Supporting Information

Schmidt et al. 10.1073/pnas.1119894109

SI Results and Discussion

Model Quality and Overall Structural Aspects. For all four protein copies in the asymmetric unit (rmsd 0.42–0.65 Å), the main chain of $WaaA_{AAE}$ as well as most of the side chains are well defined by electron density. A somewhat weaker side-chain electron density exists for residues 214–238 and 246–249. For the substrate-free and the CMP-bound structure, 99.4% and 99.1% of the amino acid residues, respectively, are in the allowed regions of the Ramachandran plot. The molecules A and C as well as B and D are related to each other by a twofold noncrystallographic symmetry (NCS) axis and are arranged in a parallel, face-to-face manner. Extensive molecular contacts between two N-terminal domains contribute to most of the surface area buried by facing $WaaA_{AAE}$ molecules (about 1,000 $Å^2$ from a total of 1,500 $Å^2$ per molecule). The interface contains two NCS-related polyethylene glycol (PEG) molecules with three to six ethylene glycol units as well as citrate (CMP-bound structure) or glycerol (substrate-free structure) molecules contacting residues F3–L6, R8, F53, P55, L76, F101, W102, P103, K121, and I125. In contrast to the PEG molecule, citrate and glycerol molecules bind outside functionally important $WaaA_{AAE}$ regions. In the substrate-free structure, we observed |Fo|-|Fc| difference density (8 σ above the mean) close to the side chain of H213, which was interpreted as a nickel ion after calculation of an anomalous-difference density map from data collected at the nickel absorption-edge wavelength of 1.48482 Å. Consequently, freshly purified $WaaA_{AAE}$ was subjected to high-resolution inductively coupled plasma mass spectrometry to analyze the presence of divalent cations such as Mg^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+}, Fe^{2+}, and Ni^{2+}. In accord with our interpretation of the structural data, by far the strongest enrichment was evident for Ni^{2+}, displaying a signal >60-fold above the background. The Ni^{2+}:$WaaA_{AAE}$ stoichiometry was about 1:4, and thus the binding site seems to be only partially occupied. Comparison of the substrate-free and CMP-bound structures revealed that the Ni^{2+} ion is located more than 9 Å away from the phosphorous atom of a bound CMP, a distance that is too far for the Ni^{2+} ion to play a role in catalysis.

As mentioned above, $WaaA_{AAE}$ forms homodimers in the crystal. Nonetheless, the biological unit most probably is a monomer, as described for many other GT-B members (1, 2). This conclusion is based on several structural and functional considerations. First, the interface region buries a PEG molecule in the potential acceptor-substrate binding site of each monomer, indicating a poor surface complementarity. Second, because functionally important regions are inaccessible, the observed dimer is not compatible with acceptor-substrate binding and membrane association. We assume that in the absence of membranes the hydrophobic effect as a key driving force for protein association (3) triggered the formation of dimers in our $WaaA_{AAE}$ preparation over a longer period. Finally, it appears very unlikely that the observed dimeric assembly allows the conformational rearrangements crucial for glycosyl transfer. It has become widely accepted that GT-B enzymes can adopt an open empty and a closed substrate-bound conformation, the latter of which brings donor- and acceptor-substrate binding sites into close proximity. The required movement of the C-terminal toward the N-terminal domain of $WaaA_{AAE}$ is not possible under the constraints of the homodimer that seems to hold the enzyme in the open conformation.

Detailed Structural Comparison of the CMP-Binding Modes in $WaaA_{AAE}$ and GT-B Sialyltransferases. The C-terminal domain of $WaaA_{AAE}$ in complex with CMP was compared with the equivalent domains of all currently available sialyltransferase structures displaying the GT-B fold, i.e., PM0188 (sialyltransferase from *Pasteurella multocida*) (4–6), Δ16psp26ST (α2,6-sialyltransferase from *Photobacterium* sp. JT-ISH-224) (7) and ΔNpp23ST (α/β-Galactoside α2,3-sialyltransferase from *Photobacterium phosphoreum*) (8) of GT-family 80, as well as Δ29NST (lipooligosaccharide sialyltransferase from *Neisseria meningitidis*) (9) of GT-family 52 (Fig. S2). The C-terminal domains of $WaaA_{AAE}$, PM0188, Δ16psp26ST, ΔNpp23ST, and Δ29NST superimpose with rmsd values in the range between 2.8 and 3.1 Å for 89–100 aligned Cα atoms, with the CMP molecules bound in matching areas. Loops connecting the C termini of β-strands with the flanking helices (cβ2, cβ4, and cβ5 in $WaaA_{AAE}$; cβ1, cβ2, and cβ3 in the sialyltransferases) and an α/β/α motif, overlapping with the α/β/α motif described for MurG (2), (cα3, cβ5, and cα4 in $WaaA_{AAE}$; cα2/cα3, cβ4, and cα3/cα4 in the sialyltransferases) form the CMP-binding sites (Fig. S2). Hydrophobic residues that sandwich the cytosine base are located in the first α-helix of the α/β/α motif and its preceding loop (F247 and L250 of $WaaA_{AAE}$) or in the second α-helix of the α/β/α motif and the loop region following cβ2 (P312 and L357 of PM0188; P406 and L451 of Δ16psp26ST; P318 and V361 of ΔNpp23ST) or can be found in the first α-helix of the α/β/α motif and the loop following cβ2 (P281 and I299 of Δ29NST) (Fig. S2). The proline residues of the highly specific CMP-binding HP sialyl motif of GT-B sialyltransferases seem to form CH–π bonds with their respective cytosine moiety, whereas the orientation of F247 in the $WaaA_{AAE}$-CMP complex suggests the existence of a π–π stacking interaction (10, 11). In all structures, the cytidine, adopting the anti-conformation (O4-C1-N1-C2 torsion angles: −111.6° in $WaaA_{AAE}$, −170.1° in PM0188, −161.8° in Δ16psp26ST, −157.4° in ΔNpp23ST, and −153.1° in Δ29NST), is buried within a deep pocket. In most enzymes ($WaaA_{AAE}$, PM0188, Δ16psp26ST, and ΔNpp23ST), the volume of this pocket matches the size of a pyrimidine base. Hydrogen-bonding interactions between the N4 amide group of the cytidine and main-chain carbonyl groups of residues forming the pocket may contribute to nucleotide specificity (5) and explain the high affinity of $WaaA_{AAE}$ for CMP (12). It is noteworthy that the ribose pucker of the donor substrate in $WaaA_{AAE}$ acquires the C2′-endo–C3′-exo conformation as opposed to C2′-exo–C3′-endo in the sialyltransferases. We assume this difference is a direct consequence of the spatial location of the glutamate residue that contacts the ribose hydroxyls. The respective glutamate is located in either the second ($WaaA_{AAE}$ E276) or the first α-helix (PM0188 E338; Δ16psp26ST E432; ΔNpp23ST E342; Δ29NST E300) of the α/β/α motif.

Interactions between the phosphate group of CMP and the protein occur predominantly at the N-terminus of the second α-helix of the α/β/α motif, the adjacent loop region, and the loop that follows cβ2 (Fig. S2). In $WaaA_{AAE}$, the main-chain amide of N273 as well as side-chain atoms of R212 contact the CMP phosphate group, whereas the sialyltransferases use serine side-chain and main-chain hydrogen bonding (PM0188 S355 and S356; Δ16psp26ST S449 and S450; ΔNpp23ST S359 and S360; Δ29NST S322 and G323) and the histidine residue of the HP sialyl motif (PM0188 H311; Δ16psp26ST H405; ΔNpp23ST H317; Δ29NST H280).

$WaaA_{AAE}$ Tryptophan Quenching. Tryptophan-fluorescence measurements of wild-type protein revealed an increase in the total fluorescence intensity upon addition of lipid IV_A (Fig. 5*B*). In

contrast, the fluorescence intensity of the W102A variant decreased slightly in the presence of lipid IV_A. Because neighboring hydrophobic and acidic amino acid residues may influence tryptophan fluorescence, we also tested variant E100A as an internal control. Upon addition of lipid IV_A, the increase in fluorescence intensity for variant E100A was reduced slightly compared with that of the wild-type protein. Because an increase in tryptophan-fluorescence intensity most likely points toward a more hydrophobic microenvironment of the side chain, even in the absence of a clear blue shift (13, 14), residue W102 is very likely to interact with the hydrophobic part of lipid IV_A.

SI Materials and Methods

Crystallization and Data Collection. Crystals of $WaaA_{AAE}$ were grown by the hanging-drop vapor-diffusion method using Nextal drop guard plates (Qiagen). Equal volumes of protein (10–18 mg/mL) and reservoir solution [100 mM Tris·HCl (pH 8.5), 35–40% (vol/vol) PEG 400, 200 mM Na-citrate, 50 mM β-mercaptoethanol] were mixed and equilibrated against 900 μL reservoir solution. Crystals were obtained at 19 °C within 2 wk and grew to maximum dimensions of $0.7 \times 0.2 \times 0.1$ mm^3. The crystals were mounted directly in a 100-K nitrogen gas stream before data collection. Diffraction data were collected to a maximum Bragg spacing of 2.0 Å at the I911-3 beamline of the MAX-lab National Laboratory for Synchrotron Radiation synchrotron (Lund, Sweden). The crystals belonged to the monoclinic space group $P2_1$ with unit-cell parameters a = 132.50 Å, b = 45.82 Å, c = 144.07 Å, and β = 97.23°. $WaaA_{AAE}$ complexes with CMP were obtained by soaking crystals in 100 mM Tris·HCl (pH 8.5), 36% (vol/vol) PEG 400, 200 mM Na-citrate, and 10 mM CMP for 4 d. Diffraction data were collected to a resolution of 2.4 Å on crystals flash-cooled in a 100-K nitrogen gas stream at beamline I911-2 (MAX-lab National Laboratory for Synchrotron Radiation, Lund, Sweden). Several different anomalous datasets were collected at beamline BL14.2 at the Berlin Electron Storage Ring Society for Synchrotron Radiation (BESSY; Berlin, Germany) and European Molecular Biology Laboratory beamline X12 at Deutsches Elektronen-Synchrotron (DESY) (Hamburg, Germany). For preparation of Ir^{3+}-derivatized crystals, K_3IrCl_6 was added to a final concentration of 10 mM to droplets containing $WaaA_{AAE}$ crystals. Following incubation at 19 °C for 4 d, anomalous data were collected to a maximum resolution of 3.1 Å at the Ir^{3+} absorption edge peak wavelength of 1.10584 Å. All datasets were processed with the programs MOSFLM (15) or XDS (16) and scaled using the program SCALA (17).

Structure Determination and Refinement. Experimental phase information was obtained using the single isomorphous replacement with anomalous scattering (SIRAS) approach in AUTOSHARP (18). Electron-density modification using SOLOMON (19) within AUTOSHARP revealed the boundaries of four protein molecules, each having two clearly distinguishable domains and several right-handed α-helices. The assignment of the helices in Coot (20) allowed the identification of initial NCS parameters with LSQKAB (21). NCS masks for the two domains were produced in NCSMASK (22) using the structure of the GT-B enzyme T4 β-glucosyltransferase (PDB ID code 1BGT) (23) as a template. For subsequent refinement of NCS parameters in IMP (24), the two domains were treated independently to allow for slightly different domain orientations of the four protein copies in the asymmetric unit. NCS averaging using MAPROT (22) revealed additional secondary structure elements. The density was improved markedly by iterative cycles of partial model building in Coot, calculation of partial model phases in SFALL (22), phase combination with experimental phases in SIGMAA (25), and solvent flattening, histogram matching, and NCS averaging in DM (26), including phase extension to 2.8 Å at the later stages of model building. After the assignment of 90% of the main chain and 60% of all side chains, one round of simulated annealing refinement was carried out in CNS (27). Model refinement and building was continued without experimental phase restraints in REFMAC (28) and Coot. Strict NCS restraints were used while gradually increasing the resolution from 2.8 Å to 2.0 Å. A final round of translation-libration-screw (TLS) refinement in Phenix (29) without NCS restraints resulted in an R_{work} factor of 20.7% and an R_{free} factor of 25.4%. The $WaaA_{AAE}$ structure in complex with CMP was determined by molecular replacement with Phaser (30) using the structure of the substrate-free enzyme as a search model. The structure subsequently was refined in REFMAC to final R_{work} and R_{free} factors of 20.2% and 25.1%, respectively. Data and refinement statistics are summarized in Table S1.

Mutagenesis. Amino acid substitutions were created in $WaaA_{AAE}$ using the Change-IT multiple mutation site-directed mutagenesis kit as recommended by the supplier (USB) with plasmid pUM216 as the template. Plasmid pUM216 is a derivative of pET-16b (Novagen) carrying the *waaA* wild-type allele of *Aquifex aeolicus* (12). The site-directed mutagenesis primers are listed in Table S2. The $WaaA_{\Delta 265–269}$ deletion variant lacking the coding sequence for the ^{265}TFVNI^{269} motif was constructed by inverse PCR using pUM216 as the template and the phosphorylated primers AAEwaaA3KdoPri (5′P-TCCGCCAACTATTGCGATTTTTCC-3′) and AAEwaaA5KdoPri (5′P-GGGGGGCATAACCTTCTGGAG-3′) for amplification. The purified PCR product was self-ligated, and the ligation mixture was used to transform *Escherichia coli* XL1-Blue cells to yield the plasmid pUM216-$WaaA_{\Delta 265–269}$. All constructs were sequenced to confirm the presence of the mutations.

Analysis of Fatty Acids Bound to $WaaA_{AAE}$. A total of 850 μg highly purified $WaaA_{AAE}$ in 50 μL $WaaA_{AAE}$ buffer was diluted with 800 μL H_2O and converted to a two-phase Bligh–Dyer system (31) by the successive addition of 3 mL of a chloroform/methanol mixture (1:2, vol/vol), 1 mL chloroform, and 1 mL H_2O. After mixing and separation of the phases by low-speed centrifugation, the organic phase was analyzed for fatty-acid methyl esters and trimethylsilyl derivatives by combined GLC/MS (32).

Induction-Coupled Plasma Mass Spectrometry. For induction-coupled plasma mass spectrometry, 0.9 mL of purified $WaaA_{AAE}$ (3.5 mg/mL) was dialyzed for 36 h against 120 mL of a buffer solution consisting of 25 mM Tris·HCl (pH 8.7), 2 mM Cymal-6, and 5 mM β-mercaptoethanol. Following dialysis, the protein sample was centrifuged, and the $WaaA_{AAE}$ concentration was determined by the Bradford assay (33). Metal analysis by high-resolution inductively coupled plasma mass spectrometry was carried out by T. Huston on a Thermo Scientific Element instrument at the W. M. Keck Elemental Geochemistry Laboratory (Department of Geology, University of Michigan, Ann Arbor, MI). The postdialysis metal content of the dialysis buffer was analyzed as control.

Fluorescence Measurements. Tryptophan-fluorescence data were recorded using a Varian Cary Eclipse fluorescence spectrophotometer with excitation and emission band passes set to 10 nm. The excitation wavelength was set to 295 nm to reduce tyrosine fluorescence interference, and emission was recorded from 300–400 nm. First, 2.2 μg of protein in 0.5 mL buffer [20 mM Tris (pH 8.5), 50 mM NaCl, 5% (vol/vol) glycerol] was measured as background reference. Then, 2 μL (1.5 μg) of a lipid IV_A solution was added, the solution was stirred and remeasured, and the data were corrected by reference subtraction. Experiments were repeated three times, and the results were averaged.

1. Grizot S, et al. (2006) Structure of the *Escherichia coli* heptosyltransferase WaaC: Binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. *J Mol Biol* 363: 383–394.
2. Hu Y, et al. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci USA* 100:845–849.
3. Bahadur RP, Zacharias M (2008) The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. *Cell Mol Life Sci* 65:1059–1072.
4. Kim DU, Yoo JH, Lee YJ, Kim KS, Cho HS (2008) Structural analysis of sialyltransferase PM0188 from *Pasteurella multocida* complexed with donor analogue and acceptor sugar. *BMB Rep* 41:48–54.
5. Ni L, et al. (2006) Cytidine 5′-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. *Biochemistry* 45: 2139–2148.
6. Ni L, et al. (2007) Crystal structures of *Pasteurella multocida* sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* 46:6288–6298.
7. Kakuta Y, et al. (2008) Crystal structure of *Vibrionaceae Photobacterium* sp. JT-ISH-224 alpha2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: Catalytic mechanism and substrate recognition. *Glycobiology* 18:66–73.
8. Iwatani T, et al. (2009) Crystal structure of α/β-galactoside α2,3-sialyltransferase from a luminous marine bacterium, *Photobacterium phosphoreum*. *FEBS Lett* 583:2083–2087.
9. Lin LYC, et al. (2011) Structure and mechanism of the lipooligosaccharide sialyltransferase from *Neisseria meningitidis*. *J Biol Chem* 286:37237–37248.
10. Brandl M, Weiss MS, Jabs A, Sühnel J, Hilgenfeld R (2001) C-H...π-interactions in proteins. *J Mol Biol* 307:357–377.
11. Weiss MS, Brandl M, Sühnel J, Pal D, Hilgenfeld R (2001) More hydrogen bonds for the (structural) biologist. *Trends Biochem Sci* 26:521–523.
12. Mamat U, et al. (2009) WaaA of the hyperthermophilic bacterium *Aquifex aeolicus* is a monofunctional 3-deoxy-D-*manno*-oct-2-ulosonic acid transferase involved in lipopolysaccharide biosynthesis. *J Biol Chem* 284:22248–22262.
13. De Kroon AIPM, Soekarjo MW, De Gier J, De Kruijff B (1990) The role of charge and hydrophobicity in peptide-lipid interaction: A comparative study based on tryptophan fluorescence measurements combined with the use of aqueous and hydrophobic quenchers. *Biochemistry* 29:8229–8240.
14. Legardinier S, et al. (2009) Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin. *J Mol Biol* 389:546–558.
15. Leslie AGW, Powell HR (2007) Evolving Methods for Macromolecular Crystallography, eds Read RJ, Sussmann JL (Springer, Berlin), vol 245, pp 41–51.
16. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 26:795–800.
17. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82.
18. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
19. Abrahams JP, Leslie AGW (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* 52:30–42.
20. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
21. Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* 32:922–923.
22. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
23. Vrielink A, Rüger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J* 13:3413–3422.
24. Kleywegt GJ, Read RJ (1997) Not your average density. *Structure* 5:1557–1569.
25. Read RJ (1986) Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A* 42:140–149.
26. Cowtan K (1994) DM: An automated procedure for phase improvement by density modification. *Jnt CCP4 + ESF-EACBM News Protein Crystallogr* 31:34–38.
27. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
28. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
29. Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
30. Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* 60:432–438.
31. Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37:911–917.
32. Wollenweber H-W, Rietschel ET (1990) Analysis of lipopolysaccharide (lipid A) fatty acids. *J Microbiol Methods* 11:195–211.
33. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.

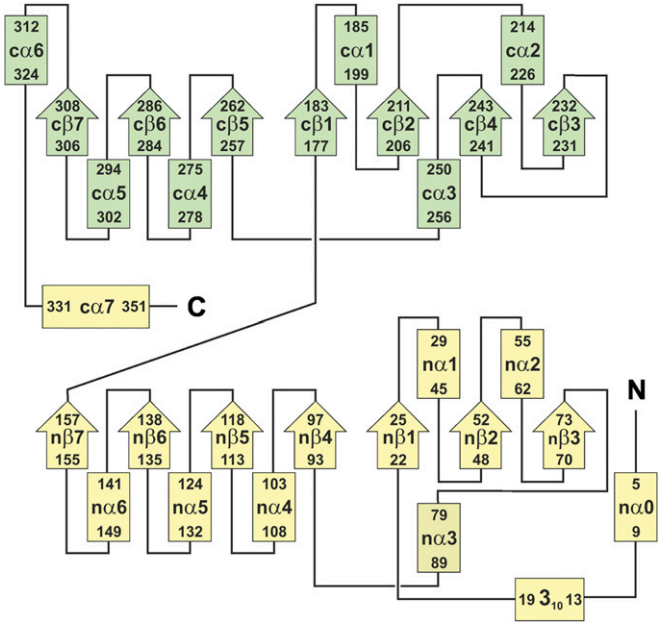

Fig. S1. Schematic representation of the topology of $WaaA_{AAE}$. The α-helices are depicted as rectangles and the β-strands as arrows. The secondary structure elements of the N-terminal domain are highlighted in yellow, and those of the C-terminal domain are in green.

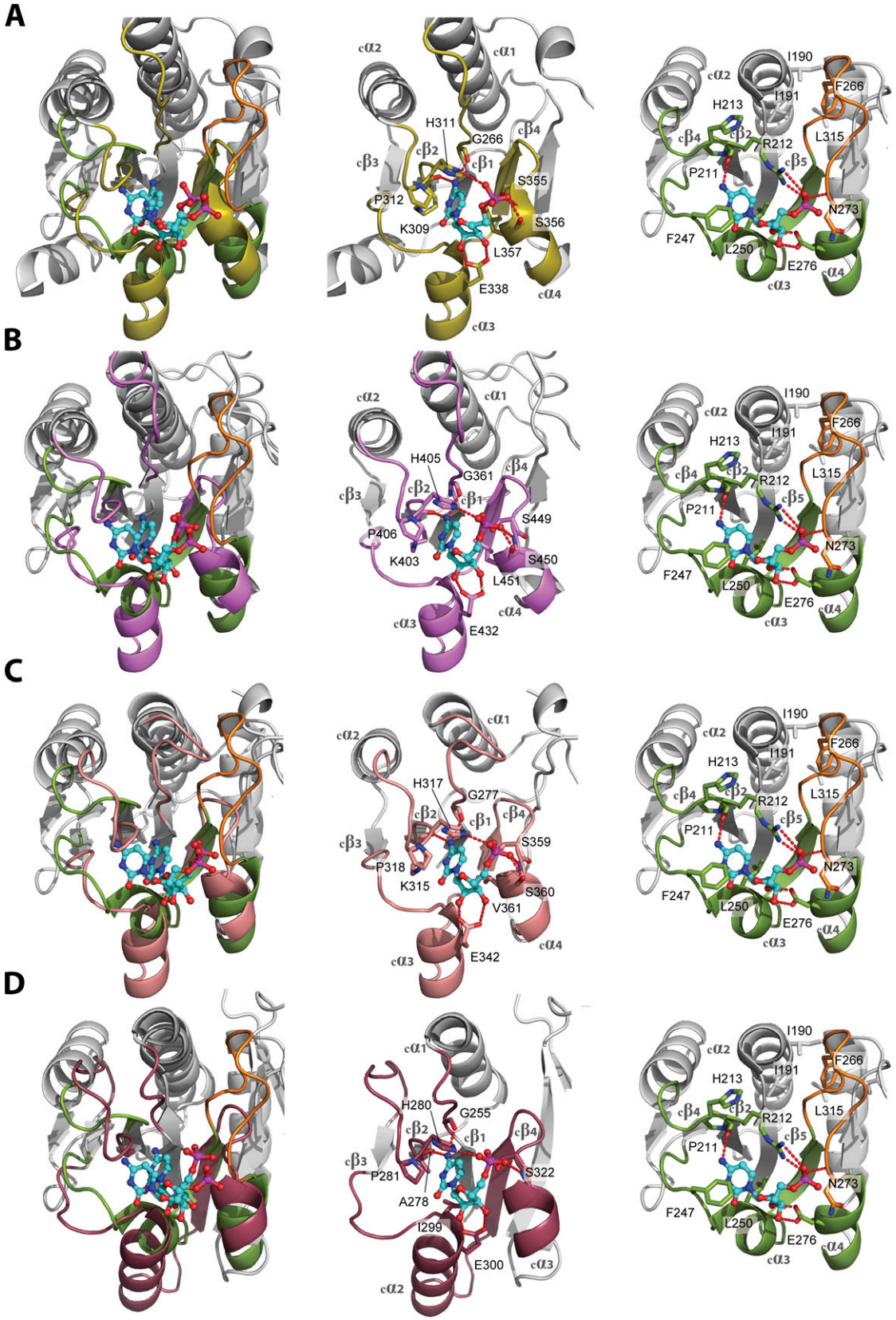

Fig. S2. CMP-binding modes in $WaaA_{AAE}$ and GT-B sialyltransferases. The CMP-binding mode as observed in $WaaA_{AAE}$ (green) was compared with the ones observed in (*A*) PM0188 (olive), (*B*) Δ16psp26ST (pink), (*C*) ΔNpp23ST (salmon), and (*D*) Δ29NST (raspberry). In each panel, the superimposition of the C-terminal domains of $WaaA_{AAE}$ and the respective sialyltransferase is shown on the left, followed by a side-by-side comparison of the two structures. Equivalent protein regions are color-coded, and amino acid residues mediating key interactions are represented as sticks. Hydrogen-bonding interactions are indicated as dotted red lines. For clarity, only the CMP moieties of the donor substrates or analogs are shown.

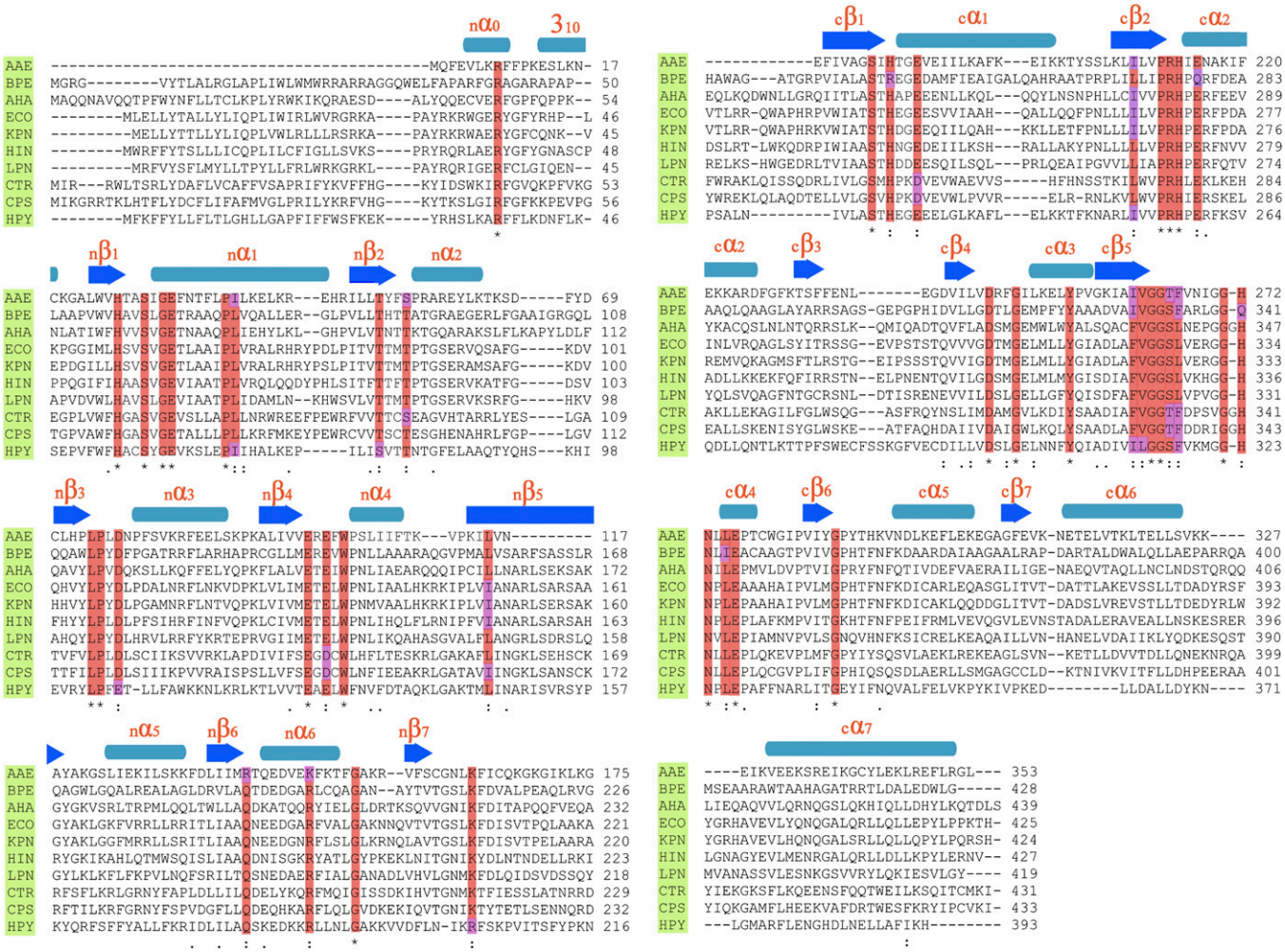

Fig. S3. Comparison of the deduced amino acid sequences of *waaA* genes using the BLOSUM62 scoring matrix for alignment. The single-letter code for amino acids is used. Invariant and similar amino acids are shaded red and magenta, respectively. Secondary structural elements of $WaaA_{AAE}$ are shown above the sequences. Species names with accession numbers in parentheses are AAE, *Aquifex aeolicus* (NP_213223); BPE, *Bordetella pertussis* (2210367A); AHA, *Acinetobacter haemolyticus* (CAB09654.1); ECO, *Escherichia coli* (AAA03745.1); KPN, *Klebsiella pneumoniae* (CAC42120.1); HIN, *Haemophilus influenzae* (NP_438812.1); LPN, *Legionella pneumophila* (CAB43071.1); CTR, *Chlamydia trachomatis* (P0C0Z9.1); CPS, *Chlamydophila psittaci* (CAA49233.1); HPY, *Helicobacter pylori* (AAD08000).

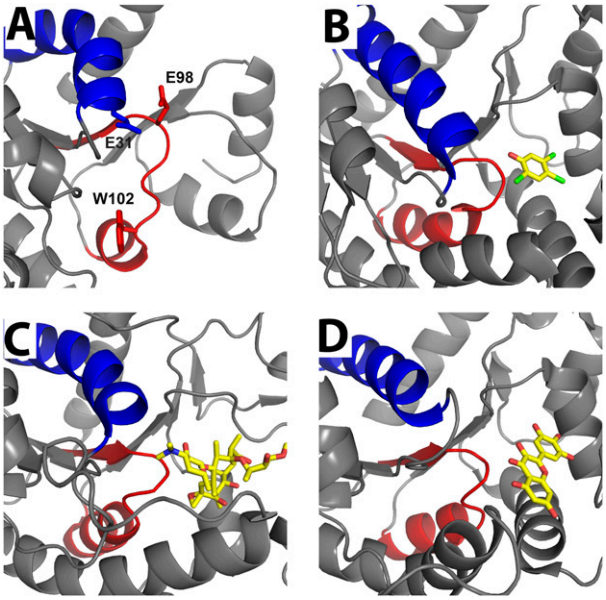

Fig. S4. Binding of acceptor substrates and acceptor-substrate analogs within the N-terminal domain of GT-B enzymes. $WaaA_{AAE}$ (*A*), UGT72B1 (*B*), OleI (*C*), and UGT78G1 (*D*) are represented as gray cartoons. Acceptor substrates and acceptor-substrate analogs (UGT72B1: 2,4,5-trichlorophenol; OleI: oleandomycin; UGT78G1: myricetin) are shown as yellow sticks on which oxygens are shown in red, nitrogens in blue, and chlorines in green. The α-helix nα1 containing the general base (E31 in the case of $WaaA_{AAE}$) is colored in blue. The β-strand nβ4, the α-helix nα4, and the connecting nβ4–nα4 loop are highlighted in red. In *B*, *C*, and *D*, the nβ4–nα4 loop is part of the acceptor-substrate binding site; this location suggests that the corresponding loop region in $WaaA_{AAE}$ (*A*) containing the highly conserved amino acid residues E98 and W102 (shown as sticks) also participates in acceptor substrate binding.

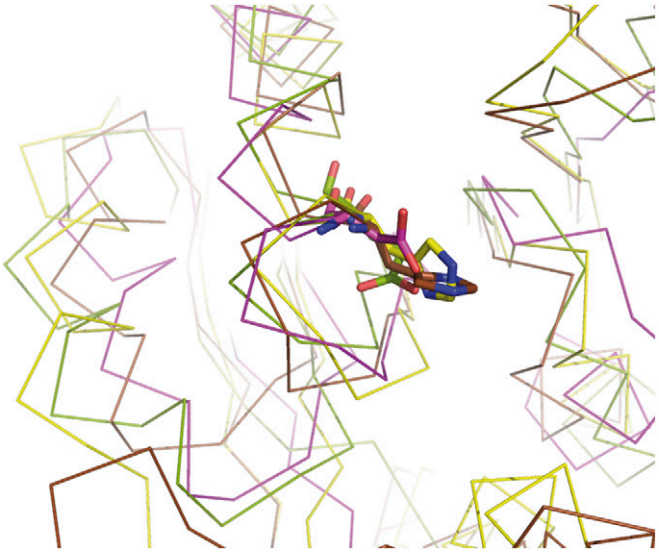

Fig. S5. Structure comparison of the N-terminal domains of $WaaA_{AAE}$ (green), WaaC (magenta), UGT72B1 (yellow), and OleI (brown). The ribbon representations of the superimposed proteins show that residue E31 of $WaaA_{AAE}$ aligns structurally with the confirmed general bases of WaaC (D13), UGT72B1 (H19), and OleI (H25). The general bases are shown as sticks.

Table S1. Data collection and refinement statistics of *A. aeolicus* WaaA

Parameter	Native $WaaA_{AAE}$*	CMP-complexed $WaaA_{AAE}$*	K_3IrCl_6-derivatized $WaaA_{AAE}$*
Beamline	MAX-lab-I911-3	MAX-lab-I911-2	BESSY-14.2
Detector	MarCCD-225	MarCCD-165	MarCCD-165
Space group	$P2_1$	$P2_1$	$P2_1$
Unit cell (Å)	a = 132.50; b = 45.82; c = 144.07	a = 131.77; b = 44.67; c = 143.64	a = 132.62; b = 44.87; c = 144.31
Unit cell (°)	α = 90.00; β = 97.23; γ = 90.00	α = 90.00; β = 97.15; γ = 90.00	α = 90.00; β = 97.27; γ = 90.00
Resolution (Å)	2.00	2.42	3.08
R_{sym}(%)†	4.2 (55.3)	4.3 (13.7)	8.4 (25.3)
I/σI‡	14.8 (2.2)	21.3 (6.8)	20.5 (5.3)
Completeness (%)	98.9 (96.7)	97.2 (88.3)	93.9 (58.8)
Redundancy	3.6 (3.3)	3.6 (3.0)	6.8 (5.5)
Anomalous redundancy	—	—	3.6 (3.1)
Reflections (n)	115974	62775	30145
R_{work}/R_{free}§	20.71/25.39	20.24/25.09	—
Atoms (n)	11858	11836	—
Rms deviations			
Bond lengths (Å)	0.008	0.015	
Bond angles (°)	1.120	1.587	

*Values in the parentheses refer to the highest-resolution shell.

†$R_{sym} = (\sum|I_{hkl} - \langle I_{hkl}\rangle|)/(\sum I_{hkl})$, where the average intensity $\langle I_{hkl}\rangle$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection.

‡I/σI is the mean reflection intensity divided by the estimated error.

§$R_{work} = ||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is equivalent to R_{work} but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.

Table S2. Site-directed mutagenesis primers

Primer*	Sequence†
E31A FWD	5′-TTCACACAGCAAGTATAGGAGCATTCAATACTTTCCTACCTATCC-3′
E98A FWD	5′-CAAAAGCCTTAATAGTAGTTGCAAGGGAATTCTGGCCCTC-3′
K162A FWD	5′-GGGTATTTTCCTGCGGGAATTTAGCATTCATTTGTCAAAAAGG-3′
R212A FWD	5′-AAACTGATACTTGTCCCGGCGCATATAGAGAACGCAAAA-3′
E276A FWD	5′-GCATAACCTTCTGGCGCCCACCTGCTGG-3′
H272A/N273A FWD	5′-AACTTTTGTAAATATAGGGGGGGCTGCCCTTCTGGAGCCCACCTGCTG-3′
E100A FWD	5′-GCCTTAATAGTAGTTGAAAGGGCATTCTGGCCCTCTTTAATAATT-3′
W102A FWD	5′-TAGTAGTTGAAAGGGAATTCGCGCCCTCTTTAATAATTTTTAC-3′
S28A FWD	5′-GCTTTATGGGTTCACACAGCAGCTATAGGAGAATTCAATACTTTCC-3′
S54A REV	5′-GTATTCCCTTGCACGGGGGGCAAAGTAAGTGAGCAGAATTC-3′
S54A/R56A REV	5′-GTCTTTAGGTATTCCCTTGCAGCGGGGGCAAAGTAAGTGAGCAGAATTC-3′
G30A REV	5′-GGATAGGTAGGAAAGTATTGAATTCTGCTATACTTGCTGTGTGAACC-3′

FWD, forward primer; REV, reverse primer.

*Primer names indicate the introduced amino acid substitutions.

†Site-specific mutations are underlined.