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.

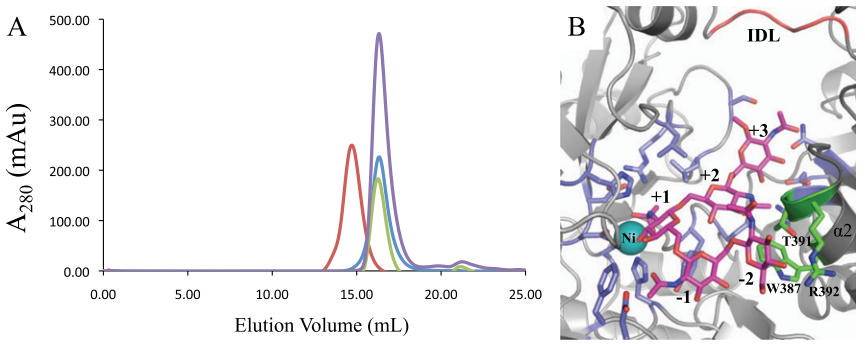


Fig. S1. (*A*) Gel filtration chromatograms of $PgaB_{22–72}$ (red), $PgaB_{22–309}$ (blue), $PgaB_{310–672}$ (green), and $PgaB_{22–309}$ plus $PgaB_{310–672}$ mixed together (purple). mAu, milliabsorbance units. (*B*) Close-up view of the top docked β-1,6-$(GlcNAc)_5$ tetrahedral intermediate (magenta, 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.

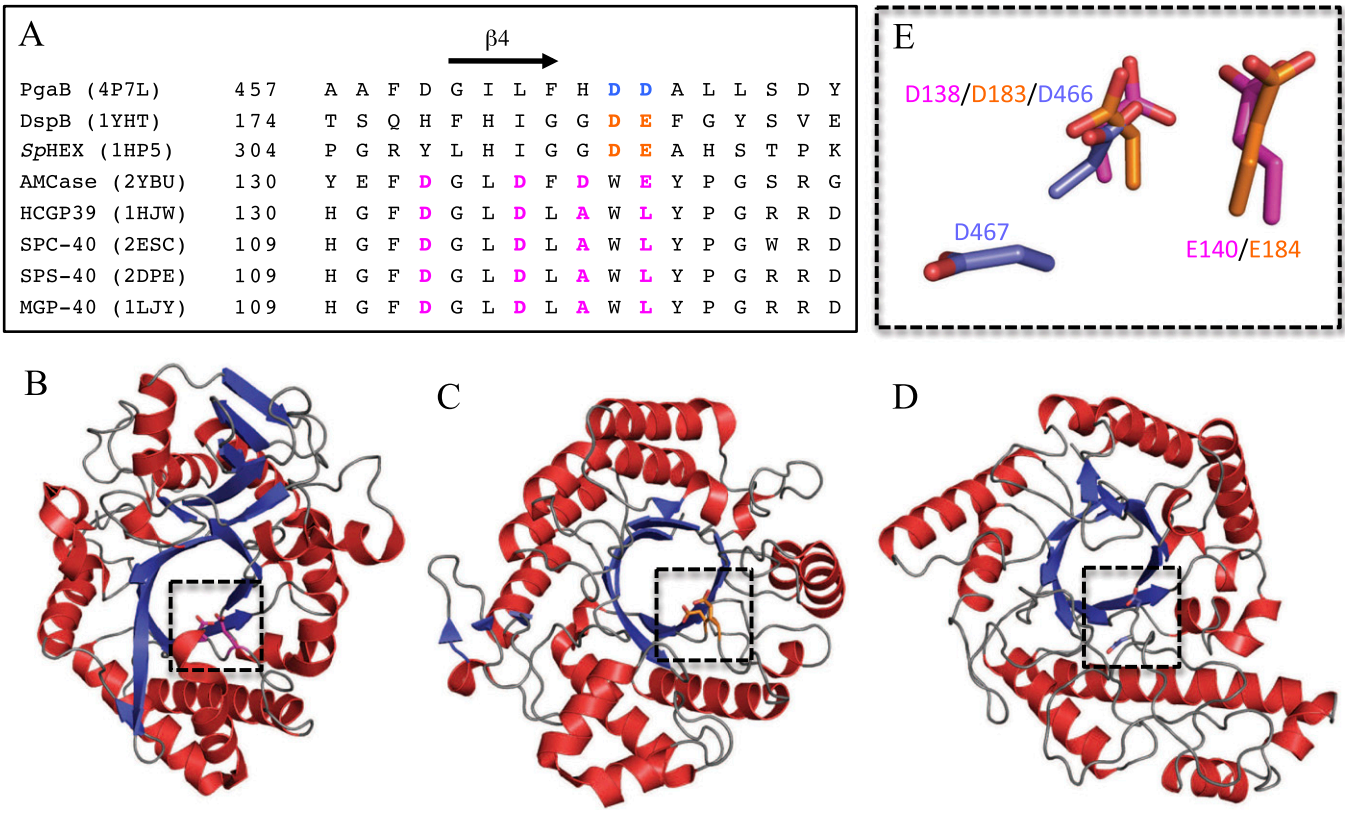


Fig. S2. Comparison of $PgaB_{310–672}$ with glycoside hydrolase family 18 (GH18) and GH20 active sites. (*A*) Sequence alignment of the catalytic motifs for GH18 (colored orange) and GH20 (colored magenta) with $PgaB_{310–672}$ (colored blue). Dispersin B (DspB) and *Streptomyces plicatus* hexosaminidase (*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*.

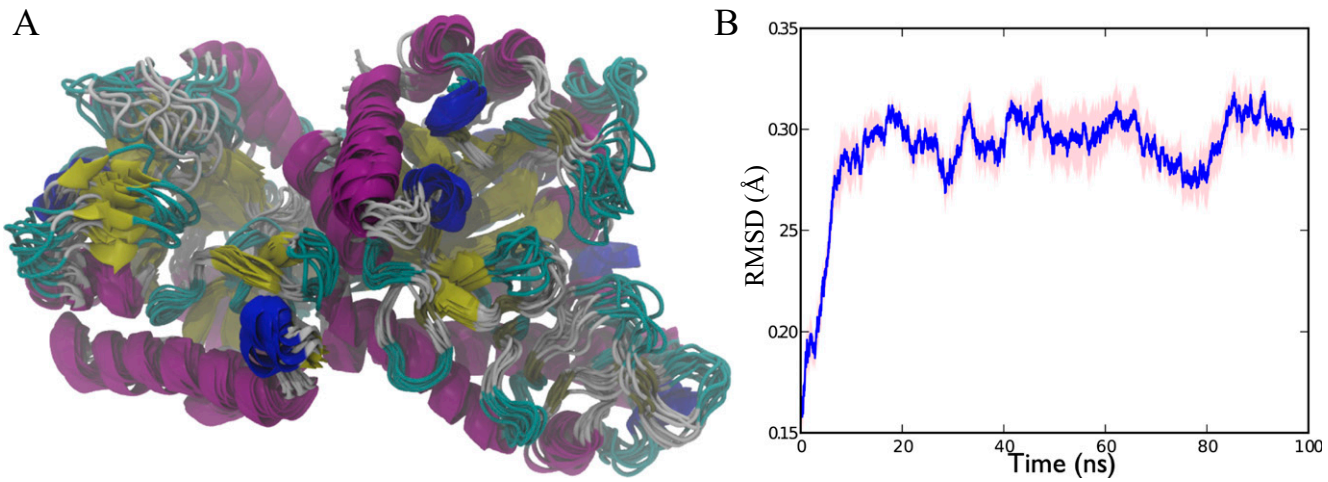


Fig. S3. Fluctuation of PgaB during MD simulations. (*A*) Fluctuations in the $PgaB_{43–667}$ structure during the MD simulations shown in a cartoon colored according to the secondary structure. Purple, α-helix; blue, 3-10 helix; yellow, β-strand; gray, β-turn; teal, loop. (*B*) Average rmsd (angstroms) 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.

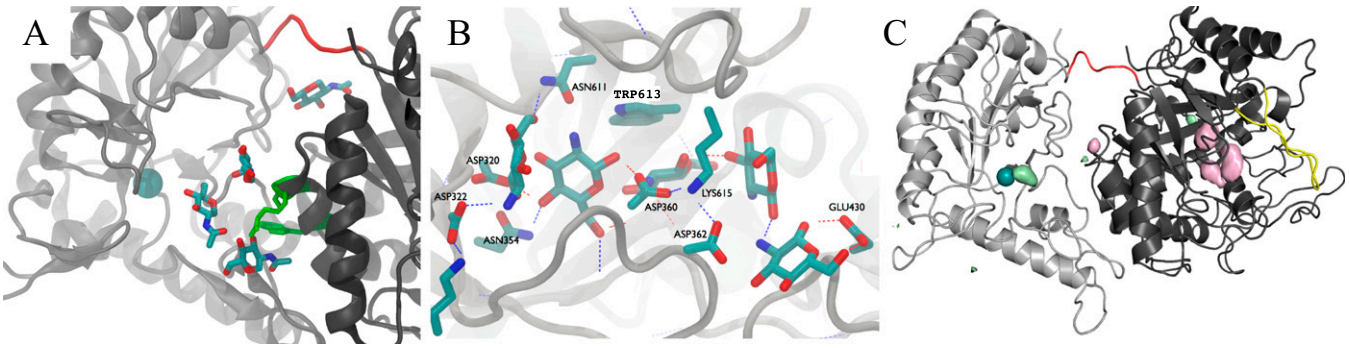


Fig. S4. (*A*) Examples of the β-D-GlcNAc (cyan, stick representation) binding modes along the N- and C-terminal domain cleft from the conserved residue patch (green) on α2 of $PgaB_{310–672}$ across the metal ion (teal) to the IDL (red). (*B*) Examples of the β-D-$GlcNH_3^+$ (cyan, stick representation) 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.

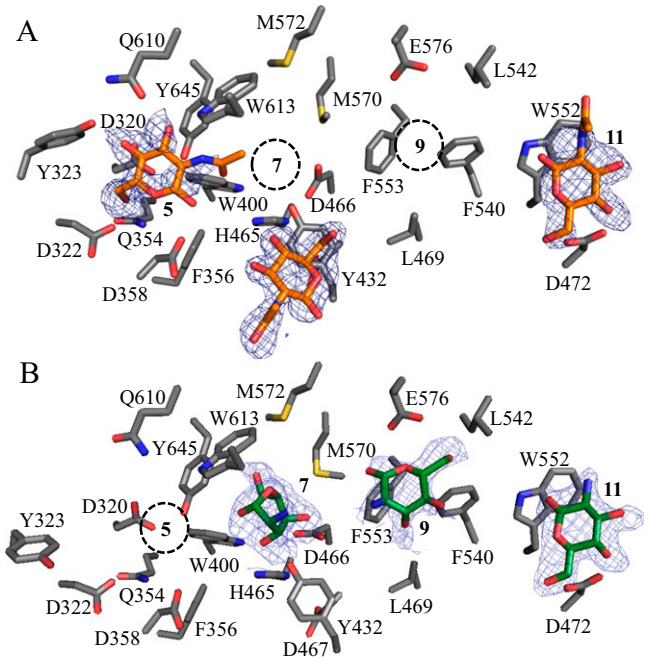


Fig. S5. Crystal structures of $PgaB_{310–672}$ in complex with GlcNAc (*A*) and GlcN (*B*). $PgaB_{310–672}$ residues, GlcNAc, and GlcN are shown in stick representation colored gray, orange, and green, respectively. Unbiased $|F_o - F_c|$ density omit maps for GlcNAc and GlcN are shown as blue mesh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

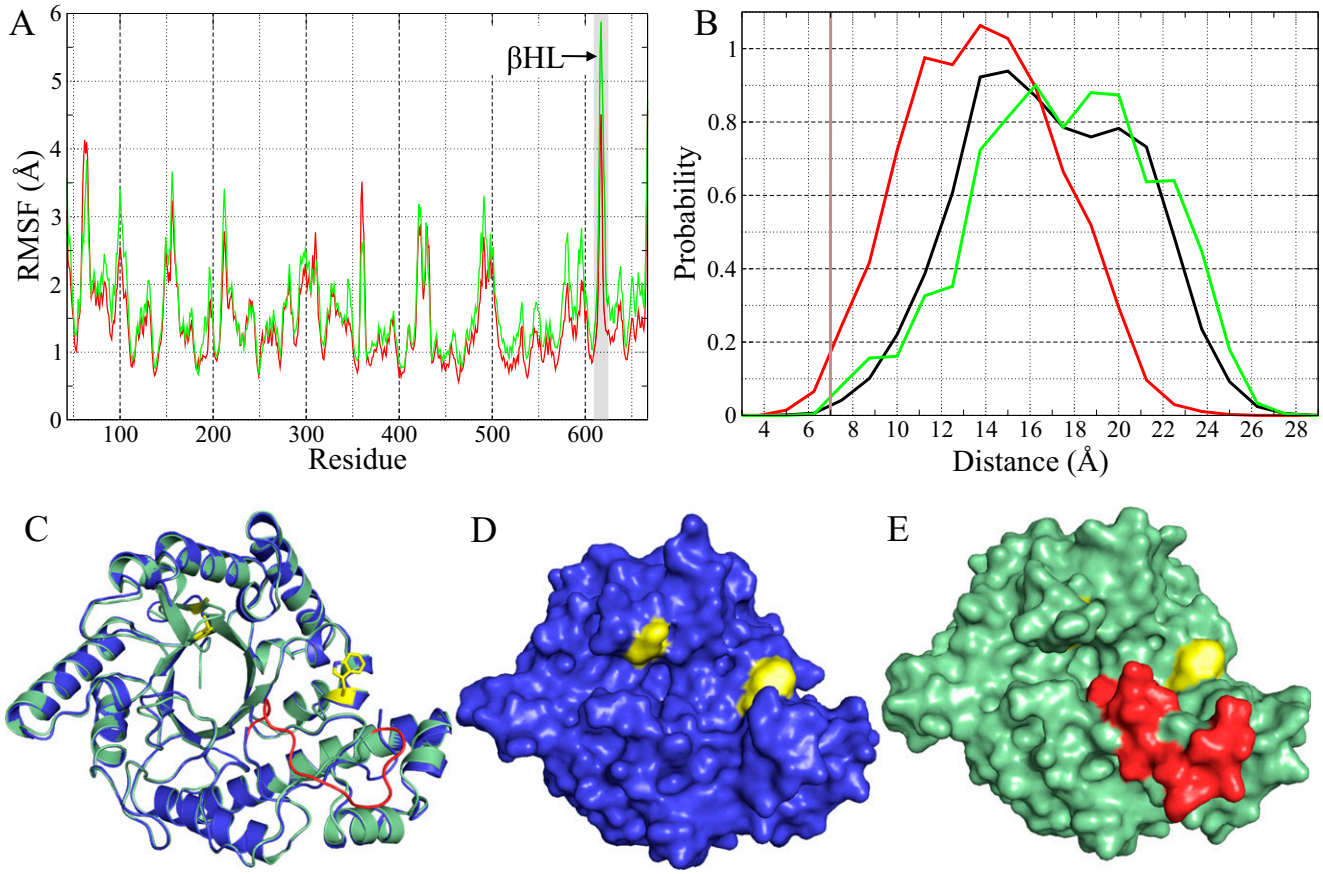


Fig. S6. β-Hairpin loop (βHL) is flexible and important for saccharide binding. (*A*) Root mean square fluctuations (RMSFs) of the β-D-GlcNAc (red) and β-D-GlcNH$_3^+$ (green) MD simulations show that the βHL is the most flexible region of the protein. (*B*) Binding of β-D-GlcNAc decreases the width 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*.

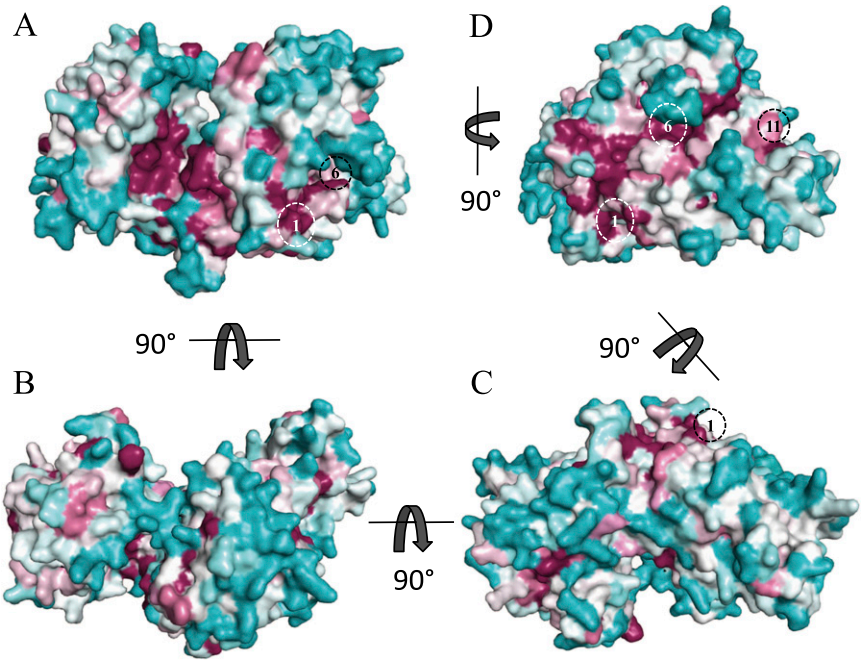


Fig. S7. Sequence conservation analysis of PgaB shown in surface representation colored from low (cyan) to high (maroon) and calculated by the ConSurf web server (1). This figure is in the same representation as Fig. 4. (*A*) De–*N*-acetylation active site and the cleft formed from the N- and C-terminal 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.

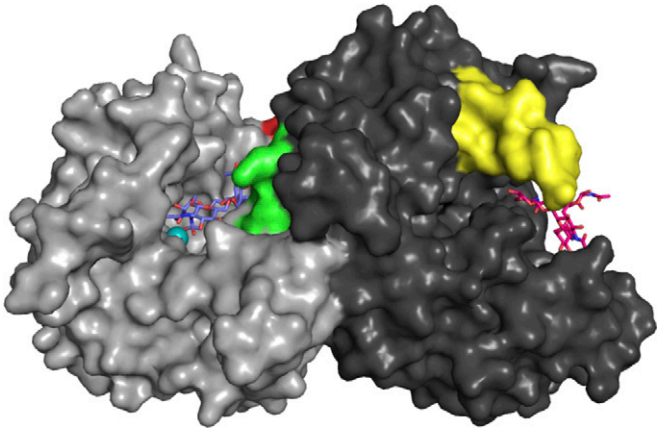
Table S2. Partial charges for the β-D-GlcNAc obtained from the Glycam Biomolecule Builder

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