Structural Analysis of the α-2,3-Sialyltransferase Cst-I from *Campylobacter jejuni* in Apo and Substrate-Analogue Bound Forms[†,‡]

Cecilia P. C. Chiu,[#] Luke L. Lairson,[§] Michel Gilbert,[||] Warren W. Wakarchuk,[||] Stephen G. Withers,[§] and Natalie C. J. Strynadka*[,#]

Department of Biochemistry and Molecular Biology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia, V6T 1Z3, Canada, Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia, V6T 1Z1, Canada, and Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, K1A 0R6, Canada

Received December 11, 2006; Revised Manuscript Received March 14, 2007

ABSTRACT: Sialic acid is an essential sugar in biology that plays key roles in numerous cellular processes and interactions. The biosynthesis of sialylated glycoconjugates is catalyzed by five distinct families of sialyltransferases. In the last 25 years, there has been much research on the enzymes themselves, their genes, and their reaction products, but we still do not know the precise molecular mechanism of action for this class of glycosyltransferase. We previously reported the first detailed structural and kinetic characterization of Cst-II, a bifunctional sialyltransferase (CAZy GT-42) from the bacterium *Campylobacter jejuni* [Chiu et al. (2004) *Nat. Struct. Mol. Biol. 11*, 163−170]. This enzyme can use both Gal-β-1,3/4-R and Neu5Ac-α-2,3-Gal-β-1,3/4-R as acceptor sugars. A second sialyltransferase from this bacterium, Cst-I, has been shown to utilize solely Gal-β-1,3/4-R as the acceptor sugar in its transferase reaction. We report here the structural and kinetic characterization of this monofunctional enzyme, which belongs to the same sialyltransferase family as Cst-II, in both apo and substrate bound form. Our structural data show that Cst-I adopts a similar GTA-type glycosyltransferase fold to that of the bifunctional Cst-II, with conservation of several key noncharged catalytic residues. Significant differences are found, however, between the two enzymes in the lid domain region, which is critical to the creation of the acceptor sugar binding site. Furthermore, molecular modeling of various acceptor sugars within the active sites of these enzymes provides significant new insights into the structural basis for substrate specificities within this biologically important enzyme class.

Glycosyltransferases catalyze the transfer of a sugar moiety from an activated sugar-phosphate donor, which can be a sugar-nucleotide or a sugar-phospholipid, onto a specific acceptor resulting in the formation of new glycosidic bonds. This class of enzyme is represented across all branches of the tree of life with vital roles in a wide variety of essential biological processes. From the Carbohydrate-Active enZYmes (CAZy)[1] database (http://afmb.cnrs-mrs.fr/CAZY/), glycosyltransferases form a large functional class (second only to glycosidases in the number of identified members) comprising thousands of proteins grouped into families based on sequence similarity (*2*). A specific class of glycosyltransferases with particular biological importance is the sialyltransferases. These enzymes use CMP-*N*-acetylneuraminic acid (CMP-Neu5Ac) as the universal donor sugar and transfer the Neu5Ac moiety onto a diverse range of complex carbohydrates found mainly on the cell surface. Within the CAZy classification, the sialyltransferases are found in families GT-29, GT-38, GT-42, GT-52, and the recently formed GT-80. Of these, only GT-42 and GT-80 currently have been characterized at the molecular level (*1*, *3*). Structures from these two families have been shown to be vastly different: GT-42 sialyltransferases adopt the GT-A glycosyltransferase fold, whereas the product of gene PM0188 from *Pasteurella multocida* adopts a fold more typical of the GT-B group. Differences in the enzyme folds of these families may be a reflection of the multifunctional nature of the *P. multocida* enzyme (α-2,3-Sialyltransferase, α-2,6-

[†] This work was funded by the Howard Hughes Medical Institute (to N.C.J.S.) and the Canadian Institutes of Health Research (to N.C.J.S., S.G.W., W.W.W.), Michael Smith Foundation for Health Research and Canada Foundation for Innovation infrastructure grant (to N.C.J.S.), the Natural Sciences and Engineering Research Council and a Human Frontiers Science Grant (to S.G.W.) and by the National Research Council of Canada (to W.W.W. and M.G.). Michael Smith Foundation for Health Research for Research Trainee scholarships to C.P.C.C. and L.L.L., and Canadian Institutes of Health Research for Doctoral Research Award to C.P.C.C. are acknowledged. Data collection was supported by the Advance Light Source, Berkeley, CA, USA.

[‡] Coordinates for the structure described in this paper have been deposited in the Protein Data Bank with codes 2P2V and 2P56.

* Corresponding author. Phone: 1-604-8220789. Fax: 1-604-8227742. E-mail: natalie@byron.biochem.ubc.ca.

[#] Department of Biochemistry and Molecular Biology, University of British Columbia.

[§] Department of Chemistry, University of British Columbia.

[||] National Research Council of Canada.

[1] Abbreviations: CAZy, carbohydrate-active enzymes; CMP-Neu5Ac, cytidine 5′-monophosphate-*N*-acetylneuraminic acid; CMP-3FNeu5Ac, cytidine 5′-monophosphate-3-fluoro-*N*-acetylneuraminic acid; EDTA, ethylenediaminetetraacetic acid; GT, glycosyltransferase; IPTG, isopropyl-β-D-1-thiogalactopyranoside; NCAM, neural cell adhesion molecule; ORF, open reading frame.

10.1021/bi602543d CCC: $37.00
Published on Web 05/23/2007

sialyltransferase, α-2,3-sialidase, and an α-2,3-trans-sialidase activity have been reported (*3*)).

Glycoconjugates displaying sialic acids play important roles in mammalian cell−cell recognition, cell differentiation, and various receptor−ligand interactions. For example, the sialyl-Lewis X glycans found on cell surface glycolipids and glycoproteins are the specific ligands responsible for leukocyte "rolling and tethering" during an inflammatory response (*4*). Other examples are poly-α-2,8-sialic acid, which is found mainly on the mammalian neural cell adhesion molecule (NCAM) implicated in numerous normal and pathological processes, including mammalian development, neuronal plasticity, and tumor metastasis (*5*, *6*). Another class of sialic acid-containing glycoconjugates are the gangliosides, which act as receptors for microorganisms and bacterial toxins, regulate cell growth and differentiation, and contribute to cell−cell and cell−matrix interactions (*7*, *8*). An additional function for sialic acid-containing glycoconjugates is the regulation of the serum half-life of some circulating proteins, where the presence or absence of terminal sialic acid is correlated with the biological half-life of these proteins (*9*−*12*). In prokaryotes, sialylated glycoconjugates can play important functions in virulence. Some bacterial pathogens that invade the mammalian host have taken advantage of the presence of sialylated glycoconjugates on the host cells. These bacteria include *Neisseria gonorrheae*, *Neisseria meningitidis*, *Escherichia coli*, *Campylobacter jejuni*, *Streptococcus sp*. and *Haemophilus influenzae*, all of which display mimics of sialylated human glycan structures on their cell surfaces, and indeed a role for these carbohydrates in pathogenesis has been demonstrated (*13*, *14*). It is thought that the presence of these carbohydrate mimics allows the pathogens to escape detection by the immune system since these molecules are not considered foreign. Further, the presence of these carbohydrates presents a physical barrier for the killing action of serum complement (*15*). Finally, it may be that certain pathogens use normal human receptors that recognize their surface carbohydrate structures as a means of aiding transmission or colonization of the host (although this mechanism remains unproven for many of these pathogens) (*16*, *17*).

The carbohydrate mimics from *C. jejuni*, which are ganglioside analogues, are an example of where this molecular mimicry sometimes elicits an immune response with detrimental effects. *C. jejuni* is the most frequently identified triggering agent of the Guillain-Barré syndrome, a post-infectious autoimmune-mediated neuropathy (*18*). Cross-reactive antibodies between *C. jejuni* lipooligosaccharides and gangliosides have been observed in the serum of Guillain-Barré patients (*19*). The cross-reactive antibodies are hypothesized to be responsible for the neuromuscular damage because gangliosides are present on the surface of nerve tissues.

Because of the prominent role of these sialylated oligosaccharides in the biology of both the pathogen and the host, the enzymes involved in their biosynthesis in bacterial pathogens and humans represent potential targets for therapeutic intervention. Collectively, a detailed understanding of enzyme structure, substrate specificity, and mechanism of action of the bacterial sialyltransferases, including differences from the mammalian sialyltransferases, could shed significant light on the fundamental properties of these enzymes as well as provide a starting point for the design of therapeutics. We have previously described the structural, biochemical, and kinetic analysis of the *Campylobacter* bifunctional (α-2,3/8-) sialyltransferase Cst-II enzyme involved in the biosynthesis of ganglioside mimics which can use either a galactoside acceptor or α-2,3-sialyl-galactoside acceptor (*1*). In this work, we describe the structural and kinetic analysis of Cst-I, a related GT-42 member. Full-length Cst-I (residues 1−430) has only 42% sequence identity to Cst-II (residues 1−291) and is best described as monofunctional, as it exhibits only an α-2,3-sialyltransferase activity and shows a high specificity for galactoside acceptors. To further our understanding of these differences as well as sialyltransferase function in general, we report here the X-ray crystal structures of the monofunctional Cst-I in both the apo and the substrate analogue-bound forms. Together with a kinetic study characterizing its enzyme activity and a proposed model for acceptor sugar binding, this work provides significant new insight into the mechanism of catalysis and substrate specificity of sialyltransferases.

EXPERIMENTAL PROCEDURES

Cloning and Overexpression. Cloning of *cst-I* (Protein sequence database #AAF13495) from *C. jejuni* OH4384 was done as previously described (*20*). A truncated version of the gene, encoding residues 1−285, was amplified by Pwo polymerase with primers 5′CTTAGGAGGTCATATGACAAGGACTAGAATGGAAAATGAAC3′ and 5′TTTAGAATGGTCGACCTAATAAAAATTAAGCATAAT3′. The PCR product was digested with *Nde*I and *Sal*I restriction enzymes and subcloned into pCWori+(-*lacZ*) vector carrying the *E. coli* maltose-binding protein MalE (Protein sequence database #P02928) without signal peptide followed by the linker GGGH and a thrombin cleavage site (IFNPRGSH). The plasmid pCWori-malE-cst-I_1-285 was subsequently transformed into *E. coli* strain AD202, and cells were grown at 37 °C in 2×YT broth with 150 μg mL^{-1} ampicillin and 0.2% (w/v) glucose. When the OD_{600} nm reached ∼0.5, a final concentration of 1.0 mM IPTG was added, and the cells were switched to 25 °C and incubated for 22 h. Cells were then harvested at 6200 × *g* for 15 min and stored at −80 °C.

Purification of the Cst-I$^{1-285}$ Product and Kinetic Assay. The cell pellet (2.4 g wet weight) was resuspended in 24 mL of 20 mM Tris-HCl pH 7.5 buffer containing 200 mM NaCl, 1.0 mM EDTA, and 5.0 mM 2-mercaptoethanol. The cells were lysed in three passes using an Emulsiflex C-5 at ∼19 000 psi. At this step, a tablet of protease inhibitor cocktail (Roche) was added to the sample. The resulting extract was ultracentrifuged at 75000 × *g* for 35 min at 4 °C. The supernatant was then dialyzed overnight against 2.0 L of 20 mM Tris-HCl pH 7.5 buffer containing 200 mM NaCl, 1.0 mM EDTA, and 5.0 mM 2-mercaptoethanol before loading it on an amylose affinity column.

A column (2.6 × 4.5 cm) packed with amylose resin (New England Biolab E8021L) was equilibrated with 125 mL of 20 mM Tris-HCl pH 7.5 buffer containing 200 mM NaCl, 1.0 mM EDTA, and 5.0 mM 2-mercaptoethanol. The overnight dialysate was added to the resin and incubated at 4 °C with rocking for 1 h. The resin was packed into a Bio-Rad open column and washed with 100 mL of equilibration buffer at a flow rate of 2.0 mL min^{-1}. The fusion protein

was eluted with 70 mL of the same buffer containing 10 mM maltose. For thrombin cleavage, the fusion protein was incubated at 4 °C overnight with 1:1000 dilution restriction grade thrombin (Sigma) in 1× thrombin cleavage buffer. The sample was then desalted via dialysis against 20 mM Tris pH 7.5 and then applied to a MonoQ HR 10/10 column (GE Healthcare) pre-equilibrated with the same buffer. Cst-I$^{1-285}$ was collected from the flow-through. MalE and the uncleaved fusion protein were removed from the column with a 20 mL gradient of 0−0.5 M NaCl in the same buffer. Kinetic analysis of the purified enzyme using lactose as cosubstrate was performed using a continuous coupled assay analogous to that previously described (*1*).

Crystallization and Data Collection. Cst-I$^{1-285}$, purified as above, was concentrated using a 15 mL Amicon Ultra-4 centrifugal filter device (10 000 MW cutoff) spun on an Allegra 21R centrifuge (Beckman Coulter). Tetragonal protein crystals started to appear when the concentration of the protein reached ~3 mg mL^{-1}. For the substrate-bound structure, both the acceptor sugar lactose and the inert donor sugar analogue CMP-3FNeu5Ac (synthesized as previously described (*21*)) were soaked into the crystals for 30 min before data collection. The crystals were transferred from 0 to 30% v/v ethylene glycol at 2.5% intervals for cryoprotection. Both apo- and substrate-bound crystals adopt the space group *I*4 with unit cell dimensions $a = b = 112.4$, $c = 58.8$ Å. For the apoenzyme, a 2.2 Å dataset was collected on a local Rigaku RU200 rotation anode equipped with OSMIC mirrors and a MAR345 image plate detector. For the CMP-3FNeu5Ac complex, a 1.7 Å dataset was collected at the Advanced Light Source beamline 8-2-1 using an ADSC Quantum Q210 2 × 2 CCD array detector. Data were processed using DENZO and SCALEPACK (*22*). Statistics for data collection and processing are summarized in Table 1.

Structure Determination, Refinement, and Modeling. The apo-Cst-I$^{1-285}$ structure was solved by molecular replacement with Molrep (*23*) using a monomer of Cst-II as the starting model. Model phases were improved by RESOLVE (*24*). The initial model was built with XtalView (*25*) using Cst-II as the template, and the final model was obtained after iterations of refinement with CNS 1.1 (*26*) and Refmac 5.0.1 (*27*). The CMP-3FNeu5Ac complex structure was solved using apo-Cst-I$^{1-285}$ structure as the starting model in the refinement. The quality of the models was analyzed with PROCHECK (*28*), and the results are summarized in Table 1. Alignment of structures was done with Align (*29*). Modeling of acceptor sugars was done using a combination of AutoDock3.0 (*30*) and manual adjustments to maximize interactions between ligands and active site residues and to position the correct oxygen in close proximity to the anomeric center of donor sugar using XtalView. Required parameter files for AutoDock3.0 were prepared using the Dundee PRODRG server (*31*). Figure 1 is prepared with ESPript (*32*), and Figures 2−6 were prepared with PyMol (http://www.pymol.org).

Table 1: Data Collection and Refinement Statistics

crystal parameters	Apo	CMP-3FNeu5Ac
Data Collection		
space group	*I*4	*I*4
diffraction statistics		
X-ray source	Rigaku 007HF Micromax	ALS 8.2.1
resolution	25−2.2	50−1.7
wavelength	1.514	1.00
total reflections	66105	429679
unique reflections	18129	40593
redundancy	3.65	10.59
completeness (%)[a]	96.5 (66.8)	99.8 (99.0)
$<I/\sigma I>$[a]	16.3 (3.5)	28.9 (2.6)
R_{sym}[a,b]	4.3 (21.4)	6.4 (52.4)
Refinement Statistics and Model Stereochemistry		
resolution (Å)	25−2.2	40−1.85
number of atoms		
protein	2129	2285
substrate	0	42
water	84	200
cryoprotectant	20	48
ion	0	1
R_{cryst}/R_{free} (%)[c]	21.2/25.0	20.2/24.1
r.m.s. deviations		
bonds (Å)	0.006	0.019
angles (°)	1.19	1.99
average B-factor (Å^2)		
protein	39.8	37.4
substrate		45.5
water	36.3	39.4
cryoprotectant	39.7	47.6
ion		34.1

[a] High-resolution shell (2.28−2.20 Å for apo form and 1.76−1.70 Å for CMP-3FNeu5Ac bound form) statistics are in parentheses. [b] $R_{sym} = \sum|(I_{hkl}) - <I>|/\sum(I_{hkl})$, where I_{hkl} is the integrated intensity of a given reflection. [c] $R_{cryst} = (\sum|F_o - F_c|)/(\sum F_o)$, where F_o and F_c are observed and calculated structure factors. 5% of total reflections were excluded from the refinement to calculate R_{free}.

RESULTS AND DISCUSSION

Sequence and Kinetic Analysis of Cst-I. Full-length Cst-I (residues 1−430) contains a predicted N-terminal sialyltransferase domain (based on sequence similarities with other members of CAZy family 42), and an additional C-terminal domain that shows localized sequence homology with two ORFs in the *cst-I* locus and three ORFs in the capsule biosynthesis locus of *C. jejuni* OH4384 but no discernible homology with any known deposited structures. Full-length Cst-I is not sufficiently soluble at the higher concentrations required for crystallization trials; thus, a truncated version was designed based on sequence alignments with the naturally shorter Cst-II, which lacks an analogous C-terminal domain. This truncated version of Cst-I (residues 1−285 or Cst-I$^{1-285}$) retains a high level of activity, showing that the C-terminal domain likely does not contribute in a significant way to the active site or the observed α-2,3-transferase activity. Although one can postulate potential roles for the apparently novel C-terminal domain based on its similarity to ORFs in the capsule biosynthesis locus, including that of cell localization and/or protein scaffolding, clearly further investigation will be required to understand its structure and its contributions to Cst-I action in vivo.

Using sialyl lactose as the acceptor at concentrations of up to 500 mM, transferase activity could not be detected above the significant ($k_{cat} = 0.1$ s^{-1}) enzyme catalyzed hydrolysis of the donor substrate. In addition, product formation could not be detected using thin layer chromatography (TLC) or capillary electrophoretic analysis with sialyl lactose as the acceptor (data not shown). These results clearly demonstrate the reaction specificity of Cst-I, with an exclusive α-2,3-transferase activity with Gal-β-R acceptor

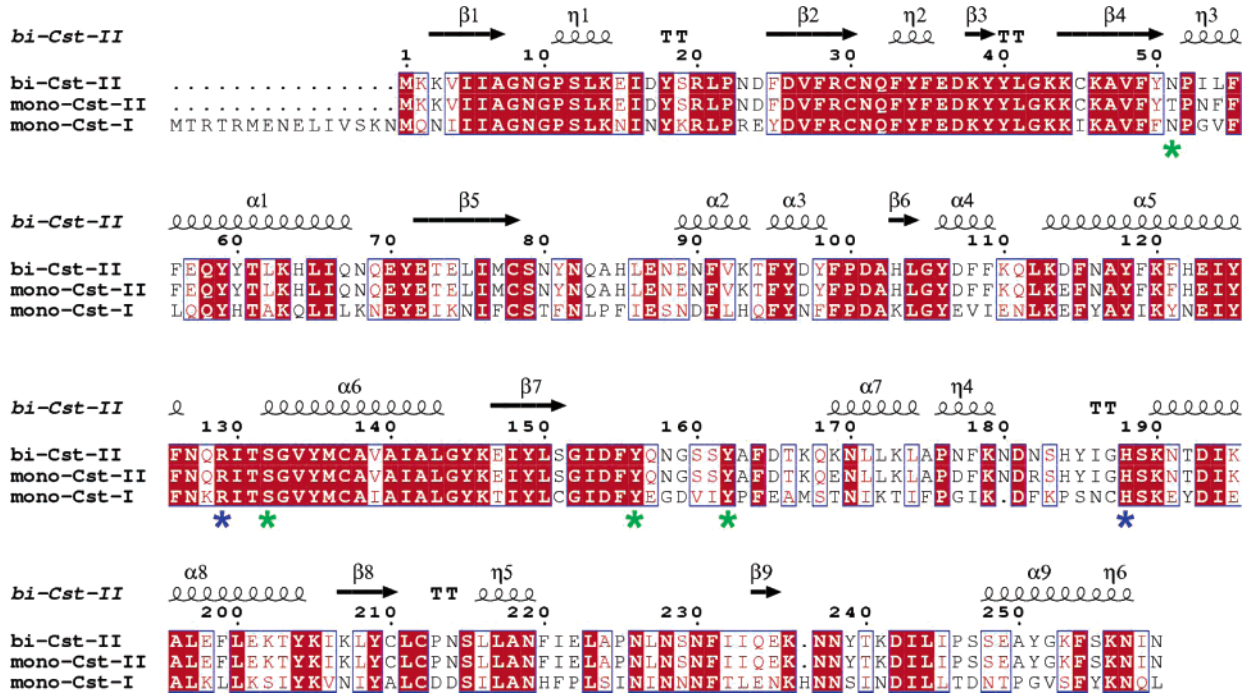

FIGURE 1: Sequence alignment of bifunctional Cst-II$^{1-259}$, monofunctional Cst-II$^{1-259}$, and monofunctional Cst-I$^{1-285}$, all in their respective truncated forms. Residues important for coordinating the donor sugar (Asn51/66, Ser132/147, Tyr156/171, and Tyr162/177) are highlighted by green asterisks. Proposed general base, His188/202, and its pKa modulator, Arg129/144, are highlighted by blue asterisks. Red box with white character shows strict identity between the sequences. Red character indicates similarity in a group, and blue frame depicts similarity across groups. Sequence alignment was generated by T-COFFEE (*38*), and final figure with secondary structures shown was prepared by ESPript (*32*).

sugar. The K_M value of the donor sugar CMP-Neu5Ac for Cst-I$^{1-285}$ was determined to be 400 μM, which is similar to the previously reported value of 460 μM for Cst-II (*1*). By contrast, monofunctional Cst-I$^{1-285}$ has a K_M value of 500 μM for the acceptor lactose, while bifunctional Cst-II has a value of 35 mM. This is a substantial difference but likely is a consequence of the bifunctional nature of Cst-II, which binds sialyl lactose with a higher affinity (K_M value of 3.5 mM) and has a higher α-2,8- than α-2,3-transferase activity.

Determination of Cst I$^{1-285}$ Structure. The structure of the catalytic domain of Cst-I$^{1-285}$ in the apo form was first solved via molecular replacement methodologies using Cst-II (PDB code 1RO7) as the starting model (there is ~53% local sequence identity between Cst-I$^{1-285}$ and Cst-II$^{1-259}$; see Figure 1). The resulting Cst-I$^{1-285}$ model is well refined, with crystallographic R/R_{free} values of 0.21/0.25 for the apoenzyme (to 2.2 Å resolution). The overall structure is a mixed α/β fold and is composed of an N-terminal Rossmann nucleotide binding domain and a C-terminal "lid" domain that encompasses the active site region of the enzyme (Figure 2). Static light scattering analysis reveals that Cst-I$^{1-285}$ forms a tetramer in solution (data not shown). Correspondingly, we observe that the predicted biological 4-fold axis of the tetramer in our structure coincides with the crystallographic 4-fold axis in the body-centered tetragonal *I*4 space group with the putative membrane interface of all four molecules aligned on the same face of the tetramer. The majority of the residues involved in oligomerization are conserved within the Cst family and are primarily aromatic and hydrophobic residues that create an extensive hydrophobic interface between the four monomers that maintains the tetrameric oligomerization state (~1300 Å^3 of buried surface).

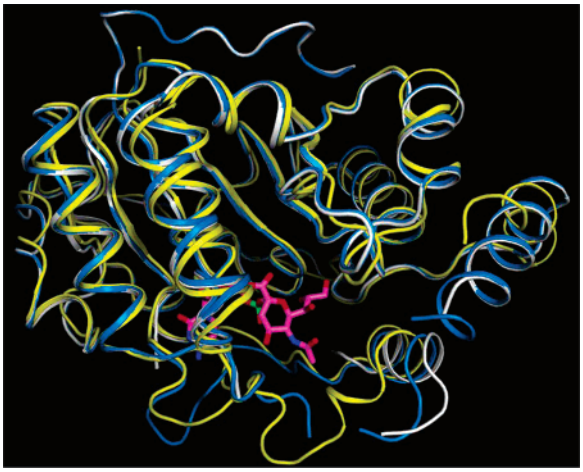

FIGURE 2: Structural alignment of the three *Campylobacter* sialyltransferases. The Cα trace of apo-Cst-I is shown in white, and the substrate-bound form is in blue, and that for Cst-II is in yellow. Donor sugar analogue is shown in magenta.

Although the overall architecture of Cst-I$^{1-285}$ is similar to that of Cst-II, with a root-mean-square deviation of 0.8 Å between 235 Cα atoms, there are major structural differences in the lid domain that folds over the active site, conformational differences that we believe dictate the consequent substrate specificity of the enzymes (Figure 2; see below).

Comparison of Structures of the apo- and Substrate-Bound Monofunctional Cst-I$^{1-285}$. The structure of Cst-I$^{1-285}$ in complex with the inert donor sugar analogue CMP-3-fluoro-*N*-acetyl-neuraminic (CMP-3FNeuAc) was solved using difference Fourier methods in the same crystal form as that of the apo Cst-I$^{1-285}$ and is well refined with R/R_{free} values of 0.20/0.24 (to 1.85 Å resolution). Cst-I$^{1-285}$ binds the CMP-3FNeuAc in the N-terminal Rossmann nucleotide binding

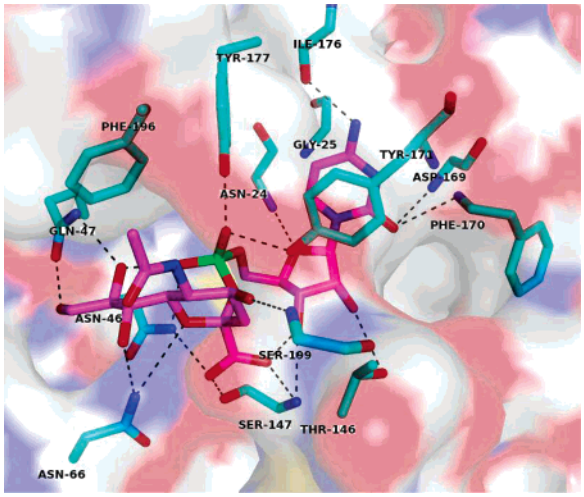

FIGURE 3: Interactions of CMP-3FNeuAc with active site residues of Cst-I$^{1-285}$. Donor analogue CMP-3FNeu5Ac is depicted with carbon atoms in magenta, nitrogen atoms are in blue, oxygen atoms are in red, and phosphorus atoms are in green. Active site residues involved in CMP-3FNeuAc binding and catalysis are labeled and shown with carbon atoms in cyan, nitrogen in blue, and oxygen in red. Dotted lines indicate hydrogen bonding.

domain. The cytidine base of CMP is tucked inside a cavity between Tyr171 and Gly25 and is stacked on the aromatic side chain of Tyr171 (Figure 3). The base also forms intermolecular hydrogen bonds with the main chain amides of Asp169, Phe170, Tyr171, and the main chain carbonyl oxygen of Ile176. The ribose moiety of the nucleotide base adopts a C3′ endo conformation with the cytidine base axial to the ring. It is held in place with hydrogen bonds to the main chain amides of Asn24 and Ser147 and the side chain OG1 of Thr146. The two free oxygens of the phosphate group form a complex hydrogen-bonding network to the side chain amide of Asn46, O8 of the sialic acid, and the side chain hydroxyl atoms of tyrosines 171 and 177. The sialic acid moiety of the donor nucleotide also forms several interactions with the active site residues. The carboxylic acid group is coordinated by both the main chain amide and side chain OG of Ser147. It also forms hydrogen bonds to ND2 of Asn66, and O3 of the ribose moiety. O4 interacts with the main chain nitrogen of Ser199, O6 and O7 of the sugar ring are within hydrogen-bonding distance of ND2 of Asn66 and O8 and O9 are hydrogen bonded, respectively, to NE2 and OE1 of the Gln47 side chain. The *N*-acetyl group forms hydrophobic interactions with the phenyl side chain of Phe196. Although the residues in the lid domain (residues 171−199) are not as well ordered as the rest of the enzyme (as indicated by their higher temperature factors), the majority of the residues can be modeled unambiguously in the electron density map. In contrast, in the apoenzyme structure where the substrate is absent, two loop regions of the lid domain (residues 171−181 and residues 194−199) are highly disordered and cannot be modeled in the electron density map. These regions include some of the key residues involved in donor sugar binding, such as the two conserved tyrosine residues, Tyr171 and Tyr177. These tyrosine residues interact with the phosphate oxygen of the nucleotide and are thought to stabilize the negative charge that would develop on the phosphate oxygens during catalysis. Indeed, a double mutation of the analogous tyrosines into alanines in Cst-II was shown to abolish the enzyme's transferase activity (*1*). Recently, Jeanneau et al. have also shown via mutagenesis and kinetic analysis that an invariant tyrosine residue in the conserved sequence motif (sialyl-motif 3) of the human sialyltransferase ST3Gal I may play a similar catalytic role (*33*).

Our comparison of the apo and substrate forms of Cst-I$^{1-285}$, the first such analysis for a family 42 sialyltransferase as the apo form of Cst-II was not amenable to crystallization, suggests that the donor substrate CMP-Neu5Ac binds initially and upon binding induces conformational changes that create the appropriate acceptor binding site. The disorder of the residues in the lid domain of the apo form effectively creates a wide open active site cavity that would serve to facilitate initial access of the CMP donor sugar to the active site. Upon binding the donor sugar, the flexible loop of the lid domain rigidifies through contacts with the donor sugar, consequently creating the acceptor binding site and minimizing potential hydrolytic effects by more effectively excluding bulk solvent from the reactive center in the closed form. Although this is the first such observation for the structurally distinct sialyltransferases, the equivalent ordering of active site residues has been observed in the structural analysis of other glycosyltransferases (e.g., SpsA (*34*), α3GT (*35*), and LgtC (*36*) to name a few).

Catalytic Residues in Cst-I. Cst-I works by a direct displacement mechanism with overall inversion of configuration at the anomeric center and the requirement of a proximal functional group that could serve as the general base to activate the acceptor sugar for the transfer to occur. His202, which is positioned immediately after the more mobile lid domain region, is highly ordered, with the lowest observed B-factors of the residues that comprise the active site. It is also conserved in Cst-II (His188) and based on the orientation of its imidazole side chain relative to the anomeric carbon of the donor sugar has been proposed as a likely candidate for the general base in the catalytic mechanism. In this role, it would extract a proton from the 3′-hydroxyl group of the incoming acceptor molecule, during its nucleophilic attack on the anomeric center of the donor CMP-Neu5Ac. Recently, several studies on human sialyltransferases (ST3Gal I (*33*), ST8Sia II and IV (*37*)) have also identified an invariant histidine residue that is critical for the catalytic activities of those enzymes. In our Cst-I$^{1-285}$ structure, we also observe a close electrostatic influence from the nearby Arg144 (5 Å). A similar interaction in the Cst-II active site was proposed to play a role in regulating the pKa of the general base histidine.

Our observations of these conserved features in Cst-I$^{1-285}$, despite many other point differences in the active site, support the previously proposed mechanism of CAZy family 42 sialyltransferases. Several additional active site residues identified from the bifunctional Cst-II structure and mutagenesis studies are conserved in Cst-I and occupy roughly the analogous position, namely, Asn31(Asn46), Asn51-(Asn66), Ser132(Ser157), Tyr156(Tyr171), and Tyr162-(Tyr177). These residues are involved in the binding of CMP-Neu5Ac. Since these catalytic players are conserved between the monofunctional Cst-I and bifunctional Cst-II variants, other critical point mutations in the active site must be responsible for the different substrate specificities of these enzymes, as is discussed further below.

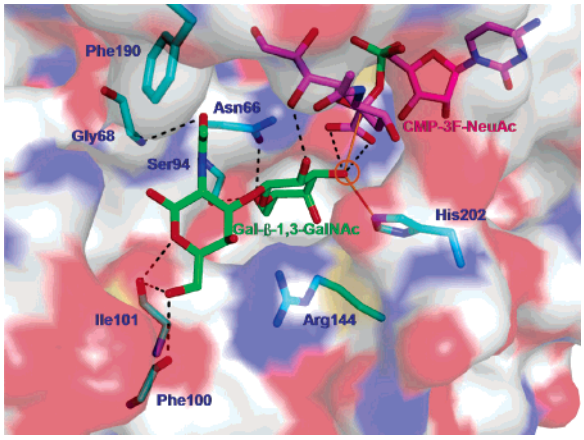

FIGURE 4: Molecular modeling of Gal-β-1,3-GalNAc into the active site of Cst-I. Active site residues are shown in cyan, donor analogue is shown in magenta, and the model of Gal-β-1,3-GalNAc is shown in green. The site at which the sialic acid would be transferred onto the acceptor, O3 of the galactose moiety, is highlighted by the orange circle. The orange line connects the imidazole ring of the proposed base His202 to the O3 of the galactose and on to the C2 anomeric center of the donor sugar.

Docking of Acceptor Sugars to the Active Site of Cst-I$^{1-285}$. Despite extensive attempts to obtain a donor−acceptor ternary complex structure via soaking of crystals with a range of acceptor analogues, no complexes were obtained. At best, we observe in the electron density map of donor/acceptor Cst-I$^{1-285}$ soaks a small amount of residual density at 1 sigma, possibly representing a very weakly occupied potential acceptor site situated on the alpha face of the donor sugar ring and in close proximity to the proposed general base His202 (Figure 4). Our density suggests that a molecule of the cryoprotectant additive ethylene glycol, which is essential for the crystallographic data collection (an extensive screen of other cryoprotectants failed), successfully competes with the acceptor sugar at this site due to the abundant amount used in cryoprotection (30% v/v). The refined model of the bound ethylene glycol indeed shows interactions with the enzyme (the hydroxyl groups hydrogen bond with O of Ile145, 2.7 Å, and OD 1 of Asn66, 2.9 Å) that may mimic that of the hydroxyl groups of the natural acceptor sugar. Using this information as a guide, the binding mode of the acceptor sugar, Gal-β-1,3-GalNAc, was modeled into the active site of Cst-I$^{1-285}$ using a combination of AutoDock3.0 and manual fitting. The final model of the bound Gal-β-1,3-GalNAc is oriented at a ~50° angle to the donor sugar ring and is sandwiched between the side chains of Asn66 and Arg144. The distance between O3 of the galactose and the anomeric C2′ of the donor sugar is approximately 3.9 Å. The imidazole ring of the proposed general base, His202, is ~3.5 Å away from the same O3 atom. Both of these distances are compatible with a potential deprotonation event by His202 and subsequent nucleophilic attack on the anomeric center of the donor sugar. Galactose is anchored in place by stacking with the amide side chain of Asn66, forming hydrophobic interactions with this key residue that we postulated previously (in Cst-II) plays a key role in defining the substrate specificity of the enzyme. O2 of the galactose is within hydrogen-bonding distance of the O7′ of the donor sugar and O4 and O6 of the acceptor are within hydrogen-bonding distance to the side chains of Glu102 and Arg144, respectively. The GalNAc portion of the acceptor is situated in a more solvent-exposed environment than the galactose. The only potential hydrogen bond to enzyme is between O4 of GalNAc and the main chain carbonyl group of Ile101. There is also a potential set of hydrophobic interactions between the *N*-acetyl group and the side chain of Phe190. The smaller number of specific interactions of GalNAc to the enzyme would explain the tolerance of the enzyme for other galactose derivatives as acceptor sugars (for example, the readily available lactose substrate used in our kinetic assays). Modeling analysis of lactose into the active site of Cst-I$^{1-285}$ shows low energy forms are able to bind in a similar orientation and with retention of similar interactions to active site residues as predicted for the Gal-β-1,3-GalNAc.

Even though sialyl lactose is not an acceptor for Cst-I, it was also used in modeling analysis for this enzyme. After several docking attempts, all the resulting models show an overall higher estimated free energy of binding (close to zero or positive in the range of −0.5 to 1.4 kcal/mol) as compared to Gal-β-1,3-GalNAc or lactose (~ −4 kcal/mol). Furthermore, the program failed to generate a model with O8 of the sialic acid moiety close to the anomeric center of donor sugar in a correct orientation (with the end of the sugar chain pointing toward the core of the enzyme). Collectively, these results provide a likely explanation for the inability of this enzyme to bind sialyl lactose tightly or in the correct orientation for the sialylation to occur.

Docking of Acceptors to Bifunctional Cst-II and Comparison to Cst-I. Although the overall structures of Cst-I$^{1-285}$ and Cst-II are similar, the two enzymes have fundamental differences in substrate specificity. Cst-I can perform only a single α-2,3 transfer of sialic acid onto galactose derivatives. In contrast, bifunctional Cst-II can transfer sialic acid first onto position 3 of galactose and perform a subsequent transfer onto position O8 of the sialyl-galactose generated from the initial transfer step. Indeed, when comparing the primary sequences of the two versions of Csts, the majority of differences occur in the lid domain region which orders over the substrate to form part of the active site. This region also shows the least conservation in the crystal structures, implying that although the enzymes utilize the same set of active site residues in catalysis, their lid domain region plays the role of selecting different acceptor molecules for the reaction to proceed.

To elucidate the detailed differences in the binding of the acceptor sugar in Cst-II, docking of both Gal-β-1,3-GalNAc (α-2,3 transfer; Figure 5a) and Neu5Ac-α-2,3-Gal-β-1,4-Glc (sialyl-lactose, α-2,8 transfer; Figure 5b) to the active site of Cst-II was performed, both of which gave reasonable negative estimations for the free energy of binding. Although the galactose sugar occupies a relatively similar space with respect to the reaction center of the donor sugar and the postulated histidine general base in the two enzymes, the major difference lies in the GalNAc sugar ring. In the docked model of Cst-I$^{1-285}$, the GalNAc ring stacks against the main chain atoms of Gly68, which forms part of a β-turn composed of Asn66-Pro67-Gly68 and Val69. In Cst-II, the replacement of the Gly68 by a serine (Ser53 in Cst-II) sterically blocks GalNAc from occupying the analogous position due to the protrusion of the serine side chain hydroxyl. As a result, the *N*-acetyl group is forced closer to Tyr185 allowing favorable

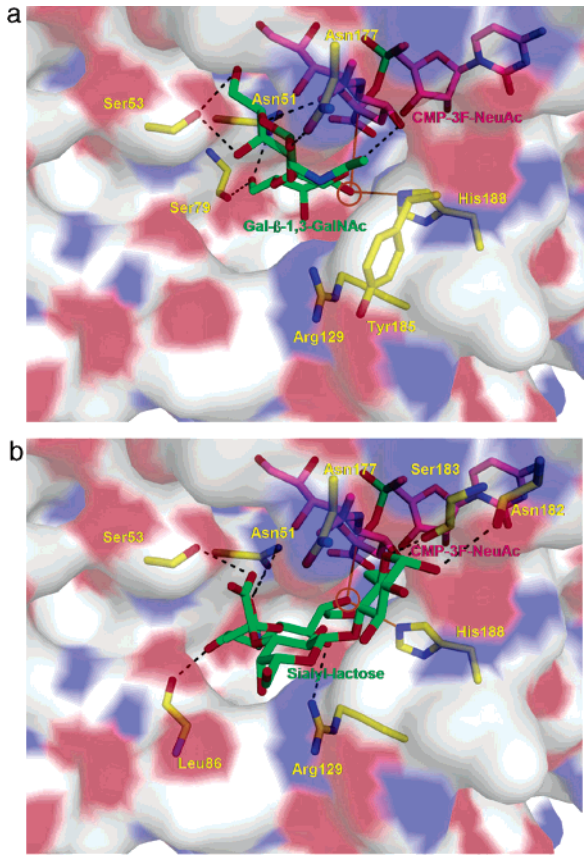

FIGURE 5: (a) Molecular modeling of Gal-β-1,3-GalNAc into the active site of Cst-II. Active site residues are shown in yellow, donor analogue is shown in magenta, and the model of Gal-β-1,3-GalNAc is shown in green. The site at which the sialic acid would transferred onto the acceptor, O3 of the galactose moiety, is highlighted by the orange circle. The orange line connects the imidazole ring of the proposed base His188 to the O3 of the galactose and on to the C2 anomeric center of the donor sugar. (b) Molecular modeling of sialyl-lactose into the active site of Cst-II. Active site residues are shown in yellow, the donor analogue is shown in magenta and is kept in the same orientation as in panel a, and the model of Gal-β-1,3-GalNAc is shown in green. The site at which the sialic acid would transfer onto the acceptor, O8 of the sialic acid moiety, is highlighted by the orange circle. The orange line connects the imidazole ring of the proposed base His188 to the O8 of the sialic acid and on to the C2 anomeric center of the donor sugar.

hydrophobic interactions with the side chain aromatic. The carbonyl oxygen of the *N*-acetyl group can also form a hydrogen bond to the NH1 of Arg129 in this position and O1 of the GalNAc is within hydrogen-bonding distance to Asn177. These predicted interactions between the enzyme and the acceptor sugar are quite different than that for Cst-I, implying that even though both enzymes employ Gal-β-1,3-GalNAc as acceptor sugars, the coordination of the terminal sugar in the active site will have significant orientational differences.

Docking of the bulkier sialyl lactose acceptor sugar involved in the α-2,8 transfer to Cst-II allows for appropriate placement of O8′ in the glycerol moiety of the sialic acid relative to the general base His188 ($\sim$3.4 Å) and the anomeric center of the donor sialic acid ($\sim$4.5 Å). In this position, Arg129 from the enzyme could also form a strong hydrogen bond to O7 of the glycerol moiety. Unlike in Cst-I$^{1-285}$ where Asn66 stacks with the galactose unit of the acceptor sugar, the equivalent Asn51 of Cst-II does not interact directly with the docked acceptor. As well, the side

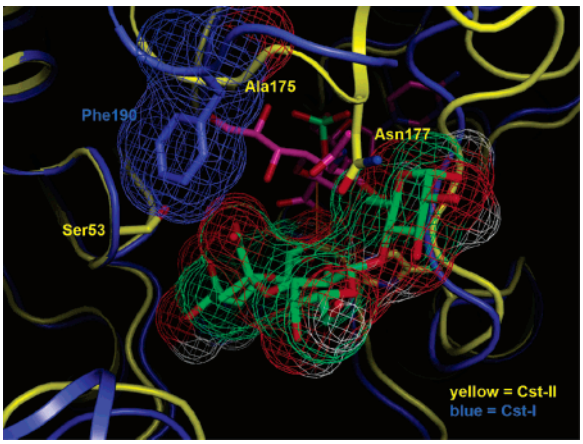

FIGURE 6: Overlap of the molecular modeling of sialyl lactose of Cst-II into the active site of Cst-I. The structure of Cst-I is shown in blue, and the structure of Cst-II is shown in yellow. The sialyl lactose model is shown in green. The mesh represents the space-filling diagram of the enclosed molecule.

chain of Ser53 again protrudes into the acceptor binding cleft and hinders close interaction of the acceptor to the β-turn of Asn51-Pro52-Ser53-Leu54 and in fact stabilizes the acceptor in the altered position through a potential interaction from its side chain hydroxyl to the β carboxyl group. It is interesting to note that the conserved Arg129 on the opposite face of the sugar ring is not suitably positioned to interact with the carboxylate of the acceptor sugar. As docked, the galactose ring of the sugar does not form many direct contacts with the enzyme active site residues but appears rather to be stabilized by intramolecular interactions such as contacts between O2 of the galactose and O7 of sialic acid, as well as O5 of galactose and O3 of the glucose. The glucose ring forms favorable vdW interactions with the amide side chain of Asn177. In the monofunctional version of Cst-II (isolated from serotype O19 (*20*)) which contains only eight amino acids difference from its bifunctional isoform (Figure 1), an aspartate residue was found at position 177 (one of the eight different amino acids). Preliminary activity data of designed point mutants shows that substitution of this aspartate residue with a neutral asparagine significantly enhances α-2,8 activity of monofunctional Cst-II (data not shown). Clearly, a negative charge at this position would disfavor the negatively charged sialic acid moiety of the acceptor sugar.

Attempts to dock the same acceptor, sialyl lactose, to the active site of the monofunctional Cst-I$^{1-285}$ provide clear indications as to the inability of this sialyltransferase variant to effectively transfer the bulkier acceptor sugar (Figure 6). Since docking of sialyl lactose specifically for Cst-I failed to yield a reasonable model, the model of sialyl lactose for Cst-II is overlapped into the active site of Cst-I. While the active site of Cst-I$^{1-285}$ can accommodate the entire substrate, the negatively charged carboxyl group of the sialyl moiety is forced into close proximity to the hydrophobic side chain of Phe190. The presence of the bulky aromatic side chain of Phe190 imposes both steric and electrostatic hindrance for the binding of this distinct carboxyl group of the sialyl-lactose acceptor. In Cst-II, the equivalent position to Phe190 is an alanine (Ala175) whose small methyl side chain would not impose the same steric and electrostatic burden on the binding of sialyl-lactose for the α-2,8 transfer.

In conclusion, although the exact role of Cst-I from GT-42 in vivo is still unknown, its structure in both apo and in CMP-3FNeu5Ac complexed forms further supports the kinetic mechanism proposed for this family of sialyltransferases. Moreover, the structure of this monofunctional version of Cst provides further understanding of the differences in substrate specificities between different isoforms of *Campylobacter* sialyltransferase within this family. Although no acceptor is observed in the active site of Cst-I^{1-285}, its structure, together with the previously solved bifunctional Cst-II structure, enables us to compare and contrast the electrostatic environment of their acceptor binding sites. Molecular modeling of different acceptors to their respective active sites allow us to identify key residues that are involved in defining substrate specificity. Although it is of utmost interest to characterize structurally and kinetically both single point and composite mutants of either Cst-I or Cst-II with switched substrate specificities, at this moderate level of sequence identity (only ~53% for the catalytic domains solved) this type of protein engineering is very difficult. Compensatory residues that surround these point differences in these active sites are not conserved. Thus, engineering of single point mutations that switch from mono- to bifunctional while maintaining kinetic competence activity is unlikely, as compensatory changes in the residues surrounding those point mutants will also be needed.

ACKNOWLEDGMENT

We thank the U.S. Department of Energy for access to data collection facilities at the ALS and the staff of ALS for provision of data collection facilities.

REFERENCES

1. 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.
2. 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.
3. Ni, L., Sun, M., Yu, H., Chokhawala, H., Chen, X., and Fisher, A. J. (2006) Cytidine 5′-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*, *Biochemistry 45*, 2139−2148.
4. Renkonen, R., Mattila, P., Majuri, M. L., Rabina, J., Toppila, S., Renkonen, J., Hirvas, L., Niittymaki, J., Turunen, J. P., Renkonen, O., and Paavonen, T. (1997) In vitro experimental studies of sialyl Lewis x and sialyl Lewis a on endothelial and carcinoma cells: crucial glycans on selectin ligands, *Glycoconjugate J. 14*, 593−600.
5. Theodosis, D. T., Piet, R., Poulain, D. A., and Oliet, S. H. (2004) Neuronal, glial and synaptic remodeling in the adult hypothalamus: functional consequences and role of cell surface and extracellular matrix adhesion molecules, *Neurochem. Int. 45*, 491−501.
6. Weinhold, B., Seidenfaden, R., Rockle, I., Muhlenhoff, M., Schertzinger, F., Conzelmann, S., Marth, J. D., Gerardy-Schahn, R., and Hildebrandt, H. (2005) Genetic ablation of polysialic acid causes severe neurodevelopmental defects rescued by deletion of the neural cell adhesion molecule, *J. Biol. Chem. 280*, 42971−42977.
7. Miljan, E. A., and Bremer, E. G. (2002) Regulation of growth factor receptors by gangliosides, *Sci STKE 2002*, RE15.
8. Lloyd, K. O., and Furukawa, K. (1998) Biosynthesis and functions of gangliosides: recent advances, *Glycoconjugate J. 15*, 627−636.
9. Elliott, S., Egrie, J., Browne, J., Lorenzini, T., Busse, L., Rogers, N., and Ponting, I. (2004) Control of rHuEPO biological activity: the role of carbohydrate, *Exp. Hematol. 32*, 1146−1155.
10. Chitlaru, T., Kronman, C., Zeevi, M., Kam, M., Harel, A., Ordentlich, A., Velan, B., and Shafferman, A. (1998) Modulation of circulatory residence of recombinant acetylcholinesterase through biochemical or genetic manipulation of sialylation levels, *Biochem. J. 336 (Pt 3)*, 647−658.
11. Kronman, C., Chitlaru, T., Elhanany, E., Velan, B., and Shafferman, A. (2000) Hierarchy of post-translational modifications involved in the circulatory longevity of glycoproteins. Demonstration of concerted contributions of glycan sialylation and subunit assembly to the pharmacokinetic behavior of bovine acetylcholinesterase, *J. Biol. Chem. 275*, 29488−29502.
12. Chitlaru, T., Kronman, C., Velan, B., and Shafferman, A. (2002) Overloading and removal of *N*-glycosylation targets on human acetylcholinesterase: effects on glycan composition and circulatory residence time, *Biochem. J. 363*, 619−631.
13. Kahler, C. M., and Stephens, D. S. (1998) Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (endotoxin), *Crit. Rev. Microbiol. 24*, 281−334.
14. Moran, A. P., Prendergast, M. M., and Appelmelk, B. J. (1996) Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease, *FEMS Immunol. Med. Microbiol. 16*, 105−115.
15. Vogel, U., Hammerschmidt, S., and Frosch, M. (1996) Sialic acids of both the capsule and the sialylated lipooligosaccharide of *Neisseria meningitis* serogroup B are prerequisites for virulence of meningococci in the infant rat, *Med. Microbiol. Immunol. (Berl) 185*, 81−7.
16. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) The lipooligosaccharides of pathogenic gram-negative bacteria, *Crit. Rev. Microbiol. 22*, 139−180.
17. Harvey, H. A., Porat, N., Campbell, C. A., Jennings, M., Gibson, B. W., Phillips, N. J., Apicella, M. A., and Blake, M. S. (2000) Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm, *Mol. Microbiol. 36*, 1059−1070.
18. Jacobs, B. C., Rothbarth, P. H., van der Meche, F. G., Herbrink, P., Schmitz, P. I., de Klerk, M. A., and van Doorn, P. A. (1998) The spectrum of antecedent infections in Guillain-Barre syndrome: a case-control study, *Neurology 51*, 1110−1115.
19. Willison, H. J., and Yuki, N. (2002) Peripheral neuropathies and anti-glycolipid antibodies, *Brain 125*, 2591−2625.
20. Gilbert, M., Karwaski, M. F., Bernatchez, S., Young, N. M., Taboada, E., Michniewicz, J., Cunningham, A. M., and Wakarchuk, W. W. (2002) The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide, *J. Biol. Chem. 277*, 327−337.
21. Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C., and Withers, S. G. (2003) *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile, *J. Am. Chem. Soc. 125*, 7532−7533.
22. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
23. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D56*, 1622−1624.
24. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr. D Biol. Crystallogr. 55 (Pt 4)*, 849−61.
25. McRee, D. E. (1999) XtalView/Xfit, A versatile program for manipulating atomic coordinates and electron density, *J. Structur. Biol. 125*, 156−165.
26. Brunger, 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. D54*, 905−921.
27. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr. D55*, 247−255.
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.

29. Cohen, G. E. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions, *J. Appl. Crystallogr. 30*, 1160−1161.
30. Morris, G. M., D. S. G., Halliday, R. H., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem. 19*, 1639−1662.
31. Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr. D60*, 1355−1363.
32. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305−308.
33. Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D. M., Harduin-Lepers, A., Delannoy, P., Imberty, A., and Breton, C. (2004) Structure-function analysis of the human sialyltransferase ST3Gal I: role of *N*-glycosylation and a novel conserved sialylmotif, *J. Biol. Chem. 279*, 13461−13468.
34. Charnock, S. J., and Davies, G. J. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms, *Biochemistry 38*, 6380−6385.
35. Boix, E., Swaminathan, G. J., Zhang, Y., Natesh, R., Brew, K., and Acharya, K. R. (2001) Structure of UDP complex of UDP-galactose:beta-galactoside-alpha-1,3-galactosyltransferase at 1.53-Å resolution reveals a conformational change in the catalytically important C terminus, *J. Biol. Chem. 276*, 48608−48614.
36. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs, *Nat. Struct. Biol. 8*, 166−175.
37. Close, B. E., Mendiratta, S. S., Geiger, K. M., Broom, L. J., Ho, L. L., and Colley, K. J. (2003) The minimal structural domains required for neural cell adhesion molecule polysialylation by PST/ST8Sia IV and STX/ST8Sia II, *J. Biol. Chem. 278*, 30796−30805.
38. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol. 302*, 205−217.

BI602543D