NMR ACTIVE SUBSTRATES FOR STRUCTURE AND FUNCTION STUDIES OF SIALYLTRANSFERASES

by

SHAN LIU

(Under the Direction of James H. Prestegard)

ABSTRACT

Novel nuclear magnetic resonance (NMR) methods have been developed for structure and function studies of sialyltransferases interacting with synthetic substrates. The presented methodology has been applied to a mammalian sialyltransferase ST6Gal-1 which catalyzes the transfer of sialic acids to the terminal galactose of carbohydrates on the cell surface. Conventional structural methods have failed to produce a structure of ST6Gal-1 because of its membrane association, native glycosylation, and poor expression in bacterial hosts. A new methodology is urgently needed to provide a precise molecular structure of ST6Gal-1 and the geometries of bound ligands. The provided information can be used in the design of specific inhibitors to regulate the distribution of sialic acids in biological systems.

NMR is a unique tool to study a broad range of proteins that resist study by X-ray crystallography. The methodology takes advantage of isotopically-labeled and nitroxide-labeled substrates which can probe geometric properties of the active site of a protein. Protein-ligand interactions are usually the key to understanding the function of enzymes. This work investigated the conformation, orientation, kinetics, and protein-contact

surfaces of bound ligands. When nitroxide-labeled substrates and isotopically-labeled protein were utilized, five phenylalanines in and near the protein active site were identified and their distances to the substrates were calculated based on NMR data. This distance-sensitive nitroxide-labeled probe also was used to derive the relative placement of substrates simultaneously present in the active site of ST6Gal-1. The above approach has improved our structural understanding of ST6Gal-1 and provided an important step toward the structure-based design of efficient inhibitors. The synthetic strategies and synthesized compounds are applicable to other sialyltransferases, and the presented NMR labeling methods will be highly valuable for studying the structures of various glycoproteins.

INDEX WORDS: nuclear magnetic resonance (NMR), sialyltransferase (ST), α-2,6-sialyltransferase (ST6Gal-1), saturation transfer difference (STD), transferred nuclear Overhauser enhancement (trNOE), residual dipolar coupling (RDC), paramagnetic perturbation, spin label, nitroxide, tetramethylpiperidine-1-oxyl (TEMPO), cytidine-5'-monophospho-N-acetyl-neuraminic acid (CMP-NeuAc), N-acetyl-lactosamine (LacNAc)

NMR ACTIVE SUBSTRATES FOR STRUCTURE AND FUNCTION STUDIES OF SIALYLTRANSFERASES

by

SHAN LIU

B.S., Beijing University of Chemical Technology, China, 2003

A Dissertation Submitted to the Graduate Faculty of The University of Georgia

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

NMR ACTIVE SUBSTRATES FOR STRUCTURE AND FUNCTION STUDIES OF SIALYLTRANSFERASES

by

SHAN LIU

Major Professor: James H. Prestegard

Committee: George F. Majetich
Jeffrey L. Urbauer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

To my Father, my Mother, and my Grandma for supporting my education and always believing in me.

ACKNOWLEDGEMENTS

I thank my research advisor, Professor James Prestegard, who guided me through this project with intelligence, optimism, and patience for the past four years. I greatly appreciate the opportunities he provided for research, public speaking, writing, teaching, traveling, and career development. He is truly my American Idol with an exemplary character and it has been my great privilege to work under him. I also thank Professor Jeffrey Urbauer for introducing me to NMR, guiding me as my advisory committee member, and helping me with job applications. I thank my other committee member, Professor George Majetich, for training me in organic synthesis, teaching me English, and providing constructive criticism regarding my performances in seminars. I thank Dr. Lu Meng for her generous supply of ST6Gal-1 and helpful research discussion. I also thank Dr. Andre Venot for detailed instruction on organic synthesis and thoughtful handcrafted New Years cards. I thank the following group members of Dr. Prestegard's group not only for their help with my research but also for cherishing our friendship: Dr. Megan Macnaughtan for teaching me NMR experiments, revising my writing, and throwing parties; Dr. John Glushka for countless conversations about NMR instruments and experiments, career planning, and job applications; Iris Tropp for editing my resume and dissertation and offering advice throughout my dissertation-writing process; Dr. Xu Wang for teaching me NMR, biochemistry, software use, and molecular modeling on a daily basis; Dr. Yizhou Liu for NMR troubleshooting and his unique sense of humor; Dr. Fang Tian for interpreting NMR pulse sequences and assistance with NMR

instrumentation; Dr. Jacque Lee for sharing research ideas; Dr. Hans Lee for preparing a plentiful amount of galactokinase; Laura Morris for her help with ST6Gal-1 modeling, computer updating and troubleshooting; Wendy Nkari for being my faithful lunchtime partner and having fun social activities together; Fei Yu for discussions of NMR experiments and tolerating my allergy-triggered sneezing while sitting next to me in the lab; Dr. Tiandi Zhuang for discussing proteins and RDCs and his wife Hong Jiang for treating me to delicious food from time to time; Dr. Anita Kishore for teaching me HPLC, sharing lunchtimes, and giving me career advice; and Dr. Lianmei Feng for being like a big sister to me. I thank the administrative staff at the CCRC, especially Beverly Chalk. I thank the members of the Boons research group for their ideas about synthesis and lending me reagents and glassware. I also thank the dozen roommates I have had by living in five different apartments, and friends Lei Chen and Lin Lin, for making my time in Athens more enjoyable. Finally, I would like to thank Li-qing Jiang, my love and my best friend, for constantly cheering me on.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Structural investigation of sialyltransferases

Sialic acids are a group of nine-carbon monosaccharides bearing a negative charge at physiological pH values. They usually occur at the non-reducing terminal ends of carbohydrates on the cell surface. Because of their negative charge and wide occurrence in exposed positions of cell-surface molecules, sialic acids act as key determinants of glycoconjugates that mediate a variety of biological phenomena, such as cell-cell recognition, cell adhesion, and various receptor-ligand interactions [1; 2; 3]. The biosynthesis of sialylated glycoconjugates is catalyzed by a family of glycosyltransferases (GT) named sialyltransferases. To date, 20 members of mammalian sialyltransferases have had their cDNA cloned [4]. However, there is very little knowledge about the fundamental structure-function aspects of these sialyltransferases because of difficulties in expressing these membrane-associated proteins in eukaryotic systems. In the Carbohydrate-Active enZymes (CAZy) database, sialyltransferases are classified into families GT-29, GT-38, GT-42, GT-52 and GT-80 based on their sequence similarities [5]. The family GT-29 is the only family containing mammalian sialyltransferases. Currently four bacterial sialyltransferases from GT-42 (Cst I and Cst II) and GT-80 (Δ24PMST1 and Δ16psp26ST) have been characterized at the molecular level [6; 7; 8; 9; 10]. Unfortunately, bacterial sialyltransferases do not share overall sequence similarities with mammalian sialyltransferases which have several unique conserved sialylmotifs [11].

The research described in this dissertation investigates the structure and function of the mammalian alpha-2,6-sialyltransferase (ST6Gal-1) through novel nuclear magnetic resonance (NMR) methods. ST6Gal-1 is a member of the family 29 glycosyltransferases as defined by the CAZy database. ST6Gal-1 catalyzes the transfer of sialic acid residues with an α-2, 6- linkage to terminal galactose residues of type 2 (Galβ1-4GlcNAc)-containing oligosaccharides and glycans of glycolipids and glycoprotein on mammalian cell surface. Variation in the amount of sialic acid incorporated during this transfer is correlated with the status of colon cancer and brain tumors and immune regulation in the human body [12; 13]. Despite its significant biological role, the precise molecular structure of ST6Gal-1 is poorly understood. Its native glycosylation, status as membrane protein, and poor expression in bacterial hosts, have contributed to the lack of structural characterization. It would be important to know the structure of ST6Gal-1, particularly the active site, in order to design efficient inhibitors to regulate the distribution of sialic acids on mammalian cell surface. In this dissertation novel NMR methods combined with synthetic strategies have been developed to determine the binding geometries of the ligands and map the active site of ST6Gal-1.

1.2 Limited methods for structural analysis

Protein structures are most commonly provided by X-ray crystallography (Table 1.1). However, there are some systems that resist crystallization. Glycoproteins, including sialyltransferases, are relatively intractable for crystallization because of their flexible and heterogeneous glycan chains and variability in glycosylation site occupancy [14; 15]. One indicator of the difficulty is occurrence in the RCSB protein data bank (PDB). When the keyword “glycoprotein” is searched in the PDB, only 1686 hits were found out

of 41687 protein structures as of Aug 12, 2008. An even smaller number of structures actually give coordinates for attached glycans.

Table 1.1. Current released entries from PDB Statistics as of Aug 12, 2008

Experimental Method	Molecule Type				
	Proteins	Nucleic Acids (NA)	Protein/NA complexes	Others	Total
X-ray	41687	1067	1913	24	44691
NMR	6465	814	138	7	7424
Electron Microscopy	129	11	47	0	187
Others	90	4	4	2	100
Total	48371	1896	2102	33	52402

NMR is a tool uniquely suited for the study of glycoproteins and the broad range of systems that are difficult to crystallize. NMR determines both the structures and dynamics of biomolecules and complements the static pictures obtained by X-ray crystallography [16; 17; 18]. With NMR, biomolecules can be studied in aqueous solution where the native function of the protein takes place, and molecular mobility can be detected at an atomic level. NMR experiments can provide site-specific information about protein motion spanning a range of time scales, and this information can shed light on the kinetics of protein-ligand and protein-protein interactions as well as structures [19; 20; 21; 22]. Structures solved by NMR are based on time-averaged information on molecules, but when motional amplitudes are small, NMR experiments yield a set of restraints on geometrical parameters like angles of bonds and distances between pairs of NMR active spins that can be used to define the structure of a biomolecule.

Currently, 14% of the protein structures in the PDB have been determined by NMR, yet the capabilities and applications of NMR continue to evolve. When the new resolution enhancing experiment, Transverse Relaxation Optimized Spectroscopy

(TROSY) was introduced, it broke the traditional size barrier (40 KDa) and made it possible to assign resonances on proteins in the 100 KDa range [23]. In the year of 2000 an NMR-derived structure for an 82 kDa-MW protein was determined using a series of 4D TROSY-based triple resonance experiments along with a 4D NOE data set [24]. The sensitivity of NMR has also been improving tremendously over the past dozen of years. Expanded NMR theories, uniform/sparse isotopic labeling, high-temperature superconductive detection coils, cold-probe technology, higher magnetic field strength (21 Tesla in the 900 MHz spectrometers), and techniques for sample preparation, have, in combination, enhanced the precision and accuracy of NMR measurements, and contributed to the increased popularity of applying NMR to biomolecular samples.

Traditionally, the NMR structure determination of proteins has relied on the detection of large numbers of nuclear Overhauser enhancement (NOE). Uniform labeling with ^{13}C, ^{15}N, and ^{2}H has provided the improved resolution necessary for the assignment of sufficient numbers of resonances needed to conduct an NOE-based structure determinations [25; 26]. This type of labeling is usually accomplished by expressing proteins in bacterial hosts which efficiently utilize low-cost ^{13}C- and ^{15}N-enriched substrates. However, there are many proteins for which bacterial expression is not practical. For example, proteins that require glycosylation for proper folding or full activity are normally expressed in eukaryotic hosts that require supplementation with a nearly complete set of labeled amino acids, some of which are very expensive [27; 28]. This is not a small class of proteins. Approximately 50% of mammalian proteins may be glycosylated [29]. Therefore, it is necessary to consider alternate sources of structural data that can be obtained with unlabeled proteins, or proteins with the sparse labels introduced by

expressing proteins in eukaryotic cells grown in media supplemented with selected sets of isotopically-labeled amino acids [30].

1.3 NMR strategies and theory

1.3.1 Structural characterization of substrates bound to proteins with fluorinated and isotopically-labeled NMR probes

Generally speaking, protein-ligand interactions are the key to understanding the biological function of proteins. However, understanding these interactions requires a precise knowledge about the binding events at an atomic level. Some NMR Experiments are designed for this purpose. They include the transferred nuclear Overhauser enhancement (trNOE), saturation transfer difference (STD), and residual dipolar coupling (RDC) measurements. These experiments can be used to observe the resonance signals of ligands and deliver a geometric map that describes how ligands bind to a protein receptor. NMR can also provide insights into the binding strength and kinetics of these ligands in protein complexes [31].

STD and trNOE experiments are designed to monitor the resonances of ligands rather than those of proteins which are difficult to study if the proteins are large or if they are not isotopically labeled. These types of experiments do not require isotopically labeled samples and use only substoichiometric amounts of target proteins, such as 50 μM of protein (with 1 mM of ligand).

The STD technique was first applied to study a nucleotide ligand binding to a muscle protein myosin (MW 470 KDa) in 1979, and not until the late 1990s, was this technique widely used in drug discovery [32; 33]. The principle behind STD is simple. The saturation of the protein is transferred to the bound ligand(s) which is exchanged into

solution where its saturation is detected. For macromolecules, proton magnetization is very efficiently transferred throughout the molecule itself and molecules in close proximity. Therefore, if some part of the protein is selectively saturated (i.e. the magnetization is attenuated), this saturation will be efficiently transferred throughout the whole protein and also to the ligands located within the protein binding site. The following figure 1.1 further illustrates the rationale of STD experiments. A train of Gaussian-shaped pulses are applied to a small region of protein resonances away from ligand resonances. The regional saturation quickly diffuses through the whole protein as well as the ligand in contact with the protein. If the ligand does not interact with the protein, its resonances are not affected by the protein saturation. Depending on the proximity of the ligand protons to the protein, there are saturation differences within the ligand molecule. The bound ligand carrying this information is exchanged into solution where it is detected by an NMR spectrometer. By subtracting a spectrum with irradiation at a frequency well off all proton resonances, an NMR spectrum is obtained in which the intensities of ligand signals are reduced in proportion to the proximities of ligand protons to protons on the protein surface.

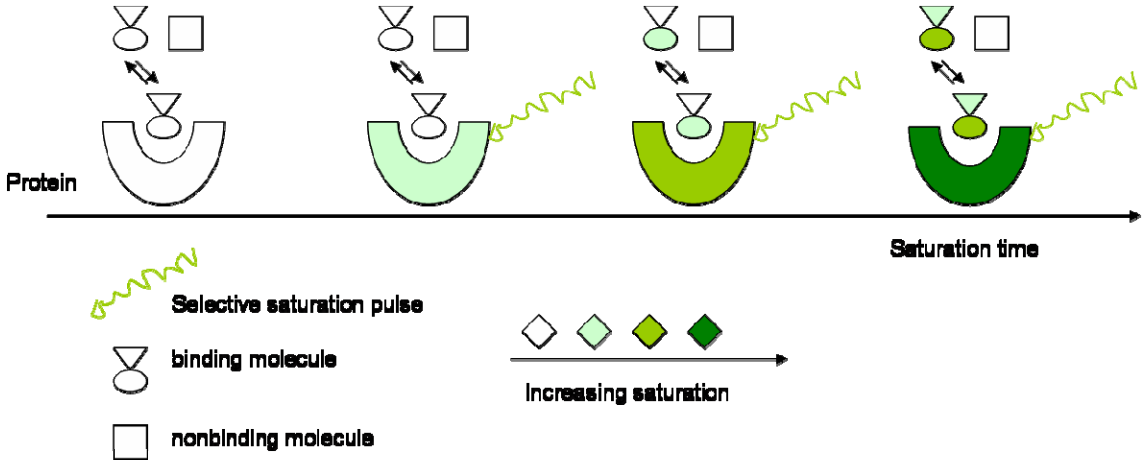


Figure 1.1. Illustration of the STD process.

The principles of trNOE were first described thirty years ago [34; 35] and since then, the trNOE method has become a commonly used tool for determining structures of the bound ligand in protein complexes [31]. Transferred NOE experiments demonstrate the substrates' binding to the protein target and measure the intramolecular distances of substrates in their bound state. Transferred NOE experiments are performed through a selective 1D version and zero-quantum filtered 2D NOESY with mixing time of 50 ms-2 s. They are basically regular NOESY experiments but applied to a protein-ligand system in dynamic exchange in which the ligand is present in excess (10-30 X). The theory underlying trNOE experiments is that NOEs from large and small molecules have opposite signs with the process establishing the NOE for the large molecule being far more efficient. Low- or medium-molecular-weight molecules (MW 1000-2000), including the substrates used for our ST6Gal-1 studies, have a short correlation time (τ_c). They exhibit positive NOEs, zero NOEs, or very small negative NOEs depending on the molecular weight, the shape of the molecule, and the magnetic field strength. Large molecules, however, exhibit strongly negative NOEs because of their long correlation time. This accounts for the driving force for the transfer of the NOE when a small-molecule ligand is bound to a large molecule (e.g. protein receptor). The bound ligand behaves as part of the large molecule and adopts the corresponding strong negative NOEs, in other words, the trNOEs.

The ligands are not tightly bound and they exchange fast ($K_D > 10^{-6}$) with the excess ligands free in solution. The free ligands maintain the information acquired in the bound state for a given period of time (determined by its relaxation time) and their signals are detected before the system reverts to its initial state. Consequently, the

binding of a ligand to a protein receptor can be easily distinguished by comparing the sign and size of NOEs in the spectra for a free ligand and a ligand-protein system. The intensities of crosspeaks in 2D NOESY spectra can be converted to distance constraints between pairs of ligand protons and these can be used to optimize its conformation in bound state. The observed NOE is the weighted average of the free and bound states. However, because of the linear dependence of the magnetization transfer rates on correlation times for large systems, the NOEs from the bound state still dominate the observed average with the strong negative NOEs adopted from the protein. Spin diffusion and on-off rates can complicate the interpretation of trNOE data as reported in Chapter 2, but in balance, trNOEs are very powerful structural probes.

Traditional NOE experiments may not yield adequate data to completely define the bound geometry of a ligand because NOEs can be detected only for protons lying within 5 Å [23]. When the ligand is a carbohydrate, the number of NOE constraints can be further reduced because the extensive hydrogen-bonding network mediates interactions with the ligand. For example, in the case of our sugar nucleotide, no NOEs were observed between the sialic acid protons and the nitrogenous base or ribose protons. Therefore, we rely on the third method, the measurement of residual dipolar couplings (RDCs) of the ligand, to obtain direct long-range angular orientation information complementary to NOE-derived distance constraints. Here is an equation illustrating the theory of RDC [36].

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \qquad (1)$$

Equation 1 shows how the dipolar couplings between pairs of spin ½ nuclei are dependent upon θ, the angle between the magnetic field and the spin–spin interaction vector. The one-bond distance between interacting spins, r, is assumed fixed by bonding geometry and known. Therefore, θ becomes the primary variable. The dipolar coupling averages to zero in isotropic solutions due to random tumbling of the molecules but, in an NMR spectrum acquired in aligned media, it appears as additional contributions to the splittings of the resonance lines, analogous to scalar (J) couplings. By measuring D_{ij}, for sufficient vectors in a geometrically well-defined fragment such as our sialic acid or nucleotide, we can determine the orientation of each fragment in the coordinate system of a common alignment frame. Therefore, bond-vector orientations located at opposite ends of a molecule will be related to each other through a common alignment frame, resulting in long-range angular orientation information. Weakly-aligned media, usually liquid crystalline, can be used to make RDCs observable in solution. Typical aqueous media used for this purpose are phospholipid bicelles, alkyl-ethyleneglycol-alcohol bilayers (e.g. C12E5-hexanol) and bacteriophage [37; 38; 39]. In these aligned media, the solute molecules experience anisotropic orientational distribution through collisional or electrostatic interactions which keep their dipolar coupling from averaging to zero. The collected coupling data can be decomposed to give information regarding the strength of alignment and orientation of a molecular fragment. The extracted information can be used for orienting different rigid components within a molecule, or studying the relative orientation of the molecules in complexes such as a ligand bound to a protein. A novel approach described in Chapter 2 involves the application of RDCs to an isotopically labeled substrate.

Fluorine NMR has been used extensively on ligand-protein complexes for substrate screening [40], stoichiometry measurement [41], and binding affinity assessment [42]. Subsequent data analysis with linear or curve fitting also can yield the inhibitor dissociation constant K_I, an essential parameter for evaluating and comparing systems [43]. The following factors contribute to the merits of ^{19}F NMR observation. ^{19}F is a spin ½ nucleus with a 100% natural abundance, a gyromagnetic ratio 94% of that for protons, and a sensitivity 83% of that for protons. These three prosperities make the detection of ^{19}F NMR as easy as that for standard proton NMR. A fluorine nucleus is a second row element that has a partially-occupied orbital with non-zero angular momentum and therefore, fluorines in molecules have a large chemical shift range compared to protons (~300 ppm for organofluorine) and a high sensitivity to the local chemical environment. Moreover, ^{19}F does not occur naturally in proteins and thus ^{19}F NMR allows selective observation of ligands without background interference.

1.3.2. Characterization of the protein active sites with spin-labeled NMR probes

Proteins with sparse isotope labels can be produced in eukaryotic hosts by using isotope-labeled forms of specific amino acids, but normally labeled sites are then not close to one another. Therefore structural analysis requires information from experiments other than nuclear Overhauser effects. One source applicable to this type of labeling relies on paramagnetic perturbation of labeled sites with nitroxide carrying probes that are either substrate mimics or are covalently attached to the protein itself.

There have been a number of examples using paramagnetic constraints to examine distances between substrates and protein residues detected by NMR [44; 45; 46; 47; 48]. The spin-labeled moiety that we chose is the commonly used and chemically stable

TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl). Its nitroxide group contains an unpaired electron that drastically increases the relaxation rate of protons and other magnetic nuclei. This effect is relatively long-range because the electron magnetic moment is 658 times larger than the proton magnetic moment and even larger when compared to most other nuclei. This effect is also easily measured; since the spin relaxation times are shortened (in particular T_2), the protein resonances are broadened and the crosspeak intensity in most multidimensional experiments is reduced. The general procedure for measuring paramagnetic relaxation enhancement is to first acquire a spectrum of protein with spin-labeled ligand in its oxidized form, and then again after reduction of the nitroxide radical with ascorbate. The peak intensity differences in the oxidized and reduced spectra should be directly attributable to the paramagnetic relaxation enhancement effect and simply related to r^{-6}. These effects are observable up to 20Å or more in most protein ^{1}H-^{15}N HSQC experiments, which allows distances between the nitroxide spin and nuclei giving rise to the cross-peaks to be calculated for sites throughout the protein.

1.4 Application to sialyltransferases and their ligands

All mammalian sialyltransferases are type II transmembrane glycoproteins with short N-terminal cytoplasmic tails (which are not essential for catalytic activity), a 16-20 amino acid transmembrane domain, a stem region (20 to 200 amino acids), and a large C-terminal catalytic domain [49]. The overall amino acid length of these enzymes varies from 300 to 600 residues. The catalytic domain of mammalian sialyltransferases contains three highly conserved amino acid sequences termed sialylmotifs L (large), S (small), and VS (very small). The L motif participates in the binding of the sugar donor CMP-NeuAc; S motif participates in the binding of both donor and acceptor substrates; VS is also

involved in catalytic activity. A new motif (III) existing between S and VS motifs has been reported [4]. The involvement of certain residues within the L motif has been clearly demonstrated and this information is very useful considering all sialyltransferases share the same donor CMP-NeuAc. However, sialyltransferases use different acceptors. Therefore, a common S motif involved in the binding of acceptors is not expected.

The catalytic domain of ST6Gal-1 was cloned and expressed in mammalian cells as early as 1987 [50], and it is this domain that we focus on here. Whereas refolding of small amounts of insoluble proteins expressed in *E. coli* has been reported recently [51], mammalian cell expression remains the preferred means of producing characterizable amount of active sialyltransferases. The ST6Gal protein used in this research was expressed as a soluble enzyme, lacking the cytoplasmic tail and transmembrane domain (Figure 1.2).

> ST6Gal-1_rat HEK293 cells (truncated from membrane)

KSMHHHHHHHHKDPSTYSKLNPRLLKIWRNYLNMNKYKVSYKGPGPGVKFSV
EALRCHLRDHVNVSMIEATDFPFNTTEWEGYLPKENFRTKVGPWQRCAVVSSA
GSLKNSQLGREIDNHDAVLRFNGAPTDNFQQDVGSKTTIRLMNSQLVTTEKRFL
KDSLYTEGILIVWDPSVYHADIPKWYQKPDYNFFETYKSYRRLNPSQPFYILKPQ
MPWELWDIIQEISADLIQPNPPSSGMLGIIIMMTLCDQVDIYEFLPSKRKTDVCYY
HQKFFDSACTMGAYHPLLFEKNMVKHLNEGTDEDIYLFGKATLSGFRNIRC

Figure 1.2. The catalytic domain sequence of ST6Gal-1. The domain has totally 321 amino acids, giving it a molecular mass of 38 kDa. F- phenylalanine, 16 of which were labeled in ^{15}N as discribed in Chapter 3. The conserved sialylmotifs: L-motif; S-motif; III-motif; VS-motif.

So far only four sialyltransferases of bacterial origin have been deposited in the RCSB Protein Data Bank. They include the 2,3-/2,8-sialyltransferase (CstII) from *C. jejuni* (OH4384) [7], the 2,3-sialyltransferase (Cst-I) from *C. jejuni* [6], the 2,3/6-sialyltransferase (Δ24PmST1) from *P. multocida* strain P-1059 [9; 10] and, most recently, the 2,6-sialyltransferase (JT-ISH-224) from *Photobacterium* sp. [8]. These sialyltransferase structures have been produced in complex with the inactive donor analog, CMP3FNeuAc, or the product, CMP; in two cases a potential acceptor, lactose, also has been included. Unfortunately, there are no overall sequence similarities between these bacterial sialyltransferases and the mammalian sialyltransferases. The sequence identities for ST6Gal-1 to CstII, Cst-I, Δ24PmST1 and JT-ISH-224 are 11%, 14%, 15% and 16% respectively, which makes it questionable to use these structures in the investigation of the structure and mechanism of our target ST6Gal-1.

A long-term goal for NMR studies of ST6Gal-1 would be to provide a detailed structure of ST6Gal-1, but even without this structure, substantial and significant information about this protein can be obtained by locating the active site of ST6Gal-1 and characterizing the bound geometry of its donors and acceptors and the geometric relation of these in their bound state. In this dissertation I investigated the structure and function of ST6Gal-1 and its associated ligands using NMR active substrates. The NMR methods employed in this research focus on bound ligands rather than the complete protein structure and provided an indirect view of the active site of ST6Gal-1.

Chapter 2 describes the structural characterization of substrates bound to ST6Gal-1 using fluorinated and isotopically-labeled probes. Ligand binding is critical for the catalytic function of proteins. However, no information is currently available on the

binding geometry and kinetics for the substrates of ST6Gal-1. Also, the differences in reactions catalyzed by the bacterial sialyltransferases, and their broader acceptor specificities, make it risky to draw conclusions from the positions of substrates in the bacterial enzymes about active site geometries for ST6Gal-1. In response to this limitation, we performed trNOE experiments to determine bound geometries of some ligands, STD experiments to identify ligand-protein contact surfaces, RDC measurements to calculate the orientation of ligand components using a synthesized ^{13}C-labeled donor, and ^{19}F NMR studies to assess the binding kinetics of a fluorinated donor analog.

Chapter 3 addresses the characterization of the active site of ST6Gal-1 using spin-labeled probes. The paramagnetic TEMPO derivative perturbs NMR resonances of the protein in a manner dependent on the inverse sixth power of the distance between the unpaired electron of the nitroxide moiety in TEMPO and the nuclei giving rise to the NMR resonances [52; 53]. The synthesized spin-labeled substrates were demonstrated to be able to identify resonances from active site residues and produce distance constraints on a ^{15}N phenylalanine labeled version of ST6Gal-1. Sixteen NMR observable sites resulted from expressing ST6Gal-1 in mammalian HEK 293 cells grown in media containing ^{15}N labeled phenylalanine, and the phenylalanine sites were ranked based on their distance to the labeled donor ligand [47]. Then related paramagnetic strategies were used to locate the position of the sugar acceptor relative to the donor analog (labeled in ^{13}C) while both were simultaneously bound to ST6Gal-1.

The results of *in vivo* labeling of sialyltransferase products using a ^{13}C-labeled donor are presented in Chapter 4. Gangliosides are the major sialic acid-bearing glycoconjugates in the nerve cells and the brain [54]. They act as receptors for

microorganisms and bacterial toxins, regulate cell growth and differentiation, and participate in cell-cell and cell-matrix interactions [55]. By developing methods for detecting gangliosides *in vivo,* the accumulation of gangliosides related to a number of diseases can be monitored. For this study, neural stem cells were chosen as the target because they produce a significant number of gangliosides during their growth. Peracetylated sialic acids enriched in ^{13}C were synthesized and then fed to stem cells during cultivation. After harvesting the cells, NMR experiments were conducted to detect generated intermediates *in vivo*. Then the gangliosides were extracted, separated, and purified, and a quantitation of sialic acid incorporation was attempted using NMR and mass spectrometry.

Chapter 6 summarizes the results of these experiments and their implications for studying the structure and function of sialyltransferases and other glycosyltransferases.

References

1. Angata, T. & Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chemical Reviews* **102**, 439-469.

2. Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M. A., Samyn-Petit, B., Julien, S. & Delannoy, P. (2001). The human sialyltransferase family. *Biochimie* **83**, 727-737.

3. Kelm, S. & Schauer, R. (1997). Sialic acids in molecular and cellular interactions. *International Review of Cytology - a Survey of Cell Biology, Vol 175* **175**, 137-240.

4. Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D. M., Harduin-Lepers, A., Delannoy, P., Imberty, A. & Breton, C. (2004). Structure-function analysis of the

human sialyltransferase ST3Gal I - Role of N-glycosylation and a novel conserved sialylmotif. *Journal of Biological Chemistry* **279**, 13461-13468.

5. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology* **328**, 307-317.

6. Chiu, C. P. C., Lairson, L. L., Gilbert, M., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. J. (2007). Structural analysis of the alpha-2,3-sialyltransferase cst-i from Campylobacter jejuni in apo and substrate-analogue bound forms. *Biochemistry* **46**, 7196-7204.

7. Chiu, C. P. C., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. J. (2004). Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. *Nature Structural & Molecular Biology* **11**, 163-170.

8. Kakuta, Y., Okino, N., Kajiwara, H., Ichikawa, M., Takakura, Y., Ito, M. & Yamamoto, T. (2008). Crystal structure of Vibrionaceae Photobacterium sp JT-ISH-224 alpha 2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition. *Glycobiology* **18**, 66-73.

9. Ni, L. S., Chokhawala, H. A., Cao, H. Z., Henning, R., Ng, L., Huang, S. S., Yu, H., Chen, X. & Fisher, A. J. (2007). Crystal structures of Pasteurella multocida sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* **46**, 6288-6298.

10. Ni, L. S., Sung, M. C., Yu, H., Chokhawala, H., Chen, X. & Fisher, A. J. (2006). Cytidine 5 '-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from Pasteurella multocida. *Biochemistry* **45**, 2139-2148.

11. Takashima, S. (2008). Characterization of mouse sialyltransferase genes: Their evolution and diversity. *Bioscience Biotechnology and Biochemistry* **72**, 1155-1167.

12. Dall'Olio, F. & Chiricolo, M. (2001). Sialyltransferases in cancer. *Glycoconjugate Journal* **18**, 841-850.

13. Hennet, T., Chui, D., Paulson, J. C. & Marth, J. D. (1998). Immune regulation by the ST6Gal sialyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4504-4509.

14. Lopez-Jaramillo, F. J., Perez-Banderas, R., Hernandez-Mateo, F. & Santoyo-Gonzalez, F. (2005). Production, crystallization and X-ray characterization of chemically glycosylated hen egg-white lysozyme. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* **61**, 435-438.

15. Baker, H. M., Day, C. L., Norris, G. E. & Baker, E. N. (1994). Enzymatic Deglycosylation as a Tool for Crystallization of Mammalian Binding-Proteins. *Acta Crystallographica Section D-Biological Crystallography* **50**, 380-384.

16. Korzhnev, D. M. & Kay, L. E. (2008). Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts of Chemical Research* **41**, 442-451.

17. Malliavin, T. E. (2006). Quantitative analysis of biomolecular NMR spectra: A prerequisite for the determination of the structure and dynamics of biomolecules. *Current Organic Chemistry* **10**, 555-568.

18. Mittermaier, A. & Kay, L. E. (2006). Review - New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224-228.

19. Datta, A. K., Sinha, A. & Paulson, J. C. (1998). Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates. *Journal of Biological Chemistry* **273**, 9608-9614.

20. Fielding, L. (2007). NMR methods for the determination of protein-ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* **51**, 219-242.

21. Marintchev, A., Frueh, D. & Wagner, G. (2007). NMR methods for studying protein-protein interactions involved in translation initiation. *Translation Initiation: Reconstituted Systems and Biophysical Methods* **430**, 283-+.

22. Pellecchia, M. (2005). Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. *Chemistry & Biology* **12**, 961-971.

23. Tugarinov, V., Hwang, P. M. & Kay, L. E. (2004). Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annual Review of Biochemistry* **73**, 107-146.

24. Tugarinov, V., Kay, L. E., Ibraghimov, I. & Orekhov, V. Y. (2005). High-resolution four-dimensional H-1-C-13 NOE spectroscopy using methyl-TROSY,

sparse data acquisition, and multidimensional decomposition. *Journal of the American Chemical Society* **127**, 2767-2775.

25. Moseley, H. N. B., Monleon, D. & Montelione, G. T. (2001). Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B* **339**, 91-108.

26. Wider, G. (2005). NMR techniques used with very large biological macromolecules in solution. *Nuclear Magnetic Resonance of Biological Macromolecules, Part C* **394**, 382-398.

27. Lustbader, J. W., Birken, S., Pollak, S., Pound, A., Chait, B. T., Mirza, U. A., Ramnarain, S., Canfield, R. E. & Brown, J. M. (1996). Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: An advance toward rapid determination of glycoprotein structures. *Journal of Biomolecular NMR* **7**, 295-304.

28. Mao, H. Y., Gunasekera, A. H. & Fesik, S. W. (2000). Expression, refolding, and isotopic labeling of human serum albumin domains for NMR spectroscopy. *Protein Expression and Purification* **20**, 492-499.

29. Varki, A. (1993). Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology* **3**, 97-130.

30. Bruggert, M., Rehm, T., Shanker, S., Georgescu, J. & Holak, T. A. (2003). A novel medium for expression of proteins selectively labeled with N-15-amino acids in Spodoptera frugiperda (Sf9) insect cells. *Journal of Biomolecular Nmr* **25**, 335-348.

31. Meyer, B. & Peters, T. (2003). NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie-International Edition* **42**, 864-890.

32. Akasaka, Y., Kizu, M., Aoike, A. & Kawai, K. (1979). Endoscopic Management of Acute Gastric Anisakiasis. *Endoscopy* **11**, 158-162.

33. Mayer, M. & Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **38**, 1784-1788.

34. Balaram, P., Bothnerb.Aa & Breslow, E. (1972). Localization of Tyrosine at Binding-Site of Neurophysin Ii by Negative Nuclear Overhauser Effects. *Journal of the American Chemical Society* **94**, 4017-&.

35. Balaram, P., Bothnerb.Aa & Dadok, J. (1972). Negative Nuclear Overhauser Effects as Probes of Macromolecular Structure. *Journal of the American Chemical Society* **94**, 4015-&.

36. Prestegard, J. H., Al-Hashimi, H. M. & Tolman, J. R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quarterly Reviews of Biophysics* **33**, 371-424.

37. Gaemers, S. & Bax, A. (2001). Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. *Journal of the American Chemical Society* **123**, 12343-12352.

38. Prestegard, J. H., Bougault, C. M. & Kishore, A. I. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews* **104**, 3519-3540.

39. Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society* **122**, 7793-7797.

40. Tengel, T., Fex, T., Emtenas, H., Almqvist, F., Sethson, I. & Kihlberg, J. (2004). Use of F-19 NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Organic & Biomolecular Chemistry* **2**, 725-731.

41. Gerig, J. T. (1998). Fluorine NMR. *Biophysics Textbook On-line* **NMR section chapter 24**, 23-24.

42. Peng, J. W. (2001). Cross-correlated F-19 relaxation measurements for the study of fluorinated ligand-receptor interactions. *Journal of Magnetic Resonance* **153**, 32-47.

43. Fielding, L. (2003). NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry* **3**, 39-53.

44. Jain, N. U., Venot, A., Umemoto, K., Leffler, H. & Prestegard, J. H. (2001). Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands. *Protein Science* **10**, 2393-2400.

45. Johnson, P. E., Brun, E., MacKenzie, L. F., Withers, S. G. & McIntosh, L. P. (1999). The cellulose-binding domains from Cellulomonas fimi beta-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. *Journal of Molecular Biology* **287**, 609-625.

46. Kuliopulos, A., Westbrook, E. M., Talalay, P. & Mildvan, A. S. (1987). Positioning of a Spin-Labeled Substrate-Analog into the Structure of Delta-5-3-

Ketosteroid Isomerase by Combined Kinetic, Magnetic-Resonance, and X-Ray-Diffraction Methods. *Biochemistry* **26**, 3927-3937.

47. Liu, S., Venot, A., Meng, L., Tian, F., Moremen, K. W., Boons, G. J. & Prestegard, J. H. (2007). Spin-labeled analogs of CMP-NeuAc as NMR probes of the alpha-2,6-sialyltransferase ST6Gal I. *Chemistry & Biology* **14**, 409-418.

48. Macnaughtan, M. A., Kamar, M., Alvarez-Manilla, G., Venot, A., Glushka, J., Pierce, J. M. & Prestegard, J. H. (2007). NMR structural characterization of substrates bound to N-acetylglucosaminyltransferase V. *Journal of Molecular Biology* **366**, 1266-1281.

49. Paulson, J. C. & Colley, K. J. (1989). Glycosyltransferases - Structure, Localization, and Control of Cell Type-Specific Glycosylation. *Journal of Biological Chemistry* **264**, 17615-17618.

50. Weinstein, J., Lee, E. U., Mcentee, K., Lai, P. H. & Paulson, J. C. (1987). Primary Structure of Beta-Galactoside Alpha-2,6-Sialyltransferase - Conversion of Membrane-Bound Enzyme to Soluble Forms by Cleavage of the Nh2-Terminal Signal Anchor. *Journal of Biological Chemistry* **262**, 17735-17743.

51. Hamamoto, T., Kurosawa, N., Nakaoka, T., Lee, Y. C., Kawasaki, M. & Tsuji, S. (1993). Expression of Mouse Beta-Galactoside Alpha-2,6-Sialyltransferase in an Insoluble Form in Escherichia-Coli and Partial Renaturation. *Glycoconjugate Journal* **10**, 236-237.

52. Berliner, L. J. (1976). *Spin labeling theory and applications*, Academic Press, New York.

53. Likhtenshtein, G. I. (1976). *Spin labeling methods in molecular biology*, Chapter 2, Wiley, New York.

54. Vyas, A. A. & Schnaar, R. L. (2001). Brain gangliosides: Functional ligands for myelin stability and the control of nerve regeneration. *Biochimie* **83**, 677-682.

55. Lloyd, K. O. & Furukawa, K. (1998). Biosynthesis and functions of gangliosides: recent advances. *Glycoconjugate Journal* **15**, 627-636.

CHAPTER 2

GEOMETRIES AND KINETICS STUDIES OF BOUND SUBSTRATES BY FLUORINATED AND ^{13}C-LABELED PROBES

Liu, S., L. Meng, K. W. Moremen, and J. H. Prestegard. To be submitted to *Biochemistry*

2.1 Introduction

Defining the geometry and kinetics of ligands bound to proteins provides a useful starting point for studying protein function and designing drug molecules that can alter protein function. Nuclear magnetic resonance (NMR) methods that focus on the properties of bound ligands rather than complete protein structures can provide a view of some of the most critical regions of protein systems, including the active sites of enzymes[1]. Traditional transferred NOE (trNOE) and saturation transfer difference (STD) methods can afford the active site geometry without the need for a structure determination of the protein. Residual dipolar coupling (RDC) measurements of ligands can provide a relatively independent set of long-distance structural constraints for determining bound ligand geometry. Fluorine NMR is a convenient and powerful tool to assess the kinetics and binding affinity of substrates with the incorporation of fluorinated ligands. In the following experiments, we employed these NMR methods to gain a first insight into the geometry and kinetics of substrates positioned in the active site of the sialyltransferase ST6Gal-1.

Transferred NOE and STD experiments are common methods for determining bound conformations and binding motifs of substrates bound to protein [1; 2; 3; 4]. Transferred NOE experiments are able to demonstrate a substrate's binding to a macromolecule and provide distance restraints between proximate pairs of protons in the substrate itself. When ligand molecules bind to large receptor proteins, NOEs undergo drastic changes leading to the observation of strong negative trNOEs as opposed to the normally weak and positive NOEs of small ligands. Because of the dominance of bound-ligand properties in the measured NOE, these negative NOEs can be converted to

distance restraints for proximate pairs of protons on the ligand. In STD experiments, the binding epitope of the substrates can be determined by measuring saturation transfer differences from the protein to the individual protons of the substrates when the protein is saturated. In 1D ^{1}H experiments, a cascade of Gaussian-shaped pulses is applied to the aliphatic protons of a protein. The resultant regional saturation quickly diffuses through over 70% of the protein by intramolecular spin diffusion and then to the ligand protons by intermolecular spin diffusion. Due to fast exchange the saturated ligands are transferred into solution where they are detected. By subtracting this spectrum from a spectrum with the protein irradiated at a frequency well off proton resonances, an NMR spectrum is obtained in which the intensities of ligand signals are reduced in proportion to the proximity of ligand protons to protons on the protein surface.

The trNOE and STD experiments were conducted to identify binding geometries and binding epitopes of donor substrates of ST6Gal-1 and its acceptor LacNAc (structures shown in Figure 2.1). When performing trNOE and STD experiments on the natural donor CMP-NeuAc, we noticed that CMP-NeuAc could be hydrolyzed into CMP and NeuAc at a much faster rate when ST6Gal-1 was present. To overcome the problem of CMP-NeuAc hydrolysis and stabilize the conformational states of the enzyme, the donor mimic CMP-3FNeuAc was synthesized and used as a non-hydrolysable competitive inhibitor of sialyltransferases. The CMP-3FNeuAc proved to be a useful probe for the catalytic mechanism of sialyltransferases Cst I, Cst II and Δ24PMST1. Here, CMP-3FNeuAc was prepared in a two-step enzymatic synthesis from N-acetylmannosamine (ManNAc), 3-fluropyruvate, and cytidine triphosphate (CTP) in overall 77% yield. CMP-3FNeuAc was synthesized using CMP-NeuAc synthetase to

react CTP with 3-fluoroNeuAc which was prepared from the reaction of 3-fluoropyruvate and ManNAc catalyzed by NeuAc aldolase [5]. The purification of CMP-3FNeuAc was done by centrifugation, ethanol precipitation [6] and size exclusion chromatography.

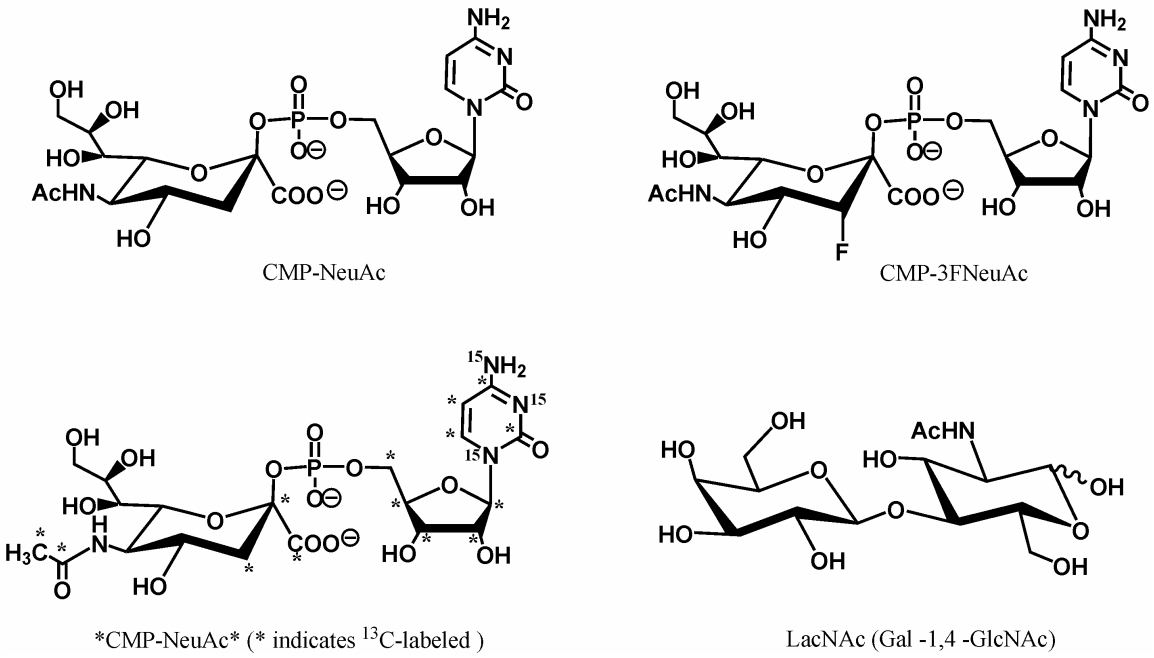


Figure 2.1. Structures of donor and acceptor substrates of ST6Gal-1

A third method used for the characterization of ligand geometry is residual dipolar coupling measurements of the ligand. NOEs can be detected only for protons lying within 5 Å [7], and when the ligand is a carbohydrate, or a sugar nuceotide donor, the number of NOE constraints across the glycosidic or phosphoester bond can be small. RDCs can provide the direct long-range angular orientation information complementary to trNOE distance restraints.

There have been a few examples of RDC measurements for fast exchanging ligands in protein complexes [8; 9; 10; 11; 12; 13]. Under ideal situations, RDCs can afford orientational information on both a protein and its ligands, namely, the relative

orientations of ligand and protein. Unlike NOEs, RDCs from the bound ligand (RDC_b) do not dominate the observed-average (RDC_{obs}). The observed RDC is the weighted average of the respective RDCs according to molar fractions in the free (F_f) and bound states (F_b; $1-F_f$), as displayed in Equation 2.

$$RDC_{obs} = F_f * RDC_f + F_b * RDC_b \quad (2)$$

Given a dissociation constant of 0.4 mM [14] for CMP-NeuAc-ST6Gal-1, we would expect about 14% of the ligand to be bound at the protein concentration used. With such a low F_b, it is clearly difficult to separate RDC_b from RDC_{obs} unless RDC_b is much larger than RDC_f . Three approaches based on protein alignment enhancement have been presented to make RDC_b larger than RDC_f [10; 12; 13]. The reported enhanced RDC_b for the ligand of Galectin-3 has been shown to be consistent with the known bound-ligand geometries. These three strategies will be applied to ST6Gal-1 to obtain the geometries of ligands bound to the protein once suitable constructs can be prepared. The procedure and ability of using RDCs to obtain the free ligand orientation is demonstrated here.

Ligand orientation and internal geometry can be determined from one bond ^{1}H-^{13}C and ^{13}C-^{13}C RDCs acquired with an f1-coupled ^{1}H-^{13}C HSQC. Obtaining these data would be greatly facilitated by using isotopically labeled ligands. For this purpose, we synthesized [$^{13}C_9$, $^{15}N_3$] CMP-beta-^{13}C-[1,2,3,10,11]-NeuAc (*CMP-NeuAc*). The labeling strategies used for *CMP-NeuAc* come from the published synthetic routes for the natural-abundance counterparts. A ^{13}C-acetyl group was introduced to the NeuAc precursor, N-[1^{13}C, 2^{13}C]-acetylmannosamine (ManNAc*), by the transfer of the ^{13}C-acetyl group of N-[1^{13}C, 2^{13}C]-acetoxy-succinimide (NAS*) to mannosamine [15]. $^{13}C_3$-pyruvate was incorporated as ^{13}Cs at the 1,2, and 3 positions of NeuAc by the enzyme

NeuAc aldolase [16]. *CMP-NeuAc* was synthesized using the enzyme CMP-NeuAc synthetase to react $^{13}C_5$-NeuAc with the commercially available [$^{13}C_9$, $^{15}N_3$]cytidine-5'-triphosphophate (CTP*) [6].

Fluorine NMR has been used extensively on ligand-protein complexes for substrates screening [17], stoichiometry measurement [18], and binding affinity assessment [19]. Subsequent data analysis with linear or curve fitting also can yield the inhibitor dissociation constant K_I, an essential parameter to evaluate the system and compare with other systems [20]. Kinetics studies of donor substrates of the protein ST6Gal-1 have not been published. Our goal was to evaluate the exchange rate of the fluorinated donor analog CMP-3FNeuAc between free and ST6Gal-1-bound states and explore the possibility of determining K_I experimentally by ^{19}F-NMR.

The chemical shifts of a fluorinated ligand free in solution and bound to a protein are normally distinguishable because the electric field, short-range contacts with the receptor and solvent, and the hydrogen-bonding possibilities experienced by the fluorine(s), are typically very different in the free and complexed states. For strong-binding ligands, their ^{19}F chemical shift changes can be as large as 8 ppm, as for 5-fluoro-L-Tryptophan bound to human serum albumin [21]. For the more common situations – fast-exchanging ligands, the chemical shift changes can be as small as 0.07 ppm, e.g. the 2-pyridinone derivatives bound to the chaperone PapD [17]. The linewidth and relaxation time of the ligand usually change significantly upon binding to the protein. When the ligand binds to a macromolecule like a protein, it assumes the long molecular correlation time of the protein and this causes line broadening. Further rate-dependent line broadening can result from the slow exchange between the free and bound states.

Normally there also is a considerable change in other relaxation parameters for spins of the ligand in the receptor-bound state. It has been reported that cross-correlated ^{19}F relaxation measurements can be used to quantitate the exchange rate of a fluorinated ligand-receptor complex and estimate the dissociation constant [19].

2.2 Experimental

2.2.1 Two-step enzymatic synthesis of CMP-3FNeuAc

ManNAc, 3-fluoropyruvate, CTP, NeuAc aldolase (EC 4.1.3.3) were purchased from Sigma-Aldrich. CMP-NeuAc synthetase (EC 2.7.7.43) was purchased from Calbiochem.

5-Acetamido-3,5-dideoxy-3-fluoro-D-erythro-L-manno-2-nonulopyranosonic acid (1). A solution of ManNAc (110.6 mg, 500 μmol), 3-fluoropyruvate sodium salt (12.8 mg, 100 μmol), sodium azide (0.3 mg) and 30 U of NeuAc aldolase in water (0.6 ml) was stirred gently at room temperature for 24 hrs (when ^{1}H-NMR showed >95% conversion of pyruvate to the product); the pH was adjusted to 7.5 by 0.1 M NaOH every few hours if necessary. The enzyme was then filtered out (Microcon, 10 kDa). The filtrate was directly loaded onto an anion exchange column (Dowex 1X-2 resin, formate form, 1 x 10 cm) and unreacted ManNAc was washed out with water (90% was recovered). Then 0.5 M formic acid was applied and the fractions containing the product as detected by TLC (1-propanol/water 7:3) were pooled and placed under vacuum to remove HCOOH. The resultant residue was lyophilized to yield 3-F_{ax}-NeuAc as a white solid (10:1 beta:alpha, 26 mg, 79%). ^{1}H and ^{13}C spectra of the product were consistent with those reported in the literature [22]. ^{13}C NMR (300 MHz, D_2O): δ 175.1 (CONH), 172.3 (COO), 95.0 (d, J = 37.7 Hz, C2), 91.5 (d, J = 188.5 Hz, C3), 70.8 (C6), 70.3 (C8), 68.2 (C7), 68.0 (d, J =

30.2 Hz, C4), 63.5 (C9), 47.5 (d, *J* = 3.0 Hz, C5), 20.22 (CH_3CO). ^{19}F NMR (300 MHz, D_2O): δ -208.3 ppm, *J* = 48.5 Hz, *J* = 29.6 Hz, external reference $CFCl_3$ (0.0 ppm). CMP3FNeuAc. Na 654.43 $C_{20}H_{31}O_{15}N_4PNa$

Cytidine 5'-phosphate, P'-(5-acetamido-3,5-dideoxy-3-fluoro-2-D-erythro-L-manno-2-nonulopyranosonic acid (2). CTP (9.5 mg, 16 μmol) and 3F-NeuAc (5.5 mg, 15 μmol) were dissolved in 400 μl of Tris buffer [90% 0.1 M Tris buffer (pH 9) and 10% 1 M $MgCl_2$ containing 0.2 mM DTT]. The pH was readjusted to 9 with 2 M NaOH (if necessary). The reaction was initiated by adding CMP-Neu5Ac synthetase (1U) and inorganic pyrophosphatase (0.4U). The reaction solution was incubated at 37°C with occasional shaking and the pH was maintained at 9 with 2 M NaOH for the first few hours. After 4 hrs, 90% of the CTP was converted to the product (^1H NMR). The solution was left to stand overnight at 37°C to complete the reaction. The mixture was filtered (Acrodisc, 0.2 μm pore size) to remove the white precipitates. The filtrate was centrifuged in a Microcon® filter (10 kDa) to remove the enzymes. After lyophilization, the product was dissolved in 200 μl H_2O and precipitated by adding ethanol (9:1 v/v), isolated by centrifugation (10,000 g; 10 minutes) and dried under vacuum. Ethanol precipitation was repeated one additional time. Finally, the product was purified on a Bio-gel P2 (extra fine) column and then lyophilized to yield CMP-β-$3F_{ax}$-NeuAc as a white fluffy solid (8.5 mg, 77 %). ^1H (600 MHz, D_2O 4.77 ppm) δ: 7.96 (d, 1 H, *J* = 7.8 Hz, H6 cytidine), 6.14 (d, 1 H, *J* = 7.3 Hz, H5 cytidine), 6.00 (d, 1 H, *J* = 4.4 Hz, H1 ribose), 4.97 (dd, 1H, *J* = 50.3 Hz, *J* = 3.0 Hz, H3 NeuAc), 4.34 – 4.30 (m, 3 H, H5 NeuAc, H3 ribose, H2 ribose), 4.26-4.22 (m, 3 H, H5 ribose, H5 ribose, H4 ribose), 4.19-4.13 (m, 2 H, H6 NeuAc, H4 NeuAc), 4.00 (d, 1 H, *J* = 7.3 Hz, H8 NeuAc), 3.92 (d,

1 H, *J* = 9.8 Hz, H9 NeuAc), 3.66 (dd, 1 H, *J* = 12.0 Hz, *J* = 6.6 Hz, H9 NeuAc), 3.49 (d, 1 H, *J* = 9.8 Hz, H7 NeuAc), 2.07 (s, 3 H, $NHCH_3$). 13C NMR (300 MHz, D2O): δ 175.1 (CONH), 171.7 (COO), 166.5 (C4 cytosine), 158.0 (C2 cytosine), 142.0 (C6 cytosine), 98.5 (dd, *J* = 37.7 Hz, *J* = 11.3 Hz, C2 NeuAc), 96.7 (C5 cytosine), 92.8 (d, *J* = 188.5 Hz, *J* = 22.6 Hz, C3 NeuAc), 89.3 (C1 ribose), 83.1 (d, *J* = 22.6 Hz, C4 ribose), 74.4 (C2 ribose), 71.8 (C6 NeuAc), 69.8 (NeuAc C8), 69.5 (C3 ribose), 68.0 (C7 NeuAc), 68.3 (d, *J* = 30.2 Hz, C4 NeuAc), 65.3 (d, *J* = 11.3 Hz, C5 ribose), 63.5 (C9 NeuAc), 47.5 (d, *J* = 3.0 Hz, C5 NeuAc), 20.22 (CH_3CO).

2.2.2 Four-step chemoenzymatic synthesis of ^{13}C-labeled CMP-NeuAc

N-[1^{13}C, 2^{13}C]-acetoxy-succinimide (NAS) ^{15}.* N-hydroxysuccinimide (NHS) (325 mg, 2.82 mmol) was dried overnight under vacuum before use. At 0^{o}C Et_3N (313 mg, 3.10 mmol) and $^{13}C_2$- acetyl chloride (250 mg, 3.11 mmol) was slowly injected into a suspension of NHS in freshly-dried CH_2Cl_2 (10 ml) under argon. The mixture was stirred at room temperature for 2 hrs and cooled down to 0^{o}C. Ethyl ether (10 ml) was injected and the formed precipitate was filtered out. After the solvent was evaporated, a white solid (492 mg) was collected and used for the next step without further purification.

N-[1 ^{13}C, 2 ^{13}C] - acetyl-mannosamine (ManNAc).* D-mannosamine hydrochloride (325.48 mg, 1.51 mmol) was dissolved in H_2O (2.8 ml) and then THF (1.2 ml) was added. At 0^{o}C NAS* (200 mg, 1.26 mmol) was added in portions while the pH was adjusted to 7.0 by NaOH (2 M). The resulting clear yellow solution was stirred at room temperature for 1/2 hr and monitored by silica gel TLC ($CHCl_3$, CH_3OH, H_2O, 65:35:5). After THF was evaporated, the mixture was applied on a small column of Dowex 50 (H^+ form) and eluted with water to remove the excess mannosamine

hydrochloride. The fraction (containing HCl, pH 4) was pooled and was adjusted to pH 5 by Dowex 1*8 (HCO_3^- form) resin. The Dowex 1*8 resin was filtered out and the filtrate was lyophilized to a clear colorless gummy residue (250.4 mg). The residue was dissolved in minimum amount of ethanol and applied onto a column of Iatrobeads (8 g) eluted first with $CHCl_3$, CH_3OH 80:20 and then with $CHCl_3$, CH_3OH and H_2O 65:35:5. The fractions containing ManNAc* were pooled and weighed at 215.4 mg (yield 77%) after evaporating the solvent.

N-[1 ^{13}C, 2 ^{13}C] –acetyl–beta- [1 ^{13}C, 2 ^{13}C, 3 ^{13}C] -neuraminic acid (NeuAc)* [16]. ManNAc* (118.5 mg, 0.53 mmol), ^{13} C_3-pyruvate sodium (11.0 mg, 0.097 mmol), and sodium azide (0.4 mg) were dissolved in 700 µl of H_2O. After the pH was adjusted to 7.5 with 0.25 M NaOH, BSA (1 mg) and Neu5Ac aldolase (0.8 U) were added to the solution. The reaction mixture was stirred at room temperature for 18 hrs (monitored by ^{1}H-NMR) and was maintained at pH 7.5. After Microcon® centrifugal filtration to remove the proteins, the filtrate was applied on an anion-exchange Dowex 1X-2 column (formate form) and first eluted with water to recover unreacted ManNAc* and then 0.5 N formic acid (TLC, 1-propanol / water 7:3). The fractions containing Neu5Ac* were pooled and HCOOH was evaporated and finally lyophilized to give the white fine powder of Neu5Ac* (19.4 mg, yield 65%).

*Cytidine-$^{13}C_9$, $^{15}N_3$-5'-monohospho-N-[1^{13}C, 2^{13}C]-acetyl-beta-[1^{13}C, 2^{13}C, 3 ^{13}C] neuraminic acid (*CMP-NeuAc*)* [6]. Cytidine-$^{13}C_9$, $^{15}N_3$-5'-triphosphate (8.75 mg, 17.7 µmol) and NeuAc* (5.3 mg, 16.9 µmol) were dissolved in 600 µl of 0.1 M Tris buffer (pH 9.0, containing 0.1 M $MgCl_2$ and 0.2 mM DTT). The reaction was initiated by adding yeast inorganic pyrophosphatase (0.4 U), CMP-Neu5Ac synthetase (1 U) and

incubated at 37°C. The solution gradually turned cloudy; 4 µl of 2 M NaOH was added to bring the pH from 8.5 back to 9.0; the pH did not change after ½ hr. The reaction was monitored by TLC and ^{1}H-NMR. After 2 hrs, the precipitates, presumably $Mg_3(PO4)_2$, was filtered by a 0.2 µm-pore-size membrane. Then the filtrate was centrifuged in a Microcon® filter (10 KDa) at 4°C (to slow down the hydrolysis of product). The filtrate was lyophilized and the residue was dissolved in 100 µl water. When 900 µl of ethanol was added, precipitates (the product and salt) formed immediately. The mixture was refrigerated for a few hours before the supernatant was discarded. The ethanol precipitation (ethanol: water 9:1 v/v) was repeated twice to remove Tris and DTT completely. The crude product was purified by the Bio-gel P2 column eluted with water at 4°C. *CMP-NeuAc* was obtained as 9.5 mg of fluffy white solid with an 86% yield. m/z 636.4, $C_{20}H_{30}N_4O_{16}PNa$. ^{1}H NMR (600 MHz, D_2O): δ 8.12 (d, 1 H, *J* = 182.6 Hz, H6 cytosine), 6.06 (d, 1 H, *J* = 174.7 Hz, H5 cytosine), 5.99 (d, 1 H, *J* = 169.8 Hz, H1 ribose), 4.48 (d, 1H, *J* = 150.1 Hz, H3 ribose) , 4.43 (d, 1 H, *J* = 151.2 Hz, H2 ribose), 4.35(d, 3H, *J* =148.2 Hz, H4 ribose, H5 ribose, H5 ribose), 4.13-4.07 (m, 2H, H6 NeuAc, H4 NeuAc), 3.96 – 3.88 (m, 2 H, H5 NeuAc, H8 NeuAc), 3.86 (dd, 1 H, *J* = 11.5 Hz, *J* = 4.0 Hz, H9 NeuAc), 3.62 (dd, 1 H, *J* = 11.5 Hz, *J* = 6.5 Hz, H9 NeuAc), 3.44 (d, 1 H, *J* = 10.0 Hz, H7 NeuAc), 2.62 (d, broad, 1H, *J* = 132.5 Hz, $H3_{eq}$ NeuAc), 2.15 (dd, 3H, *J* = 128.5 Hz, *J* = 6.0 Hz, NHCH_3), 1.74 (d, broad, 1H, *J* = 125.5 Hz, $H3_{ax}$ NeuAc). 1H-13C NMR: δ 144.0(cytosine6), 97.0(cytosine5), 89.5(ribose1), 69.5(ribose3), 75.0(ribose2), 83.0(ribose4), 64.5(ribose5), 71.5(NeuAc6), 67.0(NeuAc4), 51.5(NeuAc5), 69.5(NeuAc 8), 63.0(NeuAc9), 69.0(NeuAc7).

2.2.3 Alignment Media Protocol, edited based on the protocol by Anita Kishore

C12E5 (pentaethylene glycol monododecyl ether) was purchased from Sigma. The recommended molar ratio of C12E5: hexanol is 0.87 to make a solution of 8% (w/v) C12E5. However, the ratio of C12E5 and hexanol was optimized experimentally by observing deuterium splittings and visual appearance of the liquid status. We used 8% C12E5 and found the ratio of C12E5 to hexanol to be 0.83.

25 µl C12E5 was well dissolved in 250 ul buffer (10% D_2O included). 9 µl hexanol was added in 3 µl increments and vortexed well after each addition. The solution went from clear to milky, turbid, translucent and viscous w/ lots of bubbles (upon vortexing). This phase does align in the magnet but could be biphasic and heterogeneous. If the solution becomes cloudy, milky, and very fluid, it indicates that too much hexanol must have been added and the nematic phase has been passed. C12E5 is viscous so careful pipetting technique is essential. Once formed the mixture is stable over a wide pH range. ^2H splitting at 25^oC is usually 23-26 Hz but may change upon the addition of a biomolecule for alignment.

2.2.4 Collection of NMR spectra

1) STD & trNOE

NMR samples were prepared in deuterated 12 mM phosphate buffer (containing 200 mM NaCl) at pH 6.5 with a protein concentration of 0.3 mM. The amounts of ligand used were respectively, a 25-fold excess of CMP-NeuAc, a 5-fold excess of CMP-3FNeuAc, and a 50-fold excess of LacNAc. The spectra were acquired with the Varian BioPak pulse sequences on Varian INOVA spectrometers operating at 600 or 800 MHz (^1H frequency). The trNOE experiments were performed at 17^oC using zero-quantum

filtered 2D NOESY with a mixing time of 100 ms-500 ms and a recycling rate of 1.2 sec. The STD experiments followed the saturation transfer 1D protocol (DPFGSE 1D) without a $T_{1\rho}$ filter or solvent saturation, and used a 100 Hz Gaussian saturation pulse train of 450 ms centered at 0.8 ppm in the aliphatic region. The off-frequency spectra were recorded by irradiating at +24 ppm. The recycling delay was 2.5 sec. The data were acquired at 25^{o}C for CMP-NeuAc and CMP-3FNeuAc and 5^{o}C for LacNAc.

2) ^{19}F-NMR studies on the CMP-3FNeuAc-ST6Gal-1 complex

^{15}N-ST6Gal-1 was obtained from the Moremen laboratory at the CCRC, UGA. It was purified following expression in Pichia pastoris as a product fully enriched in ^{15}N. The purification procedure is outlined in the unpublished data by N. Tejwani and K. W. Moremen. The protein sample for NMR was prepared in pH 6.8 phosphate buffer or pH 6.5 Bis-tris buffer. The low magnetic field NMR (282 and 470 MHz) was chosen to prevent CSA effects from broadening the ^{19}F signals. The resonance from a fluorocarbon polymer used in NMR probe construction is located at -110.0 ppm. It is very intense and broad, distorting the baseline. The T_1 for the fluorine of CMP-3FNeuAc was measured as 0.6 sec at 300 MHz and the 90 degree pulse width was measured as 16 μs. For the 1D experiments, we used zero delay time, 75^{o} pulse, 0.8 sec acquisition time and about 20,000 scans (overnight) for the protein-ligand samples. Chemical shifts are corrected with the external standard to $CFCl_3$ (0.00 ppm).

2. 3 Results and discussion

2.3.1 Bound conformation from transferred NOE

Transferred NOEs can give useful information on the conformations of the individual bound ligands. Strong negative NOEs were observed from samples having a

large excess of CMP-NeuAc in the presence of ST6Gal-1 (Figure 2.2). This clearly indicates the dominance of NOEs by contributions from the ligands in bound state. In Figure 2.2, the circled crosspeak connects the H6 proton of the cytosine ring and the H3 proton of the ribose. This NOE of CMP-3FNeuAc also was strong in the absence of the protein, but it had the opposite sign. This observation in both free and bound states supports the dominant population of the anti conformation about the cytosine-ribose glycosidic bond in which the cytosine H6 and ribose H3 protons are in close proximity. Given that the distance of cytosine H5 and cytosine H6 is 2.5 Å, the distance between cytosine H6 and ribose H3 was calculated as 2.3 Å in protein-free solution and 2.6 Å in protein based on their respective crosspeak intensities. This is expected from its minimum energy structures and the conformations observed in other crystal structures. Unfortunately, there were no cross-peaks detected between the nucleotide and the sugar. However, the absence of cross-peaks is consistent with the long distances between nucleotide and sugar protons seen in the preferred low energy states and the unfolded conformation typically seen in the crystal structures.

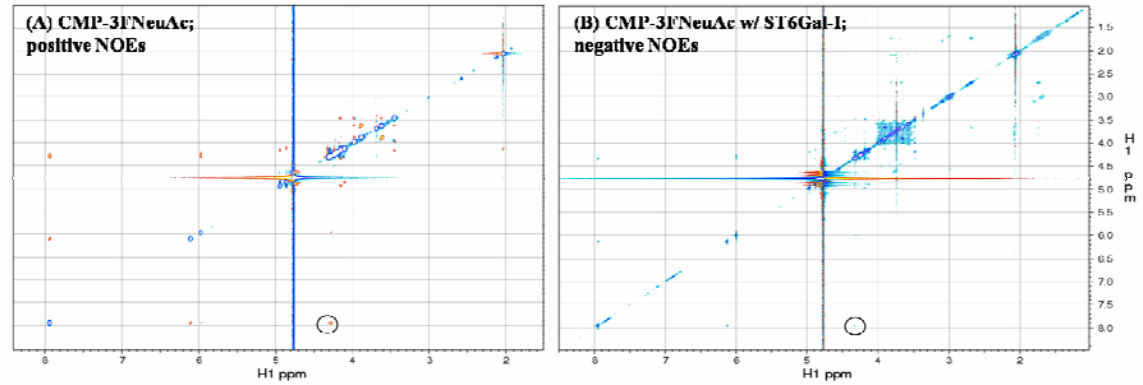


Figure 2.2. ^{1}H-^{1}H 2D NOESY spectra of CMP-3FNeuAc (0.3 mM) without (A) and with (B) ST6Gal-1 (1.5 mM) at a mixing time of 500 ms. Red indicates a positive NOE and blue indicates a negative NOE.

Transferred NOE data on the LacNAc acceptor were difficult to collect because of the weak binding to ST6Gal-1. NOEs collected at 25^{o}C were in fact positive and difficult to distinguish from NOEs seen for LacNAc in the absence of the protein. The dominant NOEs obtained in a 1D NOE experiment, in which the anomeric proton resonance of galactose was saturated showed the expected intra-ring connectivity to the axial H3 and H5 protons as well as a poorly resolved, but distinguishable connection to the trans-glycosidic GlcNAc H4 proton. The latter observation is consistent with LacNAc favoring the low energy ϕ = 123.26 (C1-O4-C4'-C5'), ψ= -64.33 (O5-C1-O4-C4') conformation in solution. To make effects of binding to LacNAc observable the temperature was lowered to 17°C where LacNAc in free solution showed minimal NOEs (actually a mix of positive and negative NOEs, possibly due to a mix of contributions from overall anisotropic motions and internal motions). As depicted in Figure 2.3, the addition of ST6Gal-1 now produces some clear negative NOEs (to Gal H3) and some negative contributions that cancel residual positive contributions at (Gal H5 and GlcNAc H4). This is consistent with maintenance of a dominant population for the minimum energy structure in the bound complex as well.

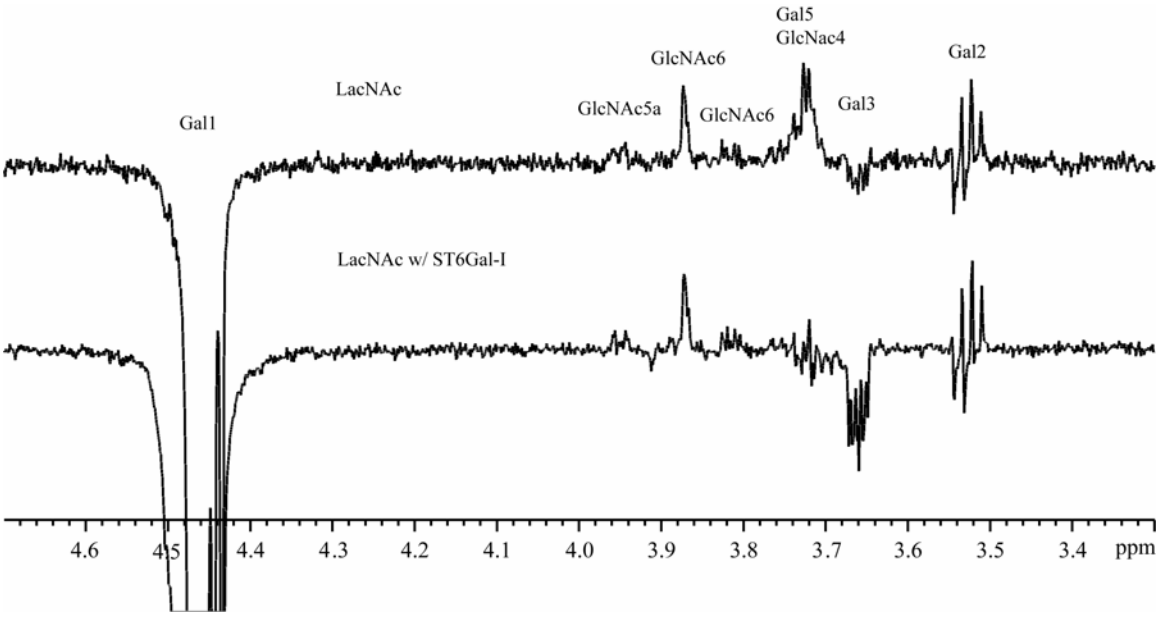


Figure 2.3. 1D NOESY spectra of LacNAc (4.5 mM) without (top spectrum) and with (bottom spectrum) ST6Gal-1 (0.3 mM) acquired at 17^{o}C

2.3.2 Binding epitopes from saturation transfer difference

Saturation transfer difference (STD) experiments were used to identify the binding epitopes of both the donor and the acceptor. Saturation transfer is most efficient when protein protons (connected by spin diffusion to a general protein proton pool) are in close proximity to ligand protons. This can indicate what parts of the ligand are in closest contact with the protein surface. Although the results were qualitatively the same, STD spectra for CMP-NeuAc were of higher signal to noise ratio than those for CMP-3FNeuAc (Figure 2.4). The derived STD percentages (scaled to the most perturbed peak – H1 of the ribose) of CMP-NeuAc show the clear preferential enhancement of the nucleotide base and adjacent parts of the ribose ring. This observation is consistent with the evidence that CMP is an efficient competitive inhibitor of ST6Gal-1. The less

efficient magnetization transfer to the NeuAc residue is consistent with the necessity of having this space to accommodate the acceptor. The binding epitope of CMP-NeuAc demonstrates that CMP serves as a recognition motif for ST6Gal-1 that has affinity for all CMP-sugar donor substrates. Appropriate positioning for the sugar portion may provide specificity for reaction [23].

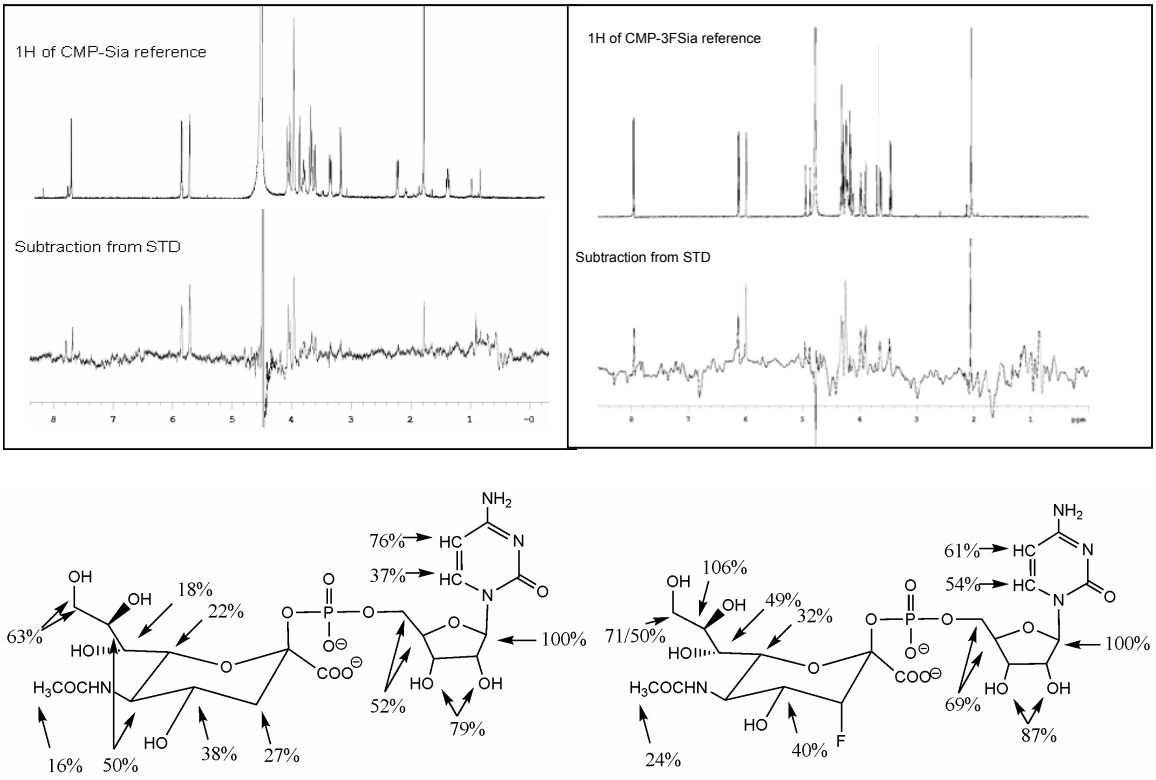


Figure 2.4. The STD spectra and the derived STD percentages for CMP-NeuAc (left) & CMP-3FNeuAc (right) bound to ST6Gal-1. STD percentages were calculated by determining individual signal intensities in the reference spectrum of the ligand and in the subtraction spectrum (subtracting STD spectrum of ST6Gal-1 from STD spectrum of the ligand with ST6Gal-1), and normalized using the largest STD effect (in this case, H1 of the ribose). The STD effect for H3ax of NeuAc was not measured accurately due to low S/N and hence only H3eq of NeuAc was shown.

An STD effect for LacNAc could only be detected when the ratio of the ligand to the protein was 35-fold or greater and the sample was cooled down to 5°C. Figure 2.5 shows the difference spectrum obtained when the aliphatic region (0.8 ppm) of the protein was saturated. There clearly was saturation transfer from the protein to the ligand, but the effects were fairly small and uniformly distributed over all protons of the LacNAc moiety. This can be the result of a binding mode where protein protons to ligand magnetization transfers are inefficient compared to intra-ligand transfers.

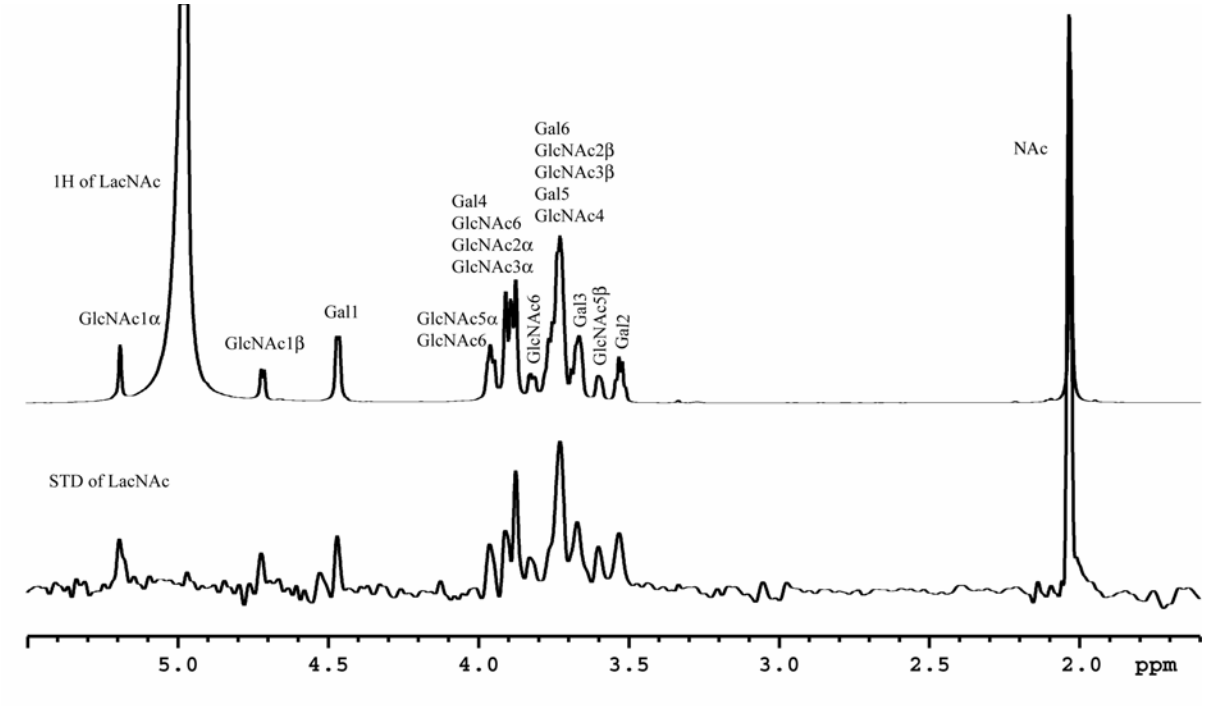


Figure 2.5. 1H and STD spectra of LacNAc. The ^{1}H NMR of LacNAc was processed with an exponential weighting function of 5 Hz to mimic the broad linewidth of LacNAc in the STD spectrum which was generated by subtracting the STD spectrum of ST6Gal-1 from the STD spectrum of LacNAc with ST6Gal-1.

There is also some concern that apparent STD effects could result from direct saturation through the long wings of Lorenzian lines associated with each proton. A

control was run by saturating the aromatic region (7.0 ppm) as opposed to the aliphatic region (0.8 ppm) of ST6Gal-1. A similar pattern was obtained except for the intensity of the N-acetyl resonance (1.9 ppm, closest to the aliphatic saturation frequency) which dropped by 3- fold. The addition of CMP-3FNeuAc did not significantly increase the STD of LacNAc. A reduction in temperature enhanced the saturation efficiency for ST6Gal-1 as expected for a medium-size protein [25].

2.3.3 RDC measurements of *CMP-NeuAc*

RDC data on a bound complex of CMP-NeuAc-ST6Gal-1 have not yet been collected, but the preliminary data of *CMP-NeuAc* in solution illustrate the acquisition procedures. Ligand orientation and internal structure can be determined using one bond ^{1}H-^{13}C and ^{13}C-^{13}C RDCs acquired with an f1-coupled ^{1}H-^{13}C HSQC experiment. To facilitate acquisition of these data a ^{13}C labeled analog was synthesized (Figure 2.6).

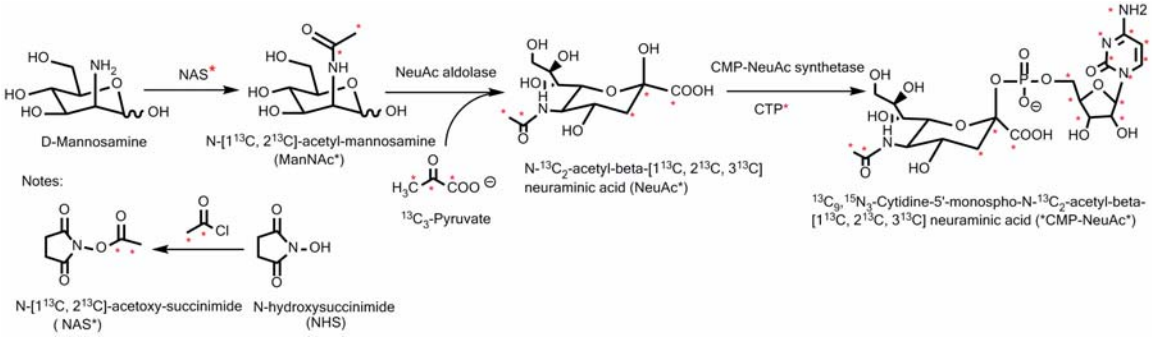


Figure 2.6. Chemoenzymatic synthesis of ^{13}C-labeled CMP-NeuAc

N-hydroxysuccinimide and ^{13}C-acetyl chloride were combined to produce NAS*, which was further reacted with D-mannosamine to make ManNAc* [15]. NeuAc aldolase was then used to react ^{13}C-ManNAc with ^{13}C-pyruvate to produce ^{13}C-[1,2,3,10,11]-NeuAc [16]. *CMP-NeuAc* was synthesized using CMP-NeuAc synthetase to react $^{13}C_5$-

NeuAc and CTP* [6]. The simplified purification procedure for *CMP-NeuAc* consisted of enzyme removal, ethanol precipitation, and size exclusion chromatography. Precautions, including operation at 4°C and avoiding low pH, were taken during purification to slow down the hydrolysis of the CMP-NeuAc product.

With a ^{13}C-enriched ligand a significant number of couplings can be measured within a fairly short acquisition time. For *CMP-NeuAc* we expected to see 7 pairs of ^{1}H-^{13}C coupled and 12 pairs of ^{13}C-^{13}C coupled peaks from CMP, in addition to 3 pairs of ^{1}H-^{13}C coupled and 3 pairs of ^{13}C-^{13}C coupled peaks from NeuAc (Figure 2.7).

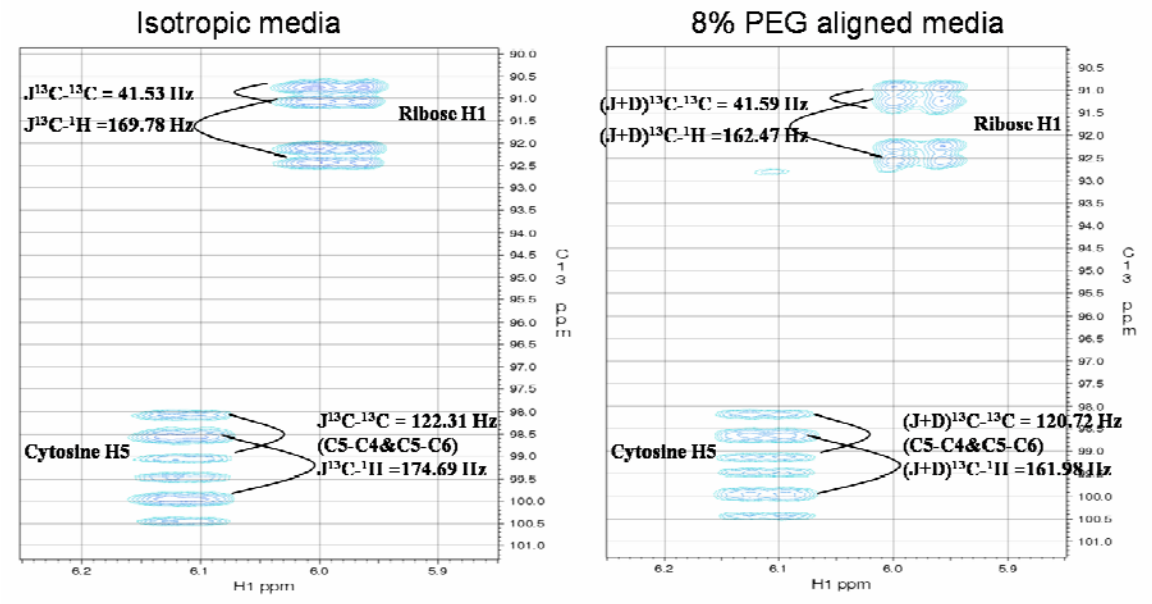


Figure 2.7. F1-coupled ^{1}H-^{13}C HSQC spectra of ^{13}C-labeled CMP-NeuAc (regional expansion). It illustrates RDC measurements on cytosine H5 (D_{CH} = -12.71 Hz) and ribose H1 (D_{CH} = -7.31 Hz). Each spectrum was collected for 1 hour at 298 K on 600 MHz (^{1}H frequency). The samples were prepared using 12 mM phosphate buffer saline (PBS) at pH 7.0 with and without the aligned media PEG (pentaethylene glycol monododecyl ether). The program Anita_nlinlS was used to fit the resonance lines.

We summarize the RDCs of *CMP-NeuAc* in the absence of ST6Gal-1 in Table 2.1. The RDCs of *CMP-NeuAc* in the presence of ST6Gal-1 also were collected one time in 8% PEG but differences for the ligand RDCs with and without protein proved too small to confidently extract bound state values. This is likely due to the relatively low fractions bound and the similarity in level of orientation for the free ligand and protein. Experiments will require enhancement of protein alignment by methods such as those recently used for Galectin-3.

Table 2.1. Summary of ^{1}H-^{13}C and ^{13}C-^{13}C RDCs of ^{13}C-labeled CMP-NeuAc in PBS

	Isotropic	C12E5	D_{C-H}		Isotropic	C12E5	Dc-c
Protons	C-H (Hz)	C-H (Hz)	(Hz)	Carbons	C-C (Hz)	C-C (Hz)	(Hz)
cytosineH6"	182.56	168.25	-14.31	C5"-C6"	67.94	68.06	0.12
cytosineH5"	174.69	161.98	-12.71	C5"-C6"&C4"	122.31	120.72	-1.59
riboseH1'	169.78	162.47	-7.31	C1'-C2'	41.53	41.59	0.06
riboseH3'	150.08	147.29	-2.79	C3'-C2'&C4'	75.06	74.69	-0.37
riboseH2'	151.25	147.88	-3.37	C2'-C1'&C3'	78.50	80.19	1.69
riboseH4'	149.21	150.96	1.75	C4'-C3'&C5'	79.87	79.37	-0.50
riboseH5'ab	147.19	155.03	7.84	C5'-C4'	41.63	40.81	-0.82
siaH3eq	130.66	132.30	1.64	C5'-C4'	41.63	40.31	-1.32
siaH3ax	131.66	132.31	0.65	C3-C2	40.38	41.16	0.78
acetyl CH3	128.17	130.06	1.89	C2-C1	40.19	39.25	-0.94
				CO-CH3	49.88	49.56	-0.32

The program, REDCAT (Residual Dipolar Coupling Analysis Tool) [24], was used to assess ligand alignment and the validity of a minimum energy structure in the free solution state. A PDB structure of CMP-NeuAc was input along with corresponding

experimental RDCs. A set of principle order parameters and a set of Euler angles for transforming the PDB file to the principal alignment frame were returned in the format of a new PDB file which was visualized in Figure 2.8. The resultant Sauson-Flamsted projection of direction distributions for axes of allowed alignment frames, gives and indication of the precision in determination of orientation of ligand alignment. The back-calculated RDCs and are in a good agreement with the experimental RDCs.

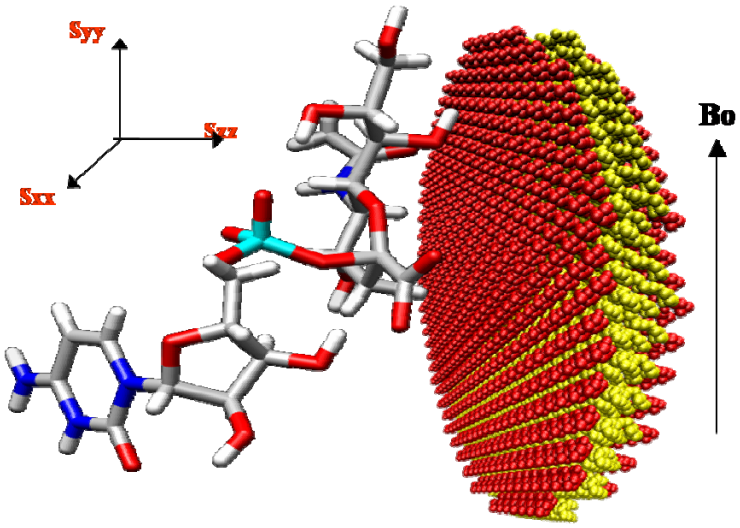


Figure 2.8. Preferred orientation of CMP-NeuAc in aligned media calculated by REDCAT with experimental RDCs. The structure is visualized using Chimera. The structure is actually four-fold degenerate because of insensitivity to 180° rotation about any of the three axes.

CMP-NeuAc can be hydrolyzed by itself in acidic or even neutral solution and it slowly degrades into CMP and NeuAc (40% degradation by 2 months at 4^{o}C). ST6Gal-1 dramatically expedites this hydrolysis process by transferring NeuAc to the acceptor water. As shown in Figure 2.9, almost 50% of CMP-NeuAc is decomposed by 10 hours and CMP-NeuAc is hydrolyzed completely at the end of 40 hours. Lowering the

temperature such as to 4°C slows down the hydrolysis process by approximately 2 fold. For our RDC measurements, the acquisition of each HSQC spectrum only took one hour during which 85% of the CMP-NeuAc stayed intact and dominated the detected signals. This short acquisition period also limited precision and the ability to detect differences in the presence and absence of the protein. Synthesis of ^{13}C-labeled 3FNeuAc [25; 26; 27] has been considered for preparing ^{13}C-labeled CMP-3FNeuAc for future RDC experiments.

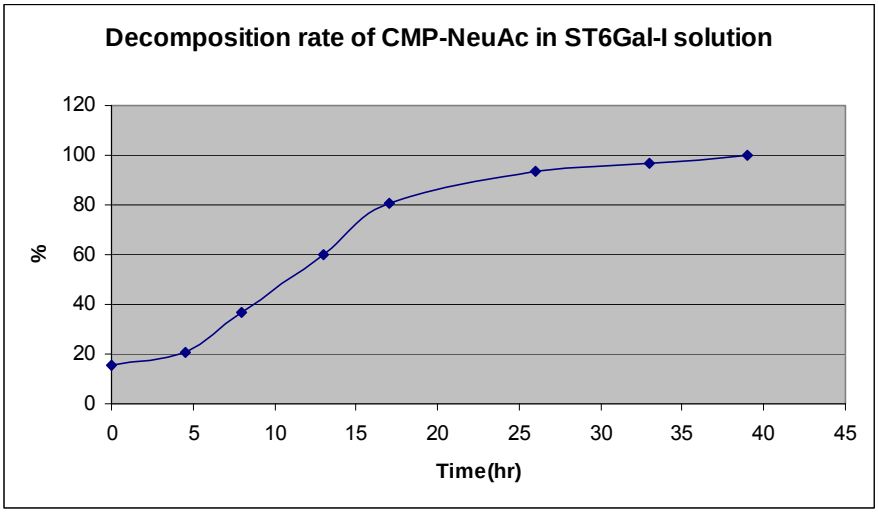
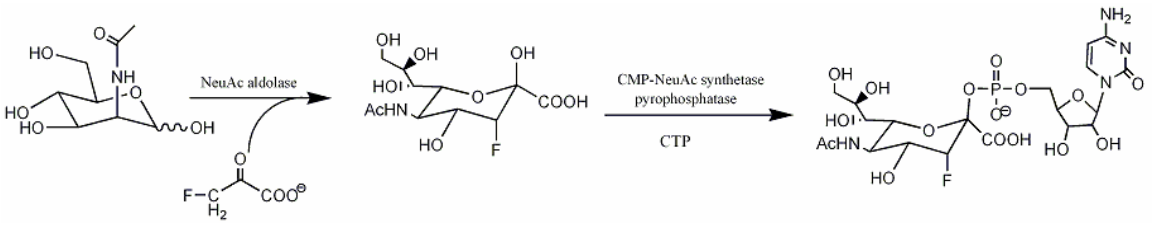


Figure 2.9. Decomposition rate of CMP-NeuAc in the presence of ST6Gal-1 at 25°C.

2.3.4 ^{19}F-NMR studies of CMP-3FNeuAc - ST6Gal-1 complex

In principle, ^{19}F NMR can provide a sensitive probe of protein-ligand interactions. For this reason a ^{19}F analog of sialyl-CMP was synthesized in a convenient two-step method as shown in Figure 2.10.



(1) (2)

Figure 2.10. Enzymatic synthesis of CMP-3FNeuAc

The introduction of an electronegative fluorine atom at a site adjacent to the anomeric center on neuraminic acid effectively slows down turnover by destabilizing the oxocarbenium ion-like transition state for the reaction [28]. Moreover, it stabilizes the labile sugar donor against spontaneous hydrolysis. Here, CMP-3FNeuAc was enzymatically synthesized from ManNAc, 3-fluropyruvate, and CTP. Neu5Ac aldolase (>24 U/ml) has high specific activity to react ManNAc with 3-fluoropyruvate to produce 3-fluoro-NeuAc in mainly beta-form. It was initially thought that CMP-NeuAc synthetase could not tolerate 3FNeuAc in the coupling reaction [26]. However, a more recent paper reported that CMP-NeuAc synthetase does efficiently catalyze the synthesis of CMP-3FNeuAc directly from 3FNeuAc and CTP. We verified the latter conclusion and CMP-3FNeuAc was obtained in 77% yield with this method.

An IC_{50} for CMP-3FNeuAc was measured as 1.44 mM based on radio-assay. By the formula $(IC_{50})*K_M/ (K_M+[S])$, K_I was calculated as 0.65 mM which implies that CMP-3FNeuAc binds weakly to ST6Gal-1. We performed simulations and determined the protein concentration $[P]_t$ and ligand concentration $[L]_t$ to be used in order to achieve a balance between the fraction of bound ligand F_b and the concentration of bound ligand $[L]_b$. For a single-intermediate model of competitive inhibition $E + I \leftrightarrow EI$, $[L]_b$ is mainly determined by $[P]_t$ and the maximum $[L]_b$ is only about 0.1 times $[P]_t$ based on the estimated K_I value (Figure 2.2). Using high $[P]_t$ would not be feasible or beneficial because the ST6Gal-1 supply was limited and this protein tends to dimerize above 0.3 mM. Hence we had few options for improving the detection of the NMR signal of the bound ligand due to its weak binding feature.

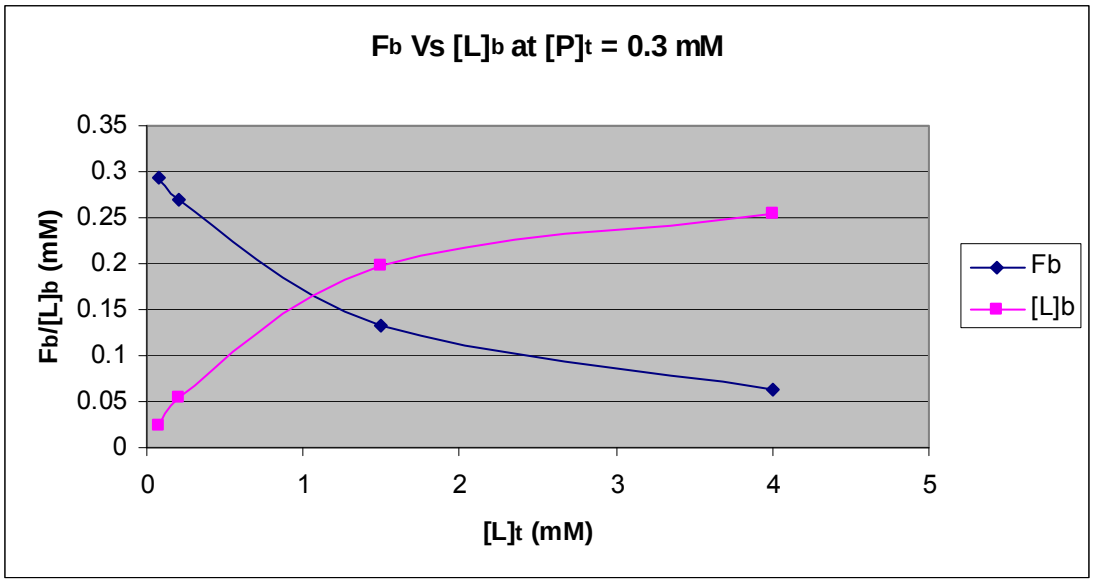


Figure 2.11. Relationship of the fraction of bound ligand F_b and the concentration of bound ligand $[L]_b$ at a constant protein concentration $[P]_t$

We conducted 1D ^{19}F experiments with $[L]_t$ at 0.08, 0.2, 1.5 mM at $[P]_t$ 0.3 mM (F_b =0.13-0.29, $[L]_b$=0.02-0.20), $[L]_t$ 0.15 mM at $[P]_t$ 0.53 mM (F_b=0.22, $[L]_b$=0.34 mM) and measured the differences in chemical shifts and linewidth for the resonance of CMP-3FNeuAc in ST6Gal-1 solution. With the ^{15}N-ST6Gal-1 sample, we further estimated the exchange rate by monitoring the protein resonance changes upon the addition of CMP-3FNeuAc. We also verified a simple two-state model for the ligands by measuring the relaxation of CMP-3FNeuAc through the titration experiments using $[L]_t$ 0.2 mM- 4 mM at $[P]_t$ 0.2 mM with the ratios ranging from 1:1 to 20:1 (ligand over protein).

1) Chemical shift changes for ^{19}F and ^{1}H resonance of CMP-3FNeuAc

Given the anticipated ^{19}F shifts of 1-10 ppm for tight binding, we would expect shifts of the order of 0.1 ppm to be observed for the weak binding system of CMP-3FNeuAc-ST6Gal-1. During the experiments, the maximum chemical shift change of 0.02 ppm for the ^{19}F resonance of CMP-3FNeuAc was observed at a 5:1 ligand ratio to

protein and there was no further change after this point when excess ligand was added. The chemical shift change of 0.02 ppm is negligible because the resonance linewidth is as big as 0.04 ppm. For proton resonances of CMP-3FNeuAc the chemical shift difference upon the addition of ST6Gal-1 is less than 0.0025 ppm while the linewidth is 0.01 ppm. Such small chemical shift differences may reflect a slow exchange situation in which the resonance of bound ligand is too broad to observe and the resonance of the free ligand was seen very intense.

Line shapes of exchanging ligands can be predicted based on simple exchange rate models [29]. The appearance of resonances in the commonly accepted three exchange regimes are described as following. At slow exchange (10 sec^{-1}), two resonance lines are observed. As the exchange rate increases (100 sec^{-1}), the resonance lines broaden with each acquiring a line width contribution on the order of the reciprocal of the lifetime in that state. This means the line representing the lower population state is always broader. When the exchange rate (450 sec^{-1}) is of the order of the chemical shift separation between the two sites, the lines become very broad and begin to coalesce or even disappear in the baseline noise. Increasing the exchange rate above the coalescence point pushes the system into fast exchange (1000 sec^{-1} to 5000 sec^{-1}) where a single signal is observed at the weighted average resonance frequency for the two sites according to their molar fractions in the free and bound states.

Protein-bound CMP-3FNeuAc constitutes a very small fraction of the total ligand concentration and even in the slow exchange regime has a relatively short life time. Because of the binding and exchange broadening effects, the resonances of CMP-3FNeuAc in its bound state can be too broad to detect. As the temperature rises from 15

^{o}C to 35 ^{o}C, the observed CMP-3FNeuAc peak becomes broader (Figure 2.12), which indicates that the free and bound ligands were in slow exchange. Raising the temperature expedited the exchange from slow to medium and this process caused line broadening. Given this mechanism it is possible to derive a bound state life time using the formula $T_2^{-1} = T_{2A}^{-1} + R_{AB}$ from the literature[30]. A represents the off-site, B represents the on-site, R_{AB} represents the transition rate between on and off site, and the value of T2 is equal to $(\Delta\nu^* \pi)^{-1}$ in which $\Delta\nu$ is the linewidth. Given that 25% of CMP-3FNeuAc is in bound state (estimated from its dissociation constant), and a broadening of 10 Hz at 35°C, the bound state life time would be 126 s^{-1}.

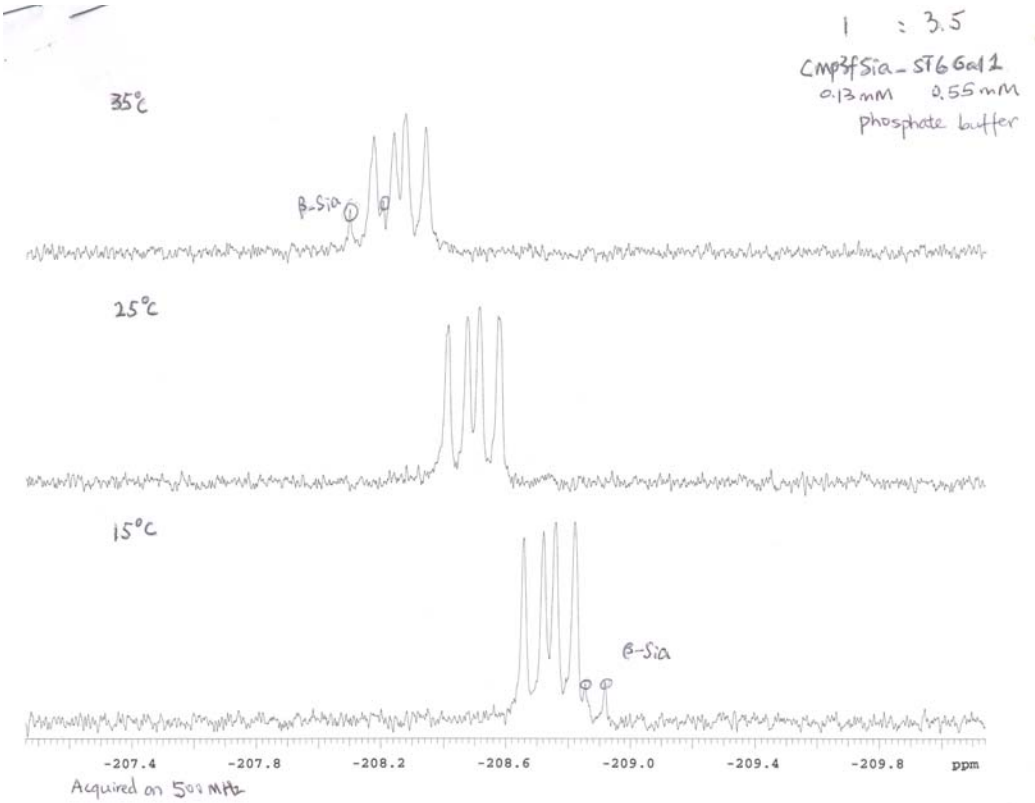


Figure 2.12. 1D ^{19}F NMR spectra of the CMP-3FNeuAc-ST6Gal-1 complex acquired at different temperatures (15^{o}C, 25^{o}C and 35^{o}C). The linewidth gets broader as the temperature increases. The small amount of impurity shown was identified as β-NeuAc.

2) Exchange rate confirmation from HSQC spectra of ^{15}N-ST6Gal-1

Uniformly ^{15}N-labeled ST6Gal-1 (pichia pastoris) was used to verify that the CMP-3FNeuAc was actually binding and to further evaluate the exchange rate. When a five-fold molar excess of CMP-3FNeuAc was added to U-^{15}N ST6Gal-1 (66% bound if K_I 0.65 mM used), it was observed in the HSQC spectrum that a few peaks weakened and broadened dramatically and a number of protein resonances split into two (Figure 2.13). The splitting, broadening and attenuation of protein resonances are features of the slow exchange of free and bound ligands as well as signs of protein dynamics changing upon CMP-3FNeuAc binding.

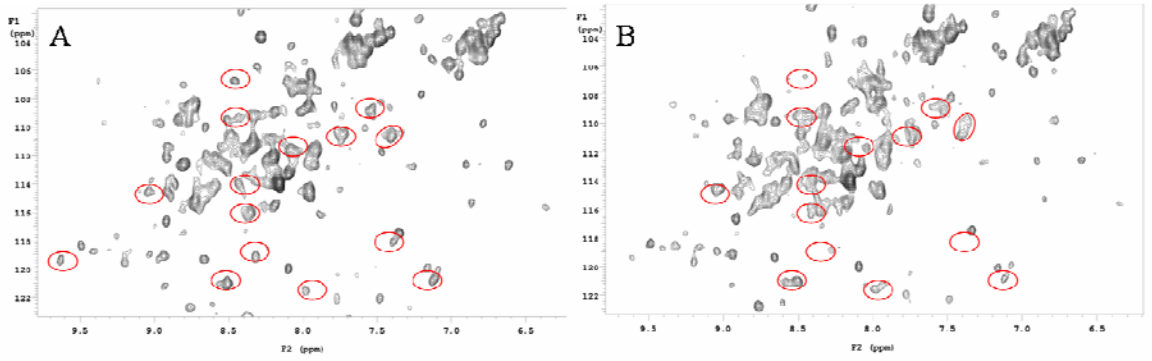


Figure 2.13. Comparison of ^{1}H-^{15}N HSQC plots of (A) U-^{15}N ST6Gal-1 before and (B) after the addition of CMP-3FNeuAc. The crosspeaks experiencing substantial changes are circled.

3) T_2 measurements of acetyl protons of CMP-3FNeuAc

T_2 measurements of acetyl protons on the ligand were also used to verify binding in a simple two-state model. One would expect resonances unaffected by exchange broadening to adopt spin relaxation contributions proportionate to bound populations and weighted by expected bound state relaxation.

We measured T_2 of acetyl protons of CMP-3FNeuAc under a variety of conditions using the cpmgt2 sequence of T_2 relaxation measurement in the Varian Bio-Pak. The results are presented in Table 2. The difference of proton frequencies at the on and off sites is expected to be very small or even indistinguishable, especially for groups such as acetyl protons which do not interact strongly with ST6Gal-1. The peak we measured seems to be the sum resonances of free and bound ligand. As shown in the Table 2.2, the T_2 of the acetyl protons of CMP-3FNeuAc decreases when CMP-3FNeuAc binds to ST6Gal-1. This demonstrates that CMP-3FNeuAc adopts the longer correlation time of ST6Gal-1. When an excess of CMP-3FNeuAc was added, T_2 adopted a simple ascendant trend, indicating there is no apparent secondary binding of CMP-3FNeuAc to ST6Gal-1.

Table 2.2. T_2 values of acetyl protons of CMP-3FNeuAc in free solution and in ST6Gal-1 solution with the indicated ligand-protein ratio.

Sample	T_2 (sec) at 25 ^{o}C
CMP3FNeuAc in phosphate buffer (pH 6.8)	0.451 ± 0.007
CMP3FNeuAc in D_2O (pH 6.0)	0.465 ± 0.001
CMP3FNeuAc/ ST6Gal-1 = 1:1 in Bis-tris (pH 6.5)	0.122 ± 0.012
CMP3FNeuAc/ ST6Gal-1 = 5:1 in Bis-tris (pH 6.5)	0.242 ± 0.008
CMP3FNeuAc/ ST6Gal-1 = 10:1 in Bis-tris (pH 6.5)	0.299 ± 0.006
CMP3FNeuAc/ ST6Gal-1 = 20:1 in Bis-tris (pH 6.5)	0.343 ± 0.003

In summary, for the ^{19}F experiments, no significant chemical shift change was detected for the observed CMP-3FNeuAc resonance. However, accelerating exchange by raising the temperature broadened the resonance, which suggested that CMP-3FNeuAc is in slow exchange between free and bound states; the slow exchange phenomenon also was confirmed when the resonance from several protein residues interacting with the ligand was observed to split in the HSQC spectrum of ^{15}N-ST6Gal-1. From these data, we concluded that the CMP-3FNeuAc-ST6Gal-1 complex is in a slow exchange regime. The resonance of bound CMP-3FNeuAc could not be observed because of its low population under relatively weak binding conditions and the more extensive exchange broadening that occurs for minor population species in slow exchange.

2.4 Conclusion

We have investigated the structural interactions of donor and acceptor substrates within the active site of ST6Gal-1. The traditional transferred NOE (trNOE) and saturation transfer difference (STD) methodologies yielded an imprint of the active site geometry without the need for a structure determination of the protein. We employed trNOE experiments to verify the substrates' binding to ST6Gal-1 and confirmed the trans-conformation about the cytosine-ribose glycosidic bond of CMP-NeuAc and CMP-3FNeuAc as well as population of a low energy conformation about the glycosidic bond of LacNAc in the presence and absence of ST6Gal-1. Our STD data for CMP-NeuAc is consistent with the enzymatic reaction mechanism in that cytidine binds more tightly to ST6Gal-1 than NeuAc, the leaving group. By using ^{13}C-labeled CMP-NeuAc we acquired a set of residual dipolar couplings (RDCs) and calculated the preferable orientation and conformation for CMP-NeuAc in free solution. RDC data on a bound

complex will be collected with the same principle to determine bound ligand geometry when ST6Gal-1 alignment is enhanced with the methods tested on Galectin-3 [10; 12; 13]. Fluorine NMR is a convenient and powerful tool to assess the kinetics and binding affinity of substrates with the incorporation of fluorinated ligands. With simple 1D ^{19}F-NMR experiments, we characterized the slow exchange and weak binding properties for the CMP-3FNeuAc-ST6Gal-1 system. All of the above data provide a framework for understanding how the substrates interact with the active site of the sialyltransferase ST6Gal-1. The strategies we developed will also be valuable for probing the structures of other glycoproteins.

References

1. Macnaughtan, M. A., Kamar, M., Alvarez-Manilla, G., Venot, A., Glushka, J., Pierce, J. M. & Prestegard, J. H. (2007). NMR structural characterization of substrates bound to N-acetylglucosaminyltransferase V. *Journal of Molecular Biology* **366**, 1266-1281.

2. Mayer, M. & Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **38**, 1784-1788.

3. Meyer, B. & Peters, T. (2003). NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie-International Edition* **42**, 864-890.

4. Siebert, H. C., Jimenez-Barbero, J., Andre, S., Kaltner, H. & Gabius, H. J. (2003). Describing topology of bound ligand by transferred nuclear Overhauser effect

spectroscopy and molecular modeling. *Recognition of Carbohydrates in Biological Systems Pt A: General Procedures* **362**, 417-434.

5. Watts, A. G. & Withers, S. G. (2004). The synthesis of some mechanistic probes for sialic acid processing enzymes and the labeling of a sialidase from Trypanosoma rangeli. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **82**, 1581-1588.

6. Knorst, M. & Fessner, W. D. (2001). CMP-sialate synthetase from Neisseria meningitidis - Overexpression and application to the synthesis of oligosaccharides containing modified sialic acids. *Advanced Synthesis & Catalysis* **343**, 698-710.

7. Tugarinov, V., Hwang, P. M. & Kay, L. E. (2004). Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annual Review of Biochemistry* **73**, 107-146.

8. Chen, Z. J. & Reif, B. (2004). Measurements of residual dipolar couplings in peptide inhibitors weakly aligned by transient binding to peptide amyloid fibrils. *Journal of Biomolecular Nmr* **29**, 525-530.

9. Jain, N. U., Noble, S. & Prestegard, J. H. (2003). Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings. *Journal of Molecular Biology* **328**, 451-462.

10. Seidel, R. D., Zhuang, T. D. & Prestegard, J. H. (2007). Bound-state residual dipolar couplings for rapidly exchanging ligands of His-tagged proteins. *Journal of the American Chemical Society* **129**, 4834-4839.

11. Shimizu, H., Donohue-Rolfe, A. & Homans, S. W. (1999). Derivation of the bound-state conformation of a ligand in a weakly aligned ligand-protein complex. *Journal of the American Chemical Society* **121**, 5815-5816.

12. Zhuang, T. D., Lee, H. S., Imperiali, B. & Prestegard, J. H. (2008). Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. *Protein Science* **17**, 1220-1231.

13. Zhuang, T. D., Leffler, H. & Prestegard, J. H. (2006). Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3. *Protein Science* **15**, 1780-1790.

14. Wang, X. F., Zhang, L. H. & Ye, X. S. (2003). Recent development in the design of sialyltransferase inhibitors. *Medicinal Research Reviews* **23**, 32-47.

15. Heidlas, J. E., Lees, W. J., Pale, P. & Whitesides, G. M. (1992). Gram-Scale Synthesis of Uridine 5'-Diphospho-N-Acetylglucosamine - Comparison of Enzymatic and Chemical Routes. *Journal of Organic Chemistry* **57**, 146-151.

16. Aubin, Y. & Prestegard, J. H. (1993). Structure and Dynamics of Sialic-Acid at the Surface of a Magnetically Oriented Membrane System. *Biochemistry* **32**, 3422-3428.

17. Tengel, T., Fex, T., Emtenas, H., Almqvist, F., Sethson, I. & Kihlberg, J. (2004). Use of F-19 NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Organic & Biomolecular Chemistry* **2**, 725-731.

18. Gerig, J. T. (1998). Fluorine NMR. *Biophysics Textbook On-line* **NMR section chapter 24**, 23-24.

19. Peng, J. W. (2001). Cross-correlated F-19 relaxation measurements for the study of fluorinated ligand-receptor interactions. *Journal of Magnetic Resonance* **153**, 32-47.

20. Fielding, L. (2003). NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry* **3**, 39-53.

21. Jenkins, B. G. & Lauffer, R. B. (1990). Detection of Site-Specific Binding and Co-Binding of Ligands to Human-Serum-Albumin Using F-19 Nmr. *Molecular Pharmacology* **37**, 111-118.

22. Nakajima, T., Hori, H., Ohrui, H., Meguro, H. & Ido, T. (1988). Synthesis of N-Acetyl-3-Fluoro-Neuraminic Acids. *Agricultural and Biological Chemistry* **52**, 1209-1215.

23. Blume, A., Angulo, J., Biet, T., Peters, H., Benie, A. J., Palcic, M. & Peters, T. (2006). Fragment-based screening of the donor substrate specificity of human blood group B galactosyltransferase using saturation transfer difference NMR. *Journal of Biological Chemistry* **281**, 32728-32740.

24. Valafar, H. & Prestegard, J. H. (2004). REDCAT: a residual dipolar coupling analysis tool. *Journal of Magnetic Resonance* **167**, 228-241.

25. Bollejones, E. W. (1954). Occurrence of Indole-3-Acetic Acid in Laminae of Hevea-Brasiliensis. *Nature* **173**, 127-128.

26. Burkart, M. D., Vincent, S. P. & Wong, C. H. (1999). An efficient synthesis of CMP-3-fluoroneuraminic acid. *Chemical Communications*, 1525-1526.

27. Lerman, O. & Rozen, S. (1983). Acetyl Hypofluorite, a New Moderating Carrier of Elemental Fluorine and Its Use in Fluorination of 1,3-Dicarbonyl Derivatives. *Journal of Organic Chemistry* **48**, 724-727.

28. Burkart, M. D., Vincent, S. P., Duffels, A., Murray, B. W., Ley, S. V. & Wong, C. H. (2000). Chemo-enzymatic synthesis of fluorinated sugar nucleotide: Useful mechanistic probes for glycosyltransferases. *Bioorganic & Medicinal Chemistry* **8**, 1937-1946.

29. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. (1996). Protein NMR spectroscopy: principles and practice. *Academic Press* **chapter 5**, 295-297.

30. Carrington, A. & McLachlan, A. D. (1967). Introduction to magnetic resonance. *Harper & Row, publishers* **chapter 12**, 208.

CHAPTER 3

NMR CHARACTERIZATION OF THE ACTIVE SITE OF ST6GAL-1 BY SPIN-LABELED SUBSTRATES

---

[1] Liu, S., A. Venot, L. Meng, F. Tian, K. W. Moremen, G-J. Boons, and J. H. Prestegard. 2007. *Chemistry & Biology (Cell)*. 14 (4): 409-418
Reprinted here with permission of publisher.

[2] Liu, S., L. Meng, K. W. Moremen, and J. H. Prestegard. To be submitted to *Biochemistry*

3.1 Introduction

Structural determination of proteins by NMR methods has traditionally relied on the detection of large numbers of nuclear Overhauser effects (NOEs). Uniform labeling with ^{13}C, ^{15}N, and ^{2}H provides the improved resolution necessary for the assignment of sufficient numbers of resonances to conduct an NOE-based structure determination [1; 2]. The labeling is usually accomplished by expressing proteins in bacterial hosts which efficiently utilize low-cost ^{13}C- and ^{15}N- enriched substrates. However, this method is not practical for ST6Gal-1 which requires glycosylation for proper folding or full activity and affords low yield in bacterial media. ST6Gal-1 is normally expressed in eukaryotic hosts that require supplementation with a nearly complete set of labeled amino acids, some of which are very expensive [3; 4]. It is therefore important to consider alternate sources of structural data that can be used with the sparse labels introduced by expressing proteins in eukaryotic cells grown on media supplemented with selected sets of less expensive ^{15}N labeled amino acids [5]. One source applicable to this type of labeling relies on paramagnetic perturbation of labeled sites on introducing nitroxide carrying probes that are either substrate mimics or are covalently attached to the protein itself [6; 7; 8; 9; 10; 11; 12; 13; 14]. Here we present probes that mimic natural substrates that bind to the active site of ST6Gal-1. The sugar donor mimicked is CMP-NeuAc and the sugar acceptor mimicked is N-acetyl-lactosamine (LacNAc). We first used spin-labeled donor and acceptor analogs to identify the labeled amino acids in the protein active site and map distances to these amino acid sites. Then related paramagnetic strategies are used to characterize the position of the sugar acceptor relative to the donor analog while both are simultaneously bound to ST6Gal-1.

The essential component of the nitroxide spin label we chose is the chemically stable moiety, TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl). The nitroxide group contains an unpaired electron that drastically increases the relaxation of neighboring protons due to the large magnitude of the electron magnetic moment. The interactions between electron and proton magnetic moments extend to 20 Å, a distance much longer than for proton-proton NOEs which are limited to distances of 6 Å or less [1; 2]. In more quantitative terms, when the nitroxide spin-labeled compound binds to the protein target, the nuclei of protein residues experience enhanced spin relaxation with a $1/r^6$ dependence on the distance, r, between the nitroxide group and the nuclear site [6; 7; 8; 9; 10; 11; 12; 13; 14]. Resonances are broadened, longitudinal spin relaxation times are shortened, and cross-peak intensity in most multidimensional experiments is reduced.

One method we used for detecting spin relaxation enhancements depends on intensity loss of cross-peaks in ^{1}H-^{15}N HSQC spectra of proteins [6; 11; 12]. The loss is primarily due to the transverse relaxation of protons during INEPT transfers to nitrogen and back, and the intensity loss can be directly correlated with the inverse sixth power of the distance between the nitroxide group and the protons giving rise to each cross-peak. The general procedure is to acquire a spectrum of the protein with the spin-labeled ligand in its oxidized paramagnetic form, then reduce the nitroxide to a diamagnetic species with an ascorbate salt, and repeat the spectral acquisition under as identical conditions as possible. The difference in cross-peak intensities between the oxidized and reduced spectra should be directly attributable to the effects of the nitroxide electron spin. Cross-peaks from protons in the ligand binding site are easily identified, and distances to protons in more remote sites can be estimated.

Similar distance-dependent perturbations can be measured for the ligand in lieu of the protein. The protein sample does not need isotope labeling, and sites in the ligand (usually ^{1}H-^{13}C) can be detected at natural abundance when rapid exchange exists and a sufficiently high concentration of the ligand can be used. If necessary, it is also relatively easy and inexpensive to isotopically enrich sites in the ligand with ^{13}C. In principle, relaxation perturbation of an exchanging ligand in the presence of a second exchanging ligand carrying a nitroxide group can be measured from any number of relaxation parameters. There are certain advantages in measuring longitudinal relaxation times (T_1) rather than the transverse relaxation time (T_2) often measured in protein systems. One of the reasons for this choice is that exchanging ligands usually suffer some resonance broadening due to intermediate exchange rates upon binding to the protein. This exchange broadening effect can be difficult to separate from the paramagnetic T_2 changes. As long as the exchange is fast compared to T_1, these rates are not affected by the exchange process. In choosing between T_1s of protons or carbons, T_1s of carbon have an advantage because T_1s of protons in large complexes suffer from spin diffusion.

In developing appropriate paramagnetic analogs, we used CMP-NeuAc (Figure 3.1), the sugar donor in the sialylation reaction catalyzed by ST6Gal-1, as a model. In theory, the spin label may be introduced to either the nucleotide base or the sugar ring. Previous studies involving the design of sialyltransferase inhibitors show that the cytidine nucleotide portion is more critical for tight binding to sialyltransferases than the neuraminic acid residue [15]. Nucleotide base-modified CMP-NeuAc analogs are not tolerated by α–2, 6-sialyltransferase while sugar-modified analogs with derivatization of the neuraminic acid residues at the C5, C8 or C9 positions show affinities for ST6Gal-1 [16].

Substitution of the proton at C3 with fluorine leads to another potentially useful inhibitor of sialyltransferases [17]. Therefore, we initiated studies in which TEMPO, the spin label nitroxide moiety, would be linked to the phosphate of CMP through its C4 carbon, taking the place of the neuraminic acid in CMP-NeuAc. The 6-membered ring of TEMPO mimics the 6-membered ring of the neuraminic acid. The additional methyl groups at C2 and C6 in TEMPO sit in the positions occupied by C4 substituents and the C7-C8 extension of neuraminic acid, and the nitroxide replaces the C5 acetyl group of neuraminic acid. To further examine the logic of this replacement, CMP-TEMPO was modeled into the crystal structure of CstII in which CMP-3FNeuAc occupies the sugar donor site. There were no severe steric conflicts identified. However, there was a noticeable absence of a carboxyl group that apparently interacts with positively-charged amino acids in the binding site. Therefore we also synthesized CMP-4carboxyTEMPO (Figure 3.1) in which a carboxyl group is attached to C4 of TEMPO and effectively mimics the carboxyl attached to C2 of neuraminic acid. CMP-4carboxyTEMPO is demonstrated here to be a good mimic of CMP-NeuAc and a useful structural probe for ST6Gal-1.

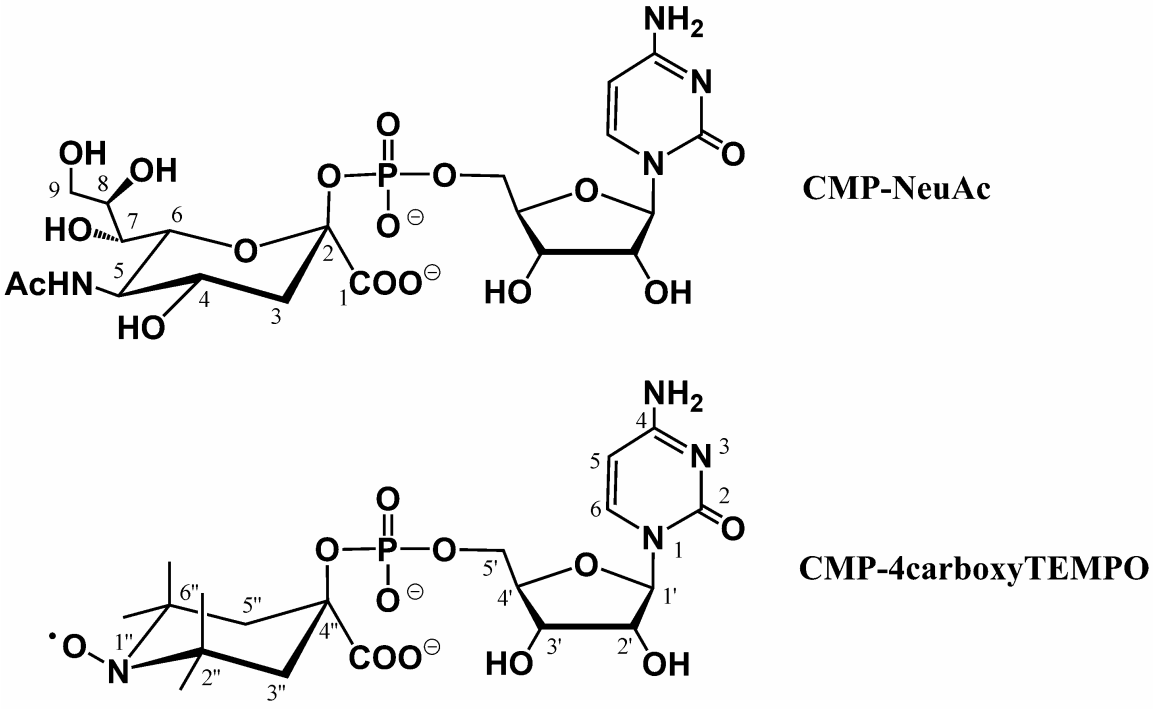


Figure 3.1. Structure of natural donor CMP-NeuAc and its analog CMP-4carboxyTEMPO

In what follows, a spin-labeled donor analog, CMP-4carboxyTEMPO, and a spin-labeled acceptor analog, LacNAc-TEMPO [11], were used to observe changes in ^{15}N-phe-labeled ST6Gal-1 resonances. Then CMP-4carboxyTEMPO was used to perturb resonances of a non spin-labeled acceptor, LacNAc. Distances between the nitroxide of the donor and the acceptor were estimated based on carbon T_1 changes for the acceptor in the presence and absence of the spin-labeled donor. A ^{13}C-labeled version of LacNAc was incorporated to improve the precision of relaxation rate measurements which were conducted via indirect detection of carbons through protons in arrayed HSQC experiments.

3.2 Experimental

3.2.1 Chemical synthesis of CMP-TEMPO (1) and CMP-4carboxyTEMPO (2)

General procedures. Chemicals and dry solvents were purchased from Aldrich (USA). The commercial 4-oxo-TEMPO was purified by chromatography (silica gel, hexanes: EtOAc, 82:18). All moisture-sensitive reactions were performed under an atmosphere of dry argon. Reactions were monitored by TLC on silica gel Kieselgel 60 F_{254} (Merck) and visualized under UV light and by charring with a ceric ammonium molybdate reagent. All evaporations were performed under appropriate reduced pressure at bath temperature < 30 ^{o}C. Chromatography was performed on silica gel Merck 70-230 mesh or, as indicated, on Iatrobeads 6RS-8060 (60 μm, Bioscan), or for reverse phase on a Waters preparative C18 WAT020594 column (125 Å, 55-105 μm). All NMR spectra of the TEMPO derivatives were acquired on Varian 300, 500 or 600 MHz NMR spectrometers, in the indicated solvents, after addition of phenyl hydrazine as a reducing agent to the nitroxide sample [18]. Mass spectra were obtained on an Applied Biosystems Voyager-DE MALDI-TOF mass spectrometer using DHB (2, 5-dihydroxybenzoic acid) as a matrix.

2-Cyanoethyl 1-oxyl-2,2,6,6-tetramethylpiperidin-4-yl N,N-diisopropyl phosphoramidite (4) [19]. Diisopropylethylammine (297 mg, 2.30 mmol) was added to a solution of 4-hydroxy-TEMPO **3** (198 mg, 1.15 mmol) and 2-cyanoethyl N,N-diisopropylchlorophosphoramidate (407 mg, 1.72 mmol) in dry CH_2Cl_2 (5 ml) cooled at 0 ^{o}C in an argon atmosphere. The mixture was then stirred for 1 hour at room temperature until TLC (hexanes: EtOAc 3:1) indicated the completion of the reaction. The solvent was co-evaporated with toluene and the residue was purified by flash chromatography (8

g Iatrobeads, hexanes: EtOAc: NEt_3 4:1:0.2%) to afford **4** (164 mg, 38%) as an orange syrup (R_f 0.7, hexanes: EtOAc 3:1). No NMR data are available for **4** due to the rapid decomposition observed in the NMR tube after addition of phenyl hydrazine. MS data are not available for **4** due to its instability in the matrix.

2-Cyanoethyl (2',3'-O,N^4-triacetylcytidin-5'-yl) 1-oxyl-2,2,6,6-tetramethylpiperidinyl-4-yl phosphate (6) [20]**.** Compounds **4** and **5** [21] were separately dried by coevaporating three times with dry toluene. Tetrazole (1 mL of a 3% solution in MeCN) was slowly syringed into a mixture of phosphoramidite **4** (160 mg, 0.43 mmol) and the cytidine **5** (125 mg, 0.38 mmol) in MeCN (2 mL) at 0 ^oC. The reaction was followed by TLC ($CHCl_3$: MeOH 95:5) and seemed complete after stirring for about 1 hour at room temperature. After addition of 3 drops of triethylamine, the solvents were co-evaporated with toluene providing an orange material (0.30 g) which was used in the next step without purification. A 5.5 M solution of tBuOOH in decane (0.47 mL) was slowly added to the crude material (0.27 g) dissolved in MeCN (4 mL) while stirring in ice. The reaction was followed by TLC ($CHCl_3$: MeOH 95:5). After 1 hour, the mixture was diluted with EtOAc and washed with a saturated $NaHCO_3$ solution followed by brine. After drying over $MgSO_4$, the solvent was evaporated and the residue was purified by chromatography (5 g Iatrobeads, $CHCl_3$: MeOH 97:3) to afford an orange product **6** (160 mg, 64%) as a 1:1 diastereoisomeric mixture (R_f 0.4, $CHCl_3$: MeOH 95:5). ^1H NMR (500 MHz, CD_3OD) for the reduced form: δ 8.13 (d, J = 7.0 Hz, 1H, H-6), 8.11 (d, J = 7.5 Hz, 1H, H-6), 7.48-7.40 (m, 2H, 2 x H-5), 6.03-6.02 (m, 2H, 2 x H-1'), 5.52-5.44 (m, 4H, 2 x 2H of H-2'& H-3'), 4.74-4.68 (m, 2H, 2 x H-1"), 4.48-4.41 (m, 4H, 2 x 2H of H-4' & H-5'), 4.40-4.33 (m, 2H, H-5'), 4.30 (p, J = 6.5 Hz, 4H, 2 x CH_2CN), 2.95-2.86 (m, 4H, 2

x OC<u>H$_2$</u>CH$_2$CN), 2.18 (s, 6H, 2 x CH$_3$CO), 2.10 (d, J = 2 Hz, 12H, 4 x CH$_3$CO), 2.10-2.01 (m, 4H, 2 x 2H of H-3" & H-5"), 1.70-1.61 (m, 4H, 2 x 2H of H-3" & H-5"), 1.22-1.10 (m, 24H, 8 x CH$_3$); MALDI-TOF/MS: (M + Na^+) m/z 679.4, calcd for $C_{27}H_{39}N_5O_{12}PNa$ 679.2225.

CMP-TEMPO (1) [20]**.** 1,8-Diazabicyclo[5,4,0]-7-undecene (DBU) (2.7 mg, 0.018 mmol) was added to a solution of the above material **6** (10 mg, 0.015 mmol) in THF (0.8 ml) and the mixture was stirred at room temperature for 45 minutes until TLC ($CHCl_3$: MeOH 95:5) indicated the completion of the reaction. A 5.4 M solution of NaOMe in MeOH (0.028 ml, 0.15 mmol) and then a mixture of 1:2 MeOH-H_2O (1.5 ml) were added to this mixture. After stirring for 12 hours at room temperature, the mixture was lyophilized providing a residue which was then loaded onto a small column of reverse-phase C18 eluted with water to remove NaOH and NaOAc, and then run through a small column of Bio-Rad AG 50W-X8 Na^+ eluted with water to remove the DBU byproduct. After freeze drying, the recovered material was purified by size exclusion chromatography on Bio-Gel P2 (extra fine, 1.5 cm x 45 cm , water, 4 ^oC) and lyophilization to produce CMP-TEMPO **1** (5.5 mg, 73%) as a yellow powder (iPrOH:0.5 M AcOH•NEt_3 5:2, R_f 0.7; HPTLC RP-18 $WF_{254S,}$ R_f 0.2, water). ^{1}H NMR (500 MHz, CD_3OD) for the reduced form: δ 8.11 (d, J = 7.5 Hz, 1H, H-6), 5.99-5.96 (m, 2H, H-5 & H-1'), 4.49-4.38 (m, 1H, H-1"), 4.28-4.00 (m, 5H, H-3', H-5', H-2', H-4' & H-5'), 1.80-1.71 (m, 2H, H-3" & H-5"), 1.56 (td, J = 3.5, 12.0 Hz, 2H, H-3" & H-5"), 1.15 (s, 3H, CH_3), 1.13 (s, 6H, 2 x CH_3), 1.09 (s, 3H, CH_3); ^{31}P NMR (500 MHz, CD_3OD, external standard H_3PO_4 = 0.00) δ 0.72; ^{13}C NMR (500 MHz, CD_3OD) for the reduced form: δ 167.1 (C-2), 158.1 (C-NH_2), 142.5 (C-6), 96.0 (C-5), 90.7 (C-1'), 83.9 (d, J = 8.8, C-4'),

75.9 (C-2'), 70.2 (C-3'), 68.8 (d, J = 6.0, C-4''), 64.5 (d, J = 5.5, C-5'), 59.6 (C-2'' & C-6''), 46.9 (d, J = 4.0, C-3''/ C-5''), 46.8 (d, J = 4.0, C-3''/ C-5''), 32.2 (CH_3), 20.4 (CH_3); MALDI-TOF/MS for the reduced form: ($M + Na^+$) m/z 501.3, calcd for $C_{18}H_{31}N_4O_9PNa$ 501.1721.

1,4-Dihydroxy 2,2,6,6-tetramethyl-4-piperidinecarboxylic acid (9) [22]. A solution of the cyanide **8** [23] (1.07 g, 5.35 mM) in 27% HCl (4.8 mL) was stirred for 2 hours at 80 ^oC. Co-evaporation with 80 mL water under reduced pressure (bath temperature < 30 ^oC) left an acidic yellowish oily solid which was left at -20 ^oC overnight. The acidic supernatant was pipetted off and the residual solid washed three times with ether and further dried *in vacuo* to yield the crude N-hydroxyde **9** (750 mg). ^1H NMR (300 MHz, D_2O): δ 2.38 (d, J = 14.9 Hz, 2H, H-3 & H-5), 2.20 (d, J = 14.9 Hz, 2H, H-3 & H-5), 1.635 (s, 6H, 2 x CH_3), 1.428 (s, 6H, 2 x CH_3). MALDI-TOF/MS: ($M + Na^+$) m/z 240.1, calcd. for $C_{10}H_{19}NO_4Na$ 240.1206. The material **9** was taken to the next step without further purification.

4-Hydroxy-4-methoxycarbonyl-1-oxyl-2,2,6,6-tetramethylpiperidine (10). A solution of diazomethane in ether was added to a solution of crude **9** (750 mg) in MeOH (10 mL) at 0 ^oC until the solution became deep yellow. The mixture was then stirred in an open flask for 2 days [24; 25]. Formation of the nitroxide (upper spot) was followed by TLC ($CHCl_3$: MeOH 95:5) and a small amount of N-hydroxyl starting material remained untransformed. Purification by chromatography (20 g Iatrobeads, $CHCl_3$: MeOH 99:1) gave **10** which was further purified by crystallization from a mixture of ether and hexanes at -20 ^oC providing **10** (684 mg, 55%) as orange crystals (R_f 0.4, hexanes:EtOAc 3:1). ^1H NMR (300 MHz, CD_3OD) for the reduced form: δ 3.71 (s, 3H, CH_3OCO), 2.04 (d, J =

13.6 Hz, 2H, H-3 & H-5), 1.80 (d, J = 13.6 Hz, 2H, H-3 & H-5), 1.35 (s, 6H, 2 x CH_3), 1.16 (s, 6H, 2 x CH_3); MALDI-TOF/MS for the reduced form: ($M + Na^+$) m/z 254.1, calcd for $C_{11}H_{21}NO_4Na$ 254.1363.

2-Cyanoethyl 4-methoxycarbonyl-1-oxyl-2,2,6,6-tetramethylpiperidin-4-yl) N,N-diisopropylphosphoramidite (11) [19]**.** Diisopropylethylammine (168 mg, 1.30 mmol) was added to a solution of the above TEMPO derivative **10** (150 mg, 0.65 mmol) and 2-cyanoethyl N,N-diisopropylchlorophosphoramidate (231 mg, 0.98 mmol) in dry CH_2Cl_2 (2 ml) cooled at 0 ^oC in an argon atmosphere. The mixture was then stirred for 24 hours at room temperature until TLC (hexanes: EtOAc 3:1) indicated the completion of the reaction. The solvent was co-evaporated with toluene and the residue was purified by flash chromatography (5 g Iatrobeads, hexanes: EtOAc: NEt_3 4:1:0.2%) to afford **11** (240 mg, 86%) as a brown syrup (R_f 0.5, hexanes:EtOAc 3:1). ^1H NMR (300 MHz, CD_3OD) for the reduced form: δ 3.77-3.66 [m, 4H, $OC\underline{H}_2CH_2CN$ & 2 x $C\underline{H}$ $(CH_3)_2$], 3.71 (s, 3H, CH_3OCO), 2.67 (t, J = 5.9 Hz, 2H, CH_2CN), 2.25-1.87 (m, 4H, H-3 & H-5), 1.33-1.12 (m, 24H, 8 x CH_3). No MS data is available for **11** due to its rapid decomposition in the matrix.

2-Cyanoethyl 2',3'-O,N^4-triacetylcytidin-5'-yl 4-carboxy-1-oxyl-2,2,6,6-tetramethylpiperidin-4-yl phosphate (12) [20]**.** Compounds **5** [21] and **11** were dried by coevaporating two times with dry toluene. Tetrazole (0.7 ml, 0.45 M solution in MeCN) was then syringed into a mixture of phosphoramidite **11** (80 mg, 0.186 mmol) and the cytidine **5** (75 mg, 0.203 mmol) in MeCN (2.5 mL) was cooled at -40^oC. After 5 minutes the reaction mixture was brought at room temperature and stirred with one weight equivalent of activated crushed 3Å molecular sieves for 2 hours. The reaction was

followed by TLC (EtOAc: MeOH 9:1). A 5.5 M solution of tBuOOH in decane (0.8 mL) was then added directly to the crude mixture at 0 ^{o}C. The reaction was monitored by TLC (EtOAc: MeOH 9:1). After 1 hour, the mixture was diluted with EtOAc and washed with a saturated $NaHCO_3$ solution and brine. After drying over $MgSO_4$, the solvent was evaporated and the residue was purified by size exclusion chromatography on Sephadex LH-20 (2.5 x 45 cm) using MeOH as the eluent to give the phosphate **12** (53mg, 40%) as a 1:1 diastereomeric mixture (R_f 0.3, EtOAc: MeOH 9:1). ^{1}H NMR (300 MHz, CD_3OD) for the reduced form: δ 8.17 (d, J = 7.6 Hz, 1H, H-6), 8.16 (d, J = 7.6 Hz, 1H, H-6), 7.49 (d, J = 7.5 Hz, 1H, H-5), 7.48 (d, J = 7.5 Hz, 1H, H-5), 6.06 (d, J = 6.3 Hz, 1H, H-1'), 6.04 (d, J = 6.3 Hz, 1H, H-1'), 5.51-5.48 (m, 4H, 2 x 2H of H-2' & H-3'), 4.60-4.40 (m, 6H, 2 x 3H of H-4' & H-5'), 4.34 (q, J = 6.13 Hz, 4H, 2 x CH_2CN), 3.79 (s, 6H, 2 x CH_3OCO), 2.89-2.86 (m, 4H, 2 x $OC\underline{H}_2CH_2CN$), 2.22-2.06 (m, 8H, 2 x 4H of H-3" & H-5"), 2.17 (s, 6H, 2 x CH_3CO), 2.10 (d, J = 1.9 Hz, 12H, 4 x CH_3CO), 1.33-1.29 (m, 12H, 4 x CH_3), 1.17 (s, 12H, 4 x CH_3); MALDI-TOF/MS for the reduced form: ($M + Na^+$) m/z 738.1, calcd for $C_{29}H_{42}N_5O_{14}PNa$ 738.2358.

CMP-4carboxyTEMPO (2) [20]**.** 1,8-Diazabicyclo[5,4,0]-7-undecene (14.7 mg, 0.096 mmol) was added to a solution of the above material **12** (53 mg, 0.074 mmol) in THF (0.8 ml) and the mixture was stirred at room temperature for 1 hour until TLC (EtOAc: MeOH 9:1) indicated the completion of the reaction. A 5.4 M solution of NaOMe in MeOH (0.083 ml, 0.45 mmol) and then a mixture of 1:2 MeOH-H_2O (1.5 ml) were added to this mixture. After stirring for 9 hours at room temperature, the mixture was lyophilized. The residue was loaded onto a small column of reverse-phase C18, eluted with water to remove NaOH and NaOAc, and then run through a small column of

Bio-Rad AG 50W-X8 Na^+ eluted with water to remove the DBU byproduct. After freeze-drying, the recovered material was purified by size exclusion chromatography on Bio-Gel P2 (extra fine, 1.5 cm x 45 cm, water, at 4°C) and lyophilization to produce CMP-4carboxyTEMPO **2** (31 mg, 74%) as a yellow powder (iPrOH:0.5 M AcOH•NEt_3 5:2, R_f 0.5; HPTLC RP-18 WF_{254S}, R_f 0.4, water)**.** ^1H NMR (600 MHz, CD_3OD) for the reduced form: δ 8.22 (d, J = 7.8 Hz, 1H, H-6), 6.12 (d, J = 7.8 Hz, 1H, H-5), 5.97 (d, J = 3.0 Hz, 1H, H-1'), 4.39 (t, J = 5.5 Hz, 1H, H-3'), 4.32-4.24 (m, 1H, H-5'), 4.24-4.18 (m, 1H, H-5'), 4.17-4.10 (m, 2H, H-2' & H-4'), 2.51 (d, J = 14.6 Hz, 1H, H-3'' or H-5''), 2.40 (d, J = 14.6 Hz, 1H, H-3'' or H-5''), 2.21 (t, J = 14.8 Hz, 2H, H-3'' & H-5''), 1.44 (d, J = 13.8 Hz, 6H, 2 x CH_3), 1.23 (s, 6H, 2 x CH_3); ^{13}C NMR (500 MHz, CD_3OD) for the reduced form: δ 181.6 (COOH), 167.2 (C-2), 158.1 (C-NH_2), 142.8 (C-6), 96.1 (C-5), 90.9 (C-1'), 84.0 (C-4'), 76.0 (C-2'), 69.6(C-3'), 64.2 (C-5'), 59.3 (C-2'' & C-6''), 45.8 (C-3'' & C-5''), 31.6 (CH_3), 22.8 (CH_3); ^{31}P NMR (500 MHz, CD_3OD, H_3PO_4 external standard = 0.00) δ 0.13; MALDI-TOF/MS for the reduced form: (M + Na^+) m/z 545.6, calcd for $C_{19}H_{31}N_4O_{11}PNa$ 545.1619.

3.2.2 Enzymatic synthesis of ^{13}C enriched LacNAc

[$^{13}C_6$] labeled GlcNAc and [$^{13}C_6$] labeled Gal were purchased from Omicron and the remaining reagents were purchased from Sigma-Aldrich. The following enzymes were used: galactokinase (EC 2.7.1.6) purified in the lab following procedures given in [26], galactose-1-phosphate uridyltransferase (EC 2.7.7.12) from Sigma-Aldrich and galactosyltransferase (EC 2.4.1.90) from Calbiochem.

A solution (1.5 ml) containing 50 mM Tris-HCl, 5 mM Mg^{2+}, 20 mM ATP and 20 mM Gal or [$^{13}C_6$] Gal at pH 7.9 was prepared in a 5 mm NMR tube [27]. Ten units of

galactokinase were added and the reaction was carried out in this tube heated to 85^{o}C for 40 minutes. After cooling the reaction mix to room temperature, one unit of Gal-1-phosphate uridyltransferase was added together with 40 mM UDP-glucose. After 40 minutes incubation at 37^{o}C, one unit of galactosyltransferase was added with 5 mM $MnCl_2$ and 10 mM GlcNAc or [$^{13}C_6$] GlcNAc, and the pH was adjusted to 7.5. The reaction mixture was left standing overnight at 37^{o}C. The crude product was purified sequentially by Centricon filtration of enzymes, anion exchange chromatography on Dowex 1X-2 (formate form, 1 x 10 cm, water) to remove the phosphates, size exclusion chromatography on Bio-gel P2 (extra fine, 1.5 x 45 cm, water) to separate the salts and finally lyophilization to collect 4.9 mg of white solid (yield 86%).

3.2.3 Purification of reduced CMP-4carboxyTEMPO

The oxyl radical of CMP-4carboxyTEMPO was reduced to a hydroxyl group by the addition of 3 μl phenylhydrazine to 2.7 mg of CMP-4carboxyTEMPO in 0.5 ml methanol. After 20 minutes' gentle swirling, the solvent was evaporated on a rotary evaporator. The following steps were mostly done under argon: adding 2-3 ml ether; vortexing; decanting; repeating twice until the supernatant was clear; adding methanol to dissolve the residue and evaporating the solvent; repeating the ether wash twice; and co-evaporating the residue with methanol to remove the remaining ether. The dried sample weighed 2.1 mg. ^{1}H NMR was taken to verify the reduction and complete removal of phenylhydrazine.

3.2.4 Production of ^{15}N-Phe ST6Gal-1

The production of ^{15}N-phe ST6Gal-1 was carried out in the Moremen laboratory at the CCRC by Dr. Lu Meng. An NH_2-terminal His-tagged form of the rat ST6Gal-1

catalytic domain (residues 97-403 from GenBank P13721) [28] was expressed in mammalian HEK293 cells in the pEAK expression vector (Edge Biosystems, Gaithersburg, MD). Transfected cells were grown at 37 ^{o}C in DMEM medium supplemented with 10% FBS and 1 μg/ml puromycin for antibiotic selection. When the cells had grown to 70% confluency, they were subcultured at a 1:5 ratio and grown to confluency in DMEM media depleted in Phe and supplemented with two-fold excess (132 mg/L) ^{15}N-Phe, 10% dialyzed FBS and 1 μg/ml puromycin. Following growth to confluency the culture media was harvested and ST6Gal-1 was purified sequentially over Phenyl Sepharose, Ni-NTA, and Superdex-75 columns. The purified protein was stored in 20 mM Bis-Tris, pH 6.5, 200 mM NaCl buffer. A detailed description of the recombinant ST6Gal-1expression in HEK293 cells, ^{15}N-Phe labeling, and purification will be published separately [Meng, L., Glushka, J., Stanton, L.H., Fang, T., Collins, R.E., Wiley, G., Gao, Z., Prestegard, J.H. and Moremen, K.W., unpublished data].

3.2.5 Activity and inhibition assays

The activity of ST6Gal-1 was determined by Dr. Lu Meng in the Moremen laboratory using a radioactive assay based on the transfer of [^{14}C] sialic acid from a CMP-[^{14}C] sialic acid donor to an N-acetyllactosamine acceptor [29]. The buffer used in the assay was sodium cacodylate (50mM, pH 6.5) containing 0.1% (v/v) Triton X-100 and 0.2 mg/ml BSA, and assays were performed at 37 ^{o}C for 15 minutes. The sialylated product was isolated by ion-exchange chromatography on Dowex 1-X 8 (0.8 x 3 cm) columns, and counted with a scintillation counter (BECKMAN LS 5000TD).

For IC_{50} measurements, reaction mixtures consisted of CMP-NeuAc (2.1 mM), CMP-[^{14}C] NeuAc (9.2 nM) as a tracer, LacNAc (4.4 mM), ST6Gal-1(11.1 nmol), and

the inhibitor CMP-TEMPO (0-17 mM) or CMP-4carboxyTEMPO (0-6 mM). For Ki measurement on CMP-4carboxyTEMPO, another series of reaction mixtures was prepared with CMP-[^{14}C] NeuAc (9.2 nM), LacNAc (4.4 mM), ST6Gal-1(11.1 nmol), CMP-NeuAc (9-171 μM), and inhibitor (0, 0.8, 1.6, 2.4 and 3.2 mM) in a total volume of 60 μl. A K_I was extracted using a double reciprocal Lineweaver-Burk plot.

3.2.6 Collection of NMR Spectra

1) Paramagnetic perturbation on ^{15}N-Phe ST6Gal-1 [30]. NMR samples were prepared in 10% D_2O / 90% Bis-tris buffer (20 mM; containing 200 mM NaCl) at pH 6.5 with a protein concentration of 0.3 mM. The protein itself is soluble to a higher level, but it appears to form dimers or higher oligomers above this concentration resulting in a severe loss of signal. The spectra were acquired on a Varian INOVA spectrometer operating at 900 MHz for protons. The spectrometer was equipped with a triple resonance cryogenic probe having z pulsed gradient coils. Spectra were run with the fast HSQC sequence supplied as part of the Varian BioPak. Data were typically acquired at 25^{o}C with 11990 Hz as SW, 1800 Hz as SW1, 48 indirect t1 complex points, 514 direct t2 complex points, an acquisition time of 0.043 second, and a recycling rate of 1.5 seconds. Each spectrum represents a total acquisition time of about 3 hours using 64 scans at each t1 point.

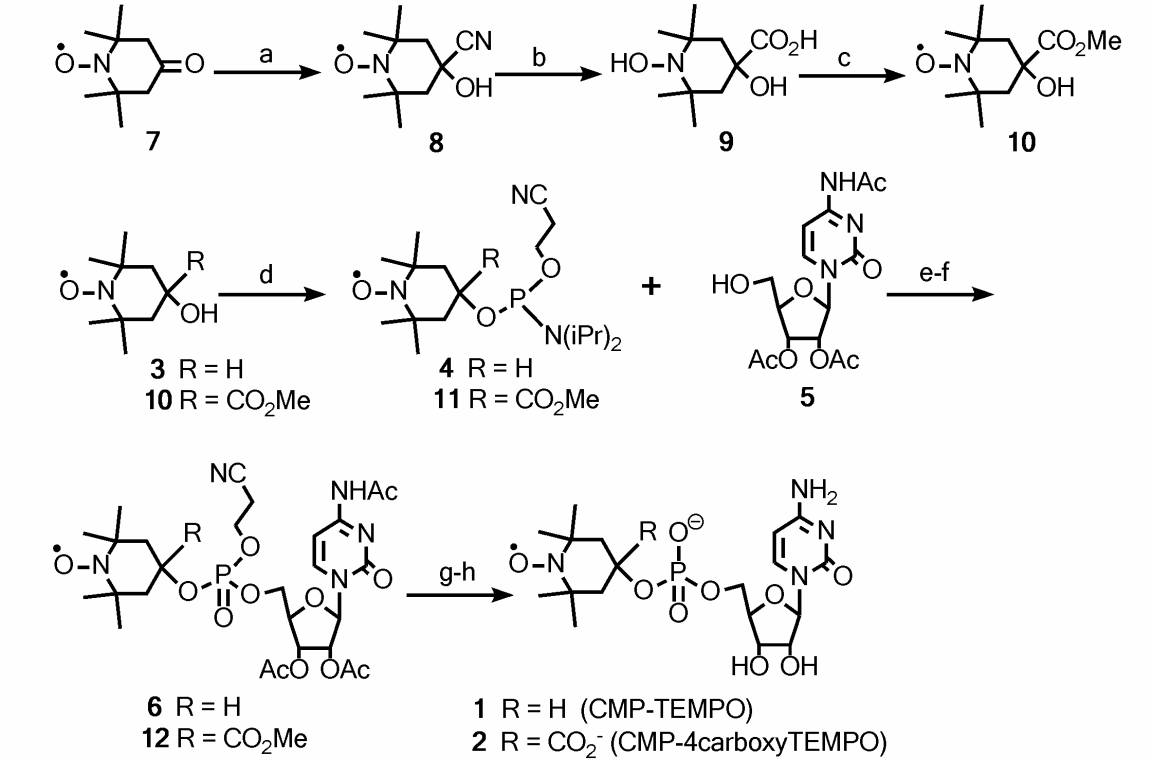
2) T_1 relaxation measurements on the carbons of ^{13}C- LacNAc. NMR samples were prepared in 10% D_2O/90% Bis-Tris buffer (20 mM; containing 200 mM NaCl)) at pH 6.5 with the concentration of ST6Gal-1 at 0.3 mM. The spectra were acquired on Varian INOVA spectrometers operating at 800 MHz (for protons). The spectra were run with the ^{13}C T1 relaxation mode of the gChsqc sequence in the Varian BioPak. Data were

typically acquired at 25^{o}C with 6828 Hz as SW, 7038 Hz as SW1, 64 indirect F1 complex points, 819 direct F2 complex points, an acquisition time of 0.12 second, a recycling rate of 1.2 seconds, and the arranged delays after the extra Z storage at 0.01, 0.08, 0.15, 0.25, 0.4, 0.6, 0.9, 1.3 seconds. Each set of the arrayed HSQC spectra represent a total acquisition time of about 9 hours for 16 scans.

3.3 Results

3.3.1 Synthetic routes for nucleotide-TEMPO derivatives

The synthesis of TEMPO-phosphate derivatives of steroids [31; 32], TEMPO-modified adenosine diphosphate derivatives [33; 34], and the synthesis of TEMPO- modified di- and tri-deoxy adenylates [35] have all been reported in the literature. While these reports provide important precedents, the present synthesis of two TEMPO derivatives of CMP-NeuAc **1** and **2** (Figure 3.2) are based upon known procedures for the synthesis of CMP-NeuAc and its analogs [19; 20; 21; 36].



Key: (a) 30% H_2SO_4/NaCN, THF, <15°C. (b) 27% HCl, H_2O, 80°C. (c) CH_2N_2/Et_2O, MeOH. (d) $(iPr)_2NP(Cl)OCH_2CH_2CN$, $Et(iPr)_2N$, CH_2Cl_2. (e)1H-Tetrazole, MeCN, -40°C. (f) t-BuOOH, MeCN. (g) DBU, THF. (h) NaOMe, MeOH-H_2O.

Figure 3.2. Synthesis of TEMPO analogs of CMP-NeuAc

The synthesis of **1** started from the available precursor **3**, while the synthesis of **2** required the preparation of intermediate **10.** Compound **10** was obtained from the commercial 4-Oxo-TEMPO **7** through the known cyanohydrin **8** [23]. Hydrolysis of the nitrile group in strongly acidic conditions [22] provided the acid **9,** in which the nitroxide had been reduced. After removal of excess HCl, the crude product was transformed into its methyl ester with diazomethane, and slowly re-oxidized by air [24; 25] to the nitroxide **10**, which was then purified. The easily prepared phosphoramidite derivatives **4** and **11** [19] were reacted with the primary hydroxyl of the known cytidine derivative **5** [21] in the presence of tetrazole [20]. The obtained intermediate phosphites were directly oxidized with tBuOOH to provide the phosphates **6** and **12** respectively, which were then purified [20].

As expected, compound **12** was obtained in a lower yield than compound **6**, likely due to the effect of steric hindrance in the intermediate **11**. Careful two-step deprotection of **6** and **12** followed an established procedure [20] to provide **1** and **2**.

3.3.2 Determination of binding affinities for spin-labeled analogs

An IC_{50} was measured as 8.7 ± 0.1 mM for CMP-TEMPO and 15-20 mM for its reduced form, while an IC_{50} was measured as 1.7 ± 0.1 mM for CMP-4carboxyTEMPO and 3.5 ± 0.1 mM for its reduced form. With the substrate level [S] of 171 μM and a K_M of 150 μM for CMP-NeuAc, the K_I for CMP-TEMPO (radical) was calculated as 4.07 mM by the formula $[IC_{50}] * K_M/(K_M + [S])$ [37], and that for CMP-4carboxyTEMPO (radical) was calculated as 0.79 mM. A K_I for CMP-4carboxyTEMPO was also determined from a more extensive investigation of the concentration dependence of activities. The K_I was extracted using a double reciprocal Lineweaver-Burk plot and determined to be 0.75 ± 0.05 mM.

The improvement in affinity of roughly a factor of 5 upon the addition of the carboxyl group is consistent with our predictions which were based on the expected favorable interactions between the carboxyl group and residues in the active site of ST6Gal-1. The affinity for CMP-4carboxyTEMPO was in a range that is useful for observing paramagnetic-enhanced relaxation.

3.3.3 Distance-dependent line broadening of ^{15}N-HSQC spectra of ST6Gal-1

ST6Gal-1 can be prepared with ^{15}N enrichment in the amide sites of specific amino acids [Meng, L., Glushka, J., Stanton, L.H., Fang, T., Collins, R.E., Wiley, G., Gao, Z., Prestegard, J.H. and Moremen, K.W., unpublished data]. An HSQC spectrum of a 0.3 mM sample prepared with all phenylalanines labeled is presented in Figure 3.3A.

There are 16 phenylalanines in the sequence of the catalytic domain of ST6Gal-1 and 16 crosspeaks are identifiable in Figure 3.3A corresponding to the sixteen labeled amide sites. These crosspeaks are not yet assigned to specific sites in the sequence, but strategies are being developed to make these assignments [38]. We can identify crosspeaks belonging to phenylalanines near the binding site from resonance perturbation by the bound substrate and their analogs.

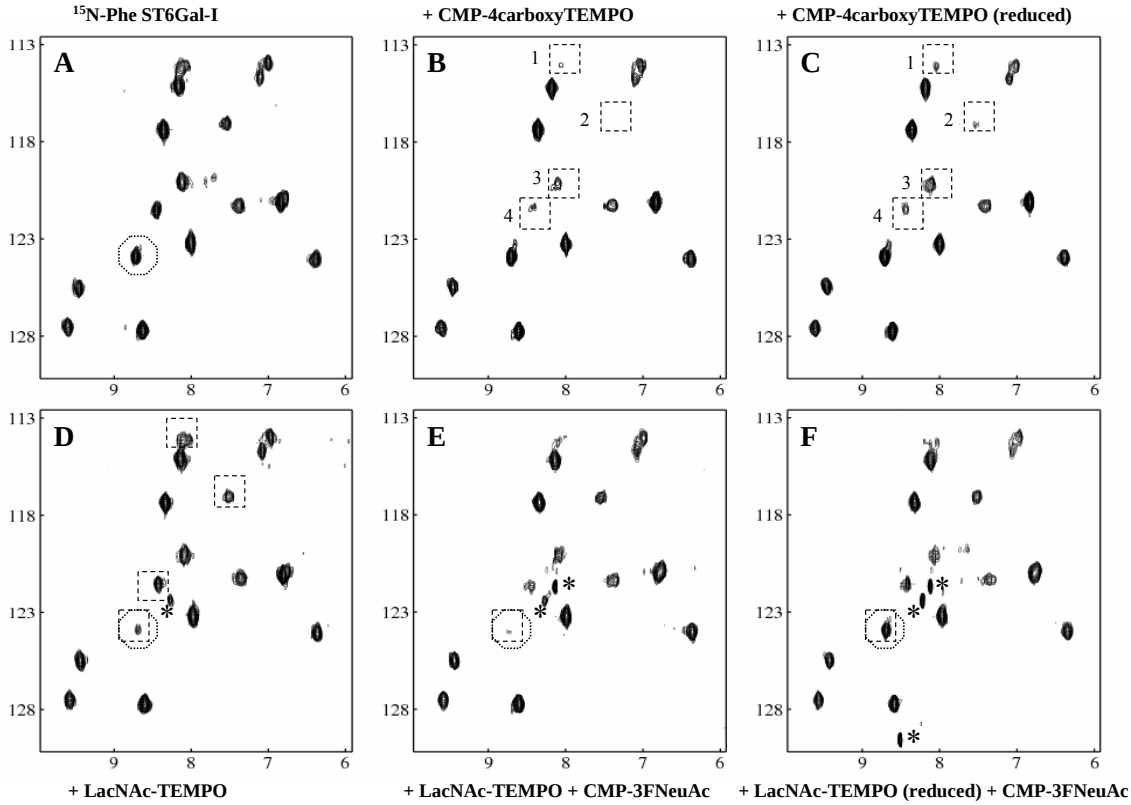


Figure 3.3. ^{1}H-^{15}N HSQC spectra of ^{15}N-Phe ST6Gal-1 in the presence of the indicated substrates. (A)^{15}N-Phe ST6Gal-1 in the absence of substrates. (B and C) ^{15}N-Phe ST6Gal-1 with a 4-fold molar excess of CMP-4carboxyTEMPO (B) before and (C) after reduction using ascorbate. (D, E and F) ^{15}N-Phe ST6Gal-1 with a 20-fold excess of LacNAc-TEMPO (D), an added 15-fold excess of CMP-3FNeuAc (E), and added ascorbate (F). The boxed peaks are broadened due the presence of spin-labeled substrates.

The circled peak is most strongly perturbed and is used to estimate structural changes resulting from the addition of CMP-3FNeuAc to ST6Gal-1. The circled peak did not change upon the sole addition of CMP-3FNeuAc to ST6Gal-1. Additional peaks that appear because of natural abundance ^{15}N-H in the NH-NeuAc moiety of CMP-3FNeuAc or the GlcNAc moiety of LacNAc-TEMPO are denoted with an asterisk.

It is known that the native substrate CMP-NeuAc can slowly be hydrolyzed by itself and the presence of ST6Gal-1 can expedite this hydrolysis. In our experiments, we employed the inert donor analog, CMP-3FNeuAc, for comparison with TEMPO analogs of CMP-NeuAc. We acquired a spectrum of ^{15}N-Phe ST6Gal-1 in the presence of a 10-fold excess of CMP-3FNeuAc [30]. Unlike many examples of cross-peak perturbation in HSQC spectra by non-paramagnetic ligands, the dominant effects were broadening or reduction in cross peak intensity rather than changes in chemical shift. This may be the result of moderately slow exchange, resulting in life-time broadening of resonances. Substantial induced movement of protein segments upon the binding of sugar nucleotides also could produce this effect and this type of motion has been suggested in the studies of several other glycosyltransferases [39].

CMP-TEMPO and CMP-4carboxyTEMPO appear to have qualitatively similar effects as CMP-3FNeuAc on the ^{15}N HSQC of ST6Gal-1 (Figure 3.3B), which indicates that these analogs very likely bind to the active site with virtually the same protein contacts as the native substrate. Consistent with the difference in binding constants, a five-fold excess of CMP-TEMPO over CMP-4carboxyTEMPO is required to produce similar levels of perturbation.

In one way, it is unfortunate that crosspeak broadening and intensity decreases are seen in the CMP-3FNeuAc spectrum, rather than crosspeak shifts. Multiple contributions to peak broadening can complicate the conversion of broadenings to distances between nitroxide- and phenylalanine-labeled sites for the CMP-4carboxyTEMPO spectrum. Fortunately, it is clear that the peaks are more severely perturbed in the presence of CMP-4carboxyTEMPO binding, even with reduced binding to the protein because of its lower binding constant. Also, there is one peak that completely disappears in the presence of CMP-4carboxyTEMPO that was not perturbed by CMP-3FNeuAc. Hence, there is certainly nitroxide-enhanced transverse relaxation present in the presence of CMP-4carboxyTEMPO.

Extraction of nitroxide-dependent paramagnetic effects can be simplified through the use of a better control spectrum. This can be provided using the fact that paramagnetic perturbation can be removed by reducing the nitroxide to a hydroxylamine. In Figure 3.3C the perturbed resonances of the protein partially reappear when the nitroxide radical is reduced by ascorbate salt. Changes in intensities between the reduced and oxidized spectra (Table 3.1) can then be related to the distances between the nitroxide and labeled phenylalanine positions.

Table 3.1. Relative intensities of the perturbed crosspeaks in the ^{1}H-^{15}N HSQC (labeled in Figure 3.3) and their corresponding distance estimates. *Values are taken from the reduced form of CMP-4carboxyTEMPO at a concentration adjusted to compensate for the binding affinity difference between its oxidized form and reduced form.

peak	^{15}N-Phe ST6Gal-1	w/ spin-label	no spin-label*	distance (Å)
1	1.153±0.046	0.058±0.048	0.374±0.031	10.6-12.9
2	1.239±0.046	0.000±0.048	0.091±0.031	<17.0
3	1.571±0.046	0.562±0.048	0.921±0.031	14.3-15.5
4	1.178±0.046	0.175±0.048	0.440±0.031	12.6-14.4

Similar to CMP-4carboxyTEMPO, the spin-labeled acceptor analog LacNAc-TEMPO can be used to identify crosspeaks belonging to phenylalanines near the LacNAc-binding site based on the steep $1/r^6$ distance dependence of the effects. LacNAc-TEMPO perturbs three of the four peaks seen with CMP-4carboxyTEMPO (Figure 3.3D). This confirms the binding of the donor near the acceptor site. However, there are significant differences. For example, there is one peak that almost completely disappears in the presence of LacNAc-TEMPO (circled in Figure 3.3D) but was not perturbed by CMP-4carboxyTEMPO. Nor was this peak perturbed by the non-paramagnetic donor mimic CMP-3FNeuAc which broadened three other protein peaks [30]. A change in this particular peak did, however, occur when CMP-3FNeuAc was included along with LacNAc-TEMPO (Figure 3.3E). Integrations of the peaks show that the intensity of the circled peak decreases by an additional factor of two when comparing Figure 3.3D to 3.3E and the peak reappeared to full intensity when the nitroxide was quenched (Figure 3.3F). This could result from the enhanced binding of the LacNAc-TEMPO molecule because of the cooperative binding of the donor and acceptor, as has been suggested for a number of glycosyltransferases [40; 41; 42; 43]. Alternatively, it could result from a

conformational change in the glycosyltransferase induced upon donor binding as documented [41; 44]. We believe the latter is the explanation for the current case because it is estimated that the binding site is nearly saturated at the concentrations of LacNAc-TEMPO used, even in the absence of the sugar donor. Based on a fixed fraction of bound LacNAc-TEMPO, it is estimated that the donor brought the acceptor 6% closer to the ^{15}N-labeled Phe associated with the perturbed peak (see equation 1). This observation is consistent with a loop movement seen in Δ24PmST1 in which a buried tryptophan flips out of the protein core to redefine the acceptor binding site once the donor binds to the protein [41].

One also must control for the possibility of non-specific binding to secondary sites. The TEMPO moiety itself is quite hydrophobic and may bind to hydrophobic patches on the protein independently of active site interactions. This can lead to some additional broadening of peaks. To rule out this possibility, a spectrum in the presence of an equivalent amount of 4-carboxy-4-hydroxyTEMPO (10) was acquired. No significant intensity loss for crosspeaks was observed in the spectrum.

3.3.4 Protein-ligand distance restraints for molecular modeling

The difference in intensity of crosspeaks between HSQC spectra of proteins in the presence and absence of a spin-label ligand (actually the presence of oxidized and reduced forms) can be used to estimate distances between the position of the nitroxide label in the active site and the protons giving rise to particular cross peaks. During the transfer and refocusing periods of the experiment, the amide proton magnetization is in the transverse plane and undergoing intensity loss due to paramagnetic relaxation. A

number of authors have employed the following equation to relate the intensity loss to distances between the nitroxide and particular protons [6; 11; 12].

$$\ln (I_{nsl}/I_{sl}) = f*t*(K/r^6)*[4\tau_c + 3\tau_c*(1+\omega_H^2\tau_c^2)^{-1}] \quad (1)$$

In this equation, I_{sl} and I_{nsl} are the peak intensities (measured as peak volumes) of resonances with and without a spin label (the presence of oxidized and reduced forms). f is the fraction of protein carrying a ligand and t (9.52 ms) is the total time during the INEPT (insensitive nuclei enhanced by polarization transfer) and refocusing periods of the HSQC pulse sequence. K is a constant related to spin properties of the system ($K = 1/15*S(S+1)\gamma^2 g^2 ß^2 = 1.23 \times 10^{-32}$ $cm^6 s^{-2}$), r is the distance between the nitroxide and a site giving rise to a cross peak of interest, τ_c is the effective correlation time for tumbling of the protein, and ω_H (900 MHz* 2π) is the precession frequency for the amide proton. Frequently, a scaling factor that would relate the inverse sixth power of distance to the log of the intensity ratio is calculated empirically using a known distance in the system. This eliminates the need to independently determine factors appearing in equation 1. Unfortunately, ST6Gal-1 does not have a crystal structure, and it is not possible to provide a known distance. However, an appropriate scaling factor can be determined using the binding constant determined for the analog (f = 57%), an estimate of the correlation time from spin relaxation data (τ_c = 20 ns), and the fundamental nuclear and electron spin constants. The distances resulting from the application of equation 1 are reported in Table 3.1.

The distances that we estimated for the perturbed resonances are all less than 17 Å and well within the bounds of the 25 Å radius expected for a 40 kDa protein. Only an upper limit can be set for peak #2 because of its complete disappearance in the presence

of the spin label. Weak intensity in the presence of the reduced spin label also makes the uncertainty in the upper limit to the distance large. Best estimate distances for the other peaks cluster between 10 and 15 Å. The narrow range is the result of the steep $1/r^6$ distance-dependence for which a variation in intensity ratios by a factor of two leads to only a 6% variation in distance. In principle we can expand the range of sensitivity and improve measurements for peak #2 by collecting data with different concentrations of spin-labeled ligand.

3.3.5 Relaxation enhancements of one ligand by the other spin-labeled ligand

The relative positions of the substrates also can be probed using paramagnetic perturbations. The paramagnetic relaxation enhancement of the bound acceptor LacNAc due to bound CMP-4carboxyTEMPO, was measured by comparing the T_1 of acceptor carbons in control and in relaxation-enhanced samples. The enhancements were in turn converted into intermolecular distances between the substrates in the protein binding site. The relaxation-enhanced sample contained 0.3 mM ST6Gal-1 , 1.0 mM ^{13}C-LacNAc and a near saturating amount of CMP-4carboxyTEMPO (3.0 mM; estimated at 79% saturation of protein using a dissociation constant (K_D) of 0.75 mM [30]). The two control samples contained ST6Gal-1, the acceptor, and the equivalent amount of CMP-3FNeuAc or reduced CMP-4carboxyTEMPO. T_1s of carbons were measured from peak intensities in an arrayed ^{1}H-^{13}C HSQC experiment in which an extra z-storage, relaxation delay, and J evolution element had been inserted just before the INEPT transfer back to protons. Examples of the data are given in Figure 3.4 which depicts exponential fits to a resonance showing a significant change in T_1 (that from the C4 carbon of the galactose residue in

LacNAc) and a resonance showing minimal perturbation (that from the C6 carbon of the galactose residue in LacNAc).

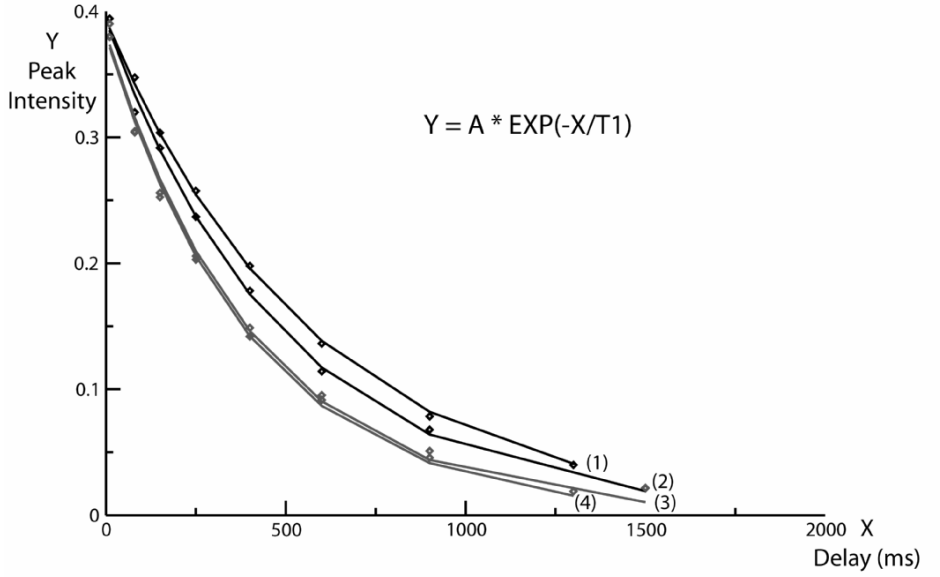


Figure 3.4. T_1 measurement examples for carbon Gal4 (black) and carbon Gal6 (gray) with and without spin-label perturbation by CMP-4carboxyTEMPO binding to the protein: (1) data for Gal4 without a spin label. The T_1 was calculated as 574 ms, (2) data for Gal4 with a spin label. The T_1 was calculated as 496 ms, (3) data for Gal6 with a spin label. The T_1 was calculated as 416 ms, (4) data for Gal6 without a spin label. The T_1 was calculated as 406 ms.

The changes in longitudinal relaxation rate (R_1, T_1^{-1}) for the sugar residues are actually small, ranging from 0.05 to 0.37 s^{-1} (Figure 3.5). This is partly due to the fact that paramagnetic effects on heteronuclear longitudinal relaxation rates by nitroxide are inherently small at high magnetic fields, and partly due to the small fraction of the acceptor bound to ST6Gal-1 while CMP-4carboxyTEMPO also was binding. With an estimated K_D value of 1.6 mM for the acceptor, the experimental sample of 0.3 mM ST6Gal-1, 1.0 mM acceptor, 3 mM CMP-4carboxyTEMPO would yield only 9% of the

acceptor bound to CMP-4carboxyTEMPO-occupied ST6Gal-1. Clearly it would be advantageous to have acceptor analogues with higher affinities for these experiments. Nevertheless, some useful data can be extracted. The enhanced relaxation rates of the Gal carbon 4 and of the α-GlcNAc carbon 6 are obviously significant. The absolute propagated standard error on average is 0.14 s^{-1} based on the standard error of 5% from the exponential fit of the peak intensity data.

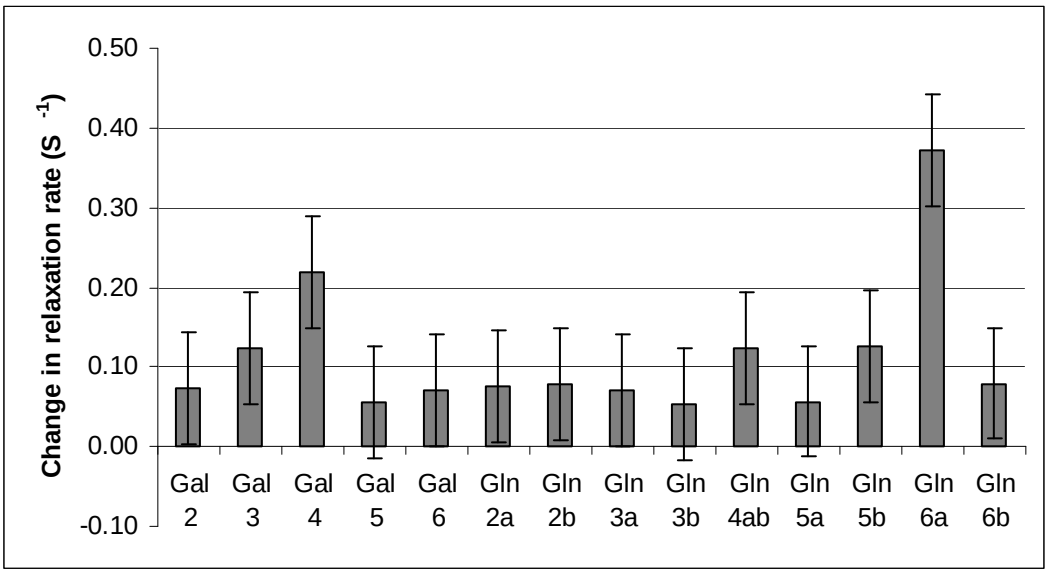


Figure 3.5. Differences in R_1 of LacNAc carbons in the control and experimental samples. All of the changes are positive and indicate paramagnetic perturbation from bound CMP-4carboxyTEMPO. The N-acetyl methyl carbons were not measured in the ^{1}H-^{13}C HSQC due to the lack of ^{13}C-enrichment in the methyl group. The data on Gal1 and GlcNAc1α and GlcNAc1β are not shown because their peak intensities were attenuated by water suppression during the acquisition of spectra.

3.3.6 Calculation of ligand-ligand distances from relaxation enhancements

Based on the R_1 changes of Gal and GlcNAc carbons between the control and relaxation-enhanced samples, the distances from the donor nitroxide to the 14 acceptor

carbons (shown in Figure 3.5) can be calculated. T_1 spin relaxation data were fit to single exponential curves using Rate Analysis in the NMRView software. The extracted relaxation rates ($R_1 = T_1^{-1}$) were then related to distances between the spin label and the site of measurement using the following equations:

$$R_{1NO}/R_{1CH} = (r_{CH} / r_{NO})^6 * (\gamma_E / \gamma_H)^2 \quad (\text{x } 0.5 \text{ for the two protons on C6 carbons}) \qquad (2)$$

$$R_{1CHobs} = (1\text{-}F_b) * R_{1CHfree} + F_b * R_{1CH} \qquad (3)$$

$$R_{1NOobs} = (1\text{-} F_b') * R_{1CHfree} + F_b' * (R_{1CH} + R_{1NO}) \qquad (4)$$

Here R_{1CH} is the relaxation rate due to protons directly bound to carbons. R_{1NO} is the relaxation rate contribution due to the presence of the nitroxide electron. $R_{1CHfree}$ is the relaxation rate contribution from the acceptor free in solution. R_{1CHobs} is the relaxation rate observed in the presence of non-spin label donor analogs and protein. R_{1NOobs} is the relaxation rate observed in the presence of CMP-4carboxyTEMPO and protein. F_b refers to the fraction of acceptor complexed with non-spin label donor analogs. F_b' refers to the fraction of acceptor complexed with CMP-4carboxyTEMPO in protein solution. The bond length of C-H (r_{CH}) used is 1.09 Å. The scaling factor (4.25×10^5) in equation (2) is based on an estimated correlation time for the protein of 20 ns and also dependent upon magnetic moments of the unpaired electron γ_E (-1.76×10^{11}) and the ^1H nuclei γ_H (2.68×10^8).

In these calculations, the internal motions of the ligand were ignored and the modulation of interactions by the tumbling correlation time for the protein was assumed to dominate the relaxation in all cases. The estimates are dependent upon the fraction of the acceptor complexed with non-spin label donor analogs, 9.2% for CMP-3FNeuAc and 7.3% for reduced CMP-4carboxyTEMPO as well as the fraction of acceptor complexed

with CMP-4carboxyTEMPO in protein solution, which was calculated as 8.9%. The distances calculated range from 8 to12 Å with 9 Å for Gal4 and 8 Å for GlcNAc6α being the shortest distances (Table 3.2). The data are consistent with the preferential placement of the donor relative to acceptor in a pattern that puts the nitroxide on the side of LacNAc where both Gal4 and GlcNAc6 are exposed.

3.4 Discussion

The above results provide a clear demonstration that nitroxide spin-labeled compounds designed to mimic natural ligands of proteins can be used to obtain structural data on proteins that are difficult to characterize by conventional methods. Two additional steps are required to produce the protein structure. First, crosspeaks must be assigned to specific sites without the aid of experiments that require uniform isotopic labeling. Appropriate methods are being developed [38], but this will take some time. Also, computational methods that use sparse constraints to produce protein structures must be improved [7; 9; 12]. However, the availability of even a few distance restraints and identification of crosspeaks from residues near the active site is an important step forward.

Prior to completion of a structure for ST6Gal-1, we can compare of our distance estimates to distances observed in the crystal structure of related molecules. The bacterial α2, 3-sialyltransferase, CstII [45] provides one such comparison. There is very low sequence homology between this protein and ST6Gal-1, but programs such as GenTHREADER [46; 47] can predict reasonable sequence alignment. Using this alignment and a model built with MODELLER [48], we can predict distances between the nitroxide of bound CMP-4carboxyTEMPO and phenylalanine amide protons. We find the closest four phenylalanine amide protons to be at 6.5, 7.8, 11.8, 16.1 Å from the nitroxide of

CMP-4carboxyTEMPO. Considering that one entry is given only as an upper limit, these distances are not far off from those listed in Table 1. In principle, the distances in Table 1 can be used to improve modeling of the ST6Gal-1 structure once sequential assignments of crosspeaks to specific sequential sites in the protein are available.

We also can compare the observed distances to the rather extensive sequence analysis that has been done for ST6Gal-1 [49]. Large (L) and small (S) and very small (VS) conserved sequence regions have been implicated in binding of the sugar donors and acceptors. These regions contain 4 phenylalanines (1 in the S-motif and 2 in L-motif and 1 in VS-motif). This number compares favorably with the number of crosspeaks perturbed in our experiments. Thus, our observations provide some of the first structural validation of sequences predicted to be involved in substrate binding for ST6Gal-1.

Based on the data presented, several conclusions can be made about the placement of donor and acceptor in the binding site of ST6Gal-1. Even in the absence of a structure for the protein, in principle, this placement can provide insights into a mechanism of action and donor/acceptor modifications that might alter this action. First, from data on perturbation of a set of phenylalanine ^{1}H-^{15}N HSQC crosspeaks using a spin-labeled acceptor analog we can make some qualitative statements about binding site occupancy and relative placement of donor and acceptor. The fact that the acceptor analog perturbs three of the four peaks that were previously observed to be perturbed by a spin-labeled donor analog suggests that they clearly bind in the same general region on the protein surface. However, the fact that one HSQC cross peak, not perturbed by the spin-labeled donor analog, was perturbed by the spin-labeled acceptor analog suggests the reducing end of LacNAc (where the spin label is bound) is well removed from contact

with the donor sugar (TEMPO in the donor analog). This is as it should be in a complex poised for addition of the donor sugar to the non-reducing end of the acceptor. These observations will become more useful as cross peaks in HSQC spectra of ST6Gal-1 are assigned and a structure determined.

In principle, related paramagnetic relaxation enhancement (PRE) experiments using a spin-labeled donor analog and observing perturbations of HSQC peaks from an acceptor, allow for the determination of approximate distances between carbons of the acceptor and the nitroxide of the donor, and construction of a more complete model for the placement of the donor and acceptor in the ST6Gal-1 binding site. Perturbations are, however, weak and broadly distributed through the acceptor except for two well-separated carbons, Gal-4 and GlcNAc-6. This may suggest that a distribution of sites is occupied in which the spin label can make close approaches to either site. This would require, at a minimum, exposure of surfaces on the acceptor carrying these two sites.

Figure 3.6 depicts a model in which the Gal-4 and GlcNAc-6 sites are exposed. The structure of α-LacNAc shown here is the minimal energy conformer generated by Glycam Biomolecular Builder [50] (also http://glycam.ccrc.uga.edu/). The color code indicates the measured proximity of the LacNAc carbons to the nitroxide of CMP-4carboxyTEMPO with red for closest (gal4, GlcNAc6), yellow for second closest (Gal3, GlcNAc 4, GlcNAc5), and green for the furthest (Gal5, Gal6, GlcNAc2, GlcNAc3). If we assume that the less perturbed side is protected by binding to the protein and represents more conserved points on the acceptor, the model is consistent with the pattern of the groups accepting/rejecting modifications in the specificity studies of the ST6Gal-1

acceptors [24]. The model also compares favorably to the binding orientation of α-lactose in the crystal structure (2IHZ) of the α-2,3/6 sialyltransferase, 24PMST1 [12].

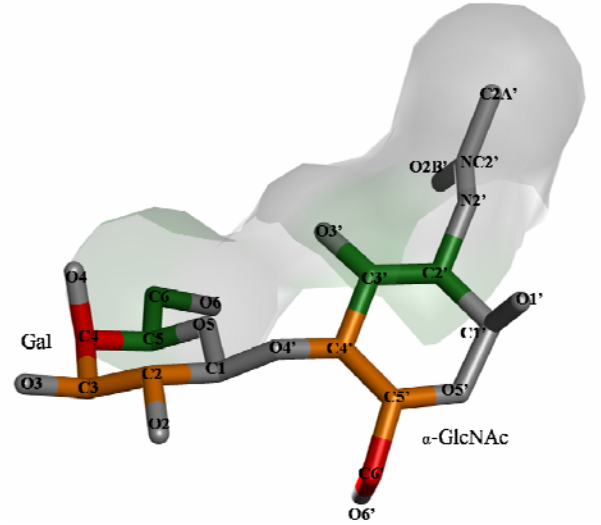


Figure 3.6. Relative position of α-LacNAc in the active site of ST6Gal-1. The color code indicates the measured proximity from the α-LacNAc carbons to the nitroxide of CMP-4carboxyTEMPO: red for closest (gal4, GlcNAc6), yellow for second closest (Gal3, GlcNAc 4, GlcNAc5), and green for the furthest (Gal5, Gal6, GlcNAc2, GlcNAc3).

Extending these pictures to a single model can be done using the close approach distance estimates for Gal-4 and GlcNAc-6 carbons. If there are simply two equally populated forms the distances for each would be adjusted down by a factor of $(1/2)^{1/6}$ (8 and 7 Å respectively). In Figure 3.7 the two models were constructed allowing the O6 oxygen of the galactose to approach the C2 carbon of the donor mimic at about 4 Å as it might in a transition state, and placing the Gal-4 or GlcNAc-6 carbons at 8 Å and 7 Å respectively. All other carbons were placed at greater distances to the nitroxide of the donor mimic to be in agreement with the lower-level paramagnetic perturbation observed during the relaxation measurements. The shaded areas represent the protein surface that has close contacts with the cytosine base as deduced from STD and surface exposure

results. Since the donor analog is not active, this structure might represent the pre-transitional state for substrates of ST6Gal-1.

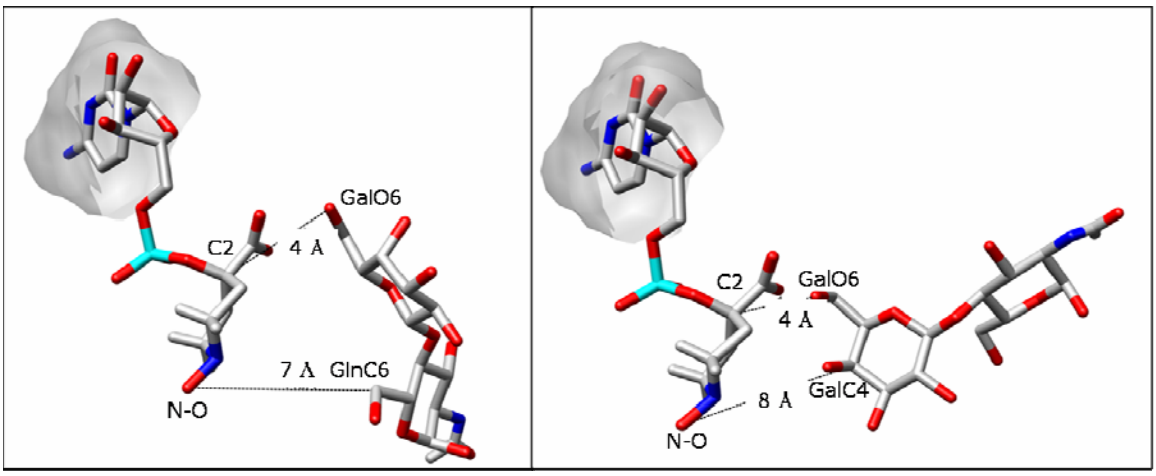


Figure 3.7. Models for the relative position of CMP-4carboxyTEMPO and LacNAc in the binding site of ST6Gal-1. The dashed lines depict distance constraints coming from PRE data and an α-(2,6) addition by the 6-oxygen of the LacNAc acceptor.

We can compare the above models to structural and functional data in the literature. There are two structures of a bacterial sialyltransferase reported in the literature and deposited in the PDB, both are for the 2, 3/6 sialyltransferase Δ24PmST1 from *Pasteurella Multocida* (2IHZ and 2IY8). Both have the inactive donor, CMP-3FNeuAc, bound, but 2IHZ has α-lactose and the 2IY8 has β-lactose. The positioning of the lactose relative to the donor differs slightly, but in each case the alignment seems more appropriate for an SN2 substitution by the 3-oxygen on the activated C2 carbon of the sialic acid mimic rather than a substitution by the 6-oxygen. In both cases the C2-O3 distance is long for a O3 attack on the activated C2 (5.4 and 5.5 Å respectively), but the C2-O6 distance is much longer (6.6 and 7.1 Å). There is the possibility that the binding orientation of α-lactose in 2IHZ might represent the bound conformation for alpha-2,6-

activity [41]. In the 2IHZ structure, rotating the torsion angle around the C5-C6 bond of Gal can move the O6 of Gal from 6.6 to 5.1 Å away from the anomeric C2 of donor. This distance might be shortened even further if the lactose disposition shifts slightly at low pH and the fluorine atom is absent in the natural donor. Given the ambiguity between 2-3 and 2-6 additions, one would expect some deviations between these structures and the structure for ST6Gal-1. However, a comparison is worth discussing. In the structure of 2IHZ with the substrates superimposed by CMP-4carboxyTEMPO and α-LacNAc, Gal C4 is the second closest (after Gal 3 carbon) to the nitroxide of CMP-4carboxyTEMPO, and GlcNAc C6 is at the shortest distance to the nitroxide of CMP-4carboxyTEMPO compared with the other carbons of GlcNAc (Table 3.2). These distances are relatively consistent with the calculated data in the paramagnetic relaxation experiments.

Table 3.2. Distances from α-LacNAc carbons to the nitroxide of CMP-4carboxyTEMPO superimposed in the structure of Δ24PmST1 (2IHZ) in which its substrates CMP-3FNeuaAc and α-lactose were both present.

	Gal						GlcNAc					
2IHZ	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
(Å)	7.7	6.9	5.7	6.5	7.5	8.7	12.3	11.7	10.7	9.6	10.4	9.6

While the structure presented in this chapter cannot be considered of sufficient resolution for detailed mechanistic modeling, it does provide a stepping stone for future work. In particular, the sequential assignments of ^{15}N-Phe ST6Gal1 resonances in the HSQC spectra should add the key amino acids to the model, work with higher affinity

acceptors should improve the sensitivity of paramagnetic probes, and orientation-sensitive data such as ligand RDCs [51; 52; 53] should greatly refine the orientation of ligands in the models. With such improvements, we can expect a better structural understanding of ST6Gal-1 and a basis for structure-based design of specific inhibitors. The strategies we developed should also find important applications for probing the structures of other glycosyltransferases.

References

1. Moseley, H. N. B., Monleon, D. & Montelione, G. T. (2001). Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B* **339**, 91-108.

2. Wider, G. (2005). NMR techniques used with very large biological macromolecules in solution. *Nuclear Magnetic Resonance of Biological Macromolecules, Part C* **394**, 382-398.

3. Lustbader, J. W., Birken, S., Pollak, S., Pound, A., Chait, B. T., Mirza, U. A., Ramnarain, S., Canfield, R. E. & Brown, J. M. (1996). Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: An advance toward rapid determination of glycoprotein structures. *Journal of Biomolecular NMR* **7**, 295-304.

4. Mao, H. Y., Gunasekera, A. H. & Fesik, S. W. (2000). Expression, refolding, and isotopic labeling of human serum albumin domains for NMR spectroscopy. *Protein Expression and Purification* **20**, 492-499.

5. Bruggert, M., Rehm, T., Shanker, S., Georgescu, J. & Holak, T. A. (2003). A novel medium for expression of proteins selectively labeled with N-15-amino acids in Spodoptera frugiperda (Sf9) insect cells. *Journal of Biomolecular Nmr* **25**, 335-348.

6. Battiste, J. L. & Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365.

7. Gaponenko, V., Howarth, J. W., Columbus, L., Gasmi-Seabrook, G., Yuan, J., Hubbell, W. L. & Rosevear, P. R. (2000). Protein global fold determination using site-directed spin and isotope labeling. *Protein Science* **9**, 302-309.

8. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease .1. Paramagnetic relaxation enhancement by nitroxide spin labels. *Journal of Molecular Biology* **268**, 158-169.

9. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease .2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *Journal of Molecular Biology* **268**, 170-184.

10. Jahnke, W. (2002). Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chembiochem* **3**, 167-173.

11. Jain, N. U., Venot, A., Umemoto, K., Leffler, H. & Prestegard, J. H. (2001). Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands. *Protein Science* **10**, 2393-2400.

12. Johnson, P. E., Brun, E., MacKenzie, L. F., Withers, S. G. & McIntosh, L. P. (1999). The cellulose-binding domains from Cellulomonas fimi beta-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. *Journal of Molecular Biology* **287**, 609-625.

13. Kuliopulos, A., Westbrook, E. M., Talalay, P. & Mildvan, A. S. (1987). Positioning of a Spin-Labeled Substrate-Analog into the Structure of Delta-5-3-Ketosteroid Isomerase by Combined Kinetic, Magnetic-Resonance, and X-Ray-Diffraction Methods. *Biochemistry* **26**, 3927-3937.

14. Schmidt, P. G. & Kuntz, I. D. (1984). Distance Measurements in Spin-Labeled Lysozyme. *Biochemistry* **23**, 4261-4266.

15. Wang, X. F., Zhang, L. H. & Ye, X. S. (2003). Recent development in the design of sialyltransferase inhibitors. *Medicinal Research Reviews* **23**, 32-47.

16. Dufner, G., Schworer, R., Muller, B. & Schmidt, R. R. (2000). Base- and sugar-modified cytidine monophosphate N-acetylneuraminic acid (CMP-Neu5Ac) analogues - Synthesis and studies with alpha(2-6)-sialyltransferase from rat liver. *European Journal of Organic Chemistry*, 1467-1482.

17. Watts, A. G. & Withers, S. G. (2004). The synthesis of some mechanistic probes for sialic acid processing enzymes and the labeling of a sialidase from Trypanosoma rangeli. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **82**, 1581-1588.

18. Lee, T. D. & Keana, F. W. (1975). Insitu Reduction of Nitroxide Spin Labels with Phenylhydrazine in Deuteriochloroform Solution - Convenient Method for Obtaining Structural Information on Nitroxides Using Nuclear Magnetic-Resonance Spectroscopy. *Journal of Organic Chemistry* **40**, 3145-3147.

19. Kondo, H., Ichikawa, Y. & Wong, C. H. (1992). Beta-Sialyl Phosphite and Phosphoramidite - Synthesis and Application to the Chemoenzymatic Synthesis of Cmp-Sialic Acid and Sialyl Oligosaccharides. *Journal of the American Chemical Society* **114**, 8748-8750.

20. Kajihara, Y., Ebata, T., Koseki, K., Kodama, H., Matsushita, H. & Hashimoto, H. (1995). Efficient Chemical Synthesis of Cmp-Neu5ac and Cmp-(Neu5ac-Alpha-2-]-8neu5ac). *Journal of Organic Chemistry* **60**, 5732-5735.

21. Chappell, M. D. & Halcomb, R. L. (1997). Synthesis of CMP-sialic acid conjugates: Substrates for the enzymatic synthesis of natural and designed sialyl oligosaccharides. *Tetrahedron* **53**, 11109-11120.

22. Avi, M., Fechter, M. H., Gruber, K., Bela, F., Pochlauer, P. & Griengl, H. (2004). Hydroxynitrile lyase catalysed synthesis of heterocyclic (R)- and (S)-cyanohydrins. *Tetrahedron* **60**, 10411-10418.

23. Wong, L. T. L., Schwenk, R. & Hsia, J. C. (1974). New Synthesis of Nitroxyl Radicals of Piperidine and Tetrahydropyridine Series. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **52**, 3381-3383.

24. Dulog, L. & Wang, W. (1992). A Bisnitroxyl Dipeptide. *Liebigs Annalen Der Chemie*, 301-303.

25. Galera, E., Szykula, J. & Zabza, A. (1987). Insect Growth-Regulators .19. Synthesis of Farnesyl-Type Juvenoids with a Tetramethylpiperidinoxyl Marker. *Liebigs Annalen Der Chemie*, 777-780.

26. Verhees, C. H., Koot, D. G. M., Ettema, T. J. G., Dijkema, C., De Vos, W. M. & Van Der Oost, J. (2002). Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon Pyrococcus furiosus. *Biochemical Journal* **366**, 121-127.

27. Gilhespymuskett, A. M., Partridge, J., Jefferis, R. & Homans, S. W. (1994). A Novel C-13 Isotopic Labeling Strategy for Probing the Structure and Dynamics of Glycan Chains in-Situ on Glycoproteins. *Glycobiology* **4**, 485-489.

28. Weinstein, J., Lee, E. U., Mcentee, K., Lai, P. H. & Paulson, J. C. (1987). Primary Structure of Beta-Galactoside Alpha-2,6-Sialyltransferase - Conversion of Membrane-Bound Enzyme to Soluble Forms by Cleavage of the Nh2-Terminal Signal Anchor. *Journal of Biological Chemistry* **262**, 17735-17743.

29. Sadler, J. E., Beyer, T. A., Oppenheimer, C. L., Paulson, J. C., Prieels, J. P., Rearick, J. I. & Hill, R. L. (1982). Purification of Mammalian Glycosyltransferases. *Methods in Enzymology* **83**, 495-496.

30. Liu, S., Venot, A., Meng, L., Tian, F., Moremen, K. W., Boons, G. J. & Prestegard, J. H. (2007). Spin-labeled analogs of CMP-NeuAc as NMR probes of the alpha-2,6-sialyltransferase ST6Gal I. *Chemistry & Biology* **14**, 409-418.

31. Sosnovsky, G. & Konieczny, M. (1976). Phosphorylative Spin-Labeling of Amino-Acids and Steroids. *Synthesis-Stuttgart*, 537-539.

32. Sosnovsky, G. & Konieczny, M. (1977). Novel Transfer-Reactions of a Nitroxyl-Labeled Phosphoryl Moiety. *Synthesis-Stuttgart*, 618-619.

33. Cocco, L. & Blakley, R. L. (1979). Synthesis of a Spin-Labeled Analog of Nicotinamide Adenine-Dinucleotide Phosphate and Its Interaction with Dihydrofolate-Reductase. *Biochemistry* **18**, 2414-2419.

34. Weiner, H. (1969). Interaction of a Spin-Labeled Analog of Nicotinamide-Adenine Dinucleotide with Alcohol Dehydrogenase .I. Synthesis Kinetics and Electron Paramagnetic Resonance Studies. *Biochemistry* **8**, 526-&.

35. Claesen, C. A. A., Daemen, C. J. M. & Tesser, G. I. (1986). Synthesis of Spin-Labeled Di-2-Deoxyadenylates and Tri-2-Deoxyadenylates. *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society* **105**, 116-123.

36. Schaub, C., Muller, B. & Schmidt, R. R. (2000). Sialyltransferase inhibitors based on CMP-quinic acid. *European Journal of Organic Chemistry*, 1745-1758.

37. Voet, D. & Voet, J. G. (2004). Rates of enzymatic reactions and Enzymatic catalysis. In *Biochemistry* 3rd edit., pp. 477-529. John Wiley & Sons INC., Hoboken, NJ.

38. Feng, L. M., Orlando, R. & Prestegard, J. H. (2006). Amide proton back-exchange in deuterated peptides: Applications to MS and NMR analyses. *Analytical Chemistry* **78**, 6885-6892.

39. Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C. A., Vizitiu, D., Withers, S. G. & Rini, J. M. (2006). X-ray crystal structures of rabbit N-

acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *Journal of Molecular Biology* **360**, 67-79.

40. Boix, E., Zhang, Y. N., Swaminathan, G. J., Brew, K. & Acharya, K. R. (2002). Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase, alpha-1,3-galactosyltransferase. *Journal of Biological Chemistry* **277**, 28310-28318.

41. Ni, L. S., Chokhawala, H. A., Cao, H. Z., Henning, R., Ng, L., Huang, S. S., Yu, H., Chen, X. & Fisher, A. J. (2007). Crystal structures of Pasteurella multocida sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* **46**, 6288-6298.

42. Ramakrishnan, B., Boeggeman, E., Ramasamy, V. & Qasba, P. K. (2004). Structure and catalytic cycle of beta-1,4-galactosyltransferase. *Current Opinion in Structural Biology* **14**, 593-600.

43. Unligil, U. M. & Rini, J. M. (2000). Glycosyltransferase structure and mechanism. *Current Opinion in Structural Biology* **10**, 510-517.

44. Qasba, P. K., Ramakrishnan, B. & Boeggeman, E. (2005). Substrate-induced conformational changes in glycosyltransferases. *Trends in Biochemical Sciences* **30**, 53-62.

45. Chiu, C. P. C., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. J. (2004). Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. *Nature Structural & Molecular Biology* **11**, 163-170.

46. Jones, D. T. (1999). GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology* **287**, 797-815.

47. McGuffin, L. J. & Jones, D. T. (2003). Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**, 874-881.

48. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure* **29**, 291-325.

49. Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D. M., Harduin-Lepers, A., Delannoy, P., Imberty, A. & Breton, C. (2004). Structure-function analysis of the human sialyltransferase ST3Gal I - Role of N-glycosylation and a novel conserved sialylmotif. *Journal of Biological Chemistry* **279**, 13461-13468.

50. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L. & Woods, R. J. (2008). GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **29**, 622-655.

51. Seidel, R. D., Zhuang, T. D. & Prestegard, J. H. (2007). Bound-state residual dipolar couplings for rapidly exchanging ligands of His-tagged proteins. *Journal of the American Chemical Society* **129**, 4834-4839.

52. Zhuang, T. D., Lee, H. S., Imperiali, B. & Prestegard, J. H. (2008). Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. *Protein Science* **17**, 1220-1231.

53. Zhuang, T. D., Leffler, H. & Prestegard, J. H. (2006). Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3. *Protein Science* **15**, 1780-1790.

CHAPTER 4

AN EXPLORATION OF METABOLIC LABELING OF NEURAL STEM CELLS AND CHARACTERIZATION OF THE METABOLITES

4.1 Introduction

Sialic acids are often found as terminal sugars in oligosaccharide chains of glycoproteins and glycolipids on mammalian cell surfaces. The distribution of sialic acids and sialylated conjugates is related to a number of diseases sates including overproduction of gangliosides (gangliosidosis) in lysosomes [1]. It would be important to develop tools for detecting excess gangliosides *in vivo* and monitoring the efficacy of the treatment for lysosome storage disorder. Our goal in this project is to develop improved NMR methods to detect gangliosides *in vivo* and explore it for future use in conjunction with MRI and MRS for which preliminary data on the detection of free sialic acids *in vivo* have existed [2]. One proposed NMR approach involves using ^{13}C labeled sialic acid donors to monitor incorporation in gangliosides and storage of gangliosides in lysosomes. It would help to see if incorporation and detection could be achieved at the single cell level. We received a generous gift of 200 million human embryonic stem cells – derived neural stem cells provided by Dr. Steven Stice of the University of Georgia. We treated differentiated neural stem cells with ^{13}C-labeled sialic acid, and assessed its adsorption into stem cells and its metabolic conversion to gangliosides by NMR and mass spectrometry (MS).

There seems to be some precedent for the uptake of at least sialic acid precursors from the medium by mammalian cells, conversion into the sialic acid donor CMP-sialic acid, and transfer of sialic acid to sugar acceptors in the course of sialocongjugate production. Sialic acid is synthesized in mammalian cells by the catalyzed condensation of pyruvate and N-acetylmannosamine. Radio-labeled acetylmannosamine (6-^{3}H) has been injected into rats and used to quantitate in vivo production of brain gangliosides [3]. ^{13}C-isotopes can be used in a similar way to produce metabolic products for *in vivo* detection by NMR. In-cell detection of some products has also been achieved in *E. coli.* In particular, ^{13}C,^{15}N-labeled α-2,8- polysialic acid has been produced in *E. coli* bacteria and detected by a combination of heteronuclear nuclear magnetic resonance (NMR) [4]. Azido sialic acids have been directly introduced to the human T-cell lymphoma Jurkat cells and its metabolic conversion into the cell-surface sialoside was monitored in vivo by fluorescence spectrometry [5].

Neural stem cells are endowed with high proliferative potential capacity for self-renewal and they are the common progenitor for neurons and glial cells. They have significant clinical use for the treatment of a variety of neurodegenerative disorders and spinal cord injuries. Although previous studies of neural stem cells have focused on the functions of proteins and their genes, the glycoconjugates expressed in these cells, particularly gangliosides, have gained increasing attention because of their important biological roles. However, little is known about the exact composition and subcellular localization of gangliosides in neural stem cells [6]. Studies on composition of gangliosides expression are expected to provide unique markers for the chronological events of

neuronal differentiation and contributed to the understandings of their functional relevance [7].

Gangliosides are particularly abundant in the central nervous system and they constitute a significant fraction (6%) of brain lipids [8]. Gangliosides are ceramide oligosaccharides that include at least one sialic acid residue among their sugar groups (Figure 4.1). The lipophilic ceramide tail is anchored to the lipid bilayer and the hydrophilic carbohydrate head is exposed to the outer plasma membrane for cellular interaction. Their complex carbohydrate groups act as specific receptors for pituitary glycoprotein hormones and bacterial protein toxins. Hence gangliosides are specific determinants of cell-cell recognition and are known to play important roles in cellular growth and differentiation in the nervous system. The elevated expression of gangliosides such as G_{M2} and G_{M3}, are often found in diseased neurons and correlated with crucial neurological developments in animals and humans [9].

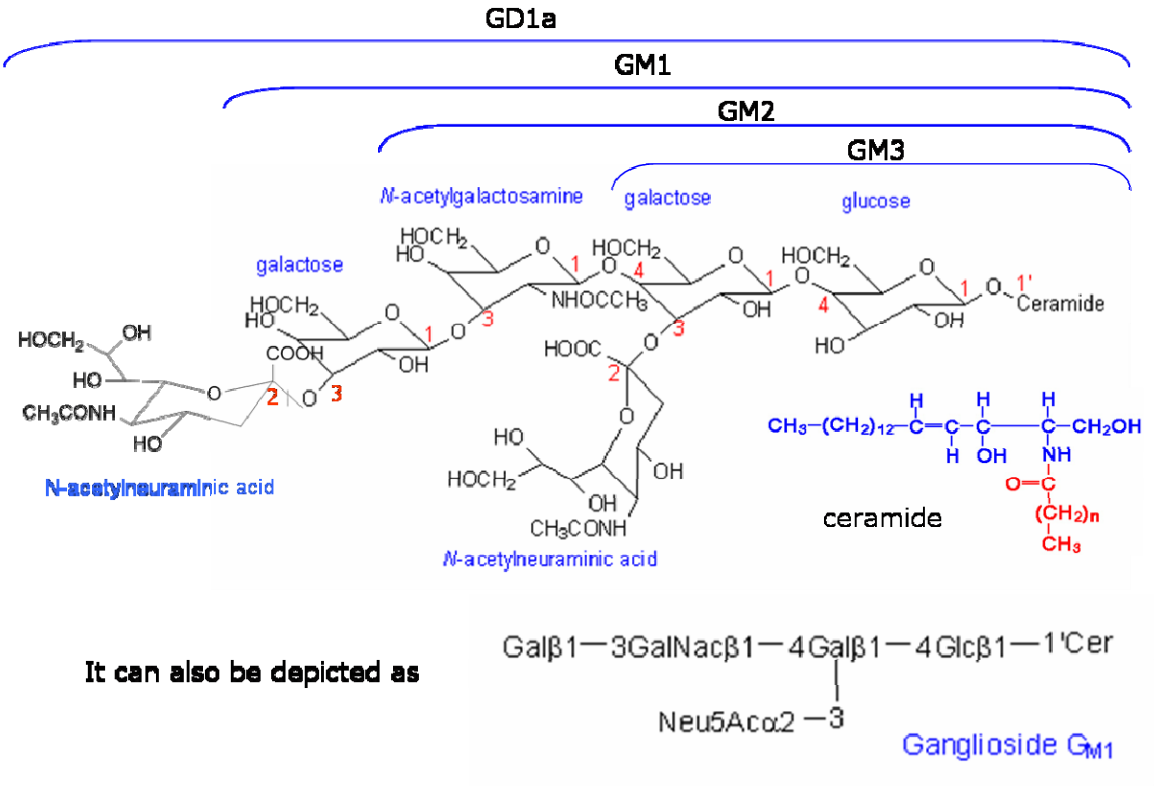


Figure 4.1. Structural formula of gangliosides

The oligosaccharide sequence and linkage sites of gangliosides can be determined by proton-proton Nuclear Overhauser Effects (NOEs) and proton chemical shifts induced by sialylation [10; 11]. Carbon chemical shift changes can be even bigger and more characteristic than proton chemical shifts. For instance, 2,3 and 2,6- linkage of sialylated oligosaccharides can be distinguished because the chemical shift of C2 in 2,6 linked sugars differs from that of its 2,3 linked counterpart by an additional contribution from the γ carbon. Since C2 is a quaternary carbon and not proton-coupled, we used a ^{1}H-detected ^{13}C-^{13}C double-quantum coherence pulse sequence to connect C2 to C3 and observe C2. This also provided an experiment that greatly suppresses natural abundant background. This background is at 1% (the natural abundance level of ^{13}C) for single quantum experiments, but becomes significant for high concentration metabolites that are

not of interest. In a double quantum spectrum the simultaneous presence of two ^{13}C carbons is required and the natural abundance background drops to 0.01%. Additionally, C1, C2 and C3 of sialic acids are very sensitive to the position of sialic acids in the oligosaccharides (OS) chain even when the sialic residues are linked to the identical sugars in the identical glycosidic linkage. One good example is G_{D1a} OS [12]. The chemical shift for C1, C2, C3 of the branched sialic acid is 175.2, 101.7, 37.5 ppm while the chemical shift for C1, C2, C3 of the terminal sialic acid is 174.3, 100.0, 39.9 ppm (standard 1,4-dioxane, 66.0 ppm in D_2O). The chemical shift difference for C3s is as big as 2.4 ppm but the rest of carbons are within 0.4 ppm difference.

For our purpose, we synthesized $^{13}C_5$-sialic acid with carbons 1, 2, and 3, and acetyl carbons, enriched in ^{13}C isotopes. It has been shown that esterification of hydroxyl groups on sugars can increase their cellular uptake by permitting passive diffusion through membranes. Their ester functional group can then be readily cleaved by cytosolic esterases in the cells for further metabolic conversion [5]. Therefore, we also synthesized acetylated $^{13}C_5$- sialic acid for comparison. We divided the immortalized stem cells (200 M) into two batches, one treated with ^{13}C-Sia (Sia-fed cells) and the other also treated with peracetylated ^{13}C-Sia (AcSia-fed cells). After harvesting the cells, we immediately took NMR snapshots of both batches of cells with the designed DQ experiments. For the AcSia-fed sample we detected the metabolic product in the pelleted cell as opposed to the external medium, and we tentatively identified this metabolite as a CMP-NeuAc derivative biosynthesized *in vivo*. To quantitate the incorporation of ^{13}C into gangliosides, we extracted the gangliosides from the stem cells and identified the ganglioside species of G_Ms and G_Ds by NMR and high performance thin layer

chromatography (HPTLC). To further determine the level of $^{13}C_5$ incorporation in gangliosides, the oligosaccharides were cleaved off the ceramide and permethylated for MS characterization.

4.2 Experimental

4.2.1 Peracetylation of ^{13}C-labeled sialic acid

N-hydroxysuccinimide and ^{13}C-acetyl chloride were combined to produce N-$^{13}C_2$-acetoxy-hydroxysuccinimide[13], which was further reacted with D-mannosamine to make N-$^{13}C_2$-acetyl-mannosamine. NeuAc aldolase was then used to react the N-$^{13}C_2$-acetyl-mannosamine with ^{13}C-pyruvate to produce ^{13}C-[1,2,3,10,11]-NeuAc [14]. This $^{13}C_5$-NeuAc (10 mg, 32 μmol) was dissolved in anhydrous pyridine (100 μl) and acetic anhydride (150 μl) at 4^{o}C. The resulting clear mixture was stirred at room temperature overnight. Methanol was added to quench the reaction and the solvent was coevaporated with toluene. The crude product (24 mg) was purified by silica gel chromatography by a column packed with 0.6 mg Iatrobeads in $CHCl_3$) and eluted sequentially with 60:1, 50:1, 40:1, 20:1 of $CHCl_3$: CH_3OH. The product in pyridium salt form was converted into the sodium salt by an AG cation exchange column eluted with water. The following lyophilization afforded 12 mg white solid (64% yield, α: β = 1:5 based on ^{1}H NMR). The product has moderate solubility in water. TLC (5:2:2 EtOAc: AcOH: H_2O) R_f 0.8 for alpha, R_f 0.7 for beta. ^{1}H NMR (300 MHz, D_2O, 4.77 ppm): 5.46 (dd, 1H, *J* = 8.1 Hz, *J* = 1.8 Hz, H7), 5.33 (sextet, 1H, *J* = 5.1 Hz, *J* = 4.2 Hz, H4), 5.20 (penta, 1H, *J* = 4.2 Hz, *J* = 3.0 Hz, H8), 4.46 (dd, 1H, *J* = 12.9 Hz, *J* = 2.4 Hz, H9), 4.24 (dd, 1H, *J* = 12.9 Hz, *J* = 4.2 Hz, H6), 4.18 (dd, 1H, *J* = 10.8 Hz, *J* = 1.8 Hz, H9), 3.99 (t, 1H, *J* = 10.5 Hz, H5), 2.89 (ddd, *J* = 84.0 Hz, *J* = 13.5 Hz, *J* = 5.4 Hz, $H3_{eq}$ for alpha), 2.65 (ddd, 1H, *J* = 84.0

Hz, *J* = 13.5 Hz, *J* = 5.4 Hz, H_{3eq} for beta), 2.12 (dt, *J* = 84.0 Hz, *J* = 12.3 Hz, $H3_{ax}$ for alpha), 2.19-1.94 (m, 15 H, O- acetyls), 2.17 (d, 3H, J = 96 Hz, NH-Ac), 2.05 (dt, 1H, *J* = 84.0 Hz, *J* = 12.3 Hz, $H3_{ax}$ for beta). ^{13}C (300 MHz, D_2O, 256 scans): 177.6 (d, *J* = 67.8 Hz, C10), 175.5 (d, *J* = 50.7 Hz, C1), 97.5 (dd, *J* = 66.3 Hz, *J* = 41.4 Hz, C2), 40.0 (d, *J* = 42.5 Hz, C3), 22.8 (d, *J* = 50.2 Hz, C11).

4.2.2 Cell cultivation

Before running the experiments, we tested the stem cells' tolerance to sialic acid. we grew 2 plates of 1 M cells in a medium containing 10-fold (20 μM) and 50-fold (100 μM) of regular sialic acids (dissolved in pH 7.0 phosphate buffered saline-PBS). During five days' exposure, both batches of cells grew normally.

The first 100 M stem cells were treated with a 50-fold excess of $^{13}C_5$-sialic acid and the other 100 M were treated with a 50-fold excess of acetylated $^{13}C_5$ -sialic acid. The cells were cultivated under published conditions [15] . It was suggested by Dr. Ron Schnaar (personal communication) that the turn-over of gangliosides is considered to be quite high – the order of a day in developing central nervous system. Therefore, the cells were harvested after 48 hours of incubation.

4.2.3 NMR of ^{13}C -labeled precursors in stem cells

Cells were harvested by scraping into 70 ml of PBS and subsequent centrifugation at 1000 rpm for 10 min into a 150 μl loose pellet. Another 75 μl of PBS was used to help transfer the pellet into a 5 mm Shigemi tube. 25 μl of D_2O were added to introduce 10% (v/v) of D_2O into the NMR sample. The sample first appeared to be a homogeneous suspension even though it settled down to two phases within 1 hour. The experiments were performed on a Varian 900 MHz spectrometer equipped with a cryogenically-

cooled probe. The INADEQUATE pulse sequence (Figure 4.2) selectively correlates double quantum (DQ) coherence of two directly-connected carbons to protons attached by one bond to one of the carbons, namely ^{13}C-^{13}C connectivity data via ^{1}H detection [16; 17].

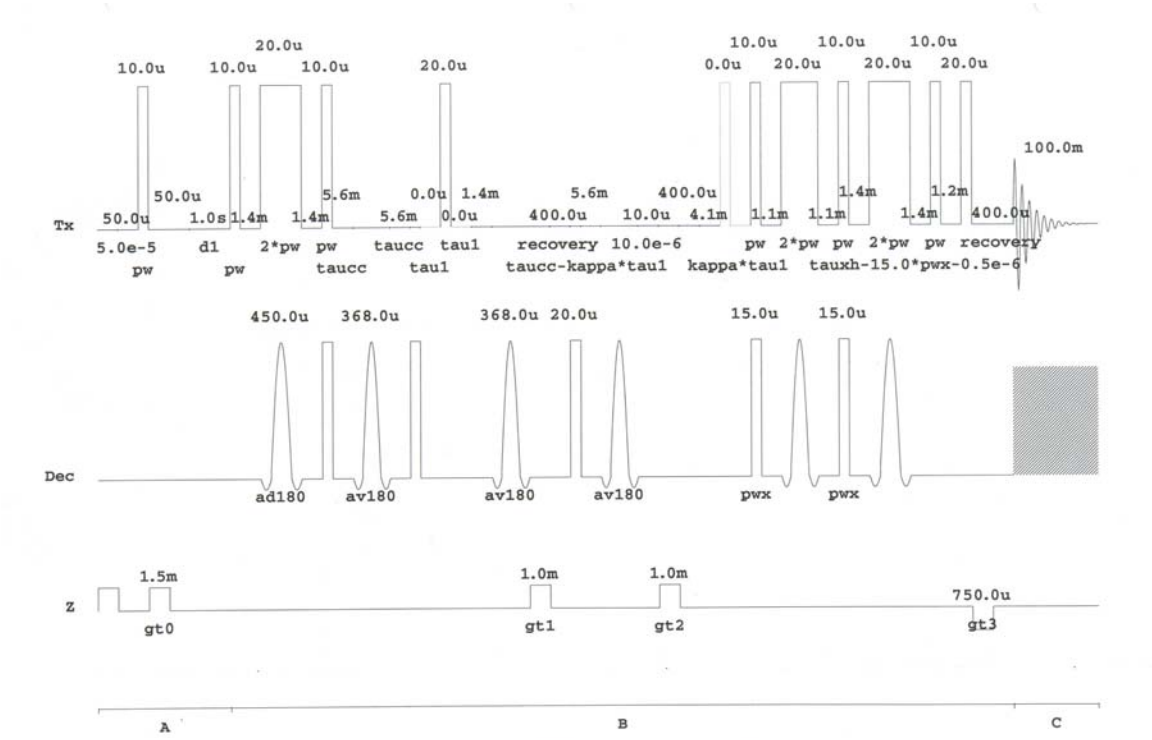


Figure 4.2. Pulse sequence of ^{1}H-detected ^{13}C-^{13}C (C2-C3) double quantum coherence displayed by Varian VnmrJ.

4.2.4 Extraction and separation of gangliosides (C-chloroform, M-methanol, W-water)

The extraction and purification methods followed the procedures used for murine neuroblastoma cells [18] and human mast cells [19]. The neural stem cells (recovered from the NMR sample) were stirred with C/M (1:1) at room temperature for 24 hours. The mixture was centrifuged at 2000 rpm for 10 minutes and the supernatant was collected. The residue was extracted twice with C/M/W (30:60:8). All the supernatant fractions were combined and the solvent was evaporated. The resulting mass of lipids was dissolved in

0.2 N methanolic NaOH, and mild saponification was carried out at 37^{o}C for 2 hr to destroy phospholipids. This was followed by neutralization with 0.2 M AcOH. The mixture was washed twice with hexane to remove fatty acids. After solvent evaporation, the residue was purified through Sephadex LH-20 column eluted with C/M (1:2). The neutral and acidic lipids were separated by DEAE Sepharose (fast flow, acetate form) with C/M/W (30:60:8), and then a gradient of C/M/0.2-0.8 M NH_4OAc (30:60:8) was applied to obtain mono-sialyl and dia-sialyl ganglioside fractions. Gangliosides were analyzed on an HPTLC plate using C/M/0.2% $CaCl_2$ (50:45:10 v/v) and specifically stained using the resorcinol-HCl reagent.

4.2.5 Preparation of oligosaccharides

Ceramide glycanase was purchased from CALBIOCHEM (Cat. # 219484). BAKERBOND spe C18 disposable reverse-phase columns were purchased from J.T.Baker (Cat. # 7020-01). Other reagents were purchased from Sigma-Aldrich.

The method followed the standard hydrolysis of gangliosides by ceramide glycanase [20]. The gangliosides (20 μg) were dissolved in 10 μl of NaOAc buffer (50 mM, pH 5.0) containing 0.2 % sodium cholate (2 mg/ml) as a detergent. Ceramide glycanase (0.1 U) and another 10 μl of NaOAc buffer were added. The mixture was incubated at 37^{o}C for 24 h. A Folch extraction was performed using C/M 2:1 to separate the oligosaccharides from the ceramide. The top layer (M/W) was collected and dried and purified by a reverse-phase C18 column eluted with 5% AcOH.

4.2.6 Permethylation

The procedure from a published protocol was followed [21]. 50% (w/w) NaOH (100 μl) was mixed well with anhydrous methanol (200 μl) and then anhydrous DMSO (4 ml).

The mixture was centrifuged briefly. The supernatant and the white foam were removed. The NaOH residue was dehydrated by washing with DMSO three times before being resuspended in DMSO (1 ml).

The following operations were carried out under Argon if possible. The prepared base (300 μl) was added to the oligosaccharide sample (~20 ug) dissolved in DMSO (200 μl). MeI (150 μl) was injected to start the reaction and the mixture was vigorously vortexed for 5 min and sonicated for 10 min. Water was added to quench the reaction. MeI was bubbled off with a very slow stream of argon until the solution appeared clear. The permethylated product was extracted by dichloromethane and washed with water and dried under an argon flow. The product was further purified by a reverse-phase C18 column sequentially eluted with water, 15% acetonitrile, and 85% acetonitrile. The fractions containing permethylated oligosaccharides were pooled and dried.

To prepare the MALD-TOF sample, the permethylated product was dissolved in 50% methanol/water and mixed with the equivalent volume of the matrix 2,5-dihydroxybenzoic acid (DHB). Mass spectra were obtained on an Applied Biosystems Voyager-DE MALDI-TOF mass spectrometer.

4.3 Results and discussion

4.3.1 *In vivo* NMR of stem cells

INADEQUATE (Incredible Natural Abundance DoublE-QUAntum Transfer Experiment) was originally designed to determine the constitution of small molecules at natural abundance [17]. Here we employed INADEQUATE experiments to improve the sensitivity of detecting small amounts of ^{13}C-labeled sialic acid in stem cells. The INADEQUATE experiment suppressed the background signals from metabolites at

natural abundance in ^{13}C and added a chemical shift component from C2 as well as C3. The C2 shift is expected to be particularly sensitive to the glycosidic linkage involving sialic acid. The resonances of C3 protons of free ^{13}C-Sia in media were observed for both the samples, but there was an additional pair of C3 proton resonances (2.45 ppm, 1.75 ppm) for the AcSia-fed stem cell sample (Figure 4.3).

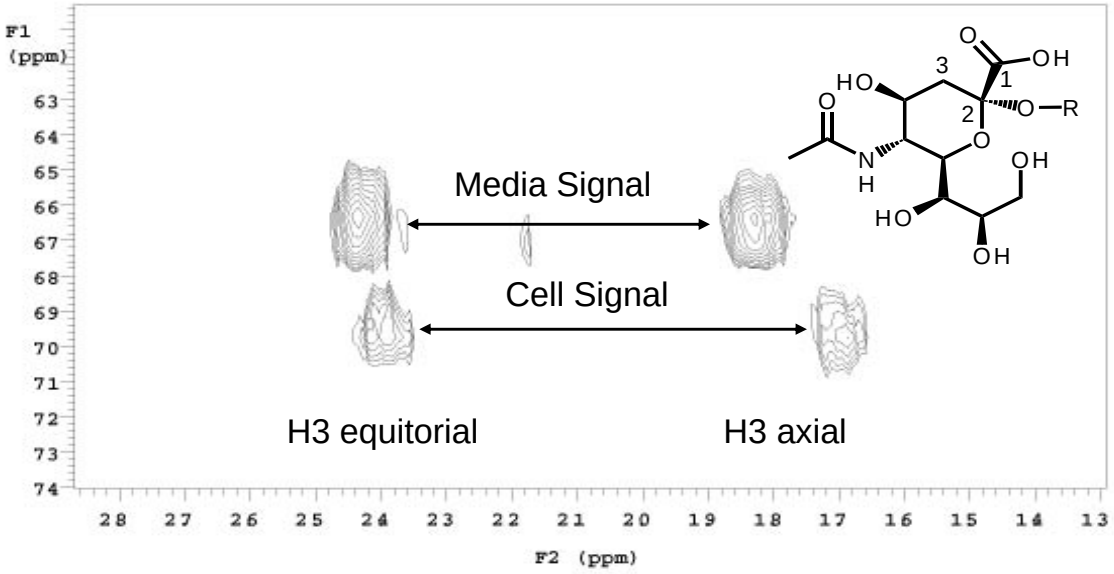


Figure 4.3. ^{1}H-detected ^{13}C-^{13}C (C2-C3) double quantum coherence spectrum for AcSia-fed stem cells. It shows the signals of the C3 methylene protons of sialic acid and its derivative.

Based on the chemical shifts (shown in Table 4.1) the new resonances most likely belong to a CMP-Sia derivative (possibly with partially acetylated Sia) which was synthesized inside the stem cells. Zero quantum spectra also were collected; they carry the additional benefit of reduced sensitivity to bulk susceptibility effects. It is noteworthy that similar internal cell-signals were not seen for the sialic acid fed cells (without acetylation). This suggests that the free sialic acids are not easily transported to the cell interior.

Table 4.1: ^{1}H chemical shifts of C3 methylene protons of sialic acid (Sia) and its derivatives in the indicated conditions

Samples (at pH 7.0)	H3eq (ppm)	H3ax (ppm)
Sia α/β	2.7/2.3	1.7/1.8
AcSia α/β	2.7/2.5	1.8/1.9
CMP-βSia	2.5	1.6
Sia in media	2.2	1.8
Sia in cell α/β	2.7 /2.2	1.6/1.8
AcSia in media	2.5	1.9
AcSia in cell	2.5	1.9
In vivo cell peaks	2.45	1.75

4.3.2 Quantitation and composition of extracted gangliosides

From ^{1}H NMR of the separated ganglioside fractions, we identified monosia-gangliosides G_Ms and disia-gangliosides G_Ds with their very characteristic olefinic, anomeric and alkyl regions [10]. Gangliosides were stained on HPTLC and quantitated using densitometry. The yield was estimated to be approximately 30 μg G_Ms and15 μg G_Ds for each batch of stem cells (100 M) with their molar ratios shown in the Table 4.2.

Table 4.2. Molar ratios of the G_Ms and G_Ds extracted from the two stem cell batches

Cells	G_{M3}	G_{M2}	G_{M1}	G_{D3}	G_{D2}	G_{D1}
AcSia-fed	1	15	5	1	1.5	15
Sia-fed	1	8	5	1	1	1.5

Because of the low concentration of gangliosides in the samples it was difficult to differentiate the specific species of G_Ms and G_Ds by NMR. Therefore, we used mass spectrometry for further analysis. The ganglioside fractions were individually treated

with ceramide glycanase to cleave off the lipid chain and yield the oligosaccharide chains which were permethylated for MS determination.

The $^{13}C_5$- ganglioside OS species with isotopic enrichment were detectable by MS in several cases (G_{M1}, G_{M2} in AcSia-cells and G_{D1}, G_{D2} in Sia-cells) despite their small amounts (Figure 4.4). However, the ^{13}C- isotopic distribution did not seem to be uniform over different types of gangliosides. G_{M2} and G_{D2} in Sia-cells have an m/z difference of 5 from their isotopic partners. We do not have an explanation for the m/z difference of 6 and 7 for the isotopic counterparts of G_{M1} and G_{D1} in AcSia-cells. We have attempted to estimate incorporation of isotopes for each species in Table 4.2. It is clear that non-acetylated sialic acid did enter the cells and was incorporated. However, the amounts incorporated with acetylated sialic acid are higher.

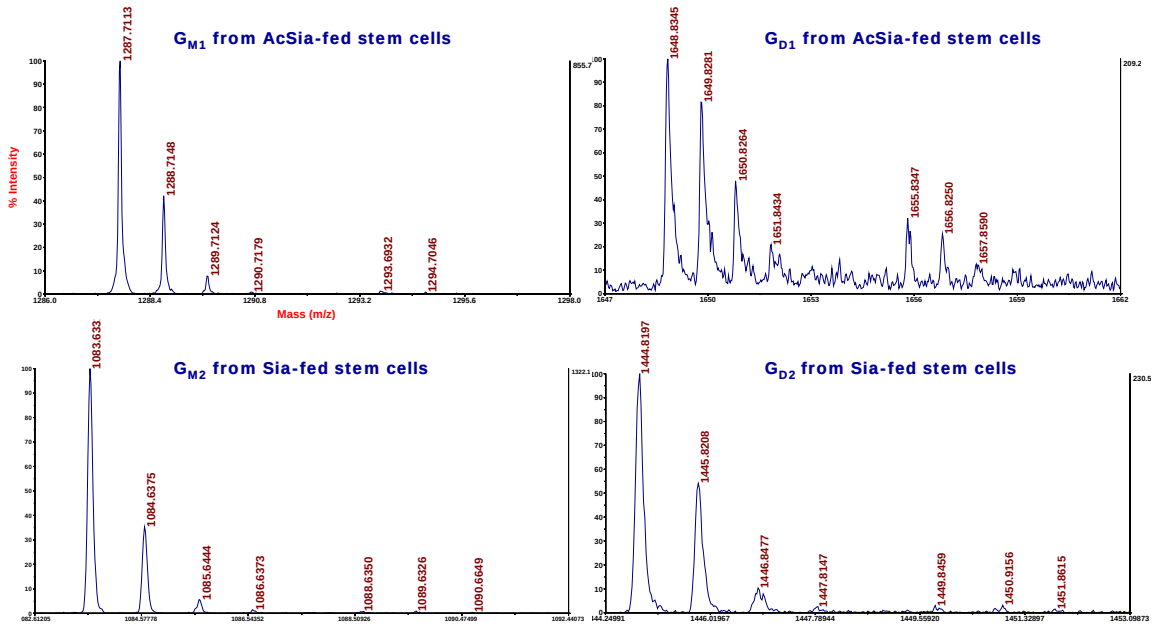


Figure 4.4. MS spectra of the oligosaccharide (OS) cleaved from the extracted gangliosides

It is an interesting fact that the types of gangliosides labeled in the two feedings are different. The expression pattern of the gangliosides could have been altered as a result of specific regulation of sialyltransferases at the transcriptional level or by intracellular signal transduction [6; 7]. For example, C17.2 cells, an immortalized neural stem cell line derived from neonatal mouse cerebellar cortex by v-myc transformation, express only a-series gangliosides, such as G_{M3}, G_{M2}, G_{M1} and G_{D1}, but not the b- and c-series due to the lack of G_{M3} sialyltransferase (ST-II) transcripts in C17 cells (Figure 4.5). The presence of acetylated sialic acid and high levels of sialic acid donor in the cells (as detected by in-cell NMR) could have altered the synthesis pathways.

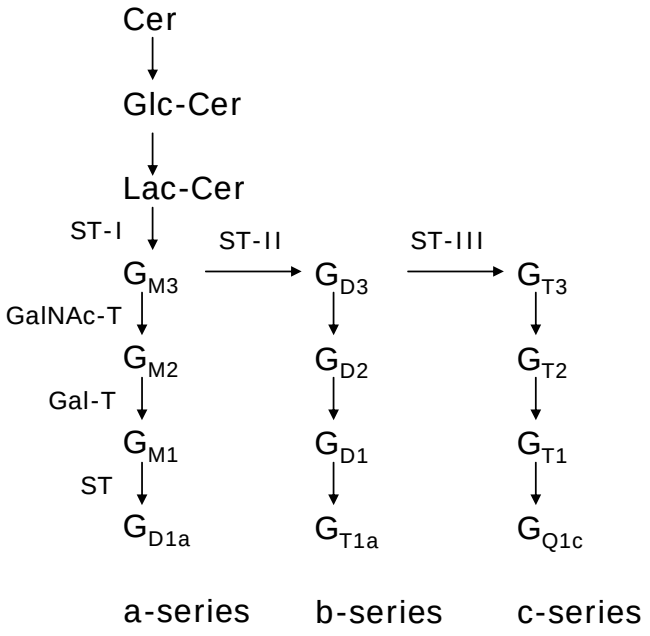


Figure 4.5 Biosynthetic pathways of major gangliosides [22]

4.4 Conclusion

Compared to sialic acid with free hydroxyl groups, the acetylated sialic acid clearly diffused more efficiently into neural stem cells and was converted into a CMP-NeuAc derivative. A small portion of this obviously went on to be incorporated in

gangliosides by intracellular biosynthesis. Less non-acetylated sialic acid entered the cells, but a higher proportion appears to have gone on to be incorporated in gangliosides. Using NMR and MS we identified the specific species of gangliosides expressed in neural stem cells and could show that the labels were differentially incorporated in these species.

Despite the low incorporation of ^{13}C-Sia into gangliosides, we obtained some preliminary results that will allow refinement of protocols for future NMR trials. An intermediate could be detected by in-cell NMR when acetylated sialic acid was fed. The use of a double quantum signal proves very effective in suppressing natural abundance signals; one just needs a higher level of label incorporation. The stages of cell differentiation should be more thoroughly investigated in order to determine the optimum phases for infusing ^{13}C-sialic acid into neural stem cells and for harvesting the cells. Additionally we intend to employ a NMR tube equipped with a simple apparatus that can replenish the buffer in the NMR tube with oxygen and keep the cells viable during NMR acquisition.

References

1. Kacher, Y. & Futerman, A. H. (2006). Genetic diseases of sphingolipid metabolism: Pathological mechanisms and therapeutic options. *Febs Letters* **580**, 5510-5517.

2. Larvaron, P., Bielicki, G., Boespflug-Tanguy, O. & Renou, J. P. (2006). Proton MRS of early post-natal mouse brain modifications in vivo. *Nmr in Biomedicine* **19**, 180-187.

3. Yamagishi, K., Mishima, K., Ohgami, Y., Iwasaki, K., Jimbo, M., Masuda, H., Igarashi, Y., Inokuchi, J. & Fujiwara, M. (2003). A synthetic ceramide analog ameliorates spatial cognition deficit and stimulates biosynthesis of brain gangliosides in rats with cerebral ischemia. *European Journal of Pharmacology* **462**, 53-60.

4. Azurmendi, H. F., Vionnet, J., Wrightson, L., Trinh, L. B., Shiloach, J. & Freedberg, D. I. (2007). Extracellular structure of polysialic acid explored by on cell solution NMR. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11557-11561.

5. Luchansky, S. J., Goon, S. & Bertozzi, C. R. (2004). Expanding the diversity of unnatural cell-surface sialic acids. *Chembiochem* **5**, 371-374.

6. Suetake, K., Liour, S. S., Tencomnao, T. & Yu, R. K. (2003). Expression of gangliosides in an immortalized neural progenitor/stem cell line. *Journal of Neuroscience Research* **74**, 769-776.

7. Liour, S. S., Kapitonov, D. & Yu, R. K. (2000). Expression of gangliosides in neuronal development of P19 embryonal carcinoma stem cells. *Journal of Neuroscience Research* **62**, 363-373.

8. Voet, D. & Voet, J. (2004). Biochemistry. *Wiley* **3rd Ed**

9. Walkley, S. U. (2004). Secondary accumulation of gangliosides in lysosomal storage disorders. *Seminars in Cell & Developmental Biology* **15**, 433-444.

10. Koerner, T. A. W., Prestegard, J. H., Demou, P. C. & Yu, R. K. (1983). High-Resolution Proton Nmr-Studies of Gangliosides .1. Use of Homonuclear Two-

Dimensional Spin-Echo J-Correlated Spectroscopy for Determination of Residue Composition and Anomeric Configurations. *Biochemistry* **22**, 2676-2687.

11. Koerner, T. A. W., Prestegard, J. H., Demou, P. C. & Yu, R. K. (1983). High-Resolution Proton Nmr-Studies of Gangliosides .2. Use of Two-Dimensional Nuclear Overhauser Effect Spectroscopy and Sialylation Shifts for Determination of Oligosaccharide Sequence and Linkage Sites. *Biochemistry* **22**, 2687-2690.
12. Sabesan, S., Duus, J. O., Fukunaga, T., Bock, K. & Ludvigsen, S. (1991). Nmr and Conformational-Analysis of Ganglioside-Gd1a. *Journal of the American Chemical Society* **113**, 3236-3246.
13. Heidlas, J. E., Lees, W. J., Pale, P. & Whitesides, G. M. (1992). Gram-Scale Synthesis of Uridine 5'-Diphospho-N-Acetylglucosamine - Comparison of Enzymatic and Chemical Routes. *Journal of Organic Chemistry* **57**, 146-151.
14. Aubin, Y. & Prestegard, J. H. (1993). Structure and Dynamics of Sialic-Acid at the Surface of a Magnetically Oriented Membrane System. *Biochemistry* **32**, 3422-3428.
15. Shin, S. J., Mitalipova, M., Noggle, S., Tibbitts, D., Venable, A., Rao, R. & Stice, S. L. (2006). Long-term proliferation of human embryonic stem cell-derived neuroepithelial cells using defined adherent culture conditions. *Stem Cells* **24**, 125-138.
16. Meissner, A., Moskau, D., Nielsen, N. C. & Sorensen, O. W. (1997). Proton-detected C-13-C-13 double quantum coherence. *Journal of Magnetic Resonance* **124**, 245-249.

17. Reif, B., Kock, M., Kerssebaum, R., Kang, H., Fenical, W. & Griesinger, C. (1996). ADEQUATE, a new set of experiments to determine the constitution of small molecules at natural abundance. *Journal of Magnetic Resonance Series A* **118**, 282-285.

18. Bieberich, E., MacKinnon, S., Silva, J., Li, D. D., Tencomnao, T., Irwin, L., Kapitonov, D. & Yu, R. K. (2002). Regulation of ganglioside biosynthesis by enzyme complex formation of glycosyltransferases. *Biochemistry* **41**, 11479-11487.

19. Ren, S. L., Kambe, N., Du, Z. M., Li, Y. L., Xia, H. Z., Kambe, M., Bieberich, E., Pozez, A., Grimes, M., Yu, R. K., Irani, A. M. & Schwartz, L. B. (2001). Disialoganglioside GD3 is selectively expressed by developing and mature human mast cells. *Journal of Allergy and Clinical Immunology* **107**, 322-330.

20. Zhou, B., Li, S. C., Laine, R. A., Huang, R. T. C. & Li, Y. T. (1989). Isolation and Characterization of Ceramide Glycanase from the Leech, Macrobdella-Decora. *Journal of Biological Chemistry* **264**, 12272-12277.

21. Anumula, K. R. & Taylor, P. B. (1992). A Comprehensive Procedure for Preparation of Partially Methylated Alditol Acetates from Glycoprotein Carbohydrates. *Analytical Biochemistry* **203**, 101-108.

22. Yanagisawa, M., Liour, S. S. & Yu, R. K. (2004). Involvement of gangliosides in proliferation of immortalized neural progenitor cells. *Journal of Neurochemistry* **91**, 804-812.

CHAPTER 5

CONCLUDING REMARKS

Sialyltransferases such as ST6Gal-1 present many challenges for X-ray crystallography and conventional NMR methods because of their membrane-associated features and required glycosylation for mammalian sialyltransferases. These properties result in crystallization difficulties for X-ray crystallography and difficulties in expressing proteins in suitable isotopically-labeled forms for NMR. The techniques developed within this dissertation, relying on unlabeled or sparsely-labeled protein and isotopically-labeled or spin-labeled ligands, have provided an important step toward structural and functional characterization of these proteins. Transferred NOE experiments demonstrated selective binding to ST6Gal-1 from the changes in the size and sign of substrate NOEs, and provided conformational information for the synthesized CMP-3FNeuAc including glycosidic bond torsion for cytosine-ribose and the distance for their proximate protons. Saturation transfer experiments identified cytidine as the binding epitope for CMP-NeuAc and the cytidine protons that have shortest distances to the protein surface. ^{19}F-NMR studies and ligand titration on CMP-3FNeuAc showed that it has a slow on-off rate when binding to ST6Gal-1. The synthesized isotopically-labeled CMP-NeuAc facilitated the acquisition of residual dipolar couplings for this compound, helping to define the relative orientation of remote parts of the molecule and setting the stage for future work on the orientation of bound donor in the ST6Gal-1 system. The synthesized nitroxide-labeled ligands identified the four phenylalanines located within or

another one near the active site of sparsely-labeled ST6Gal-1. The distances from these phenylalanines to the ligand were calculated based on the changes in resonance intensities shown in HSQC spectra. A nitroxide-labeled donor was then used to perturb the relaxation of a simultaneously bound acceptor that was synthesized in ^{13}C-labeled form. The measured relaxation changes were converted into the distances to deduce the relative placement of the interacting substrates. Finally, preliminary results were acquired for detecting one of the most important sialylated glycoconjugates - gangliosides in cells by NMR. Since there has not been a reported molecular structure of ST6Gal-1 nor this complex with its ligands, the above methods and results appear to be very valuable for studying this biologically-significant but structurally-challenging protein.

The use of sparse-labeling techniques for structural investigation is still in its infancy. However, their potential importance for difficult structural problems, such as those involving membrane proteins [1] and multi-protein complexes is clearly recognized [2; 3]. New methods for predicting structures from sparse constraints also have begun to appear. These have used sparse distance constraints from a variety of other sources, including chemical cross-linking, backbone to backbone NOEs, and paramagnetic metal ion perturbations [4; 5; 6; 7]. Distance constraints coming from the nitroxide spin-label perturbations described here can be used in a very similar way. The absence of appropriate assignment strategies for resonances from sparsely-labeled proteins has been an impediment to turning distance constraints into an accurate picture of protein structure and bound ligand geometry. However, we have been able to produce models of the way donor and acceptor ligands align in the binding site. These models map the binding site in ways that might be used as a basis for rational drug design. While the structures

presented here cannot be considered of sufficient resolution at this point, they do provide a stepping stone for future work. In particular, work with higher affinity acceptors should provide better data, and orientation-sensitive data such as ligand RDCs [8; 9; 10] should greatly improve the models of ST6Gal-1. With such improvements, one can expect a better structural understanding of ST6Gal-1 and a basis for structure-based design of specific inhibitors to regulate sialic acid distribution.

An important theme throughout this work has been the synergy between NMR spectroscopic methodology and synthetic methodology that can produce specifically labeled substrates. The synthesized substrates certainly will be applicable to other sialyltransferases. There are about 411 sialyltransferases identified in the CAZy database, all believed to use the same CMP-NeuAc sugar donor and acceptors which have structures similar to LacNAc. The synthetic approaches for the substrates, especially for the donors, are also potentially applicable to the preparation of probes for other glycosyltransferases. There are thousands of entries distributed among 87 families, but there are only a few commonly used sugar donors. This makes a limited set of spin-labeled compounds broadly applicable to a large class of structural problems. Most importantly, the sparse labeling NMR methods derived for the structural analysis of ST6Gal-1, itself a glycosylated protein, should find numerous applications to other glycoproteins in the future.

References

1. Brown, W. M., Faulon, J. L. & Sale, K. (2005). A deterministic algorithm for constrained enumeration of transmembrane protein folds. *Computational Biology and Chemistry* **29**, 143-150.

2. Tang, C. & Clore, G. M. (2006). A simple and reliable approach to docking protein-protein complexes from very sparse NOE-derived intermolecular distance restraints. *Journal of Biomolecular NMR* **36**, 37-44.

3. Ye, X. D., O'Neil, P. K., Foster, A. N., Gajda, M. J., Kosinski, J., Kurowski, M. A., Bujnicki, J. M., Friedman, A. M. & Bailey-Kellogg, C. (2004). Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. *Protein Science* **13**, 3298-3313.

4. Barbieri, R., Luchinat, C. & Parigi, G. (2004). Backbone-only protein solution structures with a combination of classical and paramagnetism-based constraints: A method that can be scaled to large molecules. *Chemphyschem* **5**, 797-806.

5. Kim, D. E., Chivian, D. & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **32**, W526-W531.

6. Li, W., Zhang, Y. & Skolnick, J. (2004). Application of sparse NMR restraints to large-scale protein structure prediction. *Biophysical Journal* **87**, 1241-1248.

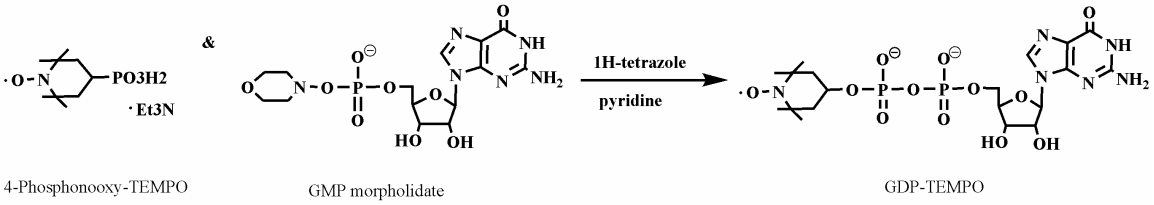
7. Qu, Y. X., Guo, J. T., Olman, V. & Xu, Y. (2004). Protein structure prediction using sparse dipolar coupling data. *Nucleic Acids Research* **32**, 551-561.

8. Seidel, R. D., Zhuang, T. D. & Prestegard, J. H. (2007). Bound-state residual dipolar couplings for rapidly exchanging ligands of His-tagged proteins. *Journal of the American Chemical Society* **129**, 4834-4839.

9. Zhuang, T. D., Leffler, H. & Prestegard, J. H. (2006). Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3. *Protein Science* **15**, 1780-1790.

10. Zhuang, T. D., Lee, H. S., Imperiali, B. & Prestegard, J. H. (2008). Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. *Protein Science* **17**, 1220-1231.

APPENDIX: SYNTHESIS OF GDP-TEMPO



All the reagents were purchased from Sigma-Aldrich. Anhydrous pyridine was used without further purification. The commercial 4-Phosphonooxy-TEMPO hydrate (pH 3.5) was only partially in the radical form (^{1}H NMR) so it was reoxidized before use. The cation-exchange resin AG 50W-X2 (H^+ form, strongly acidic) was purchased from Bio-Rad Laboratories and converted to the $Et3N^+$ salt form before use. When samples were coevaporated with anhydrous pyridine to remove residual water, argon was used to bring the pressure back to normal. Analytical thin layer chromatography was performed using silica gel 60 F254 precoated aluminum plates (Merck); compound spots were visualized by fluorescence and/or by charring after treatment with cerium molybdophosphate. Size-exclusion chromatography was performed on a column packed with Bio-Gel P-2 resin, extra fine (Bio-Rad Laboratories). ^{31}P NMR spectra were acquired on a Varian 300 MHz spectrometer in D_2O solvent. Signals at δ 7.80 ppm (GMP-morpholidate), 0.60 ppm (GMP), 0.10 ppm (TEMPO-P), -1.90 ppm (GMP-tetrazole), were referenced to 85% H_3PO_4 (δ 0.00) as external standard. All NMR spectra of the TEMPO derivatives were acquired on a Varian 500 MHz NMR spectrometer, in the indicated solvents, after addition of phenyl hydrazine as a reducing agent to the nitroxide sample. Mass spectra were obtained on an Applied Biosystems Voyager-DE MALDI-TOF mass spectrometer by using DHB (2,5-dihydroxybenzoic acid) as a matrix.

4-Phosphonooxy-TEMPO hydrate (13 mg) was loaded on the AG 50W-X2 Et_3N^+ exchange column (0.8 ml x 6 cm) and eluted with water. The solution (pH ~5) was evaporated to 1 ml, diluted with 12 ml methanol, stirred with open air at r.t. for 2 days, evaporated, and dried under vacuum to give triethylammonium TEMPO phosphate in the radical form (16 mg). The content of triethylamine was determined to be 0.7 equivalents. ^{1}H NMR (500 MHz, CD_3OD): weak signals were observed for the reduced form: δ 4.57 (broad s, 1H, H-4), 3.14(q, 7.0 Hz, 4H, O-CH2-C), 2.23 (dd, 13.5 Hz, 4.0 Hz, 1H, CHax), 2.08 (dd, 13.0 Hz, 4.5 Hz, 1H, CHax), 1.64 (t, 12.0 Hz, 1H, CHeq), 1.55 (t, 12.0 Hz, 1H, CHeq), 1.47(d, 5.0 Hz, 4H, methyl Hs), 1.29 (t, 6H, O-C-CH_3), 1.16 (m, 8H, methyl Hs),

TEMPO-$PO_3H_2Et_3N$ (12 mg, 34 μmol) was coevaporated with dry pyridine (3 x 1.5 ml) to give an orange residue. Guanosine 5′-monophosphomorpholidate 4-morpholine-N,N′-dicyclohexylcarboxamidine salt (40 mg, 55 μmol) was dissolved in 0.45 M 1H-tetrazole in ACN (0.25 ml, 113 μmol). The mixture was coevaporated with dry pyridine twice before dissolving in 0.8 ml pyridine. The resulting clear slightly yellow solution was combined with TEMPO-$PO_3H_2Et_3N$ to give a clear orange solution. After 2 hours, the reaction mixture became thick and opaque. The reaction was monitored by ^{31}P NMR in D_6-DMSO. After 4 days' stirring under argon, the reaction was diluted with water (1.5 ml) to give a clear solution which was then evaporated and coevaporated twice with water to remove pyridine. The crude product was first purified on an AG 50W-X2 Na+ exchange column and then a Bio-Gel P-2 column (1.5 x 4.5 cm) eluted with water at 2 ml/hr. The GDP-TEMPO-containing fractions were pooled and lyophilized to give a yellow fine powder (7.5 mg, 37%). ^{1}H NMR (500 MHz, CD_3OD):

no signlas were observed for TEMPO groups; for the reduced form: δ 8.05 (s, 1H, G H-9), 5.83 (d, J = 6.0 Hz, 1H, R H-1), 4.78 (m, J = 4.0 Hz , 1H, R H-2), 4.58 (m, 1H, H-4), 4.51(m, 1H, R H-3), 4.28 (m, 1H, R H-5), 4.19 (m, 2H, R H-4, R H-5), 4.15 (dd, 12.5 Hz, 4.0 Hz, 2H, 2CHax), 1.59 (td, 12.5 Hz, 4.0 Hz, 2H, 2CHeq), 1.16 (m, 12H, $4CH_3$). ^{31}P, broad, δ -11.05 (H_3PO_4, δ 0.00); R_f 0.40, 4:1 i-PrOH/0.5 M NH_4OAc; MALDI-TOF/MS for the reduced form: ($M+Na^+$) m/z 621.1, calcd for $C_{19}H_{32}N_6O_{12}P_2Na$ 621.4342.

References

Wittmann, V. & Wong, C-H (1997). 1H-Tetrazole as catalyst in phosphomorpholidate coupling reactions: efficient synthesis of GDP-fucose, GDP-mannose, and UDP-galactose. *J. Org. Chem* **62**, 2144-2147.