic peptides and their aggregates (36, 37). The ability to predict binding modes and binding sites of β-D-GlcNAc and β-D-$GlcNH_3^+$ to PgaB accurately was validated with crystal structures of $PgaB_{310–672}$ in complex with GlcNAc, GlcN, and β-1,6-$(GlcNAc)_4$ (Fig. 5). These data suggest that the spatial binding density of monosaccharides likely represents the biological binding surface of PNAG/dPNAG on PgaB (Fig. 4 and Movie S1). This approach to define carbohydrate-binding sites should be generally applicable to proteins involved in the binding or modification of exopolysaccharides.

The structural, functional, and simulation data presented herein address a number of key questions related to the biosynthesis, modification, and translocation of PNAG/dPNAG through the periplasmic space. The study provides valuable insight for synthase-dependent exopolysaccharide systems and demonstrates the utility of the brute-force MD approach to define binding sites using monosaccharides for polysaccharides with limited solubility.

Materials and Methods

$PgaB_{310–672}$ and $PgaB_{22–309}$ were subcloned into pET28a using pET28-$PgaB_{22–672}$ as a template (18). All PgaB constructs were expressed and purified as described previously with minimal modifications (8, 18). PNAG oligomers were synthesized and purified as previously outlined (8, 38). De–*N*-acetylation activity assays were performed essentially as described (8). Analytical gel filtration was conducted with PgaB constructs applied to an S-200 column. A composite structure of PgaB containing residues 43–667 ($PgaB_{43–667}$) was used during docking studies with a β-1,6-$(GlcNAc)_5$ reaction intermediate. $PgaB_{310–672}$ was crystallized using hanging-drop vapor diffusion, and the structure was solved using molecular replacement. Ligand-bound structures of $PgaB_{310–672}$ were crystallized in the presence of GlcNAc, GlcN-HCl, or β-1,6-$(GlcNAc)_6$. A PNAG oligomer binding assay was performed using intrinsic protein fluorescence quenching. MD simulations in explicit solvent were conducted on $PgaB_{43–667}$ in the apo-form, and subsequently with β-D-GlcNAc, and β-D-$GlcNH_3^+$ (partial charges are listed in Tables S2 and S3). Full experimental details are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Patrick Yip for technical assistance, Dr. Alaji Bah for help with fluorescence quenching experiments, Dr. Varvara Pokrovskaya for ^1H NMR analysis, Dr. Nilu Chakrabarti for help with parameterization of β-D-GlcNAc and β-D-$GlcNH_3^+$, Dr. Shaunivan Labiuk at the Canadian Light Source (CLS) for data collection, and Compute Canada and Consortium Laval L'Université du Québec à Montréal McGill and Eastern Quebec for providing the computational resources for the MD simulations. Research described in this paper is supported by Canadian Institutes of Health Research (CIHR) Grants 43998 (to P.L.H.), 43949 (to R.P.), and 89708 (to M.N.). D.J.L. has been supported, in part, by graduate scholarships from the University of Toronto, the Ontario Graduate Scholarship Program, and CIHR. N.C.B. has been supported, in part, by a graduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). P.L.H. is the recipient of a Canada Research Chair. The National Synchrotron Light Source beamline X29A is supported by the US Department of Energy Office of Biological and Environmental Research and the National Institutes of Health National Center for Research Resources. Beamline 08ID-1 at the CLS is supported by the NSERC, National Research Council of Canada, CIHR, Province of Saskatchewan, Western Economic Diversification Canada, and University of Saskatchewan.

1. Sutherland I (2001) Biofilm exopolysaccharides: A strong and sticky framework. *Microbiology* 147(Pt 1):3–9.
2. Vu B, Chen M, Crawford RJ, Ivanova EP (2009) Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules* 14(7):2535–2554.
3. Branda SS, Vik S, Friedman L, Kolter R (2005) Biofilms: The matrix revisited. *Trends Microbiol* 13(1):20–26.
4. Wang X, Preston JF, 3rd, Romeo T (2004) The pgaABCD locus of Escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J Bacteriol* 186(9):2724–2734.
5. Steiner S, Lori C, Boehm A, Jenal U (2013) Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein-protein interaction. *EMBO J* 32(3):354–368.
6. Itoh Y, et al. (2008) Roles of pgaABCD genes in synthesis, modification, and export of the Escherichia coli biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *J Bacteriol* 190(10):3670–3680.
7. Keiski CL, et al. (2010) AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. *Structure* 18(2):265–273.
8. Little DJ, et al. (2012) The structure- and metal-dependent activity of Escherichia coli PgaB provides insight into the partial de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine. *J Biol Chem* 287(37):31126–31137.
9. Whitney JC, Howell PL (2013) Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends Microbiol* 21(2):63–72.
10. Whitney JC, et al. (2011) Structural basis for alginate secretion across the bacterial outer membrane. *Proc Natl Acad Sci USA* 108(32):13083–13088.
11. Nishiyama T, Noguchi H, Yoshida H, Park SY, Tame JR (2013) The structure of the deacetylase domain of Escherichia coli PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family. *Acta Crystallogr D Biol Crystallogr* 69(Pt 1):44–51.
12. Morgan JL, Strumillo J, Zimmer J (2013) Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* 493(7431):181–186.
13. Omadjela O, et al. (2013) BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proc Natl Acad Sci USA* 110(44):17856–17861.
14. Pokrovskaya V, et al. (2013) Functional characterization of Staphylococcus epidermidis IcaB, a de-N-acetylase important for biofilm formation. *Biochemistry* 52(32):5463–5471.
15. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
16. Nooren IM, Thornton JM (2003) Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* 325(5):991–1018.
17. Blair DE, Schüttelkopf AW, MacRae JI, van Aalten DM (2005) Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc Natl Acad Sci USA* 102(43):15429–15434.
18. Little DJ, et al. (2012) Combining in situ proteolysis and mass spectrometry to crystallize Escherichia coli PgaB. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68(Pt 7):842–845.
19. Holm L, Rosenström P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545-9.
20. 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.
21. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* 80(5):2093–2109.
22. Sutherland TE, et al. (2011) Analyzing airway inflammation with chemical biology: Dissection of acidic mammalian chitinase function with a selective drug-like inhibitor. *Chem Biol* 18(5):569–579.
23. Manuel SG, et al. (2007) Role of active-site residues of dispersin B, a biofilm-releasing beta-hexosaminidase from a periodontal pathogen, in substrate hydrolysis. *FEBS J* 274(22):5987–5999.
24. Wagstaff JL, Sadovskaya I, Vinogradov E, Jabbouri S, Howard MJ (2008) Poly-N-acetylglucosamine and poly(glycerol phosphate) teichoic acid identification from staphylococcal biofilm extracts using excitation sculptured TOCSY NMR. *Mol Biosyst* 4(2):170–174.
25. Grachev AA, et al. (2011) NMR and conformational studies of linear and cyclic oligo-(1→6)-β-D-glucosamines. *Carbohydr Res* 346(15):2499–2510.
26. Kumar J, et al. (2006) Structure of a bovine secretory signalling glycoprotein (SPC-40) at 2.1 Angstrom resolution. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):953–963.
27. Srivastava DB, et al. (2006) Crystal structure of a secretory signalling glycoprotein from sheep at 2.0A resolution. *J Struct Biol* 156(3):505–516.
28. Srivastava DB, et al. (2007) Carbohydrate binding properties and carbohydrate induced conformational switch in sheep secretory glycoprotein (SPS-40): Crystal structures of four complexes of SPS-40 with chitin-like oligosaccharides. *J Struct Biol* 158(3):255–266.
29. Kumar J, et al. (2007) Carbohydrate-binding properties of goat secretory glycoprotein (SPG-40) and its functional implications: Structures of the native glycoprotein and its four complexes with chitin-like oligosaccharides. *Acta Crystallogr D Biol Crystallogr* 63(Pt 4):437–446.
30. Houston DR, Recklies AD, Krupa JC, van Aalten DM (2003) Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes. *J Biol Chem* 278(32):30206–30212.
31. Sun YJ, et al. (2001) The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. *J Biol Chem* 276(20):17507–17514.
32. Varela PF, Llera AS, Mariuzza RA, Tormo J (2002) Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycoproteins from Drosophila melanogaster. *J Biol Chem* 277(15):13229–13236.
33. Forman S, et al. (2006) Identification of critical amino acid residues in the plague biofilm Hms proteins. *Microbiology* 152(Pt 11):3399–3410.
34. Abu Khweek A, Fetherston JD, Perry RD (2010) Analysis of HmsH and its role in plague biofilm formation. *Microbiology* 156(Pt 5):1424–1438.
35. Fadda E, Woods RJ (2010) Molecular simulations of carbohydrates and protein-carbohydrate interactions: Motivation, issues and prospects. *Drug Discov Today* 15(15-16):596–609.
36. Li G, Rauscher S, Baud S, Pomès R (2012) Binding of inositol stereoisomers to model amyloidogenic peptides. *J Phys Chem B* 116(3):1111–1119.
37. Li G, Pomès R (2013) Binding mechanism of inositol stereoisomers to monomers and aggregates of Aβ(16-22). *J Phys Chem B* 117(22):6603–6613.
38. Leung C, Chibba A, Gómez-Biagi RF, Nitz M (2009) Efficient synthesis and protein conjugation of beta-(1→6)-D-N-acetylglucosamine oligosaccharides from the polysaccharide intercellular adhesin. *Carbohydr Res* 344(5):570–575.

Supporting Information

Little et al. 10.1073/pnas.1406388111

SI Materials and Methods

Protein Production and Purification. The plasmid pET28-PgaB$_{22–672}$ (1) was used as a template, and *pgaB*-specific primers were designed to subclone residues 310–672 and 22–309 into a pET28a expression vector (Novagen) using inverse PCR with an NdeI site and an XhoI site flanking the gene fragments. All poly-glucosamine subunit B (PgaB) variants were expressed and purified as described previously (1), with the following exceptions: (*i*) Glycerol was only included in the lysis buffer during purification, and (*ii*) PgaB$_{310–672}$ used for crystallization and poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) oligomer binding assays were purified with Tris (pH 7.0) instead of Hepes (pH 8.0) in all steps.

Preparation of PNAG Oligomers. Oligomers of PNAG from trimer [β-1,6-(*N*-acetylglucosamine [GlcNAc]$_3$)] to hexamer [β-1,6-(GlcNAc)$_6$] were prepared and purified, and their identities were confirmed as outlined previously (2, 3). PNAG oligomers were stored as lyophilized powders at room temperature and dissolved with deionized water for use in titrations and crystallization. Accurate concentrations were determined by ^{1}H NMR using dimethylformamide as an internal standard.

De–*N*-Acetylation Activity Assays. Comparison of de–*N*-acetylation activity was determined using a fluorescamine-based assay (4) and performed as described previously (2), with the following modifications: (*i*) Samples were performed in triplicate, and (*ii*) BRAND black 96-well plates were used for fluorescence measurement using a SpectraMax M2 plate reader from Molecular Devices, with an excitation wavelength (λ_{ex}) of 360 nm and an emission wavelength (λ_{em}) of 460 nm.

Analytical Gel Filtration. Concentrated protein samples (100 μM) of PgaB$_{22–309}$, PgaB$_{310–672}$, PgaB$_{22–672}$, and PgaB$_{22–309}$ mixed with PgaB$_{310–672}$ were applied to a Superdex S-200 column (GE Health) and eluted using 20 mM Hepes (pH 7.0) and 150 mM NaCl. Analysis of the protein(s) present in the peak fraction was performed by SDS/PAGE. Protein standards used to calibrate the column were ferritin, 440 kDa; aldolase, 158 kDa; conalbumin, 75 kDa; ovalbumin, 43 kDa; and ribonuclease A, 13.7 kDa.

Docking Studies. To gain insight into PNAG binding and directionality at the N-terminal domain active site, a β-1,6-(GlcNAc)$_5$ reaction intermediate (a tetrahedral geminal diol of the central sugar) was investigated using AutoDock Vina 1.1.2 (5). Receptor and ligand Protein Data Bank (PDB) files were prepared for docking using scripts provided from MGLTools 1.5.6 (6). Docking was conducted with a grid spacing of 0.37 and *xyz* of 63 × 68 × 55 using a rigid receptor with 20 poses computed. The top pose with the tetrahedral oxyanion group bound to the nickel ion had a score of −7.1 kcal/mol. The top solution was then used in another round of docking using the same grid definitions but with flexible side chains enabled within 4 Å of the bound ligand. Ten poses were computed, with the top solution having a score of −8.6 kcal/mol. This solution was used in the generation of Fig. 1*B* and calculations using the Protein Interfaces, Surfaces, and Assemblies (PISA) server.

Crystallization and Structure Determination. Purified PgaB$_{310–672}$ was concentrated to ∼200 μM and screened for crystallization conditions at 20 °C using hanging-drop vapor diffusion in 48-well VDX plates (Hampton Research) and the Midwest Center for Structural Genomics (MCSG) 1–4 sparse matrix suites (Microlytic). An initial crystallization hit was obtained in condition 36 from the MCSG-1 suite. Optimized crystals were grown in a solution containing 16–25% (wt/vol) PEG 3350 and 0.1 M bis(2-hydroxyethyl)amino-Tris(hydroxymethyl) methane (pH 6.5) by streak-seeding a 3-μL drop with equal amounts of protein and precipitant equilibrated against 200 μL of precipitant solution. PgaB$_{310–672}$ in complex with GlcNAc, glucosamine (GlcN), or β-1,6-(GlcNAc)$_6$ was crystallized in the same crystallization conditions as described above supplemented with 0.5 M GlcNAc, 0.1 M GlcN-HCl, or 25 mM β-1,6-(GlcNAc)$_6$. Rescreening PgaB$_{310–672}$ with 0.5 M GlcN-HCl in the protein solution produced crystals with P1 crystal symmetry using 30% (wt/vol) PEG 5000 monomethyl ether, 0.2 M ammonium sulfate, and 0.1 M MES (pH 6.5). Crystals were cryoprotected for 5 s in reservoir solution supplemented with 20% (vol/vol) ethylene glycol before vitrification in liquid nitrogen. Diffraction data were collected on beamline X29 at the National Synchrotron Light Source and beamline 08ID-1 at the Canadian Light Source (Table S1). The data were indexed, integrated, and scaled using HKL2000 (7), and the structures were determined by molecular replacement with PHENIX AutoMR (8), initially using the C-terminal domain of PgaB$_{42–655}$ (PDB ID code 4F9D) as a search model and subsequently using apo-PgaB$_{310–672}$ for the saccharide-complexed and P1 crystal form structures. Manual model building in Coot (9) was alternated with refinement using PHENIX.REFINE (8). Translation/libration/screw (TLS) groups were used during refinement and determined automatically using the TLSMD web server (10, 11). Structure figures were generated using the PYMOL Molecular Graphics System (DeLano Scientific; www.pymol.org/), and quantitative electrostatics were calculated using PDB2PQR (12, 13) and Adaptive Poisson–Boltzmann Solver (APBS) (14) software. Programs used for crystallographic data processing and analysis were accessed through SBGrid (15).

PNAG Oligomer Binding Assays. The binding of PNAG and chitin oligomers to PgaB$_{310–672}$ was monitored by intrinsic protein fluorescence quenching. Fluorescence measurements were carried out at 20 °C in a quartz cuvette (type no. 115F-QS; Hellma Analytics) using a PTI QuantaMaster 80 steady-state fluorometer (Photon Technology International), with a 4-nm bandwidth for both excitation and emission and a speed of 2 nm/s. Fluorescence spectra were collected between 300 nm and 400 nm with an excitation wavelength (λ_{ex}) of 288 nm, with a peak emission wavelength (λ_{em}) of 338 nm used for the calculation of the dissociation constant. Fluorescence data were collected by titrating 400 μL of 1 μM PgaB$_{310–672}$ in buffer A [20 mM Tris (pH 7.0), 150 mM NaCl] with a solution containing ∼15 mM PNAG or chitin oligomers in buffer A. Each titration was incubated for 3 min before spectra were collected and was stable up to 2 h after mixing, with maximum quenching of ∼60%. Titrations covered PNAG and chitin oligomer concentrations between 0 and 5 mM, respectively. Titrations were corrected for dilution, ligand fluorescence, and inner filter effect. All ligands had linear fluorescence over the concentration range used in this study and did not exceed 5% of PgaB$_{310–672}$ fluorescence. Equilibrium dissociation constants for PNAG oligomers were obtained by fitting the fluorescence quenching data to the single-site binding equation using nonlinear regression analysis (Prism v.6.0b; GraphPad Software): $(\Delta F/F_o \times 100) = [(\Delta F_{max}/F_o \times 100) \times [S]]/(K_d + [S])$, where $(\Delta F/F_o \times 100)$ is the percentage of fluorescence quenching relative to the initial value, F_o, after

the addition of substrate at concentration [S]. K_d is the dissociation constant for binding.

Molecular Dynamic Simulations. A composite structure of PgaB containing residues 43–667 ($PgaB_{43–667}$) was constructed by manually adding in Coot (9) residue 62 and all missing side chains in their most favorable nonclashing rotamer to $PgaB_{42–655}$ (PDB ID code 4F9D), as well as residues 610–620 and 647–667 from the $PgaB_{310–672}$ structure (PDB ID code 4P7L). All ligand and water molecules found in the crystal structures were removed from the composite structure. Histidine protonation states were assigned based on predicted pK_a values using the online software PROPKA (16–18) and histidine hydrogen-bonding geometries in the initial crystal structures. Protein and ions were modeled using the AMBER99 force field (19). Parameters for Ni^{2+} were approximated using those of Mg^{2+}. The net charge of the protein was −11e. The structure of β-D-GlcNAc was generated using the web-based Glycam Biomolecule Builder (20), and the structure of β-D-glucosammonium ($GlcNH_3^+$) was obtained from the ZINC database (21). The GLYCAM06 force field for carbohydrates (22) was used to model both β-D-GlcNAc and β-D-$GlcNH_3^+$, except for the charges of the latter, as described below. The simulation system comprised 11 Na^+ counterions, 45 molecules of free monosaccharide (either β-D-GlcNAc or β-D-$GlcNH_3^+$) at an effective concentration of 100 mM, and 19,533 and 19,991 water molecules for the β-D-GlcNAc and β-D-$GlcNH_3^+$ simulations, respectively. The initial volume of the simulation box was 713.6 nm^3. To mimic experimental conditions, 100 mM NaCl was added to the aqueous solution containing $GlcNH_3^+$. Geometry optimization of the β-D-$GlcNH_3^+$ molecule was performed using Gaussian-09 (23) with HF/6-31G*. Updated restrained electrostatic potential-derived partial atomic charges were computed for β-D-$GlcNH_3^+$ (with a net charge of +1e) by fitting to a single HF/6-31G* molecular electrostatic potential (MEP) with a restraint weight of 0.01. MEPs were computed using the charges from electrostatic potentials using a grid (CHELPG) methodology (24) with the R.E.D. III software package (25). The partial charges (Tables S2 and S3) were assigned so that the total charge of the amine group ($CHNH_3^+$) added up to +1.164e and the rest of the β-D-$GlcNH_3^+$ molecule summed to −0.164e. Aliphatic hydrogen atoms were assigned a zero partial charge for compatibility with GLYCAM06. The transferable intermolecular potential 3P (TIP3P) water model (26) was used to represent the solvent. Version 4.5.5 of the GROMACS software package (27, 28) was used to perform unrestrained all-atom molecular dynamics (MD) simulations with the stochastic dynamics algorithm using an integration time step of 2 fs and an inverse friction coefficient of 2 ps. Electrostatic interactions were calculated using particle mesh Ewald summation (29, 30), with a grid size of 0.12 nm and a Coulombic real-space cutoff of 1.1 nm. The Lennard–Jones potential was computed up to 1.2 nm using the GROMACS twin-range cutoff function with a short-range cutoff of 1.1 nm. Covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm (31).

The simulation system was first subjected to steepest descent energy minimization, which converged with a maximum force tolerance of 10 $kJ \cdot mol^{-1} \cdot nm^{-1}$, followed by a 1-ns equilibration in the canonical (NVT) ensemble using Berendsen temperature coupling (32) at 300 K with a coupling constant of 2.0. A second equilibration was performed for 1 ns in the isothermal-isobaric (NpT) ensemble using Berendsen temperature coupling and isotropic pressure coupling (32) controlled at 300 K and 1 atm, respectively. Production simulations were performed using the stochastic dynamics integrator and the Parrinello–Rahman barostat for pressure coupling (33). For all PgaB simulations (apo-form, β-D-GlcNAc, and β-D-$GlcNH_3^+$), 13 independent MD simulations of ∼130 ns were performed, yielding a total of 4.83 μs of sampling time.

Analysis Protocol. To compute spatial binding probability densities of β-D-GlcNAc and β-D-$GlcNH_3^+$, simulation frames were first fitted via rmsd alignment of the protein backbone atoms to an energy-minimized and MD-equilibrated structure. The density maps correspond to the fractional atomic occupancy of β-D-GlcNAc or β-D-$GlcNH_3^+$ molecules binned using a grid with a resolution of 1 Å. The density maps for β-D-GlcNAc and β-D-$GlcNH_3^+$ were computed using ∼165,000 time frames. The Visual Molecular Dynamics (VMD) software package (34) was used to calculate densities, and they were graphically rendered using VMD and PYMOL.

1. Little DJ, et al. (2012) Combining in situ proteolysis and mass spectrometry to crystallize Escherichia coli PgaB. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68(Pt 7):842–845.
2. Little DJ, et al. (2012) The structure- and metal-dependent activity of Escherichia coli PgaB provides insight into the partial de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine. *J Biol Chem* 287(37):31126–31137.
3. Leung C, Chibba A, Gómez-Biagi RF, Nitz M (2009) Efficient synthesis and protein conjugation of beta-(1→6)-D-N-acetylglucosamine oligosaccharides from the polysaccharide intercellular adhesin. *Carbohydr Res* 344(5):570–575.
4. Udenfriend S, et al. (1972) Fluorescamine: A reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. *Science* 178(4063):871–872.
5. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461.
6. Sanner MF (1999) Python: A programming language for software integration and development. *J Mol Graph Model* 17(1):57–61.
7. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collection in oscillation mode. *Methods Enzymol* 276:307–326.
8. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
9. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
10. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62(Pt 4):439–450.
11. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Cryst* 39(Pt 1):109–111.
12. Dolinsky TJ, et al. (2007) PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35(Web Server issue):W522–W525.
13. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32(Web Server issue):W665–W667.
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.
15. Morin A, et al. (2013) Collaboration gets the most out of software. *eLife* 2:e01456.
16. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73(3):765–783.
17. Olsson M, Søndergaard CR (2011) PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537.
18. Søndergaard CR, Olsson M (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J Chem Theory Comput* 7(7):2284–2295.
19. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117(19):5179–5197.
20. Woods Group (2005–2014) GLYCAM web. Available at http://www.glycam.com.
21. Irwin JJ, Shoichet BK (2005) ZINC—A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1):177–182.
22. Kirschner KN, et al. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J Comput Chem* 29(4):622–655.
23. Frisch MJ, et al. (2009) *Gaussian-09 Revision C.01* (Gaussian, Inc., Wallingford, CT).
24. Breneman CM, Wiberg KB (1990) Determining atom centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11(3):361–373.
25. Dupradeau F-Y, et al. (2010) The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* 12(28):7821–7839.
26. Jorgensen W, Chandrasekhar J, Madura JD, Impey RW, Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.

27. Hess B, Kutzner C, van der Spoel D (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447.
28. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
29. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092.
30. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19): 8577–8593.
31. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4(1):116–122.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684–3690.
33. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52(12):7182–7190.
34. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.

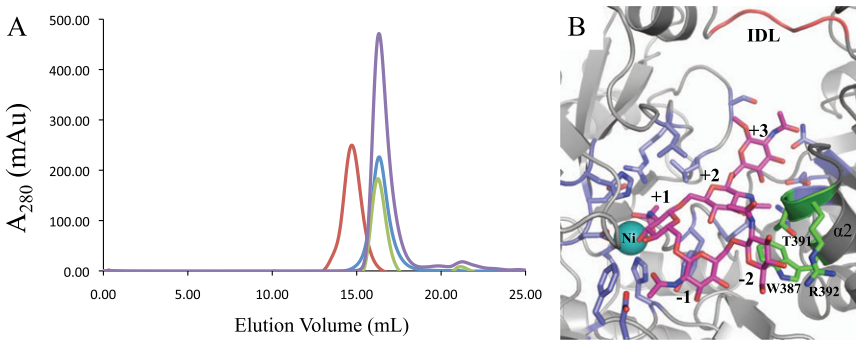


Fig. S1. (*A*) Gel filtration chromatograms of $PgaB_{22–72}$ (red), $PgaB_{22–309}$ (blue), $PgaB_{310–672}$ (green), and $PgaB_{22–309}$ plus $PgaB_{310–672}$ mixed together (purple). mAu, milliabsorbance units. (*B*) Close-up view of the top docked β-1,6-$(GlcNAc)_5$ tetrahedral intermediate (magenta, stick representation). Residues involved in ligand binding are shown as sticks colored blue, except for the conserved patch on α2 of $PgaB_{310–672}$, which is colored green. The nickel ion and interdomain linker (IDL) are colored teal and red, respectively.

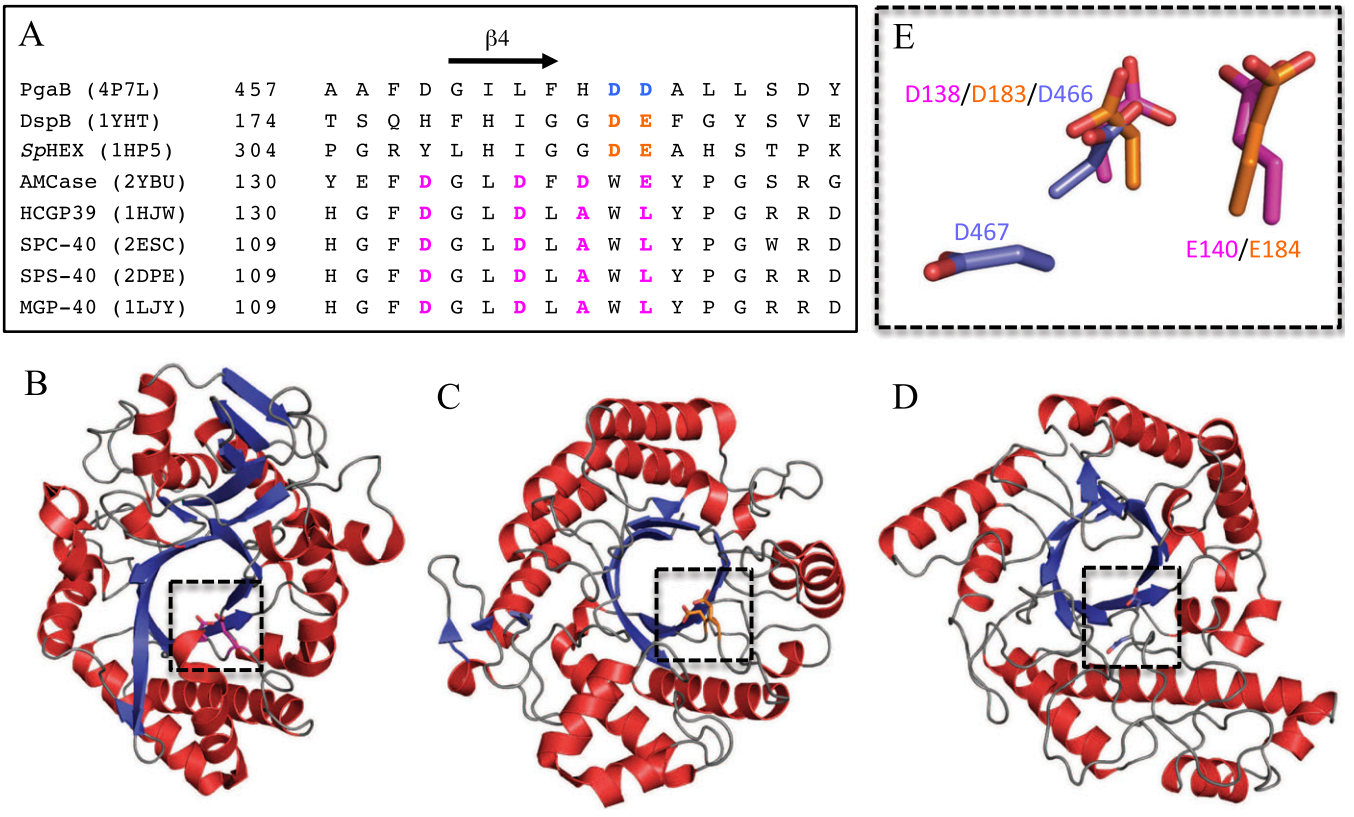


Fig. S2. Comparison of $PgaB_{310–672}$ with glycoside hydrolase family 18 (GH18) and GH20 active sites. (*A*) Sequence alignment of the catalytic motifs for GH18 (colored orange) and GH20 (colored magenta) with $PgaB_{310–672}$ (colored blue). Dispersin B (DspB) and *Streptomyces plicatus* hexosaminidase (*Sp*HEX) belong to GH20 and are catalytically active; acidic mammalian chitinase (AMCase) belongs to GH18 and is catalytically active; and human glycoprotein 39 (HCGP39), bovine secretory signaling glycoprotein (SPC-40), sheep secretory signaling glycoprotein (SPS-40), and mammary gland protein (MGP-40) belong to GH18 and are chi-lectins that bind but do not hydrolyze chitin substrates. PDB ID codes and starting residues of the alignment are supplied after the respective names. A structural comparison of AMCase (*B*), DspB (*C*), and $PgaB_{310–672}$ (*D*) is shown in cartoon representation, with α-helices, β-strands, and loops colored red, blue, and gray, respectively. (*E*) Comparison of the catalytic residues of AMCase (magenta), DspB (orange), and $PgaB_{310–672}$ (blue) after structural alignment over all Cα atoms. Boxed areas in *B–D* are insets of *E*. *E* is the superposition of the insets from *B–D*.

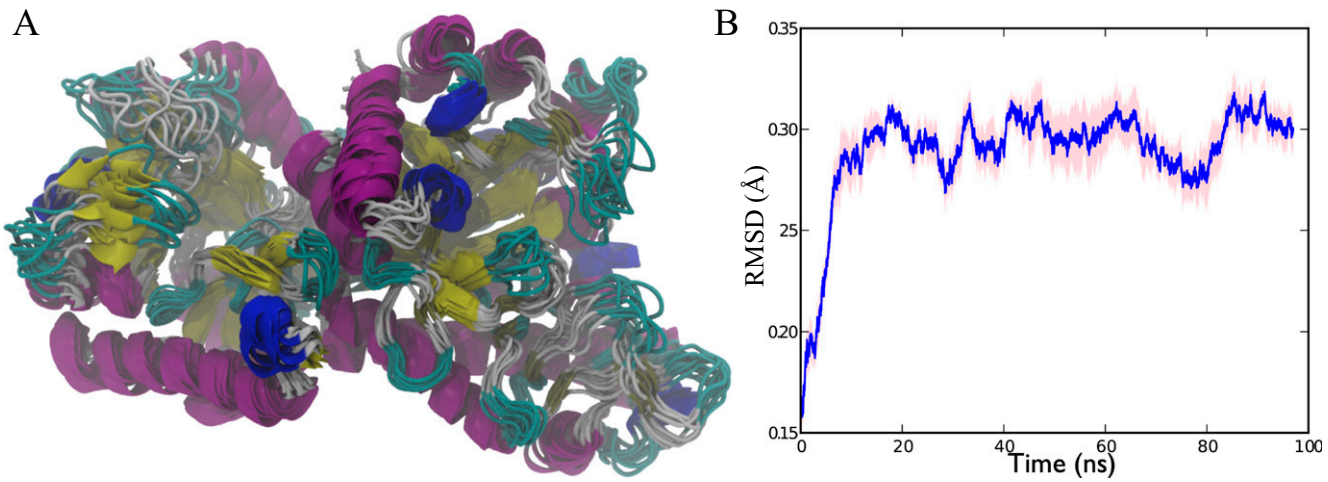


Fig. S3. Fluctuation of PgaB during MD simulations. (*A*) Fluctuations in the $PgaB_{43–667}$ structure during the MD simulations shown in a cartoon colored according to the secondary structure. Purple, α-helix; blue, 3-10 helix; yellow, β-strand; gray, β-turn; teal, loop. (*B*) Average rmsd (angstroms) of the protein from the crystal structure with the β-hairpin loop removed from the calculation. The pink-shaded region represents the error in the mean.

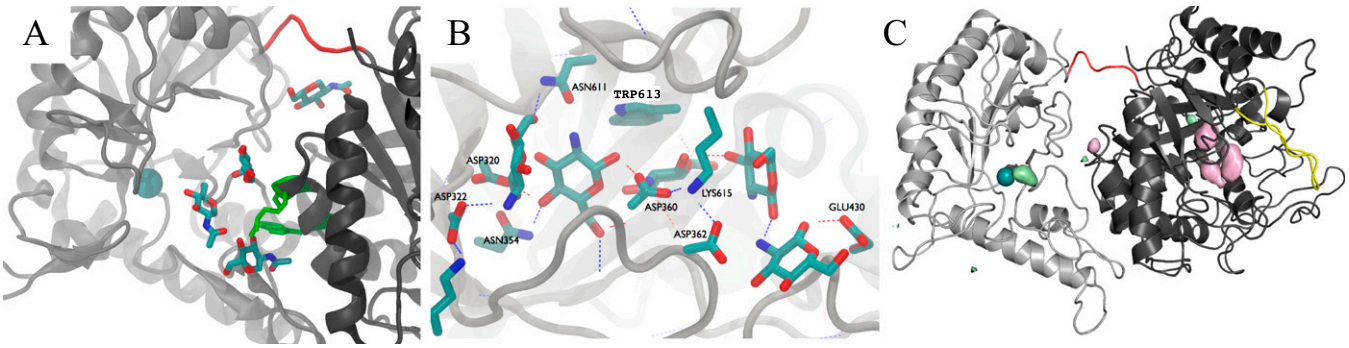


Fig. S4. (*A*) Examples of the β-D-GlcNAc (cyan, stick representation) binding modes along the N- and C-terminal domain cleft from the conserved residue patch (green) on α2 of $PgaB_{310–672}$ across the metal ion (teal) to the IDL (red). (*B*) Examples of the β-D-$GlcNH_3^+$ (cyan, stick representation) binding modes along the electronegative groove of $PgaB_{310–672}$. (*C*) Spatial probability distribution of sodium (pink) and chloride (pale green) ions depicted at an occupancy of 0.01 bound to $PgaB_{43–667}$, shown in cartoon representation with the N- and C-terminal domains colored light and dark gray, respectively. The nickel ion, IDL, and β-hairpin loop are colored teal, red, and yellow, respectively.

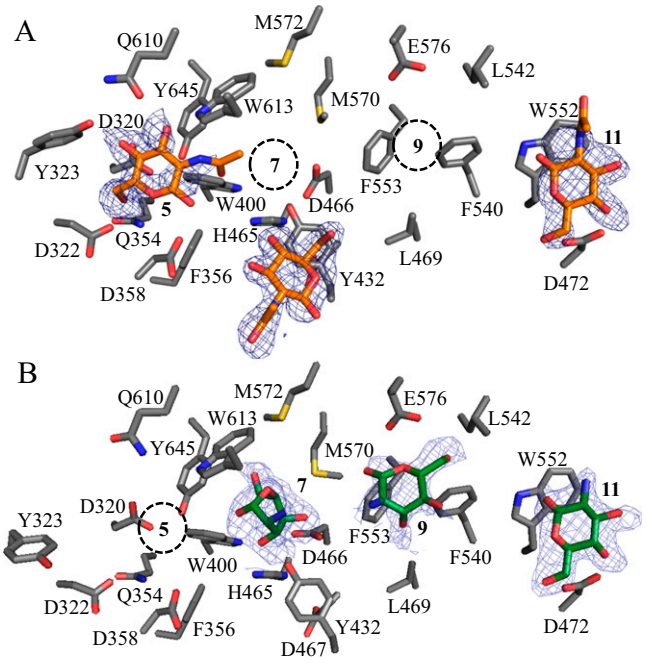


Fig. S5. Crystal structures of $PgaB_{310–672}$ in complex with GlcNAc (*A*) and GlcN (*B*). $PgaB_{310–672}$ residues, GlcNAc, and GlcN are shown in stick representation colored gray, orange, and green, respectively. Unbiased $|F_o - F_c|$ density omit maps for GlcNAc and GlcN are shown as blue mesh contoured at 2.25σ. Binding sites 5, 7, 9, and 11 are labeled, with predicted sites shown as dashed black circles.

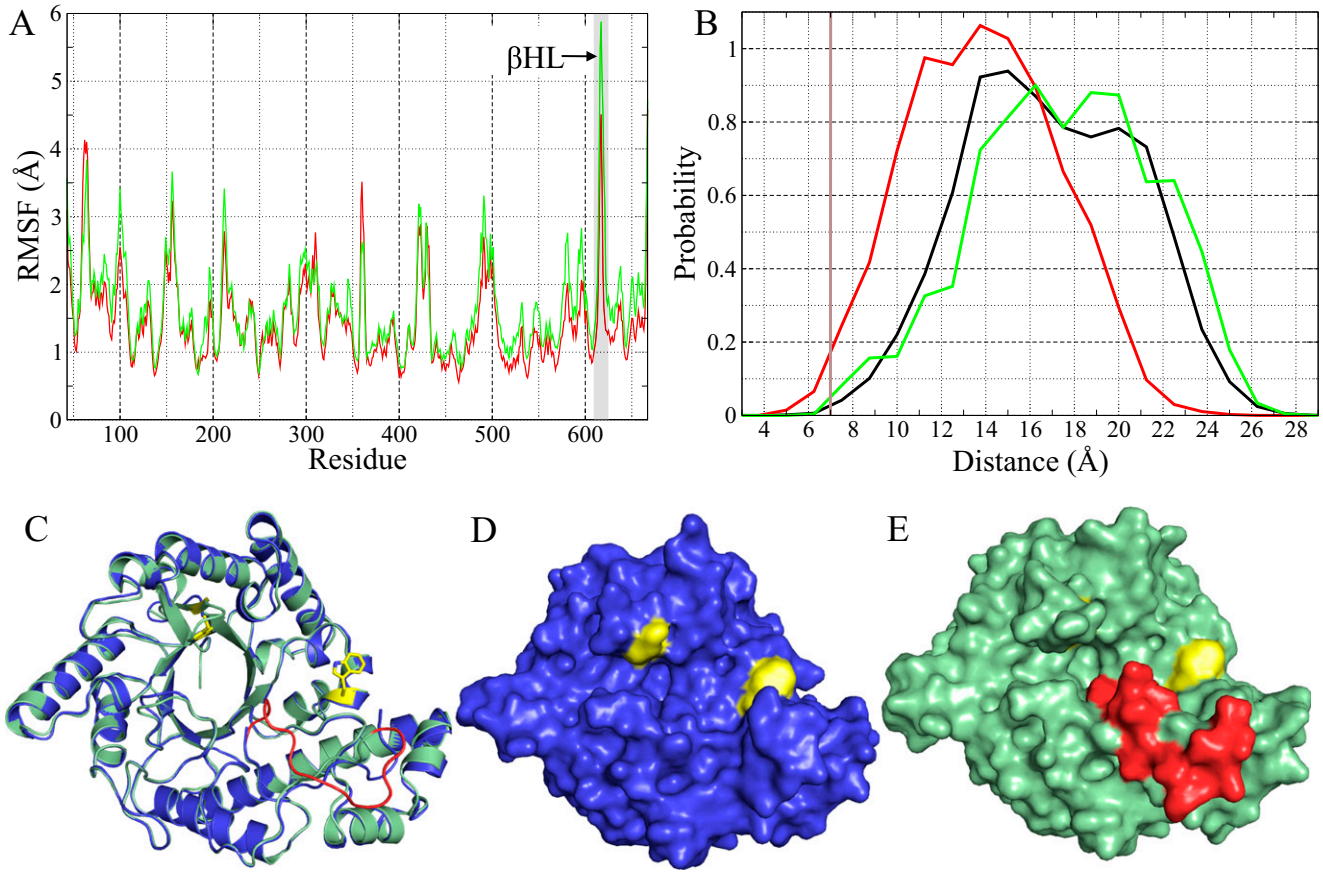


Fig. S6. β-Hairpin loop (βHL) is flexible and important for saccharide binding. (*A*) Root mean square fluctuations (RMSFs) of the β-D-GlcNAc (red) and β-D-GlcNH$_3^+$ (green) MD simulations show that the βHL is the most flexible region of the protein. (*B*) Binding of β-D-GlcNAc decreases the width of the electronegative binding groove compared with the apo- and β-D-GlcNH$_3^+$ MD simulations. Distance measurements taken at each snapshot were at the pinch point of the electronegative groove between the backbone atoms of N616 and G361. The brown line represents the closed state of the groove observed in the crystal structure of PgaB$_{310–672}$ ($P2_12_12_1$ crystal form). (*C*) Comparison of PgaB$_{310–672}$ $P2_12_12_1$ and P1 crystal forms. The P1 crystal form (blue) superimposes with the $P2_12_12_1$ crystal form (pale green) with an rmsd of 0.3 Å. The quality of the density in the P1 crystal form prevented some of the βHL residues (614–618) and a loop below subsites 7–12 (red) from being modeled. W613 and W552 are shown in stick representation and are colored yellow. Surface representations for the P1 crystal form (*D*) and $P2_12_12_1$ crystal form (*E*) are shown with the same color scheme as in *C*.

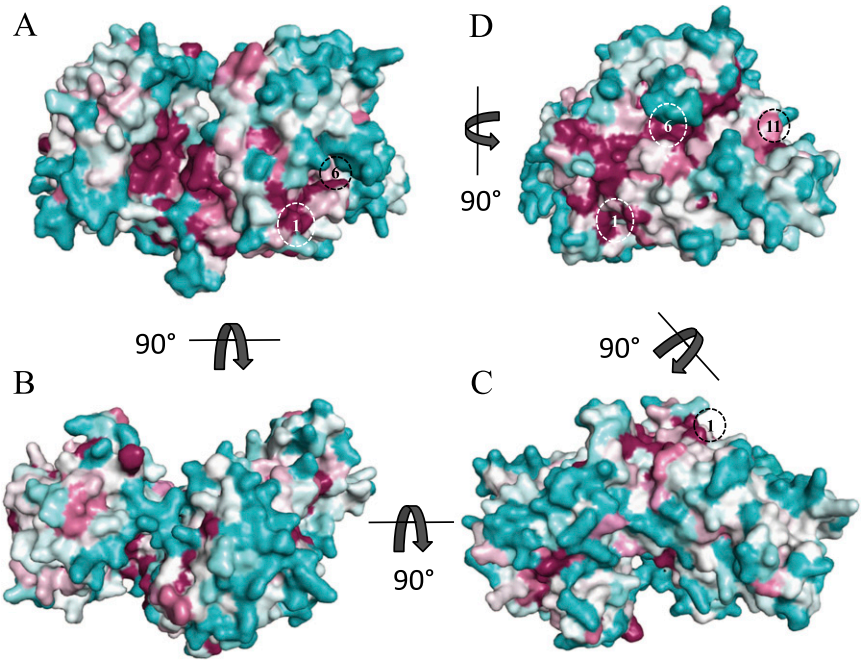


Fig. S7. Sequence conservation analysis of PgaB shown in surface representation colored from low (cyan) to high (maroon) and calculated by the ConSurf web server (1). This figure is in the same representation as Fig. 4. (*A*) De–*N*-acetylation active site and the cleft formed from the N- and C-terminal domains are highly conserved. (*B* and *C*) Region connecting the IDL to the C-terminal domain has low conservation. (*D*) C-terminal domain electronegative groove is highly conserved. In *A*, *C*, and *D*, the location of predicted binding sites in the C-terminal domain are indicated by the number in a dashed black or white circle.

1. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–W533.

Table S1. Summary of data collection and refinement statistics

	$PgaB_{310–672}$	$PgaB_{310–672}$ P1 crystal form	GlcNAc complex	GlcN complex	β-1,6-$(GlcNAc)_6$ complex
Data collection					
Beamline	NSLS X29A	NSLS X29A	CLS 08ID-1	NSLS X29A	CLS 08ID-1
Wavelength, Å	1.08	1.08	1.08	1.08	1.08
Space group	$P2_12_12_1$	$P1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters; Å, °	a = 41.8	a = 41.6, b = 54.0,	a = 42.4,	a = 42.0,	a = 42.2,
	b = 78.2	c = 86.5; α = 101.7,	b = 78.0,	b = 78.8,	b = 77.8,
	c = 97.9	β = 98.3, γ = 90.2	c = 115.3	c = 116.0	c = 114.8
Resolution, Å	50.00–1.80	50.00–1.48	50.00–1.65	50.00–1.89	50.00–1.80
	(1.86–1.80)	(1.53–1.48)	(1.71–1.65)	(1.96–1.89)	(1.86–1.80)
Total no. of reflections	400,151	370,273	648,400	444,127	362,933
No. of unique reflections	30,391	126,137	46,798	31,672	35,680
Redundancy	13.5 (12.7)	3.0 (2.9)	14.1 (13.9)	14.0 (13.7)	10.2 (10.4)
Completeness, %	97.5 (96.4)	94.1 (90.4)	99.5 (98.5)	100 (99.9)	100 (100)
Average, $I/\sigma(I)$	26.8 (5.1)	11.3 (2.4)	64.8 (5.2)	19.9 (5.0)	48.3 (5.9)
R_{merge},* %	10.4 (59.9)	5.7 (29.9)	7.7 (59.2)	13.8 (54.0)	8.8 (55.7)
Refinement					
R_{work}/R_{free}†	14.2/18.9	22.1/25.3	15.7/19.5	14.7/18.6	15.8/19.2
No. of atoms					
Protein	2,875	5,610	2,925	2,960	2,914
Ligands	5		68	36	69
Water	279	814	300	389	221
Average B-factors,‡ Å^2					
Protein	21.6	21.2	28.1	24.7	30.8
Ligands	25.8		65.2	72.0	65.4
Water	32.9	34.2	40.2	35.1	42.1
rmsd					
Bond lengths, Å	0.007	0.006	0.006	0.007	0.008
Bond angles, °	1.02	1.08	1.02	1.04	1.18
Ramachandran plot‡					
Total favored, %	99.2	99.0	98.9	98.9	98.0
Total allowed, %	100.0	100.0	100.0	100.0	100.0
Coordinate error,§ Å	0.16	0.17	0.16	0.19	0.16
PDB ID code	4P7L	4P7O	4P7Q	4P7N	4P7R

Values in parentheses correspond to the highest resolution shell. CLS, Canadian Light Source; $I/\sigma(I)$, intensity of a group of reflections divided by the standard deviation of those reflections; NSLS, National Synchrotron Light Source.

*$R_{merge} = \sum\sum | I(k) - \langle I \rangle | / \sum I(k)$, where $I(k)$ and $\langle I \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.

†$R_{work} = \sum ||F_{obs}| - k|F_{calc}|| / |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is the sum extended over a subset of reflections excluded from all stages of the refinement.

‡As calculated using MolProbity (1).

§As calculated by PHENIX (2).

1. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
2. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system formacromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.

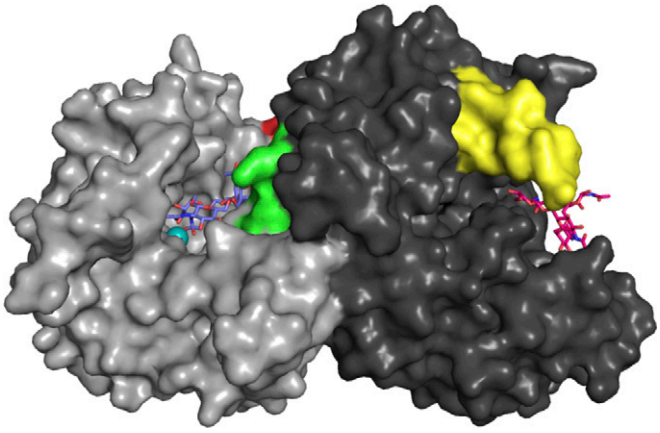
Table S2. Partial charges for the β-D-GlcNAc obtained from the Glycam Biomolecule Builder

Atom no.	Atom label	Partial charge
1	HO1	0.445
2	O1	−0.639
3	C1	0.287
4	H1	0.000
5	O5	−0.433
6	C5	0.208
7	H5	0.000
8	C6	0.289
9	H61	0.000
10	H62	0.000
11	O6	−0.689
12	H6O	0.424
13	C4	0.302
14	H4	0.000
15	O4	−0.716
16	H4O	0.436
17	C3	0.180
18	H3	0.000
19	O3	−0.681
20	H3O	0.423
21	C2	0.480
22	H2	0.000
23	N2	−0.722
24	H2N	0.300
25	C2N	0.648
26	O2N	−0.576
27	CME	0.034
28	H1M	0.000
29	H2M	0.000
30	H3M	0.000

Table S3. Computed partial charges for β-D-GlcNH$_3^+$ using CHELPG methodology

Atom no.	Atom label	Partial charge
1	HO1	0.5553
2	O1	−0.7910
3	C1	0.3296
4	H1	−0.0000
5	O5	−0.5022
6	C5	0.4140
7	H5	−0.0000
8	C6	0.2176
9	H61	−0.0000
10	H62	−0.0000
11	O6	−0.6878
12	H6O	0.4618
13	C4	0.1600
14	H4	−0.0000
15	O4	−0.6754
16	H4O	0.4543
17	C3	0.1387
18	H3	−0.0000
19	O3	−0.7678
20	H3O	0.5289
21	C2	0.6259
22	H2	0.0000
23	N3	−0.4045
24	H1N	0.3142
25	H2N	0.3142
26	H3N	0.3142

CHELPG, charges from electrostatic potentials using a grid.



Movie S1. Comparison of the MD simulation density with the docked and complex structure ligands. Surface representation of PgaB (N- and C-terminal domains colored light and dark gray, respectively) overlaid with β-D-GlcNAc (purple) and β-D-GlcNH$_3^+$ (taupe) densities and the N-terminal domain docked β-1,6-(GlcNAc)$_5$ tetrahedral intermediate (blue) and C-terminal domain β-1,6-(GlcNAc)$_4$ complex (magenta) shown in stick representation. Binding densities are depicted at occupancies of 0.15; the nickel ion is colored teal; the interdomain linker is colored red; W387, T391, and R392 are colored green; and the β-hairpin loop is colored yellow.

Movie S1