Enzymatic Basis for *N*-Glycan Sialylation

STRUCTURE OF RAT α2,6-SIALYLTRANSFERASE (ST6GAL1) REVEALS CONSERVED AND UNIQUE FEATURES FOR GLYCAN SIALYLATION*§

Lu Meng[‡1], **Farhad Forouhar**[§1], **David Thieker**[‡1], **Zhongwei Gao**[‡], **Annapoorani Ramiah**[‡], **Heather Moniz**[‡], **Yong Xiang**[‡], **Jayaraman Seetharaman**[§], **Sahand Milaninia**[§], **Min Su**[§], **Robert Bridger**[‡], **Lucas Veillon**[‡], **Parastoo Azadi**[‡], **Gregory Kornhaber**[¶‖], **Lance Wells**[‡], **Gaetano T. Montelione**[¶‖], **Robert J. Woods**[‡**2], **Liang Tong**[§3], and **Kelley W. Moremen**[‡4]

From the ‡Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, the §Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, the ¶Center for Advanced Biotechnology and Medicine, Northeast Structural Genomics Consortium, and ‖Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, and the **School of Chemistry, National University of Ireland, Galway, University Road, Galway, Ireland

Background: Specificity and enzymology of glycan sialylation is poorly understood, despite its importance in biological recognition.
Results: ST6GAL1 structure was determined, and substrate binding was modeled to probe active site specificity.
Conclusion: The structure provides insights into the enzymatic basis of glycan sialylation.
Significance: Knowledge of the enzyme structure can lead to broader understanding of enzymatic sialylation and selective inhibitor design.

Glycan structures on glycoproteins and glycolipids play critical roles in biological recognition, targeting, and modulation of functions in animal systems. Many classes of glycan structures are capped with terminal sialic acid residues, which contribute to biological functions by either forming or masking glycan recognition sites on the cell surface or secreted glycoconjugates. Sialylated glycans are synthesized in mammals by a single conserved family of sialyltransferases that have diverse linkage and acceptor specificities. We examined the enzymatic basis for glycan sialylation in animal systems by determining the crystal structures of rat ST6GAL1, an enzyme that creates terminal α2,6-sialic acid linkages on complex-type *N*-glycans, at 2.4 Å resolution. Crystals were obtained from enzyme preparations generated in mammalian cells. The resulting structure revealed an overall protein fold broadly resembling the previously determined structure of pig ST3GAL1, including a CMP-sialic acid-binding site assembled from conserved sialylmotif sequence elements. Significant differences in structure and disulfide bonding patterns were found outside the sialylmotif sequences, including differences in residues predicted to interact with the glycan acceptor. Computational substrate docking and molecular dynamics simulations were performed to predict and evaluate the CMP-sialic acid donor and glycan acceptor interactions, and the results were compared with kinetic analysis of active site mutants. Comparisons of the structure with pig ST3GAL1 and a bacterial sialyltransferase revealed a similar positioning of donor, acceptor, and catalytic residues that provide a common structural framework for catalysis by the mammalian and bacterial sialyltransferases.

Cell surface and secreted proteins and lipids commonly contain covalently attached glycan structures that interact with the extracellular environment and influence cellular physiology, pathology, and recognition (1–3). Glycan structures provide both direct interactions with binding proteins to induce biological functions (4–8) and to indirectly modulate functions through stabilization or masking of the underlying glycoconjugates (2). The diverse terminal capping residues on *N*-glycans, *O*-glycans, and glycosphingolipids are the primary mediators of these biological interactions, and negatively charged sialic acid residues are among the most prevalent and well studied terminal residues that contribute to glycan functions (9, 10).

The roles of sialic acid residues in animal systems are varied. Regulated synthesis of sialylated glycan structures is common during embryonic development and numerous disease states (11, 12). Sialic acid-binding lectins are also highly regulated in a tissue-specific manner to facilitate glycan recognition and signaling (8, 9). Elevated levels of sialylation have been noted on tumor cell surfaces (12, 13), and gene disruptions in the synthe-

FIGURE 1. **Schematic representation of the activities of the mammalian sialyltransferase subfamily members.** Mammals have 20 sialyltransferase members (CAZy GT29 (23)) that all use CMP-Neu5Ac as a sugar donor, but they are distinguished based on the acceptor monosaccharide and the position of modification. Four broad sialyltransferase subfamilies transfer to either the O6 of an acceptor Gal residue (ST6GAL subfamily, two family members), the O3 of an acceptor Gal residue (ST3GAL subfamily, six family members), the O6 of an acceptor GalNAc residue (ST6GALNAC subfamily, six family members), or the O8 of an acceptor Neu5Ac residue (ST8SIA subfamily, six family members). Additional divisions within ST3GAL and ST6GALNAC subfamilies can be seen through more detailed sequence analysis (supplemental Fig. 1)

sis of sialic acid linkages caused varied phenotypes from immune system impairment to neurological defects depending on the enzyme isoform that was ablated (14–16). The extent of sialylation can also determine clearance rates of glycoproteins from circulation (17), and pathogenic bacteria, viruses, and parasites commonly employ cell surface sialic acids as ligands for cell adhesion (2). The most notable example is influenza virus tropism based on the preferential recognition of Neu5Acα2,6Gal linkages by human viruses, whereas avian and equine viruses prefer Neu5Acα2,3Gal linkages (18).

In animal systems, sialic acids, including the most common *N*-acetylneuraminic acid (Neu5Ac), can be found in four distinct linkage types to penultimate glycans (Neu5Acα2,3Gal-, Neu5Acα2,6Gal-, Neu5Acα2,6GalNAc- and Neu5Acα2,8Neu5Ac-) (19), and these structures are highly regulated in their abundance on the cell surface and secreted *N*-linked and *O*-linked glycoproteins and glycosphingolipids (gangliosides). Synthesis of sialylated structures in mammals is catalyzed by a family of 20 sialyltransferases restricted to the CAZy GT29 family (19–23). The sialyltransferases can be further divided into at least four subfamilies based on the linkage types catalyzed and the glycan acceptors that are modified (Fig. 1) (20, 22). Even within a single sialyltransferase subfamily, there are few examples of completely overlapping acceptor specificities. Sequence identity between the subfamilies is surprisingly low (<30% identity) (supplemental Fig. 1) and restricted to clustered regions, termed sialylmotif sequences (20, 22, 24). In contrast, bacterial sialyltransferases are found in distinct CAZy families (GT30, GT38, GT42, GT73, and GT80) where they transfer mono- or polysialic acid or related KDO residues (23). Minimal sequence identity has been observed between animal and bacterial sialyltransferases, but structural similarity was observed between

some bacterial GT42 sialyltransferases and the mammalian sialyltransferases within the sialylmotif region (25).

Insights into the domain structures and specificities of the mammalian sialyltransferases have come from extensive enzymatic studies on wild type and mutant enzymes (20, 26, 27), from the crystal structures of a pig ST3GAL1 involved in the synthesis of Neu5Acα2,3Gal linkages on *O*-linked glycans (25), and a structure of human ST6GAL1 that was published while this manuscript was being completed (29). The overall domain organization among the sialyltransferases is similar (28). A single NH$_2$-terminal transmembrane domain tethers the enzyme within the membrane of the Golgi complex with the globular catalytic domain facing the lumen. A varied "stem" domain acts as a linker between the membrane anchor and the more conserved catalytic domain. Because all mammalian sialyltransferases can employ CMP-Neu5Ac as a sugar donor, the conserved sialylmotif sequences were proposed to include the donor binding site, and this was subsequently confirmed by mutagenesis studies (24) and the structure of ST3GAL1 containing bound CMP (25). In contrast, the broad differences in acceptor specificity across the sialyltransferase family were reflected in divergent sequences outside the sialylmotifs (20, 22).

In an effort to probe the structural basis of the diverse substrate specificities among the sialyltransferases, we have determined the structure of rat ST6GAL1, an enzyme that synthesizes Neu5Acα2,6Gal linkages on the termini of complex *N*-glycans. The structure revealed several conserved features with ST3GAL1 (25), the bacterial GT42 sialyltransferase, CstII (30), and the recently published structure of human ST6GAL1 (29), including the sialylmotif region involved in binding the sugar-nucleotide donor. Modeling, molecular dynamics (MD)[5] simulations, and kinetic analysis of site-directed mutants provide evidence for a broadly similar positioning of sugar donor, glycan acceptor, and catalytic base for all four enzymes. In contrast, the protein fold outside the conserved sialylmotif sequences that includes the acceptor-binding site and more than half of the remaining protein structure was quite distinct between the bacterial and mammalian enzymes with distinct substrate specificities. A structural comparison between the bacterial and mammalian sialyltransferases suggests a divergence from a common ancestor and also provides a framework for comparison with other mammalian sialyltransferases.

EXPERIMENTAL PROCEDURES

Expression and Selenomethionine Labeling of ST6GAL1

Recombinant rat ST6GAL1 was expressed by transient transfection of HEK293 suspension cultures as a soluble secreted fusion protein essentially as described previously (31). Briefly, the fusion protein coding region was designed, codon optimized, and chemically synthesized by GeneArt AG (Regensburg, Germany) and was composed of a 25-amino acid signal sequence followed by a His$_8$ tag, an AviTag recognition site for *in vitro* biotinylation (32), the "superfolder" GFP coding region (33),

---

[5] The abbreviations used are: MD, molecular dynamics; SeMet, selenomethionine; r.m.s.d., root mean square deviation; EndoF1, endoglycosidase F1; TEV, tobacco etch virus; Ni-NTA, nickel-nitrilotriacetic acid; PDB, Protein Data Bank; PNGaseF, peptide *N*-glycosidase F.
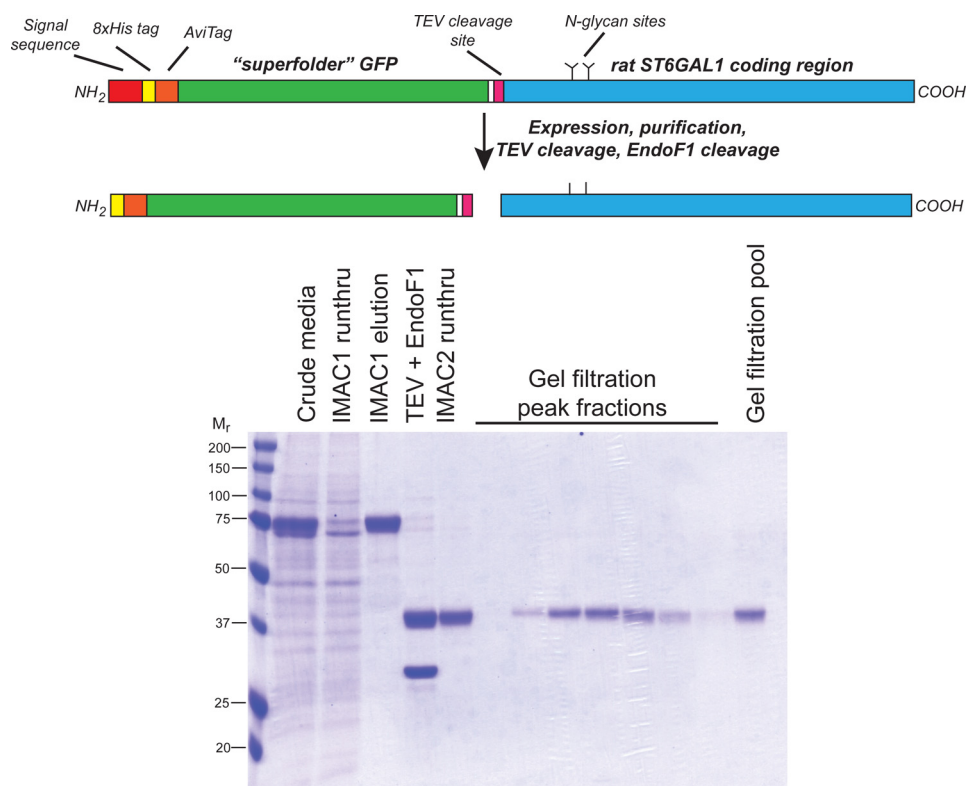
FIGURE 2. **Strategy for ST6GAL1 expression, purification, and tag/glycan cleavage.** A diagrammatic representation of the coding region for the recombinant ST6GAL1 fusion protein expression product is shown at the *top* of the figure. The expression construct encoded a fusion protein containing an NH₂-terminal signal sequence followed by an His₈ tag, AviTag, superfolder GFP, TEV protease cleavage site, and the catalytic domain of ST6GAL1 (see "Experimental Procedures" for details). Expression of the recombinant product in HEK293S cells resulted in the secretion of the fusion protein into the culture medium (*crude media, lower panel*), and subsequent Ni-NTA purification yielded a highly enriched enzyme preparation (*IMAC1 elution, lower panel*). Cleavage of the enzyme with TEV protease and EndoF1 resulted in removal of tag sequences and glycans down to a single GlcNAc residue attached to the peptide backbone (*TEV + EndoF1, lower panel*). Chromatography over Ni-NTA resulted in the elution of the ST6GAL1 catalytic domain and retention of the tag sequences, TEV protease, and EndoF1 on the column by virtue of their His tags (IMAC2 elution). The enzyme was further purified over Superdex-75, and peak fractions from the gel filtration column are shown in the *lower panel*. The final pool of purified ST6GAL1 (gel filtration pool) resulted in an ∼80% yield from the crude culture medium.

the 7-amino acid recognition sequence of the tobacco etch virus (TEV) protease (34), and residues 95-403 of rat ST6GAL1 (Uniprot P13721) (Fig. 2 and supplemental Fig. 2). The entire coding region was subcloned into the pGEn2 mammalian expression vector (31) that employs a CMV-based promoter and enhancer sequences to drive recombinant protein expression.

Wild type suspension culture HEK293 cells (FreeStyle 293-F cells, Invitrogen) or mutant HEK293S GnTI⁻ cells (35) (ATCC catalog no. CRL-3022) were maintained at $0.5–3.0 \times 10^6$ cells/ml in a humidified CO₂ platform shaker incubator at 37 °C. 293-F cells were maintained using serum-free Freestyle 293 expression medium (Invitrogen), and HEK293S cells were cultured in an equal volume of Freestyle 293 expression medium (Invitrogen) and Ex-cell 293 serum-free medium (Sigma). Transfections of the suspension cultures were accomplished by dilution to $∼1.5 \times 10^6$ cells/ml 24 h prior to transfection with fresh culture medium followed by resuspension at $∼2.5 \times 10^6$ cells/ml in fresh Freestyle 293 expression medium immediately prior to transfection. Transfections were initiated by direct addition of 4.5 µg/ml of the ST6GAL1-pGEn2 plasmid DNA and 10 µg/ml polyethyleneimine (linear 25-kDa polyethyleneimine, Polysciences, Inc., Warrington, PA) to the suspension culture (36). For transfections in 293-F cells, the cultures were diluted 1:1 with Freestyle 293 expression medium containing 4.4 mM valproic acid (2.2 mM final) 24 h after trans-

fection, and protein production was continued for a further 4–5 days at 37 °C. For HEK293S cell transfections, the cultures were diluted 1:1 with ESF serum-free medium (Expression Systems, Davis, CA) containing 4.4 mM valproic acid 24 h after transfection, and protein production was continued for a further 4–5 days at 37 °C. For metabolic labeling of HEK293S cells with selenomethionine (SeMet), cells were transfected as described above, and 12 h after transfection the media were exchanged for custom methionine-free Freestyle 293 expression medium (Invitrogen) for 6 h to deplete methionine pools, and the cultures were subsequently resuspended in methionine-free Freestyle 293 expression medium containing 60 mg/liter SeMet at density of $∼2.0 \times 10^6$ cells/ml. The protein production phase was 4–5 days at 37 °C before harvest of the conditioned medium.

Purification, Deglycosylation, and Fusion Protein Tag Removal of Recombinant ST6GAL1

The conditioned culture medium was harvested, clarified by sequential centrifugation at 1200 rpm for 10 min and 3500 rpm for 15 min, and passed through a 5-µm filter (Millipore). The medium was adjusted to contain 20 mM imidazole, 20 mM NaCl, and 3 mM sodium phosphate, pH 7.2, and loaded onto a column containing 25 ml of Ni-NTA Superflow (Qiagen, Valencia, CA) equilibrated with 20 mM HEPES, 300 mM NaCl, 20 mM imidaz-
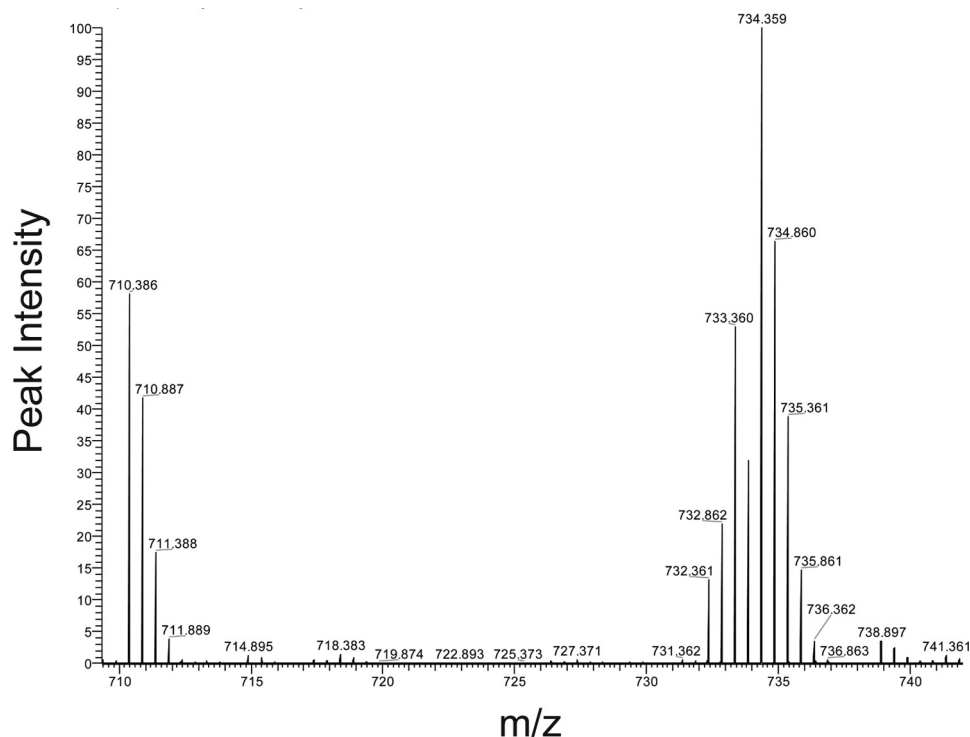
FIGURE 3. **Analysis of incorporation of selenium into a methionine-containing peptide from ST6Gal1.** SeMet-labeled ST6GAL1 was proteolytically digested and analyzed by LC-MS for the ratio of unlabeled to SeMet-labeled peptide fragments. Shown is the full MS spectra of the peptide LMNSQLVTTEKR that displays the observed mixture of a doubly charged unlabeled (major isotopic peak of 710.386 *m/z*) and labeled (major isotopic peak of 734.359 *m/z*) peptide that represents the difference between sulfur and selenium incorporation into the corresponding methionine. Note the unusual isotopic distribution of the labeled peptide due to the isotopic composition of selenium (a mixture of multiple isotopes where the abundance is $^{80}Se > ^{78}Se > ^{76}Se > ^{82}Se > ^{77}Se > ^{74}Se$). Incorporation of selenium based on these peak areas is 79.3%, and when all Met-containing peptides were utilized for calculations, the incorporation was determined to be $73 \pm 8\%$.

ole, pH 7.2 (buffer A). Following the loading of the sample, the column was washed with 150 ml of buffer A and eluted first with 60 ml of buffer A containing 50 mM imidazole, followed by 60 ml of buffer A containing 100 mM imidazole, and 250 ml of buffer A containing 300 mM imidazole. Peak fractions containing GFP fluorescence (300 mM imidazole elution) were pooled and concentrated to ~1 mg/ml using an ultrafiltration pressure cell membrane (Millipore, Billerica, MA) with a 10-kDa molecular mass cutoff. Purified recombinant TEV protease (37, 38) and endoglycosidase F1 (EndoF1) (39), both generated in *Escherichia coli*, were added at ratios of 1:40 and 1:20 relative to the GFP-ST6GAL1, respectively, and incubated overnight at room temperature to cleave the fusion protein and glycan structures. The fusion protein tag, the recombinant TEV protease, and EndoF1 all contain terminal His tags, and the resulting digestion products were passed through a 25-ml Ni-NTA Superflow column, and the eluted protein preparation was concentrated by ultrafiltration to ~3 ml. The sample was further purified on a Superdex 75 column (GE Healthcare) preconditioned with a buffer containing 20 mM HEPES, 200 mM NaCl, 60 mM imidazole, pH 7.2. Peak fractions of ST6GAL1 were collected and concentrated by ultrafiltration to 12 mg/ml for crystallization.

### Analysis of Selenomethionine Incorporation by Mass Spectrometry

Purified SeMet-labeled ST6GAL1 was reduced, alkylated, and trypsin-digested and then analyzed by LC-MS on a ThermoFisher LTQ-Orbitrap XL (Fig. 3). Total peak intensity was summed for each methionine-containing peptide in each charge state, and the ratio of incorporation was calculated by dividing the SeMet total peak intensity by the sum of both the labeled and unlabeled (sulfur-methionine) total peak intensity for each respective charge state.

Crystallization of ST6GAL1

Purified recombinant rat ST6GAL1 was screened for crystallization conditions using the high throughput crystallization facility at the Hauptman Woodward Medical Research Institute. Based on the high throughput screening results, the protein was crystallized by the microbatch method at 21 °C. The protein solution (2 μl of 12 mg/ml ST6GAL1 in 20 mM HEPES, 200 mM NaCl, and 60 mM imidazole, pH 7.2) was mixed with 2 μl of the precipitating solution consisting of 100 mM Tris, pH 9, 40% (w/v) PEG 1000, and 100 mM lithium bromide. Clusters of small crystals appeared after 1 week, which were subsequently optimized by the seeding method. The resulting larger and thicker crystals, 50 μm in size, were transferred from under oil to a 5-μl crystallization solution and then flash-frozen in liquid nitrogen for data collection at 100 K.

The crystals of rat ST6GAL1 belong to space group *C*2 with cell parameters of $a = 134.72$ Å, $b = 49.75$, $c = 86.06$ Å, and $\beta = 92.30°$. There are two molecules of ST6GAL1 in the crystallographic asymmetric unit (Fig. 4). For SeMet-labeled ST6GAL1 crystals, a single wavelength anomalous diffraction data set to resolution 2.4 Å was collected at the peak absorption wavelength of selenium at the X4A beamline of the National Syn-
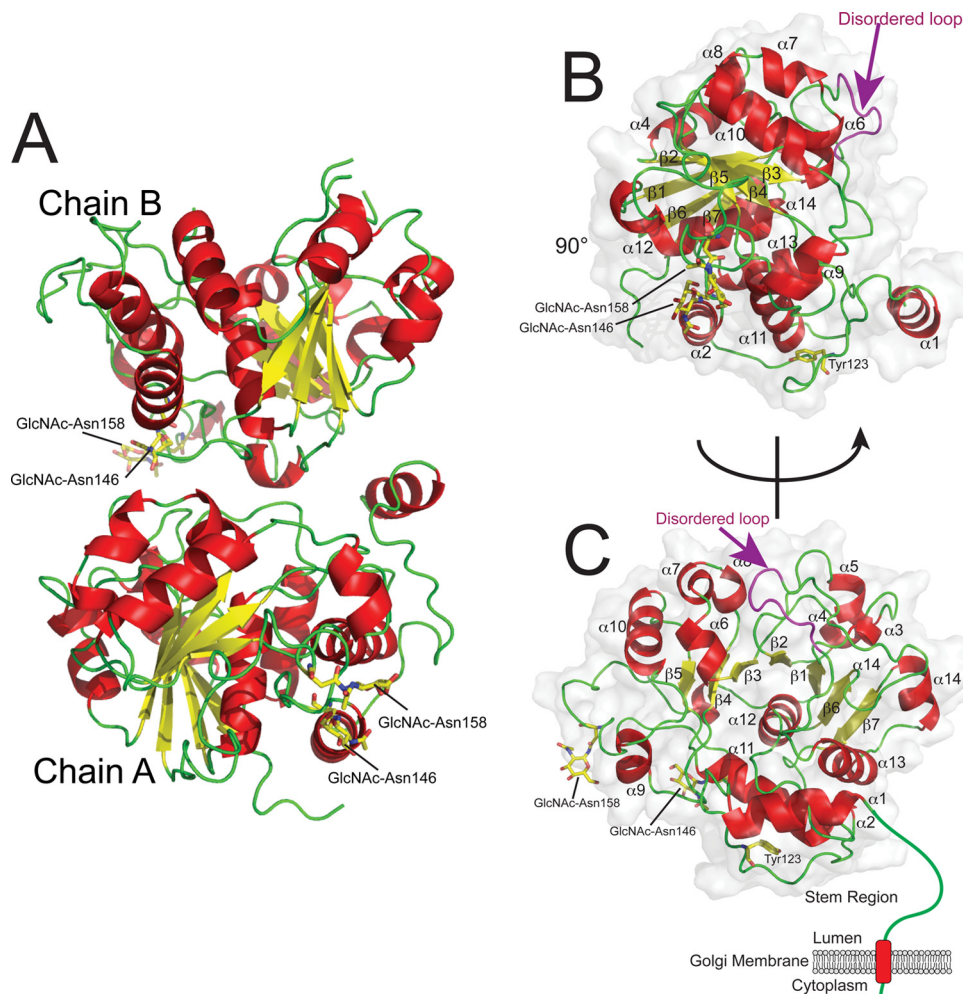
FIGURE 4. **Schematic representations of the ST6GAL1 structure.** The two monomers of ST6GAL1 in the asymmetric unit are stacked essentially head to tail with one unit (chain B) sitting adjacent to the active site region of the adjoining unit of the crystal lattice (chain A) (*A*). The electron density of chain A is more complete, with only a 6-residue segment of a disordered loop across the top of the β-sheet missing from the structure. In contrast, chain B is less complete, missing 12 residues from the disordered loop plus an additional 2 residues adjacent to the loop and 13 residues from the NH_2 terminus. The two *N*-glycan sites are indicated by an electron density of monosaccharides at the Asn-146 and Asn-158 consensus glycosylation sites as shown by the *yellow stick figures*. The overall structure of ST6GAL1 resembles the single Rossmann-like (GT-A variant 2) fold (*B* and *C*). The representation in *C* shows the seven β-strands and 14 α-helices of the catalytic domain from an end-on view of the β-sheet, and *B* is a 90° rotation for a side-on view. The NH_2 terminus of the structure is linked to a transmembrane segment through a nonconserved stem region (*green line*). The disordered segment of the polypeptide backbone at the top of the sugar-nucleotide-binding site was modeled as indicated by the *magenta line*. Rat ST6GAL1 has a sequence isoform that is generated through RNA editing of the genome-encoded Tyr-123 to a Cys residue that can influence enzyme localization and susceptibility to proteolysis (90). The structure of rat ST6GAL1 described here is the Tyr-123 isoform (shown in *yellow stick* representation) and the amino acid side chain is buried at the interface between a convoluted NH_2-terminal peptide segment and helix α11 stabilized through a hydrogen bond to the peptide bond carbonyl of Pro-126 and van der Waals interactions with α11.

chrotron Light Source. The diffraction images were processed with the HKL package (40), and 15 of 20 possible selenium sites were located with the program Shelx (41). SOLVE/RESOLVE (42) was used for phasing the reflections and automated model building, which correctly placed 20% of the residues with side chains in each protomer of the asymmetric unit. The majority of the model was manually built with the program XtalView (43) and refined by CNS (44). Noncrystallographic symmetry restraint was applied for most stages of the refinement of the structure. The data processing and refinement statistics are summarized in Table 1, and the protein is designated with a target ID of RnR367A in the Northeast Structural Genomics Consortium.

Modeling of the Ternary Complex

Preprocessing of the Enzyme—Several modifications to the protein structure were required before the computational stud-

TABLE 1
Summary of crystallographic information

Structure	Rat ST6GAL1
Maximum resolution	2.4 Å
No. of observations	153,712
R_{merge} (%)[a]	11.9% (48.6%)
No. of reflections used in refinement	40,789
Resolution range used in refinement	46.7 to 2.4
Completeness	99.1% (98.7%)
R factor	20.0% (25.4%)
Free *R* factor[b]	25.9% (32.9%)
r.m.s.d. in bond lengths	0.007 Å
r.m.s.d. in bond angles	1.3°

[a] The numbers in parentheses are for the highest resolution shell.
[b] 10% of the reflections were selected for free *R* calculation.

ies could be performed. The selenomethionine residues were converted to methionine. MolProbity (45) was used to predict possible Asn/Gln/His side chain flips, as well as preferred histidine protonation states, by optimizing hydrogen bonding net-

works within the protein. Six residues within a flexible loop were not resolved in the crystal structure; the terminal residues in this gap in the protein backbone were capped by attaching NME and ACE residues to the NH_2 and COOH termini, respectively. Although the protein has two glycosylation sites (Asn-146 and Asn-158), they are distal to the active site, by ~23 and 31 Å, respectively, and glycans at these positions were not included in the simulation.

Active Site Placement of Substrates—The donor substrate (CMP-Neu5Ac) was modeled into the active site based on the crystal structure of the sialyltransferase CstII (PDB 1RO7 (30)). The backbones of the ST6GAL1 and CstII proteins were aligned using Swiss-PDBViewer (46), and the coordinates for CMP-3-fluoro-*N*-acetylneuraminic acid (CMP-3FNeu5Ac) transferred to ST6GAL1. The ligand was then modified to create CMP-Neu5Ac by replacing the fluorine atom with hydrogen. Coordinates for the disaccharide acceptor Galβ1,4GlcNAcβ-O-Me (*N*-acetyllactosamine) were generated using the GLYCAM-Web tool (47).[6] The *N*-acetyllactosamine was docked into the binding site with AutoDock Vina (AD-Vina) (49). During docking, the glycosidic linkage and hydroxyl groups of the ligand were allowed to rotate, whereas all other components were kept fixed. The top-ranked model from the AD-Vina analysis appeared to be catalytically plausible and was employed in subsequent simulations.

Energy Minimization—The ST6GAL1·CMP-Neu5Ac·*N*-acetyllactosamine complex was neutralized with one Cl$^-$ ion, and the system solvated with TIP3P water (7,320 waters) in a truncated octahedral box, with side dimension of 74 Å. Energy minimizations and MD simulations were performed using AMBER12 (50), with the FF99SB (51), GAFF, and GLYCAM06 (47) parameters assigned to the protein, CMP-Neu5Ac, and *N*-acetyllactosamine, respectively. Prior to the calculation of partial atomic charges, the geometry of the CMP-Neu5Ac donor was optimized at the HF/3–21G* level of theory using GAUSSIAN09 (52). Partial atomic charges were computed using the RESP (53) module of AMBER12, based on fitting to electrostatic potentials computed at the HF/6 −31+ +G** level of theory. The system was minimized using the steepest descent method (1000 cycles) before switching to conjugate gradient (24,000 cycles). A two-step minimization protocol was imposed, in which initially all atoms other than the water, the hydroxyl groups in the acceptor, and protein side chains were restrained (10 kcal/mol Å^2). In the second stage, because the flexible loop near the active site was not resolved in the crystal structure, the protein Cα atoms, as well as the CMP portion of the CMP-Neu5Ac donor, were restrained, although all other atoms were allowed to relax. These stage two restraints were maintained during all subsequent MD simulations.

MD Simulation—MD simulations were performed with the pmemd.cuda version of AMBER12 (54). A cutoff for non-bonded interactions was set to 8 Å. Electrostatic interactions were treated with the Particle-Mesh Ewald algorithm (55). SHAKE was employed to constrain hydrogen-containing bonds, enabling an integration time step of 2 fs. The system was heated to 300 K under NVT conditions over 60 ps by employing the Berendsen thermostat with a coupling time constant of 1 ps and allowed to equilibrate for a total of 1 ns under NPT conditions. A post-equilibration data set was collected for 100 ns, also under NPT conditions.

Interaction Energy and MD Data Analysis—The interaction energy of the acceptor (*N*-acetyllactosamine, ligand) with the donor·enzyme complex (CMP-Neu5Ac·ST6GAL1) was computed with the single-trajectory Molecular Mechanics-Generalized Born Solvent Accessible Surface Area method (56, 57) using the MPBSA.py.MPI module (50). The simulation was divided into 5-ns bins, and average interaction energy contributions were computed from an ensemble of 100 snapshots evenly distributed within each bin. Before the analyses, all water molecules and ions were removed from each complex, and the contribution from desolvation approximated through the GB implicit solvation model (igb = 2) (58). This type of approach has been applied by us (59) and others (60, 61) for the prediction of carbohydrate-protein affinity. To minimize equilibration artifacts, only the last 80 ns of the NPT data set were further analyzed. The average interaction energy of the 80-ns trajectory was calculated using the same approach.

Structural metrics, such as r.m.s.d. plots, torsion angles, hydrogen bond properties, and coordinate averaging were generated using the cpptraj module of AmberTools13 (50). The φ- and ψ-glycosidic torsion angles were defined by the O5-C1-O4-C4 and C1-O4-C4-C5 atomic sequences, respectively. For the sake of completeness, a generous hydrogen bond definition was applied, cutoff (4.0 Å) and occupancy (>25%).

Mutagenesis and Kinetic Analysis

Active site residues were chosen for mutagenesis based on the structure of ST6GAL1, and the mammalian expression construct was used to perform site-directed mutagenesis using the QuikChange™ mutagenesis kit (Stratagene, La Jolla, CA). Mutant enzymes were generated by transient transfection of 293-F cells, and enzyme activity was determined following purification by Ni-NTA Superflow chromatography. Enzyme activity measurements were obtained through the use of a phosphatase-coupled assay utilizing colorimetric detection of inorganic phosphate with malachite green-based reagents (malachite green phosphate detection kit, R&D Systems, Minneapolis, MN). This assay method determines ST6GAL1 activity through the use of an enzyme-coupled 5′-nucleotidase (CD73) reaction to liberate inorganic phosphate from the CMP enzymatic product followed by a malachite green phosphate detection reagent (62). Assays were performed in a 50-μl reaction volume containing 100 mM MES, pH 6.5, 0.1 μg of recombinant ST6GAL1, CD73 (0.5 ng/μl), *N*-acetyllactosamine (2.4 mM for routine assays or varied from 0.2 to 8.0 mM for kinetic analysis), and CMP-Neu5Ac (0.25 mM for routine assays or varied from 0.025 to 1.0 mM for kinetic analysis). After 30 min of incubation at 37 °C, the reaction was stopped by addition of 10 μl of malachite green phosphate detection reagent A followed by incubation for 10 min at 25 °C and addition of 10 μl of reagent B. After 20 min at 25 °C, the absorbance was read using a Spectra MAX 190 spectrophotometer (Molecular Devices) at 620 nm and compared with equivalent analyses for a phosphate stan-

---

[6] From the Woods Group (2005–2013), *GLYCAM Web,* Complex Carbohydrate Research Center, University of Georgia, Athens, GA.

TABLE 2

Kinetic parameters for wild type rat ST6GAL1 and mutants using CMP-Neu5Ac as donor and *N*-acetyllactosamine as acceptor substrate

Enzyme form	GFP fluorescence[a]	Donor CMP-NeuAc			Acceptor Gal-β-1,4-GlcNAc		
		K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
		μM	min^{-1}	$mM^{-1}min^{-1}$	mM	min^{-1}	$mM^{-1}min^{-1}$
WT (293-F)[b]	1451	92 ± 10	35.6 ± 0.3	387	1.8 ± 0.3	40.3 ± 0.8	22.4
WT (HEK293S)[c]	450	93 ± 10	29.4 ± 0.4	316	1.4 ± 0.4	21 ± 1.8	15
S231A	2			No expression			No expression
S319A	8			No expression			No expression
Y119A	92			No expression			No expression
Y119F	1294	78 ± 8	23.4 ± 0.4	300	1.8 ± 0.5	18.8 ± 0.7	10.4
H367A	1463			No activity			No activity
Y366A	1314			No activity			No activity
N230A	1212			No activity			No activity
F208A	537	400 ± 60	11.3 ± 0.7	28	1.0 ± 0.1	25.8 ± 1.3	26.3
Q232A	1615	133 ± 33	6.2 ± 0.4	47	2.6 ± 0.6	8.1 ± 0.6	3.1
D271A	1374	365 ± 60	2.9 ± 0.1	8	6.5 ± 1.0	5.5 ± 0.2	0.8
Y272A	1191	220 ± 40	3.0 ± 0.1	14	3.5 ± 0.7	4.4 ± 0.2	1.3
Q354A	795	100 ± 30	25.0 ± 0.9	250	1.8 ± 0.4	38.6 ± 2.2	21.4
K355A	1232	95 ± 28	41.6 ± 0.3	438	1.1 ± 0.6	23.6 ± 3.9	21.5
C350A/C361A	1221			No activity			No activity

[a] Expression and secretion of the GFP-ST6GAL1 fusion proteins in transiently transfected HEK293 cells were determined by measuring the relative fluorescence of the recombinant proteins secreted into the media.

[b] The GFP-ST6GAL1 fusion protein was expressed in 293-F (wild type) cells and purified by Ni-NTA chromatography. The fusion protein tags and complex *N*-glycans were retained in this enzyme preparation similar to the analysis of all of the mutant enzymes.

[c] The GFP-ST6GAL1 fusion protein expressed in HEK293S cells under SeMet labeling conditions, was purified by Ni-NTA chromatography, cleaved to remove tag sequences and *N*-glycans, and further purified as described for crystallography of the enzyme catalytic domain.

dard curve. ST6GAL1 enzyme activity values (nanomoles/min) were determined at varied substrate concentrations, and kinetic data were fit using SigmaPlot (Enzyme Kinetics Module 1.2) to determine K_m and k_{cat} values.

RESULTS

Expression and Purification of ST6GAL1—Mammalian sialyltransferases are glycosylated disulfide bond-containing proteins in the lumen of the secretory pathway and generally present challenges for expression as functional, soluble enzymes in *E. coli*. We expressed the catalytic domain of ST6GAL1 in mammalian cells as a truncated fusion protein, where the endogenous NH$_2$-terminal transmembrane anchor was exchanged for a cleavable signal sequence and His$_8$-AviTag-GFP-TEV fusion protein tags (termed GFP-ST6GAL1 below), leading to secretion of the recombinant product into the culture media. Transient transfection of the GFP-ST6GAL1 construct into either 293-F or HEK293S suspension cultures resulted in high levels of enzyme secretion (~100 and ~75 mg/liter, respectively) that could be readily purified by Ni-NTA chromatography with high yield (~80% recovery based on GFP fluorescence, see Fig. 2).

The ST6GAL1 catalytic domain contains two consensus *N*-glycosylation sites, and expression in wild type 293-F cells yielded glycosylated products that were resistant to digestion with EndoF1 but sensitive to digestion by peptide *N*-glycosidase F (PNGaseF) only after prolonged digestion. The resulting digestion products were insoluble (data not shown), which is consistent with prior studies following PNGase digestion (63). Alternatively, expression of GFP-ST6GAL1 in the HEK293S (GnTI$^-$) cell line led to recombinant products containing high mannose *N*-glycans that were fully susceptible to EndoF1 digestion under native conditions. The resulting enzyme preparation had similar solubility, activity, and kinetic constants compared with the glycosylated recombinant enzyme (Table 2) suggesting that the single remaining GlcNAc residue was sufficient to maintain a soluble and functional enzyme.

A workflow was employed for enzyme isolation where large scale production (0.5–1 liter) of GFP-ST6GAL1 was followed by purification over a Ni-NTA column, concentration to ~1 mg/ml, and subsequent simultaneous digestion with recombinant His-tagged TEV protease and His-tagged EndoF1 to achieve complete cleavage of tag sequences and glycan structures. Subsequent chromatography over a Ni-NTA column resulted in direct elution of the cleaved ST6GAL1 product, while the cleavage enzymes and tags were retained on the column by virtue of their His tags. Further purification by gel filtration and concentration yielded enzyme preparations compatible with crystal screening (Fig. 2). Initial screening of crystals resulted in diffraction to ~2.5 Å. However, attempts to solve the structure by molecular replacement using the structure of pig ST3GAL1 as the probe were unsuccessful, because the sequence identity between ST6GAL1 and ST3GAL1 is only ~20% (supplemental Fig. 1).

Selenomethionine Labeling and Crystallization of ST6GAL1— As an alternative strategy to solve the structure, we performed SeMet labeling of ST6GAL1 in mammalian cells. Metabolic labeling with SeMet is a standard procedure in bacterial expression systems, but toxicity of SeMet in mammalian cells makes labeling more challenging. Several parameters and SeMet concentrations were tested for optimal labeling of HEK293S cultures analogous to the optimization of labeling in adherent HEK293 cell cultures (64). Concentrations of SeMet above 80 mg/ml were toxic and led to reduced cell viabilities and significantly decreased protein expression (data not shown). SeMet labeling during polyethyleneimine-mediated transient transfection also led to reduced cell viabilities. However, addition of a 6-h methionine starvation step 12 h post-transfection, followed by a production phase of 4–5 days in medium containing 60 mg/ml SeMet, led to GFP-ST6GAL1 expression and secretion at ~30 mg/liter, cell viabilities of ~60% at harvest, and SeMet incorporation levels of ~73 ± 8% based on multiple tryptic peptides and charge states following proteolysis and by LC-MS

analysis (Fig. 3). Using this protocol, SeMet-labeled ST6GAL1 was purified using workflows similar to those used for the unlabeled enzyme. Crystals were obtained that diffracted to 2.4 Å and the structure was solved by SAD phasing, revealing two molecules in the asymmetric unit. Repeated efforts to screen for co-complexes with substrates, substrate analogs, or inhibitors (including CMP, CMP-Neu5Ac, CMP-3FNeuAc, Galβ1,4GlcNAc-O-Me, Galβ1,4GlcNAc-β-o-nitrophenyl, and the broad sialyltransferase inhibitor JFD 00458 (65)) either by co-crystallization or soaking into pre-formed crystals were unsuccessful.

Structural Features of ST6GAL1—The fusion protein construct was designed to delete residues 1–94 of ST6GAL1 and initiate at residue 95 analogous to prior studies on the truncation and secretion of the enzyme (66). However, the electron density map of the ST6GAL1 structure starts at residue 104 in one monomer of the asymmetric unit (A-chain in Fig. 4) and residue 117 of the other monomer (B-chain) indicating that the upstream peptide segments are disordered. The overall fold of rat ST6GAL1 consists of a 7-stranded twisted β-sheet with 14 α-helical segments for the A-chain monomer (Fig. 4). The shorter NH$_2$ terminus of the B-chain monomer results in the loss of the first helix from the structure. A 6-residue segment across the top of the β-sheet (residues 355–360) is disordered in the A-chain. The equivalent disordered segment in monomer B is larger, comprising 12 residues (residues 353–366) as well as an adjacent disordered segment of residues 346–347. These regions resemble the disordered "lid" regions capping the sugar-nucleotide donor-binding site on other glycosyltransferases (67, 68), including ST3GAL1 (25), that typically become structured upon binding the sugar-nucleotide donor.

Purified ST6GAL1 eluted as a monomer by gel filtration in the final purification step, and the monomer interactions in the crystal lattice do not appear to be sufficient to reflect a biologically relevant oligomeric assembly. This contrasts with full-length ST6GAL1 that behaves as a homo- or hetero-tetrameric complex when isolated from cell cultures (69), and bimolecular fluorescence complementation indicates that both homodimerization and heterodimerization with B4GALT1 occur *in vivo* (70). Oligomers of ST6GAL1 form *in vivo* predominantly through the transmembrane or stem region segments (69), and Cys-24 in the transmembrane domain has been shown to be essential for disulfide-linked homodimer formation (71). These membrane-proximal sequences are missing from the present mammalian expression construct.

The topology of ST6GAL1 broadly resembles that of ST3GAL1 (25) and the bifunctional GT42 α2,3/8-sialyltransferase CstII (Fig. 5) (30), especially within the confines of the conserved sialylmotif sequence and the majority of the seven-stranded β-sheet. These structures conform to the glycosyltransferases GT-A (variant 2) fold (25), contrasting with the more common GT-A (variant 1) fold that also includes a single Rossmann-like domain and the GT-B structures with two Rossmann-like domains (72). Although the seven core β-strands occupy similar locations in all three proteins, the final β-strand in ST6GAL1 (β7) follows an extended helical linker allowing an antiparallel insertion at the end of the β-sheet (Fig. 4) in contrast to all-parallel arrangements in ST3GAL1 and CstII. Outside the sialylmotif region, the three proteins diverge signifi-

cantly in the helical and loop segments that compose >65% of the remaining protein structure, including the acceptor-binding site. These structural differences reflect the minimal primary sequence similarity between the proteins, including numerous insertions and deletions (Fig. 5). ST6GAL1 and ST3GAL1 contain NH$_2$-terminal extensions that contribute to the catalytic domain as well as the cytoplasmic tail, transmembrane domain, and stem regions that tether the enzyme to the Golgi membrane. In contrast, CstII contains no equivalent NH$_2$-terminal sequences, but instead it contains an extension at the COOH terminus that contributes to the catalytic domain and potentially a membrane anchoring region (73).

Three disulfide bonds are found in ST6GAL1 comprising all 6 Cys residues in the catalytic domain as predicted from peptide sequencing analysis (74). One disulfide (Cys-181 to Cys-332) is conserved in all mammalian sialyltransferases and is positioned at the base of the sialylmotif L and S and links the segments to stabilize and anchor the scaffold forming the CMP-Neu5Ac-binding site. Based on mutagenesis studies, this disulfide has been shown to be critical for catalytic activity as well as folding and transport from the endoplasmic reticulum (ER) to the Golgi complex (71, 75). The other disulfides are unique to ST6GAL1. One disulfide links the COOH-terminal Cys residue to an NH$_2$-terminal helical segment (Cys-139 to Cys-403) on the back side of the catalytic domain relative to the active site. Mutagenesis of Cys-139 or Cys-403 to eliminate the disulfide bond demonstrated that this linkage is not essential for activity, specificity, or localization of the enzyme (71, 74). The final disulfide anchors the disordered loop at the top of the sialylmotif region, because a weak electron density can be seen extending from the side chain of Cys-350 toward the predicted position of Cys-361 in the disordered loop. This disulfide bond is critical for catalytic activity both *in vitro* and *in vivo* (71, 74), but mutagenesis of the full-length enzyme does not impact transport to the Golgi complex in transfected cells (71). C350S and C361S mutants have reduced affinities for both sugar donor and acceptor (71). Two nonconserved disulfides are also found in ST3GAL1, but their positions are quite distinct from those found in ST6GAL1. CstII, by contrast, contains no disulfide bonds.

The two *N*-glycosylation sites in ST6GAL1 are clearly indicated by electron density of monosaccharides attached to the Asn-146 and Asn-158 side chains within the respective *N*-glycan consensus sites. Both monosaccharides face the solvent and are linked to loop regions at the NH$_2$ terminus of the protein on the same side of the globular domain (Fig. 4). Prior work indicated that neither glycan is required for transport of the full-length enzyme from the ER to Golgi, nor are they required for catalytic activity *in vivo* (66). However, the absence of *N*-glycans destabilizes the full-length protein in cell extracts, and no activity could be detected *in vitro*. Similarly, the glycans were required for effective folding and secretion of truncated, secreted enzyme forms, because double mutants defective in ST6GAL1 glycosylation were retained in the ER as inactive enzymes (66). Based on the positions and solvent accessibility of the glycosylation sites, the intact glycans do not appear to directly contribute to the structure of the mature protein, although they may contribute to stabilization or solubility of the
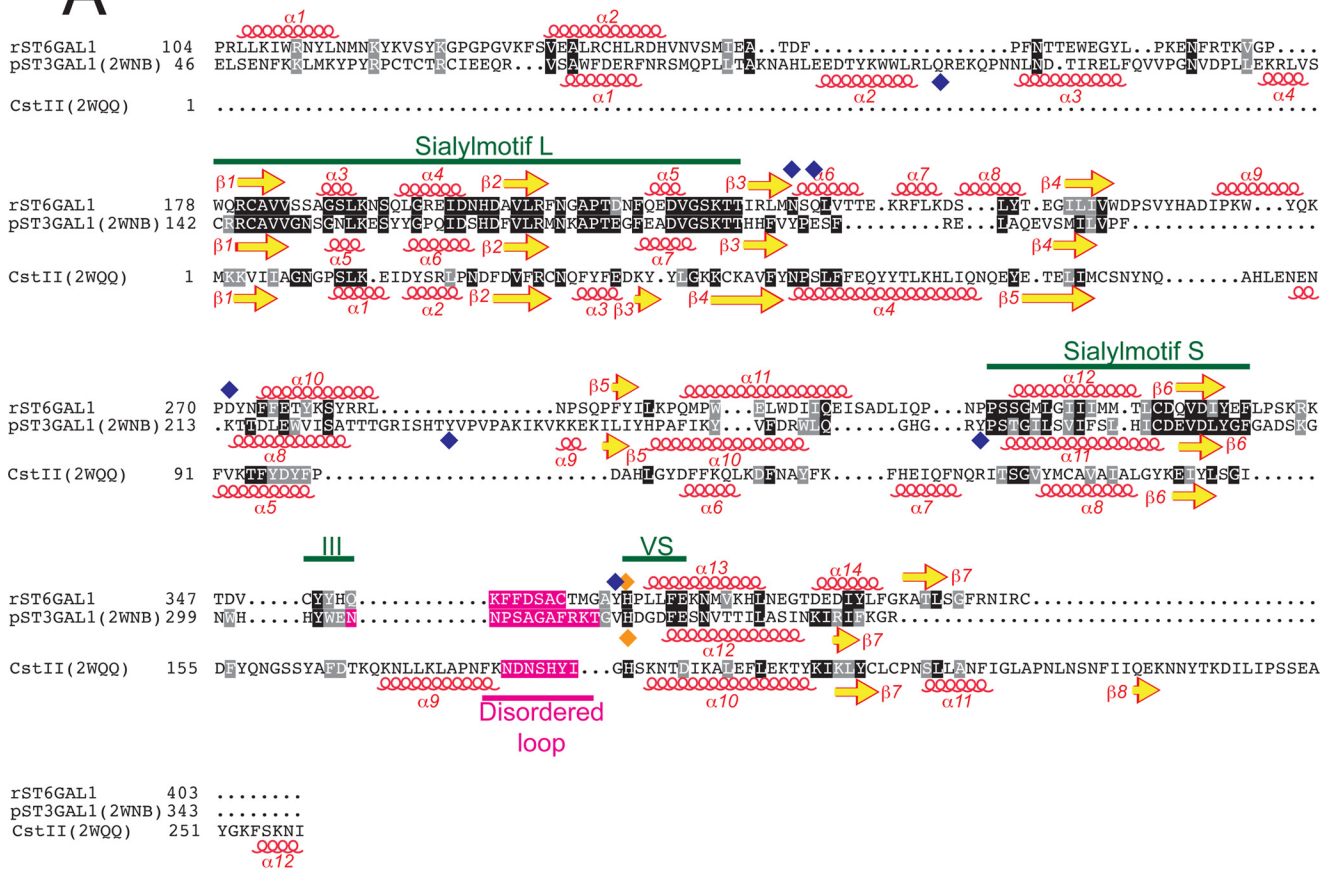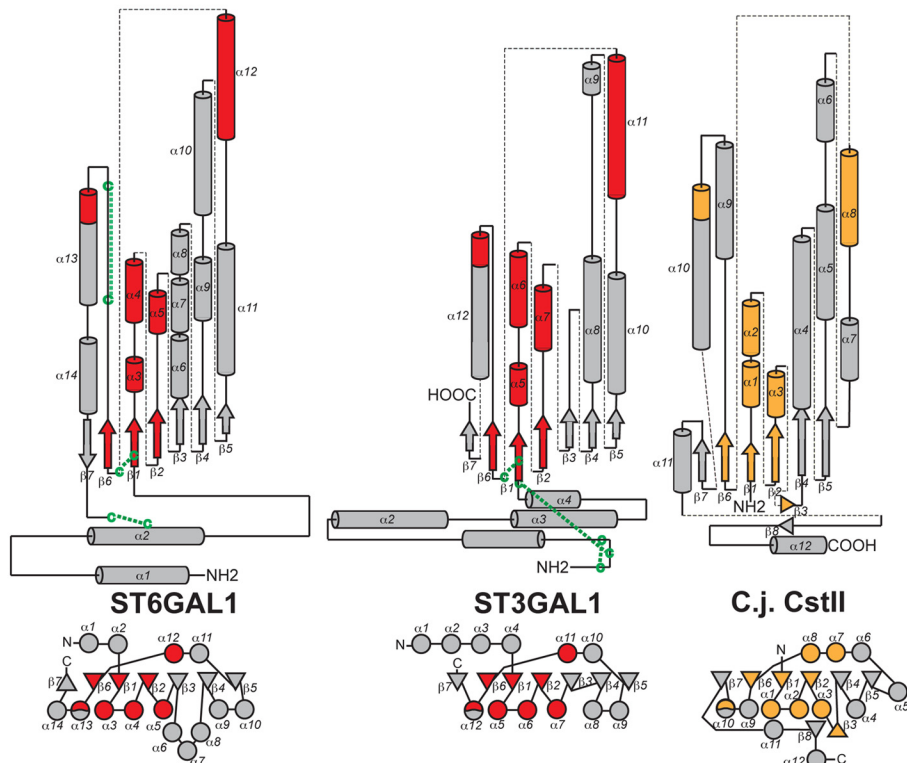
globular domain as indicated by rapid precipitation and inactivation of the PNGase-treated enzyme. A single GlcNAc residue at each site is sufficient to maintain enzyme function, because EndoF1 cleavage does not compromise enzyme solubility or catalytic characteristics (Table 2). For human ST6GAL1, a partial PNGase digestion resulted in the removal of a single *N*-glycan equivalent to Asn-158 in the rat enzyme (29), but the remaining biantennary complex *N*-glycan at the equivalent of Asn-144 in the rat enzyme was retained and contributed to contacts in the active site in the adjoining unit in the crystal lattice and helped to stabilize the loop that was disordered in the rat ST6GAL1 structure.

Sialylmotif Structures—The conserved sialylmotifs occupy one side of the globular enzyme structure with two of the β-strands (β1 and β2) coming from sialylmotif L along with adjoining linker regions containing three short helical segments. A third β-strand (β6) and a central helix (α12) are derived from sialylmotif S, and the Cys-181 to Cys-332 disulfide bond tethers the two sialylmotifs (Fig. 5). Other smaller sialylmotif regions contain the His-367 general base (sialylmotif VS) and a segment adjacent to the disordered loop capping the sugar donor-binding site (sialylmotif III, Fig. 6). Surprisingly, few of the conserved residues in the sialylmotifs face directly into the sugar donor site (Fig. 6). Of the residues that are in the equivalent CMP-binding site in ST3GAL1, only five conserved amino acids are within 4.5 Å of the nucleotide monophosphate, and only the peptide bond nitrogen of Gly-273 and the side chain of Asn-173 provide polar interactions. A conserved Phe-292 residue in ST3GAL1 interacts with a hydrophobic face of the ribose (25, 76), but the equivalent residue in ST6GAL1 is not positioned to interact with the sugar donor. The remaining residues in proximity to the CMP-binding site in ST6GAL1 are either not conserved or do not appear to interact directly with the substrate. Structural alignment of ST6GAL1 and ST3GAL1 with CstII demonstrates similar positions for the sialylmotif-like sequences in the bacterial enzyme (Figs. 5 and 6) with r.m.s.d. values of 1.71 Å (ST3GAL1 to CstII) and 1.75 Å (ST6GAL1 to CstII). The conserved residues defining the sialylmotifs largely compose the underlying scaffold of β-strands and α-helical connections that stabilize the less conserved residues that face into the CMP-Neu5Ac-binding site (Fig. 6).

Additional conserved sialylmotif features include the central 12-amino acid α-helix (α12 in sialylmotif S) that contributes to helix dipole interactions with the C1 carboxylate of the CMP-Neu5Ac donor in ST3GAL1 (25) and CstII (30), the His-367 catalytic base, and the residues that flank the disordered loop (25). These general features are fully conserved with the other mammalian and bacterial sialyltransferases (Fig. 6 and supplemental Fig. 1) forming the framework of sugar donor interactions and catalytic residues for this superfamily of enzymes.

Modeling, Molecular Dynamics Simulations of Substrate Interactions, and Comparison of Active Sites for Sialyltransferases—In contrast to the conserved features of the sugar donor site and catalytic residues, the adjoining residues of the acceptor-binding site are quite different in sequence, secondary structure, and position between ST6GAL1, ST3GAL1, and CstII. Our inability to crystallize co-complexes of ST6GAL1 with substrate analogs led us to employ molecular modeling and molecular dynamics simulations to probe the structural basis of enzyme-substrate interactions. The position of CMP-Neu5Ac was modeled using the CMP complex with ST3GAL1 (PDB 2WNB (25)) and the CMP·Neu5Ac complex with CstII (PDB 1RO7 (30)) based on the similarities of the sialylmotif domains and positions of sugar-nucleotide interactions in these structures. The *N*-acetyllactosamine acceptor was computationally docked in the acceptor site based on constraints of the Gal O6 nucleophile in proximity to the His-367 catalytic base and the C2 position of the CMP-Neu5Ac sugar donor. Initial docking was performed with the *N*-acetyllactosamine acceptor, and subsequent energy minimization and MD simulations were performed using restraints on the CMP portion of the sugar donor, because additional interactions between CMP and the missing residues of the disordered loop could not be modeled reliably. MD simulations were performed for 100 ns, and structural metrics regarding substrate dynamics and energetics of interactions were determined (Tables 3 and 4).

The *N*-acetyllactosamine ligand remained tightly associated with the acceptor site throughout the simulation. The Gal residue was relatively restricted in position, although the GlcNAc residue was more dynamic (r.m.s.d. values of 0.96 and 1.42 Å, respectively, see Fig. 7). The O6 is the only portion of the Gal residue that changed position during the simulation, and the two observed rotamers are reflected in the two conformations in the r.m.s.d. plot (Fig. 7*A*). This hydroxyl group flipped between an interaction with the His-367 catalytic base, and facing toward the C2 of the CMP-Neu5Ac sugar donor. The relative stability of the Gal residue compared with that of the GlcNAc becomes apparent in the conformational overlay of 20

FIGURE 5. Structure-based sequence alignment of rat ST3GAL1, pig ST3GAL1, and *C. jejuni* CstII and topology diagrams for ST6GAL1, ST3GAL1, and CstII. The structures of ST6GAL, ST3GAL1, and CstII were aligned using STRAP (48), and the resulting primary sequence alignment is displayed (*A*) with corresponding secondary structure elements indicated by *yellow arrows* (β-strands) or *red helices* (α-helices) displayed above (ST6GAL1) or below (ST3GAL1 and CstII) the respective sequences. Identical amino acids are highlighted in *white text* on a *black background*, and similar amino acids are indicated in *white text* on a *gray background*. Residues in ST6GAL1 that were found to interact with the Galβ1,4GlcNAc acceptor in the MD simulation studies are indicated with a *blue diamond* above the sequence. Residues in ST3GAL1 that interact with its Galβ1,3GalNAc acceptor based on the crystal structure (25) are indicated with a *blue diamond* below the sequence. The His base is indicated with the *orange diamond* in both sequences. Residues that are missing in the ST6GAL1 and ST3GAL1 structures (disordered loop, Fig. 4) are indicated with *white text* on a *magenta background*. Sialylmotifs are indicated by the *green bar above* the respective sequences. Topology diagrams for ST6GAL1, ST3GAL1, and CstII are shown in *B*. The GT-A variant 2-fold is characterized by seven β-strands linked by extended loop regions. The figure displays two types of topology diagrams. An extended diagram with *arrows* that indicate β-sheets and tubes that indicate α-helices is shown in the *upper panels* with each segment drawn to scale with the exception of the *black dotted lines* used to link the respective segments. Disulfide bonds are indicated with a *green C* and an adjacent *green dotted line*. The flattened topology diagrams (*B, bottom*) use *circles* to represent α-helical segments and *triangles* to represent β-sheets. The positions of the sialylmotif sequences are indicated by *red shading* for ST6GAL1 and ST3GAL1 or *orange shading* for the sialylmotif-like sequences in CstII. The three sialyltransferases have distinct topologies as indicated in both the flattened topology diagrams (*B, bottom*) and in the scaled topology diagrams (*B, top*), but the sialylmotifs form a core of the domain structure upon which distinctive adjoining sequences are arranged. *C.j.,* *C. jejuni.*
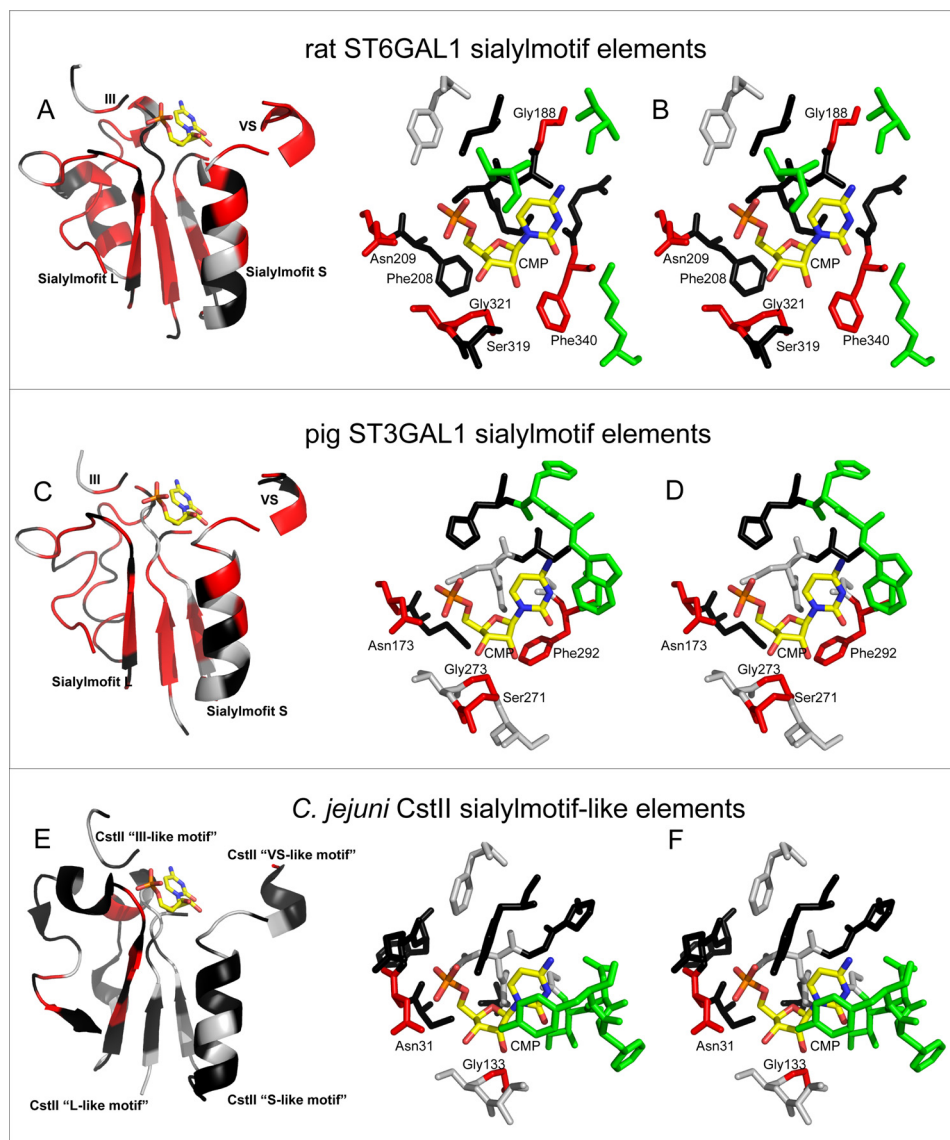
FIGURE 6. **Structural comparison of the sialylmotifs from ST6GAL1, ST3GAL1, and CstII.** The structures and sequences of ST6GAL1, ST3GAL1, and CstII were aligned as described in Fig. 5, and the structures of the sialylmotifs are displayed for the respective proteins in schematic form (*left panels*) with the CMP derived from the ST3GAL1 structure (25) in the donor site for reference (*yellow sticks*). The schematic representation of the ST6GAL1 and ST3GAL1 sialylmotifs are colored to represent identical residues (*red*), similar (*gray*), or dissimilar (*black*) residues between the two proteins. For the CstII sialylmotif structure, the schematic representation is colored for residues that are identical with both ST6GAL1 and ST3GAL1 (*red*), identical or similar to either ST6GAL1 or ST3GAL1 (*gray*), or dissimilar (*black*) to the other two sialyltransferases. The *right panels* for each respective enzyme show a stereo stick representation of the regions of the protein structures flanking the equivalent of CMP within the ST3GAL1 structure (25). Coloring of amino acid side chains is similar to those in the *left panels* with the exception of the residues that are shown in *green*, which are not part of the sialylmotif sequences. Key conserved or interacting residues are labeled. Few of the conserved (*red*) sialylmotif sequences are in direct contact with the CMP suggesting that the sialylmotif sequences create a scaffold framework for the more variable residues in the CMP-binding site.

models of the *N*-acetyllactosamine ligand (Fig. 7*C*). The range of conformations that were adopted within the simulation creates a single large cluster on a Ramachandran plot of the glycosidic torsion angles (Fig. 7*B*) with average φ and ψ values for the Galβ1,4GlcNAc linkage of $-86 \pm 10.6°$ and $-140 \pm 12.9°$, respectively. The GlyTorsion web tool (77, 78) was used to analyze the Galβ1,4GlcNAc linkage in 276 crystal structures within the PDB. A similar distribution was evident within the φ and ψ values of these experimentally derived structures, indicating that range of motion displayed for the bound acceptor in the MD is consistent with common torsion angles for the disaccharide.

The MD simulations indicated that the Gal residue of the disaccharide acceptor is held in place through a matrix of interactions, including hydrophobic stacking of Tyr-366 with the nonpolar face of the pyranose ring (accounting for 31% of binding energy) and hydrogen bonds with Asp-271, Asn-230, His-367, and Gln-232 (22, 14, 9, and 5% of total binding energy, respectively; see Table 3). The protein forms hydrogen bonds with the O2, O3, O4, and O6 of the Gal acceptor. These were generally stable interactions as indicated by hydrogen bond occupancies of >80% (Table 4) and significant contributions to the overall binding energy (>50%, Table 3). The hydrogen bonds to the side chains of Asn-230 and Gln-232 also provide

TABLE 3

Molecular Mechanics-Generalized Born Solvent-Accessible Surface Area energy[a] decomposition for the interaction between the receptor ST6GAL1/CMP-Neu5Ac and the substrate Galβ1,4GlcNAc

Receptor Residue[b]	van der Waals	Electrostatic	Polar Desolvation	Non-Polar Desolvation	Total	Percent of Total Receptor
Y366	-4.35	-2.01	1.66	-0.45	-5.16	31
D271	1.04	-17.82	13.15	-0.08	-3.71	22
N230	-0.73	-2.71	1.28	-0.11	-2.28	14
H367	-0.73	-1.48	0.79	-0.08	-1.50	9
CMP-Neu5Ac	-1.97	0.77	0.35	-0.28	-1.13	7
Q232	-1.75	-0.18	1.25	-0.20	-0.87	5
A365	-0.55	-1.78	1.62	-0.11	-0.83	5
Subtotal	-9.04	-25.21	20.10	-1.31	-15.48	93
Percent of Total	64	75	61	74	93	

Glycan Residue	van der Waals	Electrostatic	Polar Desolvation	Non-Polar Desolvation	Total	Percent of Total Glycan
Gal	-8.69	-21.97	19.65	-1.82	-12.83	91
GlcNAc	-5.32	-5.56	10.78	-1.05	-1.16	8
OMe	-0.21	-6.25	6.28	-0.01	-0.19	1

	van der Waals	Electrostatic	Polar Desolvation	Non-Polar Desolvation	Total	Percent of Total
Total (Protein)	-14.22	-33.78	33.04	-1.76	-16.72	54
Total (Ligand)	-14.22	-33.78	36.70	-2.89	-14.18	46
Total	-28.44	-67.56	69.74	-4.65	-30.91	

[a] Values are in kcal/mol. Because of the application of restraints on the Cα positions, entropic effects are not included.

[b] Only those residues that contribute more than 0.5 kcal/mol to ligand binding are listed.

stereochemical specificity for the axial position of the Gal O4, and the combined effects of the hydrophobic stacking and hydrogen bonds provide the appropriate positioning of the Gal O6 in proximity to the Nε atom of the His-367 base and the C2 position of the sugar donor. The strong energetic contributions of van der Waals and hydrophobic interactions of Tyr-366 with the Gal sugar ring (−5.2 kcal/mol) relative to total hydrogen bonding interactions (−8.4 kcal/mol) might have been over-

TABLE 4

Average intermolecular hydrogen bond lengths and occupancies

All intermolecular hydrogen bonds with occupancies greater than 25% and distances less than 4 Å.

Donor		Acceptor			
Residue	Atom	Residue	Atom	Occupancy	Distance
				%	Å
Asp-271	Oδ	Gal	HO2	100	2.7 (0.10)[a]
Asp-271	Oδ	Gal	HO3	100	2.7 (0.15)
Asn-230	Oδ	Gal	HO4	98	2.8 (0.15)
Gal	O5	GlcNAc	HO3	91	2.8 (0.15)
Gal	O3	N230	HNδ	81	3.5 (0.28)
Asp-271	Oδ	Gal	HO2	80	3.7 (0.19)
His-367	Nε	Gal	HO6	39	2.9 (0.17)
Asn-230	Nδ	Gal	HO4	39	3.8 (0.15)
Neu5Ac	O1A	N209	HNδ	81	3.0 (0.19)
Neu5Ac	O1A	S320	HOγ	72	3.2 (0.37)
Tyr-352	OH	Neu5Ac	HO9	43	3.3 (0.32)
Gal	O6	Neu5Ac	HO4	28	3.2 (0.37)

[a] Standard deviation is given in parentheses.
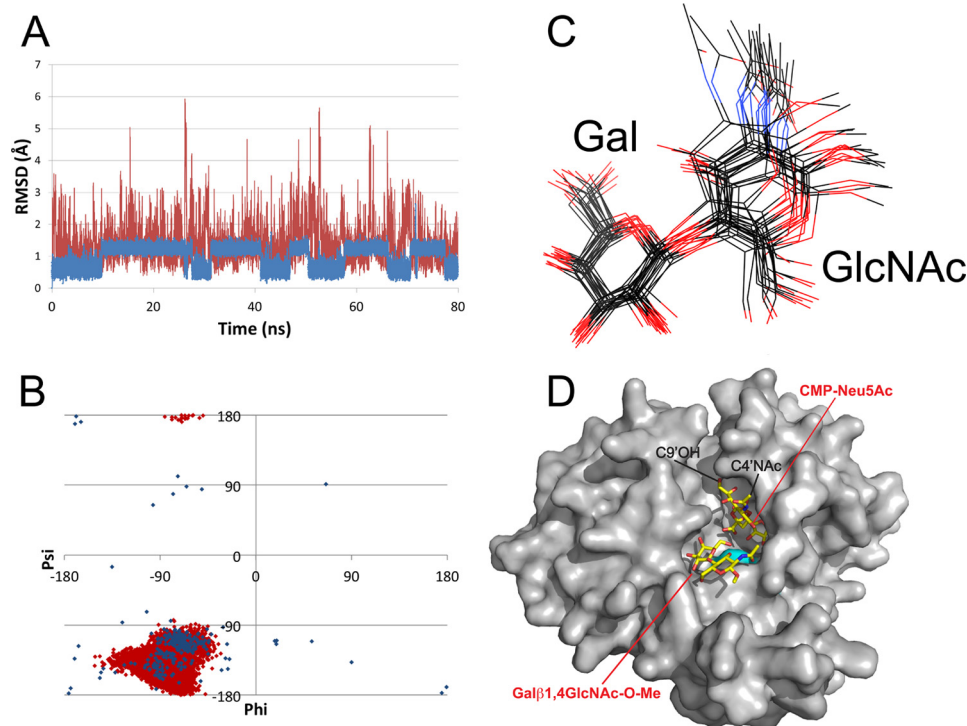
FIGURE 7. MD simulation of the acceptor disaccharide (Galβ1,4GlcNAc-O-Me) in the ST6GAL1 active site. The r.m.s.d. of the Gal (*blue*) and GlcNAc (*red*) residues across the MD trajectory are shown in *A*. The chart includes 40,000 frames spanning 80 ns of the MD simulation. The first frame served as a reference model during r.m.s.d. analysis. A Ramachandran plot of the glycosidic linkage of the *N*-acetyllactosamine acceptor in the binding site of ST6GAL1 is shown in *B*. The glycosidic linkage between Gal and GlcNAc from every frame of the 80-ns trajectory is represented on the chart by *red diamonds* based upon their Φ and Ψ angles, in degrees. *Blue diamonds* represent values of the Gal-GlcNAc torsion that have been obtained from 276 crystal structures as described under "Experimental Procedures." The 20 models of *N*-acetyllactosamine from the conformational overlay of the MD simulation are displayed as an evenly distributed set of conformations across the 80-ns trajectory and the respective monosaccharides are labeled (*C*). A surface representation of rat ST6GAL1 containing the modeled sugar nucleotide donor (CMP-Neu5Ac) and acceptor disaccharide (Galβ1,4GlcNAc-O-Me) is displayed (*D*) with a surface representation of the protein (*gray*) and stick representation for the substrates (*yellow*). A single Galβ1,4GlcNAc-O-Me structure from the MD simulation that most resembled the average of the acceptor conformations was chosen for the illustration. The position of the His-367 base is indicated by the *cyan* surface. The Neu5Ac C5′ *N*-acetyl and C9′OH are both accessible to the solvent consistent with the ability to enzymatically transfer NeuAc derivatives modified in these positions from sugar nucleotides to acceptor glycans (79, 80).

looked had the analysis not included the important contributions of desolvation energy in counter-balancing the direct electrostatic interactions of the hydrogen bond donors (Table 3). The binding energy for the Gal residue accounts for ~90% of the total interaction energy for the disaccharide. The GlcNAc residue exhibited significantly greater mobility during the simulation as indicated by the r.m.s.d. plot (Fig. 7*A*) and the diverse positions sampled during the MD simulation (Fig. 7, *B* and *C*) and displayed only a single hydrogen bond (between the GlcNAc *N*-acetyl and the peptide nitrogen of Ala-365). This interaction presumably makes a significant contribution to acceptor binding and likely accounts for the 80-fold increase in K_m for lactose as an acceptor compared with *N*-acetyllactosamine (79).

Binding interactions with the Neu5Ac residue of the sugar donor were less numerous, consistent with the primary binding affinity of sugar nucleotides to glycosyltransferases coming from the nucleotide component (80). Energetics of interaction with the nucleotide were not calculated, because this part of the molecule was restrained during the simulation. However, the modeling and energy minimization of the CMP-Neu5Ac in the sugar donor site revealed numerous interactions with the nucleotide. These include hydrogen bonds between the O2 and O3 of the ribose ring to the Oγ of the conserved helix capping Ser-319 and peptide amide nitrogen of Gly-321 within a Ser-Ser-Gly sequence. These residues are highly conserved among the sialyltransferases within an (S/T)(S/T)G consensus (Figs. 5 and 6). For ST3GAL1, the sequence is Ser-Thr-Gly, and only the Gly-273 amide interacts with the ribose. In both enzymes the first Ser of the consensus is the capping residue for the core α-helix. Additional interactions for ST6GAL1 include a hydrogen bond from O2 of the cytosine base to the NϵH$_2$ group of Lys-266. Hydrophobic stacking between Phe-208 and the C4–C5 bond of the ribose is also present. This is in contrast to the hydrophobic contact of Phe-292 with the C1 region of the ribose in ST3GAL1, which interacts from an alternative position in the binding pocket. Other polar interactions for ST6GAL1 include hydrogen bonds between Nδ2 of Asn-209 and the phosphate. These interactions result in a 3'-endo configuration of the ribose and positioning of the Neu5Ac residue with the C7–C9 glycerol side chain and C5-*N*-acetyl facing toward the solvent (Fig. 5). This orientation is consistent with a tolerance of the enzyme toward sugar donor modifications at O9 and the *N*-acetyl groups (81, 82).

Mutagenesis and Kinetic Analysis of Putative Active Site Residues—In an effort to confirm the interactions between the enzyme and the sugar donor and glycan acceptor, mutagenesis studies were performed on several residues, and kinetic constants were determined for the recombinant products. The mutants generally fell into four categories as follows: mutants that destabilized protein expression, mutants where the protein was stably expressed but was catalytically inactive, mutants that were compromised in catalytic turnover, and mutants that contained wild type enzyme activity (Table 2). Three amino acid mutations (S231A, S319A, and Y119A) led to protein destabilization. Ser-231 is the helix capping residue for the central α12 helix that contributes to helix dipole interactions with the sugar donor, and removal of this side chain might be expected to alter

enzyme stability. Hydroxyl groups of Ser-319 and Tyr-119 face toward the acceptor-binding site and provide extensive hydrogen bonds and hydrophobic interactions, respectively, to adjoining residues in the active site. The Tyr-119 side chain hydroxyl likely does not contribute directly to catalysis, because the isosteric Y119F mutant was well expressed and exhibited wild type activity.

Three single amino acid mutants (H367A, Y366A, and N230A) and one double mutant (C350A/C361A) were well expressed but had no detectable catalytic activity using *N*-acetyllactosamine as an acceptor. These results validate the assignment of the catalytic base (His-367) and key contributors to binding interactions with the acceptor Gal residue (Tyr-366 and Asn-230) in the MD simulations. The Cys double mutant eliminated the disulfide bond flanking the disordered loop, and loss of activity indicates that this disulfide plays a critical role in anchoring this loop during catalysis. Similarly, mutations in two other residues (Q232A and D271A) resulted in 7.2- and 28-fold decreases in k_{cat}/K_m values for the acceptor and 8.2- and 48.4-fold decreases in k_{cat}/K_m values for the donor, respectively. These results are consistent with their roles in hydrogen bonding to the acceptor Gal residue in the MD simulations. The effects of these latter mutations on the catalytic efficiency of both the acceptor and sugar donor were surprising but indicate that there may be cross-talk between donor and acceptor binding affinity in the active site. In contrast, the F208A mutation, the residue that provides hydrophobic interactions with the ribose ring, resulted in 13.8-fold reduction in k_{cat}/K_m values for the sugar donor but no effect on kinetic constants for the acceptor. Finally, the Y272A mutant resulted in a 17.3-fold reduction in k_{cat}/K_m values for the disaccharide acceptor and a 27.6-fold reduction in k_{cat}/K_m values for the sugar nucleotide donor. This residue resides in the acceptor site with the side chain OH facing toward the GlcNAc residue. MD simulation–studies indicate that this residue had minor contributions to hydrogen bonds, but the residue could indirectly contribute to catalysis by stabilization of other adjacent residues within the catalytic site. Mutations in two polar residues that are present in the disordered loop (Q354A and K355A) had no effect on catalytic activity suggesting that they do not directly interact with the sugar donor.

DISCUSSION

Sialylation of mammalian cell surface and secreted glycoproteins and glycolipids occurs through the action of a diverse family of sialyltransferases that act on unique subsets of acceptor glycans and create four distinct types of sialic acid linkages (Neu5Acα2,3Gal-, Neu5Acα2,6Gal-, Neu5Acα2,6GalNAc-, and Neu5Acα2,8Neu5Ac-). The enzymes use a common CMP-Neu5Ac donor and a conserved donor binding site and catalytic residues assembled from sialylmotif sequences to carry out sugar transfer. Surprisingly, comparison of the mammalian ST6GAL1 and ST3GAL1 structures with the bacterial CstII structure revealed minimal conservation in the residues directly interacting with the sugar nucleotide donor (Fig. 6). Instead, the conserved sialylmotif sequences compose the underlying scaffold of the Rossmann fold and adjoining loop regions that stabilize the residues of the donor-binding site.

These structural features of a conserved α/β scaffold and variable residues for substrate interactions are reminiscent of the NAD(P)(H) binding domains of the two-domain dehydrogenases (83). In this latter case, the underlying structural elements of the Rossmann fold are conserved, although the residues that directly interact with the nucleotide cofactor are variable, suggesting an evolution toward multiple independent strategies for binding the cofactor within the constraints of the conserved α/β scaffold (84, 85).

The structural and catalytic similarity between the bacterial and mammalian sialyltransferases suggests a common ancestor and a conserved evolutionary module composed of the sialylmotif sequences that facilitates sugar donor binding and activation of the glycan acceptor hydroxyl as a nucleophile in the S_N2 reaction mechanism. The modularity of the sialylmotif region relative to the remainder of the domain structure can be seen through the comparison of the topologies of the mammalian *versus* bacterial sialyltransferases (Fig. 5). The mammalian enzymes have extended sequences NH_2-terminal to the sialylmotif that contribute to the catalytic domain as well as membrane tethering, although the COOH terminus of the mammalian enzymes ends almost immediately after the final β-strand (25). In contrast, the NH_2-terminal end of the bacterial enzyme starts at the equivalent of sialylmotif L but contains an extended COOH-terminal sequence beyond the final β-strand that contributes to the catalytic domain and membrane tethering (30). The rapid evolution of the sequences adjoining the sialylmotifs provide for the diversity of acceptor structures and linkage specificities among mammalian and bacterial sialyltransferases. Functional GT29 sialyltransferase genes have been detected in most Metazoa, including Arthropoda and Echinodermata, but not Nematoda (21). However, related sequences have also been found in plants, algae, archaea, and a collection of viruses (23, 86). The structural similarity of the GT42 sialyltransferases from *Campylobacter jejuni* now expands the range of related sialyltransferase folds to bacteria and suggests a much broader common origin for these enzymes than was previously anticipated.

At the biochemical level, the sialyltransferases are probably the most heavily studied of all the mammalian glycosyltransferases. ST6GAL1 was the first of the mammalian sialyltransferases to be cloned (87) and biochemical (79, 88, 89), genetic (16), and cell biology (28, 66, 71, 90) studies have continued to use this enzyme as a paradigm for understanding Golgi glycosylation. Despite the experimental detail on the enzymology of the sialyltransferases, the enzymes have been a challenge to express in quantities and forms compatible with structural characterization. ST3GAL1 was successfully expressed in bacteria, and the crystal structure was determined (25), but most sialyltransferases require eukaryotic hosts to provide the required glycosylation, disulfide bonding, and chaperone systems present in the secretory pathway. We used mammalian cells for enzyme production and developed strategies for efficient transient transfection, high level expression, and workflows to remove tag sequences and heterogeneous glycan structures for structural studies. Methods were also adapted for metabolic labeling with SeMet to provide recombinant protein for crystallization and phasing of diffraction data by anomalous diffraction approaches. However, our inability to achieve co-crystallization with substrate analogs directed us toward computational studies to predict and evaluate substrate interactions.

Substrate docking and MD simulations of sugar donor and acceptor glycan interactions revealed a conserved location for the CMP-Neu5Ac interaction and a mode of binding predominantly through the CMP moiety similar to ST3GAL1 and CstII, although the details of interacting residues between the sialyltransferases varied (Fig. 6). Similar interactions with CMP were observed in the structure of human ST6GAL1 in complex with CMP (29). The acceptor sites in the sialyltransferases are more diverse, yet the general positions and angles of approach for the distinct disaccharide acceptors to the sugar donor and catalytic His residues are remarkably similar between ST6GAL1, ST3GAL1, and CstII (Fig. 8). Comparison of the acceptor complex for ST3GAL1 and ST6GAL1 reveals that the position of the penultimate HexNAc residue is similar, but the plane of the Gal monosaccharide is rotated 180°. The effect is to exchange the position of the O3 of the Gal in the ST3GAL1 acceptor site for the O6 of the Gal in the ST6GAL1 structure. Stabilization of the complex is also quite different as a consequence of the sugar flip. The disaccharide in the ST3GAL1 structure is stabilized by extensive hydrogen bonding between Tyr-269 and the axial hydroxyls of the C4′OHs of the Gal and GalNAc and hydrogen bonds to the O6 of the Gal, interactions that take advantage of the down-facing axial hydroxyls of the disaccharide acceptor. For ST6GAL1, the interactions include hydrophobic stacking of Tyr-366 with the nonpolar side of the Gal residue and a network of hydrogen bonds to the Gal O2, O3, O4, and O6, a mode of hydrophobic stacking and hydrogen bonding that is common among Gal-specific binding proteins (91–93). The results are a series of interactions that are exquisitely complementary to the respective Gal residue orientations in the active sites, to position either the O3 (ST3GAL1) or O6 (ST6GAL1) acceptor appropriately for nucleophilic attack on the C2 position of the sugar donor. Similarly, the Neu5Ac acceptor in CstII is positioned through an extensive hydrogen bond network to the C1 carboxylate, O4, *N*-acetyl, and O7, to position the O8 for nucleophilic attack similar to the other sialyltransferases. Thus, varied strategies for acceptor binding result in a common goal of providing a stereospecific transfer to a unique acceptor hydroxyl.

In contrast, the recent structure of human ST6GAL1 was found in complex with CMP and a nonsialylated biantennary complex *N*-glycan acceptor that extends from the equivalent of rat ST6GAL1 Asn-146 into the acceptor site of an adjoining ST6GAL1 molecule in the crystal lattice similar to an enzyme·acceptor complex (29). Most aspects of the protein structure are similar to the rat ST6GAL1 structure described here despite the fact that the glycosylation state, crystal contacts, and space group are entirely different. Surprisingly, the mode of binding of the Galβ1,4GlcNAc disaccharide in the acceptor site is significantly rotated by ~37° and shifted by 1.9 Å relative to the disaccharide position in the rat ST6GAL1 acceptor site identified by MD simulation (Fig. 9). The Gal residue in the human ST6GAL1 structure is similarly stacked on the aromatic ring of Tyr-369 (equivalent to Tyr-366 in the rat
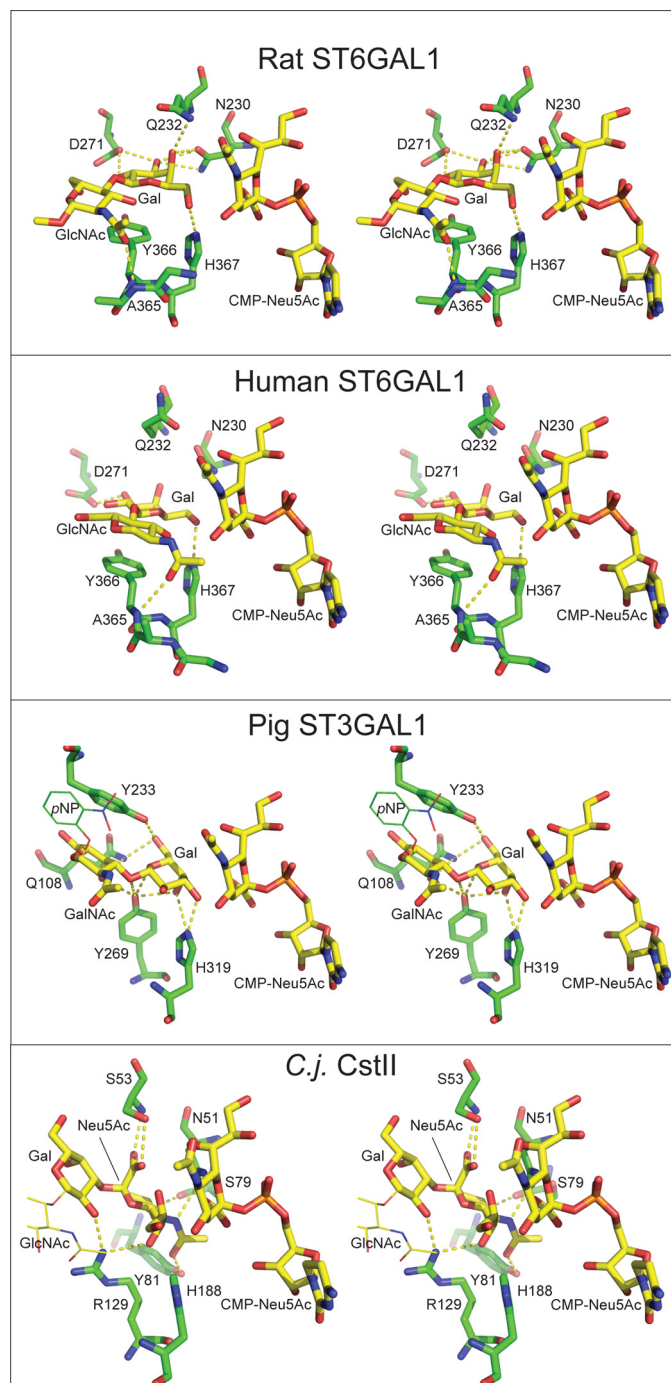
FIGURE 8. **Comparison of residues in the acceptor binding sites between rat ST6GAL1, human ST6GAL1, pig ST3GAL1, and CstII.** The structures of rat ST6GAL1, human ST6GAL1, pig ST3GAL1, and *C. jejuni* CstII (*C.j.* CstII) were aligned as described in Fig. 5, and residues interacting with the respective acceptor disaccharides were displayed in a stereo diagram. For rat ST6GAL1, a single Galβ1,4GlcNAc-O-Me structure from the MD simulation that most resembled the average of the acceptor conformations was chosen for the illustration. For human ST6GAL1 (PDB 4JS1 (29)), only the terminal Galβ1,4GlcNAc structure of the biantennary *N*-glycan was displayed. For ST3GAL1, the Galβ1,3GalNAc-α-*p*-nitrophenol structure was displayed, but the *p*-nitrophenol was displayed as *thin lines* compared with the *stick* form of the remaining disaccharide. For CstII, the Neu5Acα2,3Galβ1,4GlcNAc acceptor structure was displayed, but the reducing terminal GlcNAc was displayed as *thin lines* compared with the *stick* form of the remaining disaccharide. The equivalent of the CMP-Neu5Ac modeled in the active site of rat ST6GAL1 is shown in *stick* form for all models. The corresponding acceptor structures and CMP-Neu5Ac are shown in *yellow stick* form, and amino acid residues are shown in *green stick* form. Hydrogen bonds are indicated by the *yellow dotted lines*.
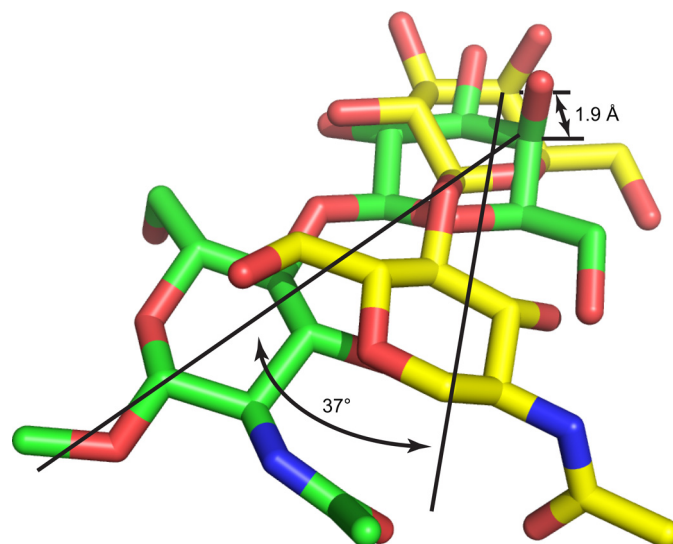
FIGURE 9. **Comparison of the bound geometry of the acceptor disaccharide in the active site of rat and human ST6GAL1.** The relative position of the disaccharide acceptors in the active sites of rat ST6GAL1 (*green stick* representation) and human ST6GAL1 (*yellow stick* representation; PDB 4JS1 (29)) were compared by structural alignment of the two proteins and display of the corresponding Galβ1,4GlcNAc acceptor glycans. Despite an r.m.s.d. of 0.63 Å over 1766 protein atoms for the two proteins, the corresponding acceptor structures were poorly aligned in the acceptor site. The C4 position of the Gal residue was offset by 1.9 Å. When the C4 atoms of the Gal residues were superimposed, the corresponding angle from the C4 Gal to the C1 position of the GlcNAc for the two disaccharides was determined to be ~37°. These data indicate a significantly different angle of approach for the two disaccharides into the respective acceptor sites.

enzyme) with hydrogen bonds to Asp-274 and His-370 (equivalent to Asp-271 and His-367 in the rat enzyme) (Fig. 8). In both enzymes the Gal O6 is within hydrogen bonding distance of the catalytic His residue and is positioned appropriately for nucleophilic attack of the CMP-Neu5Ac donor. However, several hydrogen bonds are missing between the terminal Gal residue in the acceptor binding site for the human ST6GAL1 structure as a consequence of the alternative position and angle of rotation. Hydrogen bonds to Asn-230 and Gln-232 to the O3 and O4 that define the axial Gal O4 specificity are missing, and a significant portion of the binding energy is lost. Steric clashes also appear to exist between our modeled position of the O4 of the Neu5Ac and the GlcNAc *N*-acetyl as a result of the altered angle of entry. These contacts are compensated for by additional interactions of the extended biantennary complex structure that bridges the crystal lattice. Although the empirical data of the crystal structure is compelling, the altered positions and angles of entry for the disaccharide acceptor compared with the other sialyltransferase structures suggest that the geometry of the human ST6GAL1 acceptor complex may be influenced by the structural constraints of crystal packing interactions. However, the data also indicate that there may be an unexpectedly malleable positioning of the Gal residue to act as a nucleophile during the sugar transfer mechanism.

Kinetic analysis of active site mutants was consistent with the model for substrate binding derived from the MD simulations. Mutations in the catalytic His-367 or Tyr-366 that stacks underneath the nonpolar face of the Gal residue eliminated all catalytic function. Similarly, mutation of Asn-230, a residue that displays hydrogen bonds with the Gal O3 and O4 in the
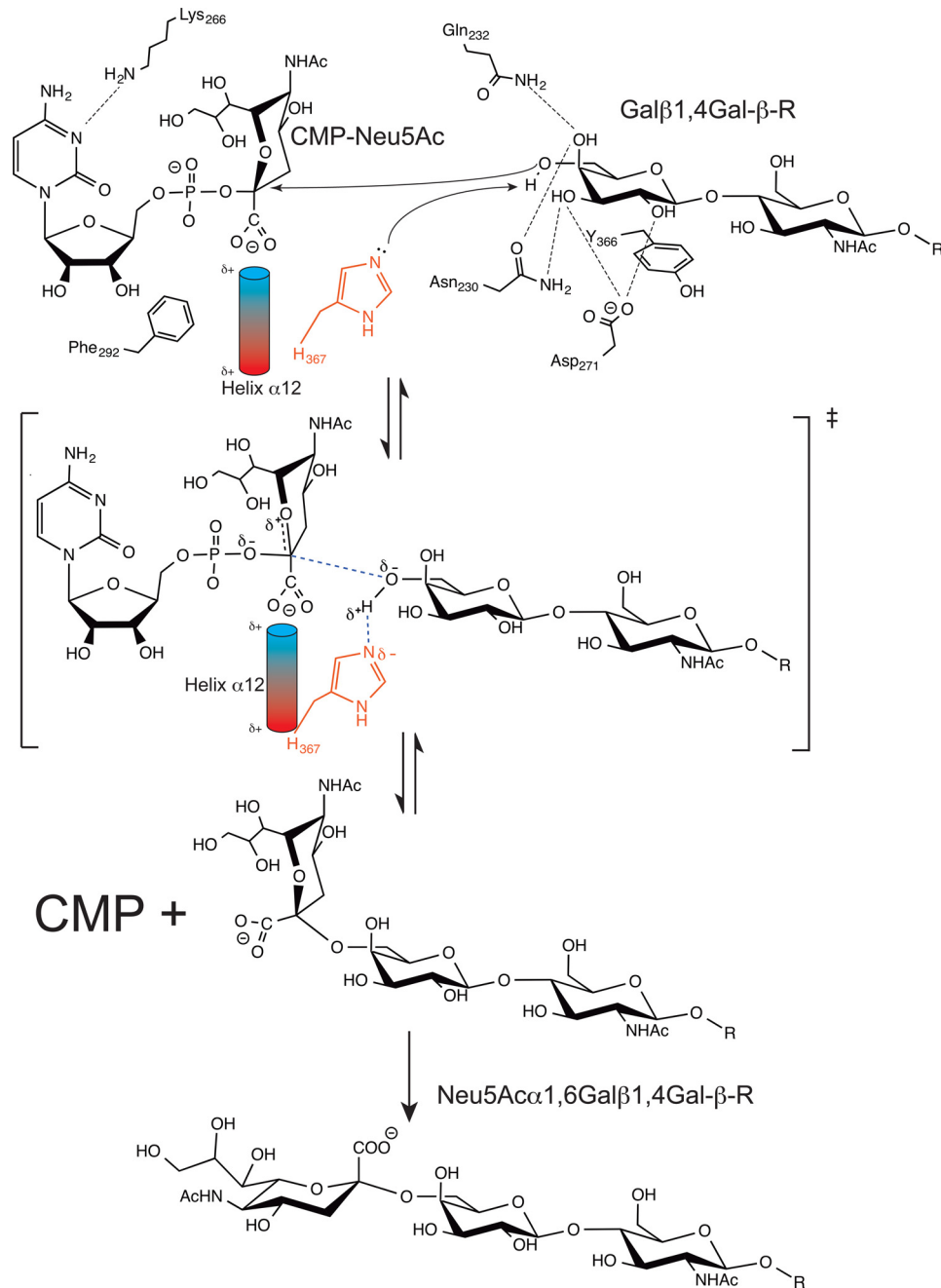
FIGURE 10. **Reaction scheme for rat ST6GAL1.** Interactions between the Galβ1,4GlcNAc acceptor, CMP-Neu5Ac, and the corresponding protein residues are indicated in the *upper panel*, including a helix dipole interaction between the C1 carboxyl of the CMP-Neu5Ac and the positive end of helix α12. His-367 acts as base to deprotonate the O6 of the terminal Gal residue for nucleophile attack of the C2 of Neu5Ac. A ring-flattened oxocarbenium ion transition state is formed, and CMP acts as leaving group in the $S_N 2$ mechanism.

MD simulation, resulted in complete loss of activity, whereas this residue does not appear to interact with either the acceptor or donor in the human ST6GAL1 structure. Other residues that contribute to binding interactions in the MD simulations (Gln-232 and Asp-271) also resulted in a significant loss in catalytic efficiency following mutagenesis. Mutations of other residues that faced into the acceptor site (Ser-319, Tyr-110, and Tyr-272) also impacted enzyme activity, but the loss of these side chains may have had indirect effects on the structure of the binding site. Disruption of the disulfide that anchors the disordered loop eliminated all catalytic activity, but two polar residues within the loop (Gln-354 and Lys-355) had no effect on enzyme activity. These latter data indicate that the stability of the loop is critical for catalysis, but these 2 residues within the loop presumably do not interact directly with the sugar donor.

The modeled complex of *N*-acetyllactosamine in the rat ST6GAL1 structure has 5 key residues that interact with the Gal residue based on MD simulations and kinetic analysis (His-367, Tyr-366, Asn-230, Gln-232, and Asp-271). These interactions result in the Gal residue contributing ~90% of the donor affinity (Table 3). With the exception of the catalytic His-367, all four of the other residues are conserved only in ST6GAL1 and

ST6GAL2, the only two mammalian enzymes that synthesize Neu5Acα2,6Gal linkages (supplemental Fig. 1). Similarly, the residues involved in acceptor binding for ST3GAL1 (Tyr-269, Tyr-233, and Gln-108) are conserved only for ST3GAL1 and ST3GAL2 despite the fact that there are four other α2,3Gal-specific sialyltransferases that fall into a separate sialyltransferase clade (supplemental Fig. 1). In addition, each of the sialyltransferase subfamilies contains distinct disulfide bonding patterns and different insertion and deletion sequences within the nonconserved regions (supplemental Fig. 1) (19–22). These data suggest that prediction of active site interactions among the sialyltransferases based on the structures of ST6GAL1 and ST3GAL1 will be quite difficult, especially for those with different linkage or acceptor specificities. In most cases, much of the specificity for the acceptor occurs through recognition of the unique stereochemistry of the nonreducing terminal monosaccharide, although more distal interactions are able to fine-tune the details of acceptor recognition. For example, ST6GAL1 prefers the terminal *N*-acetyllactosamine linkage on the α1,3Man branch of a biantennary complex type *N*-glycan 11-fold over the equivalent *N*-acetyllactosamine terminus on the α1,6Man branch (31). The complex glycan bound in the acceptor site, the human ST6GAL1 structure (29), presents a potential view of interactions that could contribute to branch specificity. However, further experimentation is required to confirm that these contacts contribute to *N*-glycan branch specificity in solution.

There are four common features of the sialyltransferase catalytic sites characterized to date (25, 29, 30, 76). First, the geometry, but not the direct interactions, within the sugar nucleotide-binding site is conserved and provides appropriate positioning of the Neu5Ac C2 adjacent to the catalytic His base. Second, each of the enzymes contains an extended central α-helix that is capped by a Ser hydroxyl side chain and adjoining Gly residue that facilitates interaction with the ribose ring of the donor and also has a helix dipole oriented toward the Neu5Ac C1 carboxyl. Third, the neighboring catalytic His is positioned to act as a general base in the catalytic mechanism (Fig. 10). Fourth, the enzymes contain a variable acceptor domain with stereochemical complementarity to the appropriate acceptor glycan structures. The concerted base deprotonation of the acceptor hydroxyl and the nucleophilic attack on the Neu5Ac C2 carbon results in the formation of a ring-flattened oxocarbenium-like transition state (94, 95) and an inverting S_N2-like direct displacement of the phosphate of CMP as the leaving group (Fig. 10). The sialyltransferases do not employ D*X*D metal ion-binding sites common among some glycosyltransferases for charge stabilization of the departing phosphate leaving group (94). However, the helix dipole could stabilize the developing phosphate charge (25, 76). A mechanism has also been proposed for a charge relay complex from the protonated catalytic His to the Neu5Ac C1 carboxylate and subsequently to the helix dipole to delocalize the protonated state of the His (29). These multiple proposed roles for the helix dipole are consistent with the conserved placement of this central helix in the sialyltransferase active site.

Thus, we present here the structure of rat ST6GAL1, the modeling of substrate complexes in the enzyme active site with energetic assessment through MD simulation, and we provide kinetic data on catalytic mutants to support the model for substrate recognition and catalysis. The studies also provide a framework for further work on other sialyltransferases that will provide a fundamental understanding of acceptor recognition and catalysis during sialylation of diverse glycan structures.

REFERENCES

1. Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448–462
2. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130
3. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (eds) (2008) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
4. Dennis, J. W., Lau, K. S., Demetriou, M., and Nabi, I. R. (2009) Adaptive regulation at the cell surface by *N*-glycosylation. *Traffic* **10,** 1569–1578
5. Yan, H., Ohno, N., and Tsuji, N. M. (2013) The role of C-type lectin receptors in immune homeostasis. *Int. Immunopharmacol.* **16,** 353–357
6. Clark, G. F. (2013) The role of carbohydrate recognition during human sperm-egg binding. *Hum. Reprod.* **28,** 566–577
7. Sato, S., Ouellet, M., St-Pierre, C., and Tremblay, M. J. (2012) Glycans, galectins, and HIV-1 infection. *Ann. N.Y. Acad. Sci.* **1253,** 133–148
8. Varki, A., and Gagneux, P. (2012) Multifarious roles of sialic acids in immunity. *Ann. N.Y. Acad. Sci.* **1253,** 16–36
9. Crocker, P. R., Paulson, J. C., and Varki, A. (2007) Siglecs and their roles in the immune system. *Nat. Rev. Immunol.* **7,** 255–266
10. Lehmann, F., Tiralongo, E., and Tiralongo, J. (2006) Sialic acid-specific lectins: occurrence, specificity and function. *Cell. Mol. Life Sci.* **63,** 1331–1354
11. Nairn, A. V., Aoki, K., dela Rosa, M., Porterfield, M., Lim, J. M., Kulik, M., Pierce, J. M., Wells, L., Dalton, S., Tiemeyer, M., and Moremen, K. W. (2012) Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. *J. Biol. Chem.* **287,** 37835–37856
12. Schultz, M. J., Swindall, A. F., and Bellis, S. L. (2012) Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metastasis Rev.* **31,** 501–518
13. Swindall, A. F., Londoño-Joshi, A. I., Schultz, M. J., Fineberg, N., Buchsbaum, D. J., and Bellis, S. L. (2013) ST6Gal-I protein expression is upregulated in human epithelial tumors and correlates with stem cell markers in normal tissues and colon cancer cell lines. *Cancer Res.* **73,** 2368–2378
14. Yang, W. H., Nussbaum, C., Grewal, P. K., Marth, J. D., and Sperandio, M. (2012) Coordinated roles of ST3Gal-VI and ST3Gal-IV sialyltransferases in the synthesis of selectin ligands. *Blood* **120,** 1015–1026
15. Angata, K., Long, J. M., Bukalo, O., Lee, W., Dityatev, A., Wynshaw-Boris, A., Schachner, M., Fukuda, M., and Marth, J. D. (2004) Sialyltransferase ST8Sia-II assembles a subset of polysialic acid that directs hippocampal axonal targeting and promotes fear behavior. *J. Biol. Chem.* **279,** 32603–32613
16. Hennet, T., Chui, D., Paulson, J. C., and Marth, J. D. (1998) Immune regulation by the ST6Gal sialyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 4504–4509
17. Fukuda, M. N., Sasaki, H., Lopez, L., and Fukuda, M. (1989) Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood* **73,** 84–89
18. Imai, M., and Kawaoka, Y. (2012) The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr. Opin. Virol.* **2,** 160–167
19. Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M. A., Samyn-

Petit, B., Julien, S., and Delannoy, P. (2001) The human sialyltransferase family. *Biochimie* **83,** 727–737

20. Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. (2005) The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology* **15,** 805–817

21. Takashima, S. (2008) Characterization of mouse sialyltransferase genes: their evolution and diversity. *Biosci. Biotechnol. Biochem.* **72,** 1155–1167

22. Patel, R. Y., and Balaji, P. V. (2006) Identification of linkage-specific sequence motifs in sialyltransferases. *Glycobiology* **16,** 108–116

23. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328,** 307–317

24. Datta, A. K., and Paulson, J. C. (1995) The sialyltransferase "sialylmotif" participates in binding the donor substrate CMP-NeuAc. *J. Biol. Chem.* **270,** 1497–1500

25. Rao, F. V., Rich, J. R., Rakić, B., Buddai, S., Schwartz, M. F., Johnson, K., Bowe, C., Wakarchuk, W. W., Defrees, S., Withers, S. G., and Strynadka, N. C. (2009) Structural insight into mammalian sialyltransferases. *Nat. Struct. Mol. Biol.* **16,** 1186–1188

26. Li, Y., and Chen, X. (2012) Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl. Microbiol. Biotechnol* **94,** 887–905

27. Dall'Olio, F. (2000) The sialyl-α2,6-lactosaminyl-structure: biosynthesis and functional role. *Glycoconj. J.* **17,** 669–676

28. Paulson, J. C., and Colley, K. J. (1989) Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J. Biol. Chem.* **264,** 17615–17618

29. Kuhn, B., Benz, J., Greif, M., Engel, A. M., Sobek, H., and Rudolph, M. G. (2013) The structure of human α-2,6-sialyltransferase reveals the binding mode of complex glycans. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1826–1838

30. Chiu, C. P., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2004) Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat. Struct. Mol. Biol.* **11,** 163–170

31. Barb, A. W., Meng, L., Gao, Z., Johnson, R. W., Moremen, K. W., and Prestegard, J. H. (2012) NMR characterization of immunoglobulin G Fc glycan motion on enzymatic sialylation. *Biochemistry* **51,** 4618–4626

32. Beckett, D., Kovaleva, E., and Schatz, P. J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8,** 921–929

33. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88

34. Tözsér, J., Tropea, J. E., Cherry, S., Bagossi, P., Copeland, T. D., Wlodawer, A., and Waugh, D. S. (2005) Comparison of the substrate specificity of two potyvirus proteases. *FEBS J.* **272,** 514–523

35. Reeves, P. J., Callewaert, N., Contreras, R., and Khorana, H. G. (2002) Structure and function in rhodopsin: high level expression of rhodopsin with restricted and homogeneous *N*-glycosylation by a tetracycline-inducible *N*-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 13419–13424

36. Backliwal, G., Hildinger, M., Chenuet, S., Wulhfard, S., De Jesus, M., and Wurm, F. M. (2008) Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. *Nucleic Acids Res.* **36,** e96

37. van den Berg, S., Löfdahl, P. A., Härd, T., and Berglund, H. (2006) Improved solubility of TEV protease by directed evolution. *J. Biotechnol.* **121,** 291–298

38. Wu, X., Wu, D., Lu, Z., Chen, W., Hu, X., and Ding, Y. (2009) A novel method for high level production of TEV protease by superfolder GFP tag. *J. Biomed. Biotechnol.* **2009,** 591923

39. Kwan, E. M., Boraston, A. B., McLean, B. W., Kilburn, D. G., and Warren, R. A. (2005) *N*-Glycosidase-carbohydrate-binding module fusion proteins as immobilized enzymes for protein deglycosylation. *Protein Eng. Des. Sel.* **18,** 497–501

40. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326

41. Sheldrick, G. M. (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. D Biol. Crystallogr.* **46,** 467–473

42. Terwilliger, T. C. (2003) Solve and Resolve: Automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37

43. McRee, D. E. (1999) XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156–165

44. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol Crystallogr.* **54,** 905–921

45. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 12–21

46. Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30,** S162–S173

47. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput Chem.* **29,** 622–655

48. Gille, C., and Frömmel, C. (2001) STRAP: editor for structural alignments of proteins. *Bioinformatics* **17,** 377–378

49. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput Chem.* **31,** 455–461

50. Case, D. A., Darden, T. A., Cheatham, I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) *AMBER 10,* University of California, San Francisco, CA

51. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65,** 712–725

52. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J., Inc., Wallingford, CT. (2009) *Gaussian 09,* Version D.01, Gaussian, Inc., Wallingford, CT

53. Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A well behaved electrostatic potential based method using charge restraints for deriving atomic charges–the Resp model. *J. Phys. Chem.* **97,** 10269–10280

54. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8,** 1542–1555

55. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092

56. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., 3rd (2000) Calculating structures and free

energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33,** 889–897

57. Srinivasan, J., Miller, J., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **16,** 671–682

58. Onufriev, A., Bashford, D., and Case, D. A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55,** 383–394

59. Kadirvelraj, R., Grant, O. C., Goldstein, I. J., Winter, H. C., Tateno, H., Fadda, E., and Woods, R. J. (2011) Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Acα2-6Galβ1-4GlcNAc human-type influenza receptor. *Glycobiology* **21,** 973–984

60. Kadirvelraj, R., Gonzalez-Outeiriño, J., Foley, B. L., Beckham, M. L., Jennings, H. J., Foote, S., Ford, M. G., and Woods, R. J. (2006) Understanding the bacterial polysaccharide antigenicity of *Streptococcus agalactiae versus Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 8149–8154

61. Bryce, R. A., Hillier, I. H., and Naismith, J. H. (2001) Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys. J.* **81,** 1373–1388

62. Wu, Z. L., Ethen, C. M., Prather, B., Machacek, M., and Jiang, W. (2011) Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology* **21,** 727–733

63. Fast, D. G., Jamieson, J. C., and McCaffrey, G. (1993) The role of the carbohydrate chains of Gal β-1,4-GlcNAc α2,6-sialyltransferase for enzyme activity. *Biochim. Biophys. Acta* **1202,** 325–330

64. Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P., and Nikolov, D. B. (2006) Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci.* **15,** 2008–2013

65. Rillahan, C. D., Brown, S. J., Register, A. C., Rosen, H., and Paulson, J. C. (2011) High throughput screening for inhibitors of sialyl- and fucosyltransferases. *Angew. Chem. Int. Ed. Engl.* **50,** 12534–12537

66. Chen, C., and Colley, K. J. (2000) Minimal structural and glycosylation requirements for ST6Gal I activity and trafficking. *Glycobiology* **10,** 531–583

67. Ramakrishnan, B., Ramasamy, V., and Qasba, P. K. (2006) Structural snapshots of β-1,4-galactosyltransferase-I along the kinetic pathway. *J. Mol. Biol.* **357,** 1619–1633

68. Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. (2005) Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem. Sci.* **30,** 53–62

69. Hassinen, A., Pujol, F. M., Kokkonen, N., Pieters, C., Kihlström, M., Korhonen, K., and Kellokumpu, S. (2011) Functional organization of Golgi N- and O-glycosylation pathways involves pH-dependent complex formation that is impaired in cancer cells. *J. Biol. Chem.* **286,** 38329–38340

70. Hassinen, A., Rivinoja, A., Kauppila, A., and Kellokumpu, S. (2010) Golgi N-glycosyltransferases form both homo- and heterodimeric enzyme complexes in live cells. *J. Biol. Chem.* **285,** 17771–17777

71. Qian, R., Chen, C., and Colley, K. J. (2001) Location and mechanism of α2,6-sialyltransferase dimer formation. Role of cysteine residues in enzyme dimerization, localization, activity, and processing. *J. Biol. Chem.* **276,** 28641–28649

72. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77,** 521–555

73. Lee, H. J., Lairson, L. L., Rich, J. R., Lameignere, E., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2011) Structural and kinetic analysis of substrate binding to the sialyltransferase Cst-II from *Campylobacter jejuni*. *J. Biol. Chem.* **286,** 35922–35932

74. Hirano, Y., Suzuki, T., Matsumoto, T., Ishihara, Y., Takaki, Y., Kono, M., Dohmae, N., and Tsuji, S. (2012) Disulphide linkage in mouse ST6Gal-I: determination of linkage positions and mutant analysis. *J. Biochem.* **151,** 197–203

75. Datta, A. K., Chammas, R., and Paulson, J. C. (2001) Conserved cysteines

76. Rakic, B., Rao, F. V., Freimann, K., Wakarchuk, W., Strynadka, N. C., and Withers, S. G. (2013) Structure-based mutagenic analysis of mechanism and substrate specificity in mammalian glycosyltransferases: porcine ST3Gal-I. *Glycobiology* **23,** 536–545

77. Lütteke, T., Bohne-Lang, A., Loss, A., Goetz, T., Frank, M., and von der Lieth, C. W. (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology* **16,** 71R–81R

78. Lütteke, T., and von der Lieth, C. W. (2009) Data mining the PDB for glyco-related data. *Methods Mol. Biol.* **534,** 293–310

79. Weinstein, J., de Souza-e-Silva, U., and Paulson, J. C. (1982) Sialylation of glycoprotein oligosaccharides N-linked to asparagine. Enzymatic characterization of a Galβ1 to 3(4)GlcNAc α2 to 3 sialyltransferase and a Galβ1 to 4GlcNAc α2 to 6 sialyltransferase from rat liver. *J. Biol. Chem.* **257,** 13845–13853

80. Blume, A., Angulo, J., Biet, T., Peters, H., Benie, A. J., Palcic, M., and Peters, T. (2006) Fragment-based screening of the donor substrate specificity of human blood group B galactosyltransferase using saturation transfer difference NMR. *J. Biol. Chem.* **281,** 32728–32740

81. Bleil, J. D., and Wassarman, P. M. (1988) Galactose at the nonreducing terminus of O-linked oligosaccharides of mouse egg zona pellucida glycoprotein ZP3 is essential for the glycoprotein's sperm receptor activity. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 6778–6782

82. Rauvolfova, J., Venot, A., and Boons, G. J. (2008) Chemo-enzymatic synthesis of C-9 acetylated sialosides. *Carbohydr. Res.* **343,** 1605–1611

83. Lesk, A. M. (1995) NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5,** 775–783

84. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossman, M. G. (1973) D-Glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance. *Proc. Natl. Acad. Sci. U.S.A.* **70,** 3052–3054

85. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250,** 194–199

86. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., eds) pp 3–12, The Royal Society of Chemistry, Cambridge, UK

87. Weinstein, J., Lee, E. U., McEntee, K., Lai, P. H., and Paulson, J. C. (1987) Primary structure of β-galactoside α2,6-sialyltransferase. Conversion of membrane-bound enzyme to soluble forms by cleavage of the NH₂-terminal signal anchor. *J. Biol. Chem.* **262,** 17735–17743

88. Paulson, J. C., Weinstein, J., and de Souza-e-Silva, U. (1984) Biosynthesis of a disialylated sequence in N-linked oligosaccharides: identification of an N-acetylglucosaminide (α2–6)-sialyltransferase in Golgi apparatus from rat liver. *Eur. J. Biochem.* **140,** 523–530

89. Weinstein, J., de Souza-e-Silva, U., and Paulson, J. C. (1982) Purification of a Galβ1 to 4GlcNAc α2 to 6 sialyltransferase and a Galβ1 to 3(4)GlcNAc α2 to 3 sialyltransferase to homogeneity from rat liver. *J. Biol. Chem.* **257,** 13835–13844

90. Ma, J., Qian, R., Rausa, F. M., 3rd, and Colley, K. J. (1997) Two naturally occurring α2,6-sialyltransferase forms with a single amino acid change in the catalytic domain differ in their catalytic activity and proteolytic processing. *J. Biol. Chem.* **272,** 672–679

91. Sujatha, M. S., Sasidhar, Y. U., and Balaji, P. V. (2004) Energetics of galactose- and glucose-aromatic amino acid interactions: implications for binding in galactose-specific proteins. *Protein Sci.* **13,** 2502–2514

92. Rini, J. M. (1995) Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 551–577

93. Drickamer, K. (1997) Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5,** 465–468

94. Breton, C., Fournel-Gigleux, S., and Palcic, M. M. (2012) Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **22,** 540–549

95. Unligil, U. M., and Rini, J. M. (2000) Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10,** 510–517