

THE DEVELOPMENT OF COMPUTATIONAL CARBOHYDRATE GRAFTING AND THE  
GLYCAM FORCE FIELD TO UNDERSTAND HOW GLYCAN STRUCTURE ALTERS  
BIORECOGNITION

by

MATTHEW BRUCE TESSIER

(Under the Direction of Robert J. Woods)

ABSTRACT

Complex carbohydrates (glycans) have long been known to play a role in immune response, regulation of cellular activity, and cell-cell interactions, to name a few. Thus the ability to model glycan structure and their interactions with other biomolecules (i.e. biorecognition) is essential to understanding and exploiting glycan functionality in the design of pharmaceutical glycomimetics, or molecules with similar properties to glycans. Over several decades, computational methods have become essential to characterizing glycan structure and bioactivity when only sparse experimental data is available. This work expands on those efforts by improving on the GLYCAM molecular mechanics force field to include a wider range of glycan structures including glycolipids and glycosaminoglycans. The GLYCAM force field was used in molecular dynamics (MD) simulations to predict the three-dimensional (3D) structures of glycans and glycoconjugates. Then, utilizing the three-dimensional glycan structure data from glycan simulations and experimental data, a virtual glycan 3D structure library was generated. In this case, the virtual library was employed to establish the first computational prediction of bulk

carbohydrate-protein specificity using a method called Computational Carbohydrate Grafting (CCG). This method has been shown to be useful in augmenting the results of experimental specificity screening and it can be used to test the specificity of glycans which are not included on the experimental arrays while providing 3D structures of protein-carbohydrate complexes. The CCG method was used to predict the binding specificity restrictions of the anti-tumor antibody JAA-F11 and provided a 3D structural rationale for its binding specificity. The development of the force field and CCG method are all part of an effort to better understand how the 3D structure of glycans impact biorecognition so as to guide the development of novel therapeutic or diagnostic glycomimetics.

INDEX WORDS: Computational carbohydrate grafting, GLYCAM, Glycosaminoglycans, Glycolipids, Lipids, Carbohydrates, Glycans, Force field parameter development, Virtual glycan array, Virtual docking, Molecular mechanics, Antibody binding, Glycan array, NMR

THE DEVELOPMENT OF COMPUTATIONAL CARBOHYDRATE GRAFTING AND THE  
GLYCAM FORCE FIELD TO UNDERSTAND HOW GLYCAN STRUCTURE ALTERS  
BIORECOGNITION

by

MATTHEW BRUCE TESSIER

B.S., The Florida State University, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial  
Fulfillment of the Requirements of the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

© 2013

Matthew Bruce Tessier

All Rights Reserved

THE DEVELOPMENT OF COMPUTATIONAL CARBOHYDRATE GRAFTING AND THE  
GLYCAM FORCE FIELD TO UNDERSTAND HOW GLYCAN STRUCTURE ALTERS  
BIORECOGNITION

by

MATTHEW BRUCE TESSIER

Major Professor: Robert J. Woods

Committee: James H. Prestegard  
Geoffrey D. Smith

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
December 2013

## ACKNOWLEDGEMENTS

I would like to acknowledge and thank the many individuals who contributed to my research throughout the years beginning with my research advisor, Professor Robert J. Woods, and my advisory committee, Professors James Prestegard and Geoffrey Smith. I would like to thank those directly involved in the work that has matured into the manuscripts presented herein as **Chapters 2-4 (Footnotes 1, 2, and 3)**. There are several mentors from Prof. Woods' research group that I would specifically like to acknowledge for their guidance early on in graduate school, namely Dr. Austin B. Yongye, Dr. Sarah Tschampel-Wittkopp, Dr. Sameer Kawatkar, Dr. Mari DeMarco, and Dr. Elisa Fadda. Very special thanks to Dr. B. Lachele Foley who has been there from the start as mentor, colleague, and friend. I'd like to acknowledge several peers for acting as a source of feedback, participating in discussions, working together on projects, and wonderful friends: Olga Charvatova, Charlisa Daniels, Oliver Grant, Jodi Hadden, Diarmuid Kenny, Dr. Christine Leoff-Dawson, Spandana Makeneni, Joanne Martin, Valerie Murphy, Anita Nivedha, Dr. Ahammadunny Pathiaseril, Kausar Samli, Amika Sood, Michael Tessier, Xiaocong Wang, and Dr. Nina Weisser. Another special thanks to Mark Baine who is a great friend and colleague, and without whom paperwork would have been a nightmare. Lastly, I would like to acknowledge the unwavering support of my family, and especially my wife, Tzafra, who has been with me throughout the whole process. My sincerest thanks to everyone who has made this work possible.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	iv
CHAPTER	
1. INTRODUCTION AND LITERATURE REVIEW .....	1
1.1 Overview .....	1
1.2 Glycan Structure .....	2
1.3 Computational Methods for Predicting Glycan Structure and Properties.....	11
1.4 Defining Protein-Carbohydrate Complexes .....	17
2. EXTENSION OF THE GLYCAM06 BIOMOLECULAR FORCE FIELD TO LIPIDS, LIPID BILAYERS AND GLYCOLIPIDS .....	23
2.1 Abstract .....	24
2.2 Introduction .....	24
2.3 Methods.....	26
2.4 Results and Discussion.....	35
2.5 Conclusions .....	50
2.6 Acknowledgements.....	53
2.7 Epilogue .....	53

3. CONFORMATIONAL ANALYSIS OF GLYCOSAMINOGLYCANS BY NMR AND MOLECULAR DYNAMICS SIMULATIONS USING PARAMETER EXTENSIONS FOR THE GLYCAM FORCE FIELD .....	55
3.1 Abstract .....	56
3.2 Introduction .....	56
3.3 Methods.....	60
3.4 Results .....	68
3.5 Discussion .....	85
3.6 Acknowledgements.....	87
4. COMPUTATIONAL SCREENING OF THE HUMAN TF-GLYCOME PROVIDES A STRUCTURAL DEFINITION FOR THE SPECIFICITY OF THE ANTI-TUMOR ANTIBODY JAA-F11 .....	88
4.1 Abstract .....	89
4.2 Introduction .....	90
4.3 Results .....	95
4.4 Discussion .....	107
4.5 Materials and Methods.....	110
4.6 Acknowledgements.....	114
4.7 Epilogue .....	114
5. CONCLUSIONS AND FUTURE DIRECTIONS .....	116
6. REFERENCES .....	118

7. APPENDIX .....	138
7.1 Conformational analysis of glycosaminoglycans by NMR and molecular dynamics simulations using parameter extensions for the GLYCAM force field .....	138
7.2 Computational screening of the human TF-glycome provides a structural definition for the specificity of anti-tumor antibody JAA-F11 .....	146

# 1. INTRODUCTION AND LITERATURE REVIEW

## 1.1 Overview

Complex carbohydrates form the basis of many biological processes vital to cellular function including biosynthetic pathway regulation, cell-cell interactions, and immune response. While it has been known that aberrant glycosylation is important in numerous disease states[1], demonstrating the role that glycan structures have on biological functionality continues to be a challenge for the combined fields of carbohydrate chemistry, biochemistry and biology, also known as glycobiology[2,3]. The need to understand this relationship has led to the development of glycan microarrays which allow screening of a protein against hundreds of glycan structures in a single experiment[4,5,6]. However, the quantity of binding partners identified using this method presents a problem for structural biochemistry. In determining the structure of a protein-ligand complex, experimental techniques depend on the isolation or synthesis of sufficient quantities of reagents before diffraction or NMR data is even collected. This provides an opportunity for computational modeling methods to reduce costs by characterizing binding interactions without the need for substantial quantities of reagents.

This work contributes to glycobiology through the characterization of the 3D structure of glycans and glycoconjugates while establishing a new method to relate these structures to biorecognition (the process by which cells respond to chemical compounds). Computational 3D structure predictions were made by expanding on the GLYCAM06 molecular mechanics force field to include new glycoconjugate structures, particularly those related to lipids (Chapter 2) and glycosaminoglycans (GAGs, Chapter 3). A modified version of the GLYCAM06 force field was

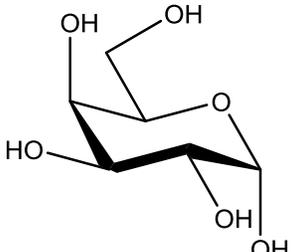
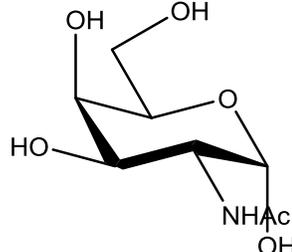
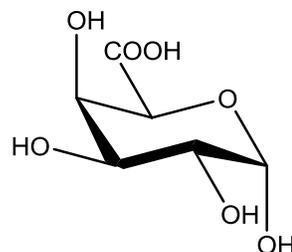
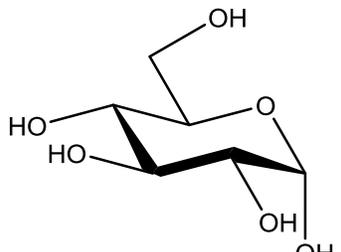
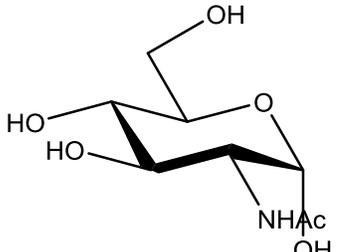
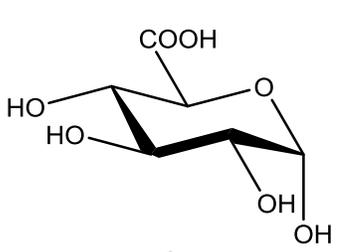
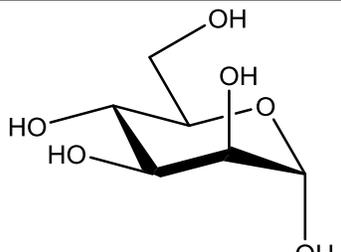
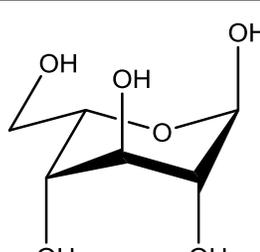
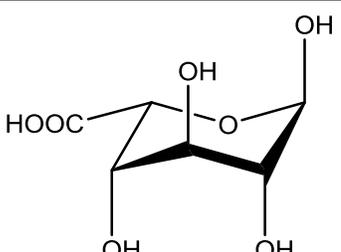
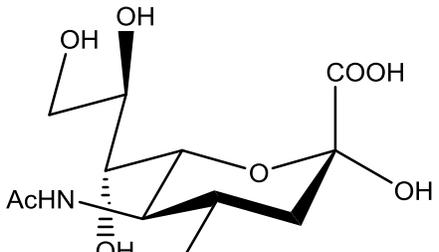
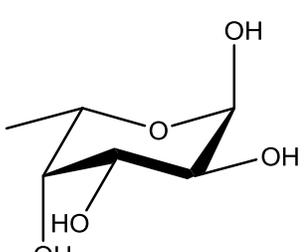
then employed in computational simulations of GAGs to expound their 3D structures and relate them to experimental observables (Chapter 3). The success of GLYCAM at defining carbohydrate 3D structure facilitated the development of a novel method used to characterize the glycan-binding specificity of an antibody-carbohydrate interaction (Chapter 4). This new technology augments the specificity information from glycan microarray screenings and provides an opportunity to screen glycans not found on experimental arrays.

## 1.2 Glycan Structure

### 1.2.1 Defining Monosaccharide and Glycan Molecules

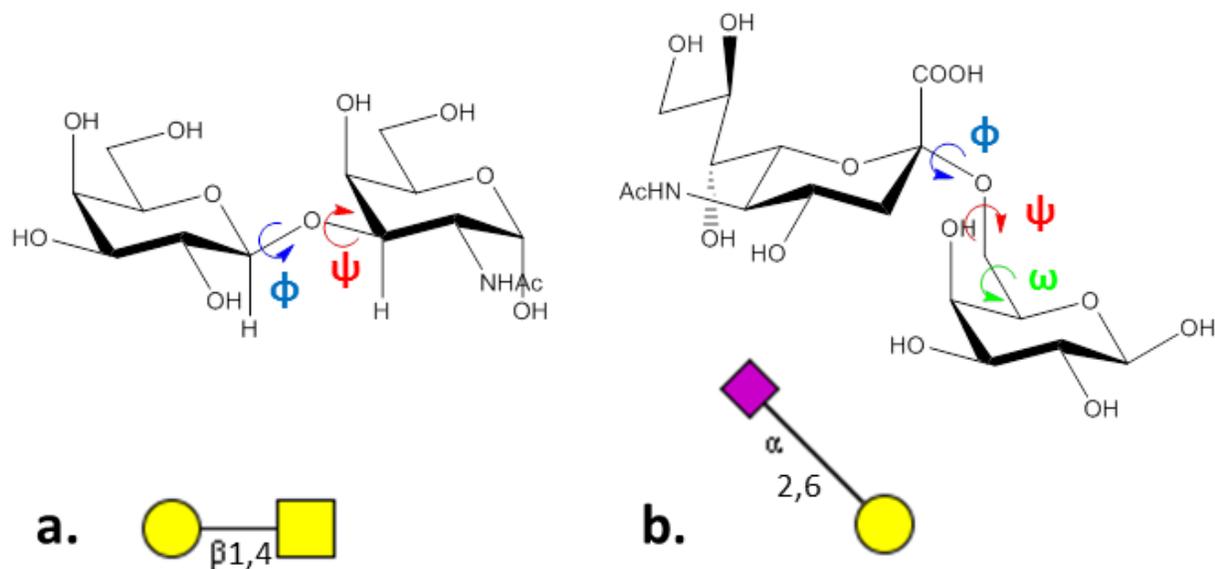
Understanding the information carrying potential of a glycan structure begins with deconstructing the polymers into monosaccharide units. These polymers contain monosaccharides composed of carbon, oxygen and hydrogen atoms which are found as heterocycles. In order to define the monosaccharide type the chiral orientations of carbons are represented in the saccharide nomenclature as *dextro-* or *laevo- rotatry (D/L)* at the stereocenter furthest from the carbonyl,  $\alpha$  or  $\beta$  at the anomeric carbon, and using a trivial name to represent the relative orientations of substituents on the remaining carbon positions[7]. Further modifications to the monosaccharide units are often found which include including functional groups like carboxylate, amine, and sulfate. Examples of the monosaccharides most relevant to this work are shown in their pyranose forms (**Table 1.1**). Additional information on monosaccharide chemical structure can be found in the many other resources such as the Essentials of Glycobiology[3,7].

**Table 1.1** Examples of monosaccharide residues which form six-membered rings. Shown from top to bottom are the chemical structure, CFG cartoon representation, chemical name, and abbreviated chemical name.

 <p style="text-align: center;">   <math>\alpha</math>-D-galactose  <math>\alpha</math>-D-Gal         </p>	 <p style="text-align: center;">             N-acetyl-<math>\alpha</math>-D-galactosamine  <math>\alpha</math>-D-GalNAc         </p>	 <p style="text-align: center;">   <math>\alpha</math>-D-galacturonic acid  <math>\alpha</math>-D-GalA         </p>
 <p style="text-align: center;">   <math>\alpha</math>-D-glucose  <math>\alpha</math>-D-Glc         </p>	 <p style="text-align: center;">             N-acetyl-<math>\alpha</math>-D-glucosamine  <math>\alpha</math>-D-GlcNAc         </p>	 <p style="text-align: center;">   <math>\alpha</math>-D-glucuronic acid  <math>\alpha</math>-D-GlcA         </p>
 <p style="text-align: center;">   <math>\alpha</math>-D-mannose  <math>\alpha</math>-D-Man         </p>	 <p style="text-align: center;"> <math>\alpha</math>-L-idose  <math>\alpha</math>-L-Ido         </p>	 <p style="text-align: center;">   <math>\alpha</math>-L-iduronic acid  <math>\alpha</math>-L-IdoA         </p>
 <p style="text-align: center;">   <math>\alpha</math>-N-Acetylneuraminic acid         </p>	 <p style="text-align: center;">   <math>\alpha</math>-L-fucose         </p>	

$\alpha$ -Neu5Ac	$\alpha$ -L-Fuc	
------------------	-----------------	--

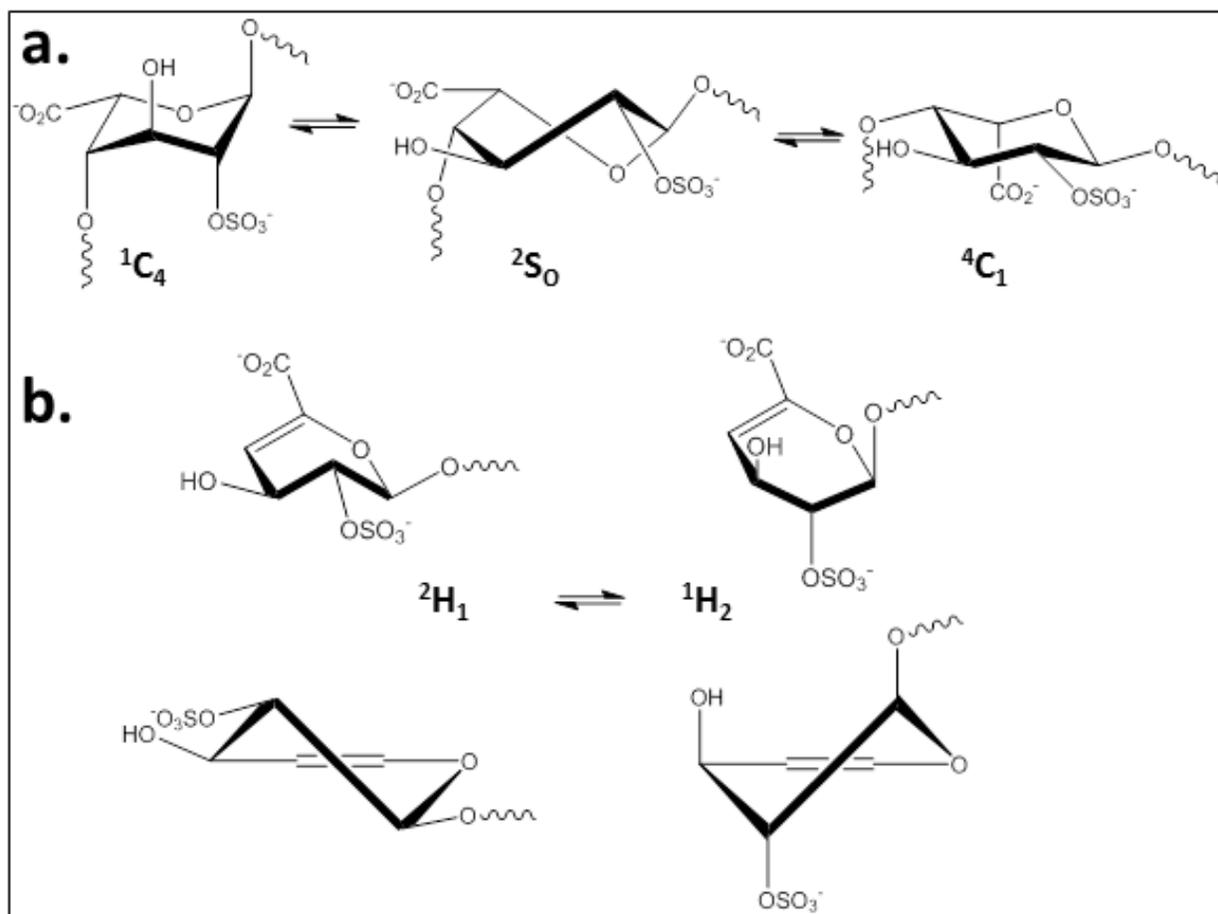
Carbohydrate polymers are formed by a covalent bond between monosaccharides units, called the glycosidic linkage, which forms between the carbon anomeric center of one sugar (non-reducing side) and an exocyclic oxygen on the other sugar (reducing side) as is shown in **Figure 1.1**. The linkage forms at least two flexible torsions called  $\varphi$  and  $\psi$  starting from the non-reducing terminus. The  $\varphi$  angle is formed on the non-reducing side of the linkage by the bond between anomeric carbon and the exo-anomeric reducing side oxygen ( $O_{\text{exo-anomeric}}$ ). The  $\varphi$  angle is adjacent and formed between  $O_{\text{exo-anomeric}}$  and the reducing side carbon atom which may be either endo- (**Figure 1.1a**) or exo-cyclic (**Figure 1.1b**) to the reducing terminal monosaccharide. When this carbon is exo-cyclic, further flexible torsions are formed which are referred to as  $\omega_x$  angle(s) where x is the number of the largest carbon atom in the rotatable torsion bond. The inherent flexibility of these torsions means that multiple stable conformations may be observed for a given glycosidic linkage. As such, the glycan three-dimensional (3D) structure should also be described by the population of these conformation(s).



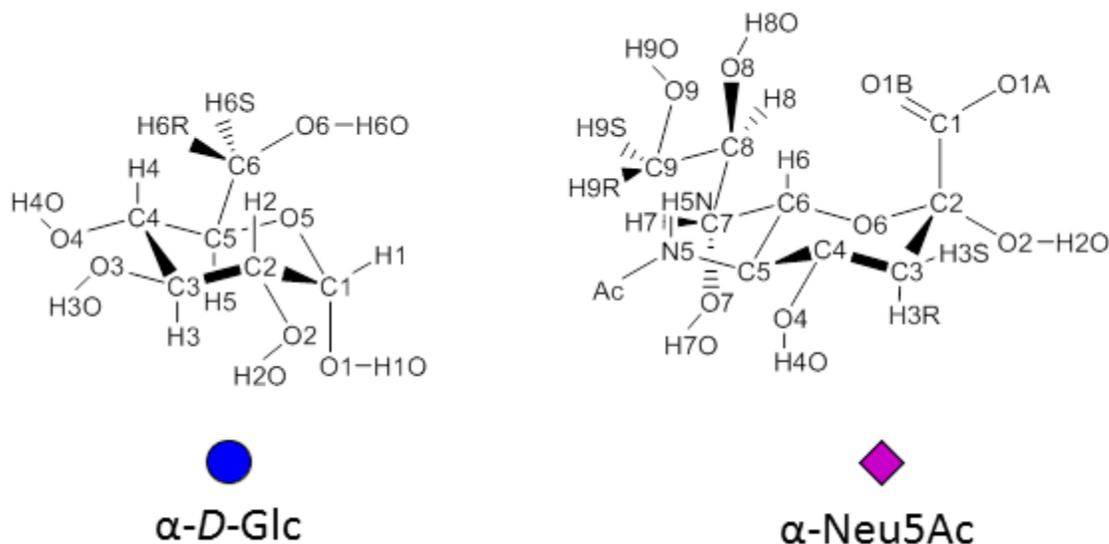
**Figure 1.1** Representations of the glycosidic linkage torsion angles  $\phi$  (blue),  $\psi$  (red), and  $\omega$  (green) for Gal $\beta$ -1,4-GalNAc $\alpha$  (a) and Neu5Ac $\alpha$ 2,6-Gal $\beta$  (b). Also shown are the cartoon representations using the extended CFG notation[8]. The non-reducing terminus of each sugar is shown on the left while the reducing terminus is shown on the right.

Most glycan pyranoside ring structures exist as a single conformation, usually in a chair-like state, however a few, such as those based on idose (idose, iduronic acid, etc.), are capable of existing as a mixture of more than one. Typically, *D*-pyranosides sugars exist as the  ${}^4C_1$  chair conformation (**Figure 1.2a**) in which carbon  $C_4$  (see **Figure 1.3** for the numbering schemes) is above the plane of atoms  $C_2$ ,  $C_3$ ,  $C_5$ , and  $O_5$  while carbon  $C_1$  is below this plane when viewed from the standard orientation such that the ring oxygen is oriented back and to the right while the anomeric carbon is the furthest atom on the right. In *L*-pyranosides this trend is reversed making the  ${}^1C_4$  chair the more stable ring shape. *L*-Idoses, particularly iduronic acid (IdoA) in glycosaminoglycans (GAGs), often favor the  ${}^1C_4$  ring state but also sample the  ${}^2S_0$ , skew-boat, and  ${}^4C_1$  states. The occupancy of each state varies considerably based on the adjacent sugars, linkages, and chemical modifications to IdoA and neighboring residues[9]. As will be shown

later in Chapter 3, the shape of the ring and sulfation pattern can alter the 3D shape of the polymer by introducing a bend in the glycan polymer (**Figure 3.4**).



**Figure 1.2** The most commonly observed ring conformations in iduronate (**a**) and  $\Delta$ 4,5-unsaturated uronate (**b**). The sampled shapes and populations depends on the glycosidic linkage, ring substituents (e.g. sulfate), and neighboring residues.

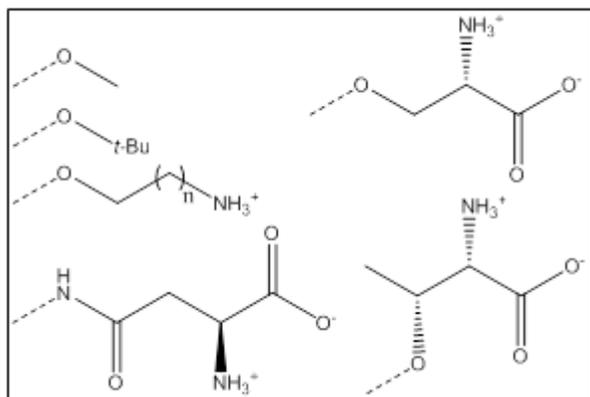


**Figure 1.3** Numbering and naming schemes for sugars. Shown are the typical hexopyranose atom name designations using glucose as a representative structure (left) and the 9-carbon N-acetylneuraminic acid (right,  $\alpha$ -Neu5Ac).

Another type of six-membered ring shape which is particularly relevant to GAGs occurs in non-reducing terminal  $\Delta$ 4,5-unsaturated uronates ( $\Delta$ 4,5-uu). These  $\Delta$ 4,5-uu sugars are formed as glycosidic cleavage products which retain the exocyclic substituent orientations of their saturated parent sugar[10]. The similarity to both the parent sugar and the catalytic complex in enzymes makes these well suited as transition state mimics in drug development[11] and substrates for glycan synthesis[12]. Unlike the saturated parent, which typically does not change ring states (with a few exceptions such as idoses),  $\Delta$ 4,5-uu rings frequently exchange states between two half-chair ( ${}^1H_2$  and  ${}^2H_1$ ) structures (**Figure 1.2b**).

The last component to the overall glycan structure is the reducing terminal substituent, or aglycone (**Figure 1.4** and **Figure 2.10**), which is typically a protein or lipid. The presence, or absence, of glycans on these biomolecules is often essential to complex biological processes used

to control cellular localization, protein degradation or biosynthetic pathways[13]. The type of aglycone and the local environment it associates with can impact glycan bioactivity by altering the surface area exposure or glycan rotamer lifetimes relative to that of a glycan free in water[14]. Additionally, an aglycone presents the saccharide with a different local chemical environment from bulk solvent (e.g. water) through variations in the electrostatic potential and interactive response. Thus the aglycone structure, while not always directly involved in glycan biorecognition, is integral in determining the 3D presentation and biological activity of the aggregate.



**Figure 1.4** Several aglycones attached to the reducing terminal sugar. The first (top left corner) is the typical *O*-methyl linker used in most molecular dynamics simulations of sugars. Just underneath is an *O*-*tert*-butyl linkage also used in simulations. The remaining are typical linkers used in glycan arrays. Also, the amino acid aglycones are found as attachments for glycans in proteins and peptides. Some lipid aglycones are shown in **Figure 2.10**.

## 1.2.2 Glycan Primary Structure

The primary sequence of glycans is formed by fitting together the monosaccharide components with the glycosidic linkage information to form linear or branched polymer sequences. Unlike peptides and proteins, glycans are capable of having non-linear sequences

making the theoretical sequence diversity much more substantial. If a library of theoretical glycans (theoretical glycome) were formed using only the eight  $\alpha/\beta$ -D-hexopyranoses then the potential number of disaccharides is 1024 while the number of possible di-peptide sequences formed with the twenty typical amino acids is 400. If this glycome is expanded to include trisaccharides then there are nearly 115,000 possible glycans where only 8000 tri-peptide combinations exist. Despite the vast theoretical glycome size, biological systems are only able to produce sequences for which the enzymatic machinery exists, thus, known mammalian glycomes are limited to around 7000 glycan sequences[15].

### **1.2.3 Glycan Secondary & Tertiary Structure**

Glycan secondary structure is defined herein as the shape formed between two saccharides (formed by the ring shapes and glycosidic linkage). The tertiary structure is the combination of all secondary structure properties which form the entire glycan polymer shape. In protein structure, the secondary structure is formed by small polypeptide fragments forming ordered structures while the tertiary structure describes how the secondary structures fit together as a single peptide chain. Compared to peptide linkages, glycosidic linkages are far more dynamic with fewer stable intra-molecular hydrogen bonds to maintain rigidity. This makes defining the tertiary structure more difficult in glycans as they frequently transition between one or more glycosidic linkage rotamers and sometimes between different ring states. Instead, the glycosidic rotamers, ring shape(s) and lifetimes of those two properties (i.e. secondary structure) are combined for each residue pairing to describe the overall shape of a glycan (i.e. tertiary structure). Additionally, the dynamics of glycan structures makes a simple description of glycan

shape difficult as the ‘average’ glycan structure may not look like any shape ever sampled by the glycan.

In spite of the dynamics of glycan structure, it is possible to identify the states that form the ensemble of 3D-structures for a glycan using experimental methods, such as nuclear magnetic resonance (NMR), or computational methods which employ multi-state sampling, such as molecular dynamics (MD) or Monte Carlo simulations. Experimental methods which rely on ordered structures like x-ray and neutron diffraction techniques are employed on free glycans which are amenable to crystallization, such as cellulose or sucrose; however, the majority of unligated glycans are experimentally characterized using solution-phase data from NMR[16]. NMR structure data is collected as an average of the states sampled over the experimental time. Glycans that sample more than one conformation on the NMR timescale require the use of 3D models along with fitting functions in order to decompose the data. Some of the NMR data collected or analyzed in this work includes three-bond  $J$ -couplings which describe the through-bond coupling of magnetic nuclei and correlates to torsion angles (described in section 3.3.5); nuclear Overhauser effects (NOEs) which describe through-space proton-proton transfer of spin polarization and correlates with inter-nuclear distances (described in section 3.3.5); and saturation transfer difference (STD) spectroscopy data which is used to measure the change in peak intensity caused by molecular magnetization transfer and correlates with distances between saturated and measured nuclei (described in section 4.5.3).

## 1.3 Computational Methods for Predicting Glycan Structure and Properties

### 1.3.1 Molecular Dynamics

Classical mechanics simulations begin with defining the molecular structure using spherical representations of atoms (atomistic), atoms combined with aliphatic hydrogens (united atom), or multiple atoms (coarse grained). Most atomistic MD simulations define the potential energy of system,  $V_{MM}$ , by splitting it into bonded and non-bonded components (**Equation 1.1**).

**Equation 1.1** Most frequently employed form of a Class I molecular mechanics force field used in molecular dynamics simulations.

$$V_{MM} = \sum_{\text{Bonds } i,j} K_{i,j}^{Bond} (r - r_{eq})^2 + \sum_{\text{Angles } i,j,k} K_{i,j,k}^{Angle} (\theta - \theta_{eq})^2 \\ + \sum_{\text{Torsions } i,j,k,l} \sum_n \frac{V_n}{2} [1 + \cos(n\varphi - \gamma_n)] + \sum_{\substack{\text{Non-Bonded} \\ i < j}} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

The bonded components include the bond length, angle bend and torsion angle terms while the non-bonded components include Lennard-Jones and Coulombic potentials. Harmonic bond potentials are used to describes the change in energy of the current bond length ( $r$ ) for atoms  $i$  and  $j$  from the equilibrium length ( $r_{eq}$ ) using a spring constant ( $K_{i,j}^{Bond}$ ). Similarly, the angle potential describes the change in energy of the bond-angle ( $\theta$ ) from the equilibrium angle( $\theta_{eq}$ ) using a spring constant ( $K_{i,j,k}^{Angle}$ ). The last bonded component describes a torsion term for non-classical (quantum) corrections to the torsion angles ( $\varphi$ ) which are not directly included by non-bonded terms or indirectly by assignment of the bonded terms. This cosine function is usually fit to minimize the difference between the quantum mechanical torsion profile

and the classical potential in small molecules using coefficients  $V_1, V_2, V_3 \dots V_n$ . Phase correction terms ( $\gamma_n$ ) may also be included but are not necessary.

The non-bonded description of van der Waals potential comes from the Lennard-Jones equation in which the interaction depends on the pairwise interaction distance ( $R_{ij}$ ) for an attractive coefficient ( $B_{ij}$ ) and a repulsive coefficient ( $A_{ij}$ ). The Coulombic potential describes the interaction energy between point charges  $q_i$  and  $q_j$  using a dielectric constant ( $\epsilon$ ).

In order to calculate the force necessary for a molecular dynamics simulation, the potential energy ( $V_{MM}$ , **Equation 1.1**) and the atomic position ( $x$ ) are integrated using Newton's 2<sup>nd</sup> law (**Equation 1.2**).

**Equation 1.2** Newton's second law used to determining force ( $F$ ).

$$F = ma = -\frac{dV_{MM}}{dx_i}$$

The change in atomic position is determined using the Verlet algorithm (**Equation 1.3**) which accounts for the change in position ( $x$ ) from time  $t$  to  $(t+\Delta t)$  and the previous position at time  $(t-\Delta t)$ .

**Equation 1.3** Basic form of the Verlet algorithm used to compute the change in position of a particle ( $x$ ) as a function of time ( $t$ ).

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \Delta t^2 a(t)$$

Velocities are not directly solved in the Verlet equation but can be determined using either the velocity Verlet or leap-frog methods. The AMBER simulation package uses the leap-

frog method which calculates the average velocity between times  $(t-\frac{1}{2}\Delta t)$  and  $(t+\frac{1}{2}\Delta t)$ . Initial velocities are defined by a Maxwell-Boltzmann distribution at the starting simulation temperature  $T$ .

The MD simulations in this work fall into the category of the canonical (NVT) or isothermal-isobaric (NPT) in which the number of particles (N) and temperature (T) are kept constant. The NVT ensemble also maintains a constant volume (V) and was employed in the MD simulations of DMPC lipid bilayers (Chapter 2). The NPT ensemble maintains constant pressure (P) and was used in the MD simulations of GAGs (Chapter 3). The weak-coupling thermostat[17] was used to control the temperature. This method rescales velocities to maintain a temperature bath using a correction coupled to a time constant. In NPT ensembles, pressure is similarly controlled using a pressure coupling constant to a target pressure, typically 1 atm[17]. The choice of which simulation model used depends on the desired results. In order to obtain experimentally-consistent results, an NVT simulation often requires pre-equilibration with a short NPT simulation to obtain the optimal volume for a given temperature and pressure. In this instance, NVT ensembles were used on lipid bilayers to avoid using a surface tension restraint to maintain the correct lipid packing density as the bilayer packing will be maintained by the constant volume. The NPT ensembles were preferred for simulations of glycans as this allows the volume dimension to vary, avoiding the need to pre-equilibrate the simulation box dimensions that would need to be done in an NVT simulation.

### 1.3.2 GLYCAM Carbohydrate Force Field

Carbohydrate force fields for MD simulations first appear in the early 1980s as simulation software and computational resources became available[18]. The first release of the GLYCAM force field was in 1995 (GLYCAM93) which produced a carbohydrate force field consistent with the existing AMBER protein force field[19]. This early release separated the torsion potentials for the anomeric carbon based the orientation (axial or equatorial) of the exocyclic anomeric oxygen. Additionally, the endocyclic oxygen was treated as a unique atom type. In this early form, partial charges were derived from the electrostatic potential (ESP) to fit all atoms while all 1,4-non-bonded interactions were scaled down to be consistent with AMBER force fields. The next release of GLYCAM in 2001 (GLYCAM 2000) focused on optimizing the orientation preferences for the exocyclic free  $\omega$ -rotamers, eliminating 1,4-scaling factors [20], and using a restrained electrostatic potential (RESP) method [21]. Subsequent modifications to GLYCAM2000 included development of a new charge model incorporating of ensemble averaged charges[22] and the inclusion of lone-pairs using a TIP5P-consistent design which is also found the subsequent version of GLYCAM[23].

The next major GLYCAM release came in 2008 (GLYCAM06) when the force field was completely rebuilt[24]. In GLYCAM06, the anomeric carbon was now a single atom type instead of two and a carbohydrate-specific  $sp^3$  carbon was introduced to distinguish GLYCAM parameters from those of other AMBER force fields. An important change was made to the treatment of partial charges as aliphatic hydrogens were now assigned a zero charge during the RESP fitting and in the final charge set. This version of GLYCAM also incorporated the development of acidic sugars instead of adopting directly from existing AMBER parameters.

The research presented in Chapters 2 and 3 describes the modifications to the GLYCAM06 carbohydrate force field. The first expansion to GLYCAM06 included new parameters to describe lipids and was the first lipid force field released in AMBER[25]. This expansion established a set of parameters which were able to reproduce a wide array of lipids with different head-groups, varying degrees of tail-group unsaturation, and cholesterol-like molecules. The parameters have the capability to model a diverse set of lipid structures so long as ensemble charge models are generated for these lipids. These parameters were seamlessly integrated within the carbohydrate parameters in GLYCAM06 making them ideal for glycolipids (glycans conjugated to lipids)[14] as well as non-lipid aglycones derived from similar fragments, such as the protonated alkyl-amine chains employed as spacers in glycan arrays[26]. The parameter-fitting methods established for unsaturated lipid tail-groups also guided the framework and initial parameter development for  $\Delta$ 4,5-unsaturated uronate parameters (Chapter 3). These force field modifications were validated using simulations of the DMPC bilayer in which bilayer bulk properties, such as thickness and head-group density, were compared to experiment. At the time it was not feasible to simulate more bilayers due to computational limitations (slow simulation speeds).

The GLYCAM06 lipids work represents a multi-author manuscript in which I am primary author, and the research was overseen by R. Woods. My contributions to the research were to the development of the force field (including charges) with training and some development provided by A. Yongye. While my work included simulating the DMPC bilayer, most post-simulation analysis and visualization of the bilayer was performed by M. DeMarco with contributions from this author in establishing protocols for analyzing the bilayer density from AMBER output.

### 1.3.3 GLYCAM Prediction of GAG Structure

GAGs are carbohydrates composed of repeating disaccharide units of a hexosamine, and another hexose that is often an anionic saccharide. These glycans have proven challenging to carbohydrate force field development efforts due to the complexity in modeling charges, ring puckering and obtaining adequate simulation sampling[27,28]. In this work, two major advancements are made along this front: 1) refining and expanding upon acidic sugar parameters (including charges), 2) and development of a transferable sulfate charge model.

The first major expansion for this parameter set targets the ring conformations of iduronates[29] which are dynamic with ring conformational changes occurring on timescales of hundreds of nanoseconds[18]. This was done by distinguishing the endocyclic oxygen atom type from the exocyclic anomeric oxygen which allows the endocyclic ring torsion rotation ( $X-O_{\text{endo}}-C_{\text{endo}}-X$ ) to decouple from the exocyclic ring torsion rotation describing the  $\phi$  angle ( $X-C_{\text{endo}}-O_{\text{exo}}-X$ ). The dynamics of  $\Delta$ 4,5-unsaturated uronates ( $\Delta$ 4,5-uu) ring conformations were also parameterized[29] as these acidic sugars have been shown to have dynamic ring shapes which change on timescales easily accessible to molecular dynamics simulations. To date, no other molecular mechanics force field exists which accurately models the populations of these states in molecular dynamics simulations[30,31].

The other major change to GLYCAM06 is the creation of a transferable charge model for N- and O- sulfate based on ensemble average charges derived from sulfated glycans. These parameters significantly reduce model building time and allow sulfated sugars to be built using the GLYCAM web-interface or scripted through its web service[32].

The combined GLYCAM06 carbohydrate/lipid parameter set, including the GLYCAM11 update for IdoA[28], were employed in simulations of five GAG disaccharides containing  $\Delta$ 4,5-uu, and two GAG tetrasaccharides (GlcA-GlcNR-IdoA2S-GlcNR-O-Me, molecule **6** R=Acetyl and **7** R=SO<sub>3</sub><sup>-</sup>) containing one IdoA each[29]. All seven GAG molecules were experimentally characterized using NMR  $J_{\text{HH}}^3$ -couplings and proton-proton NOE's and compared to properties from simulation data.

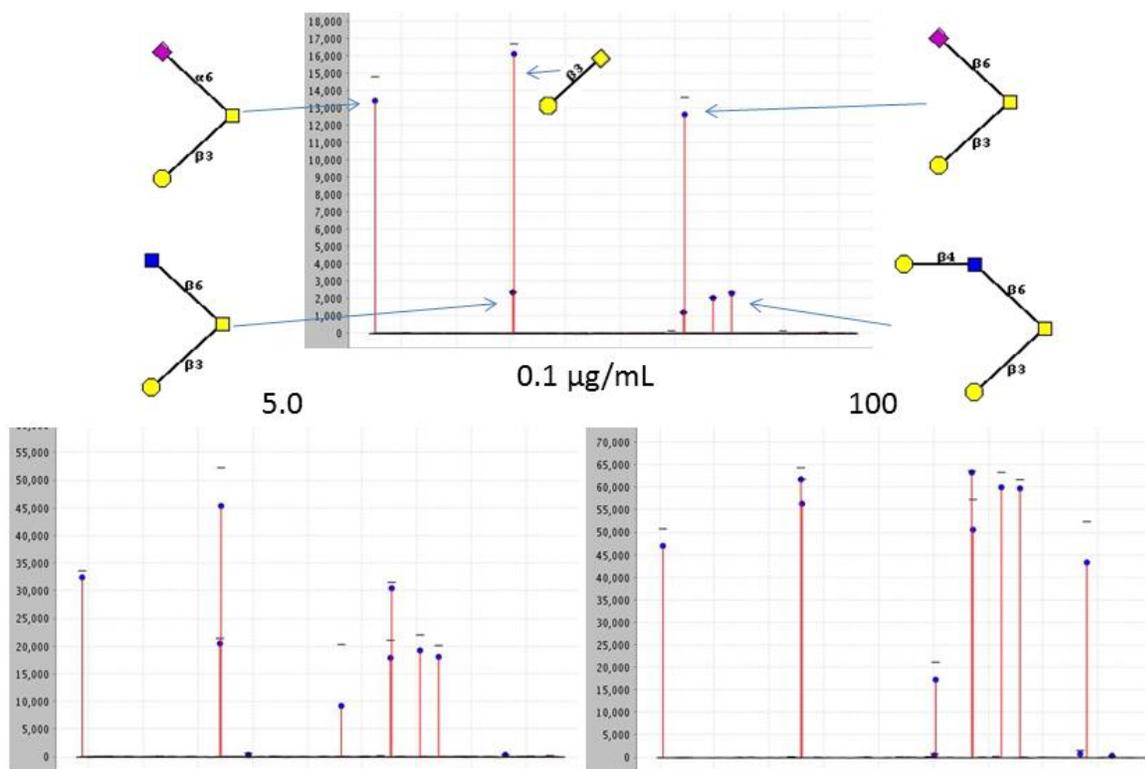
The GLYCAM06 modification and GAG simulation work also represents a multi-author manuscript in which I am primary author and the research was overseen by R. Woods. My contributions to the research were to the parameter development, validation and fitting to NMR observables with some parameter development by X. Wang. I performed all simulation work and analysis except in developing initial test simulations for  $\Delta$ 4,5-uu prior to fitting to NMR observables which was done by A.Singh. Samples of the GAG tetrasaccharides were provided by A. Venot under the supervision of G.-J. Boons. K. Pederson collected and provided the NMR observables under the supervision of J.H. Prestegard. Lastly, I performed decomposition of the NMR data to determine ring conformation populations and compared these to theoretical models.

## **1.4 Defining Protein-Carbohydrate Complexes**

### **1.4.1 Computational Prediction of Specificity with Computational Carbohydrate Grafting**

The specificity of a protein defines the its selectiveness for binding with other molecules. A highly selective protein will recognize very few non-target molecules while a promiscuous protein will tolerate a variety of non-target molecules. Experimentally, specificity

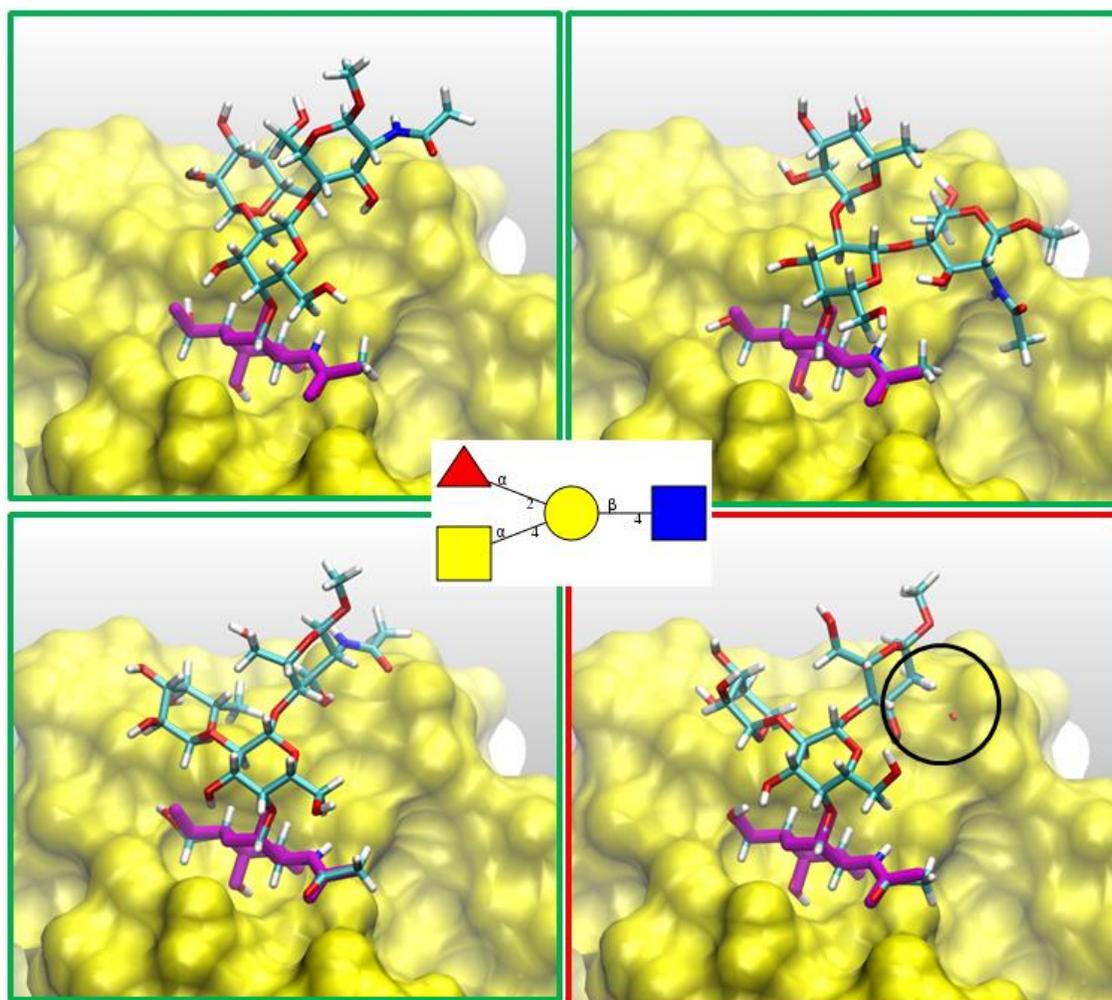
can be determined more cost effectively than affinity, relying on bulk screening methods like glycan array microarrays to test the specificity of a single protein for a variety of potential binding partners. Microarrays are surfaces on which potential ligands are covalently fixed. A protein of interest is then washed over the surface and rinsed leaving only the protein which is bound to favorable ligands. Next a fluorescent tag molecule which binds to the protein is washed over the surface, attaching to any ligated protein, before being rinsed. The binding of the protein to a particular molecule is then measured by the relative fluorescence intensity of the attached tag molecule. The process is then repeated at various protein concentrations to qualitatively determine the specificity of the protein[5]. Shown in **Figure 1.5** is the glycan microarray screening of the anti-tumor antibody JAA-F11 which is discussed in Chapter 4. Analysis of the glycan arrays provides a view of the glycan primary structure components that make up the specificity of a protein; the TF-antigen in the case of JAA-F11. This type of analysis does not explain the structural rationale for selecting these specific binding partners while excluding others containing similar ligand motifs.



**Figure 1.5** Glycan array screening data for the anti-tumor antibody JAA-F11 run at protein concentrations of 0.1, 5.0 and 100  $\mu\text{g}/\text{mL}$ . The five glycan sequences with the most specificity are shown with the TF-glycan shown inset in the middle of the 0.1  $\mu\text{g}/\text{mL}$  bar graph. Y-axes are in relative fluorescent units (RFUs).

Traditionally, crystallographic and NMR techniques have been employed to define the 3D structure of protein-ligand binding interactions. Often, when co-crystals are unavailable or NMR data is sparse, computational docking methods are employed to model the 3D complex using affinity calculations. In this work the concept of modeling this 3D complex was expanded in a novel method developed to determine specificity. This method, called computational carbohydrate grafting (CCG)[26], uses a glycan fragment in the binding site of a target protein as a scaffold for grafting on longer glycan polymers (**Figure 4.1**). Polymers are grafted onto the scaffold using each of their possible 3D conformations; then these complexes are checked for

overlaps with the protein surface. CCG uses both the putative binding partners, which do not overlap, and non-binders, which do overlap, to define specificity. This provides a rationale for observed selectivity based on secondary and tertiary glycan structure, allowing better design of glycomimetics which also fit the binding pocket. The use of protein overlap to characterize the specificity was demonstrated using the CpGH89 enzyme binding module (CBM32-5) which was co-crystallized with a monosaccharide fragment of the binding polymer. CpGH89 is an enzyme complex produced by the bacterium *Clostridium perfringens* that recognizes a very specific glycan sequence found in the gastric mucosa. Besides the catalytic domain, several carbohydrate binding sub-units, or carbohydrate binding modules (CBMs), are present in CpGH89. One of the sub-units, CBM32-5, is unique within the complex in that it recognizes a fucosylated glycan sequence while the CBMs can only recognize the non-fucosylated sequence. Grafting the 3D ensemble of structures from the fucosylated polymer onto the crystal monosaccharide revealed that a single amino acid mutation in the CBM32-5 binding module allows it to accommodate a fucose for many poses identified in an MD simulation (**Figure 1.6**)[33].



**Figure 1.6** The fucosylated tetrasaccharide ligand (middle inset) for CBM32-5 (part of the CpGH89 complex). The crystallographic protein (yellow surface) and co-crystallized GlcNAc (purple) are shown. Models from the MD simulation of the tetrasaccharide which do (red box, bottom right) and do not clash (green boxed frames, top and bottom left) are shown to emphasize how steric overlap analysis works. The black circle indicates the steric overlap.

The CCG method can also be used to characterize the general specificity of a binding protein for an entire virtual library of glycans. In this way, the computed specificity can be compared with experimental specificity-screening techniques like glycan microarray analysis. An added benefit is that this virtual library can contain more sequences than found on the experimental arrays, allowing it to identify novel glycan binding partners. The capabilities of the

CCG method were demonstrated on the JAA-F11 anti-cancer antibody with its binding antigen, the TF-antigen (Chapter 4)[26]. This resulted in CCG predicting several novel putative binding sequences which potentially lead to antibody cross-reactivity in humans. However these putative binding partners are not accessible to antibodies, thus, antigenic activity was predicted to be limited to tumor cells which is consistent with prior experimental work[34].

The computational specificity screening work also represents a multi-author manuscript in which I am primary author and the research was overseen by R. Woods. My contributions to the research were to the generation of the docked TF-antigen to JAA-F11, development of the virtual glycan array modeled after the CFG array, analysis of the CFG array specificity results, development of the CCG methodology, and analysis of CCG results including STD-NMR calculations. The creation of GLibrary3D as well as coding of the automated CCG method was completed by O. Grant. J. Heimburg-Molinaro collected the glycan array data under the supervision of D. Smith. S. Jadey isolated the JAA-F11 antibody under the supervision of K. Rittenhouse-Olson. X-ray crystallographic data was collected by A. Gulick with antibody sequencing data provided by S. Deutscher. STD-NMR data was collected by J. Glushka.

## **2. EXTENSION OF THE GLYCAM06 BIOMOLECULAR FORCE FIELD TO LIPIDS, LIPID BILAYERS AND GLYCOLIPIDS<sup>1</sup>**

---

<sup>1</sup> M.B. Tessier, M.L. DeMarco, A.B. Yongye and R.J. Woods. 2008. *Mol. Sim.* 34(3):349-364. Reprinted here with permission of the publisher.

## 2.1 Abstract

GLYCAM06 is a generalisable biomolecular force field that is extendible to diverse molecular classes in the spirit of a small-molecule force field. Here we report parameters for lipids, lipid bilayers and glycolipids for use with GLYCAM06. Only three lipid-specific atom types have been introduced, in keeping with the general philosophy of transferable parameter development. Bond stretching, angle bending, and torsional force constants were derived by fitting to quantum mechanical data for a collection of minimal molecular fragments and related small molecules. Partial atomic charges were computed by fitting to ensemble-averaged quantum-computed molecular electrostatic potentials.

In addition to reproducing quantum mechanical internal rotational energies and experimental valence geometries for an array of small molecules, condensed-phase simulations employing the new parameters are shown to reproduce the bulk physical properties of a DMPC lipid bilayer. The new parameters allow for molecular dynamics simulations of complex systems containing lipids, lipid bilayers, glycolipids, and carbohydrates, using an internally consistent force field. By combining the AMBER parameters for proteins with the GLYCAM06 parameters, it is also possible to simulate protein-lipid complexes and proteins in biologically relevant membrane-like environments.

## 2.2 Introduction

Proteins and carbohydrates often exist as components of membranes[35,36] and their functions depend in part on their orientation (presentation) relative to the membrane surface[37]. In the case of glycolipids, the lipid tail embeds in the membrane bilayer exposing the

carbohydrate head group to the surrounding aqueous environment. Although the physical and bulk structural properties of lipid bilayers, such as the compressibility modulus, bilayer thickness, or lipid head group density, are available for a wide variety of lipids[38], detailed 3D structural characterization of embedded molecules remains challenging[37].

In the case of glycolipids, NMR spectroscopy has been used to examine the presentation of the carbohydrate head group[39], and data suggest that the glycan adopts distinct conformations. However, experimental data may be insufficient to uniquely characterize the conformations and orientations populated by the carbohydrate head group relative to the membrane surface[40]. Therefore, there is a potential for computational methods to augment and complement the sparse experimental data for membrane-associated biomolecules. This potential has been demonstrated in the case of lipid bilayers (reviewed in [41]) and protein/lipid bilayer systems[42,43,44,45,46], wherein MD simulations have enhanced our understanding of the structural, functional and dynamic properties of these systems.

Here we extend the GLYCAM biomolecular force field for carbohydrates to lipids, phospholipids and glycolipids. Phospholipids are the primary component of cell membranes and their hydrophobic tails are composed of aliphatic carbon chains that may be saturated, monounsaturated, or polyunsaturated. The hydrophilic head groups are often charged or zwitterionic, containing phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, or may be linked to other types of biomolecules. Lipid classes, such as sphingolipids, phospholipids and glycolipids, combine to yield the heterogeneous and multifunctional eukaryotic plasma membrane[35]. The lipid composition of a biological membrane is non-uniform and creates functionally specific regions of the membrane, such as lipid rafts[47], that can impact the activity

of associated proteins[35]. Lipid rafts are often associated with the presence of cholesterol in the membrane[47].

The similarities in atomic composition and connectivities between carbohydrates and lipids facilitated the extension of the recently reported GLYCAM06 (glycans and glycoconjugates in AMBER) biomolecular force field[48] to lipids and glycolipids. To remain consistent with the GLYCAM06 formalism, the following criteria were used to guide its extension: 1) the new parameters should be transferable to the most common and biologically relevant lipids, lipid bilayers, and glycolipids, 2) they should be self-contained and therefore readily transferable to many quadratic force fields, 3) as few new atom types as possible should be introduced, 4) the new parameters should be compatible with all existing parameters and molecular classes in GLYCAM06, 5) the accuracy of the parameters should be rigorously assessed by application to developmental and test molecules through comparison with theoretical and experimental data, and 6) the use of 1–4 non-bonded electrostatic or van der Waals (vdW) scale factors should be avoided[49].

To illustrate the performance of the resultant parameters in a bilayer simulation, a 15 ns MD simulation of a 1,2-dimyristoyl-sn-glycero-3-phosphatidylcholine (DMPC) bilayer in explicit TIP3P water was performed. The physical properties of DMPC are well established, thus facilitating critical comparison between the MD and experimental data.

## **2.3 Methods**

### **2.3.1 Parameter Development**

Quantum mechanical (QM) geometry optimizations were performed using the GAUSSIAN98 software package[50] at the HF/6-31G\* level of theory unless otherwise noted,

while single point energies were calculated at the B3LYP/6-31++G(2d,2p) level, in accordance with the parameterisation methods outlined in GLYCAM06[48].

The AMBER 8 software package[51] was employed for all molecular mechanical (MM) and MD calculations. As recommended for carbohydrates[20] 1-4 non-bonded and electrostatic scale factors were set to unity (SCEE =1 and SCNB = 1, respectively). Atomic vdW parameters were taken from the GLYCAM/AMBER parameter set[48], which originated from AMBER[52]. Consistent with GLYCAM06, torsion rotation terms and valence harmonic force constants were generated by fitting to QM data computed for representative molecules at the B3LYP/6-31++G(2d,2p)//HF/6-31G\* level, for rotamers sampled at 30 degree increments. Torsion coefficients ( $V_1$ ,  $V_2$ , and  $V_3$ ) for the classical bond rotation terms were fitted to the quantum data without using phase shifts, employing a multi-variable least squares algorithm. Torsion contributions were explicitly defined for all constituent linkages employing representative molecular fragments. Equilibrium bond lengths and angles were selected from the best available experimental data for each representative molecule, or from closely related molecules, with preference being given to gas-phase structural data; sources and structures are listed in **Table 2.1**. The Cambridge Structural Database[53] was accessed through the ConQuest software package[54] to identify pertinent experimental structures.

Partial charges for use in the gas-phase minimisations of the small molecule parameterisation sets were computed by fitting to the molecular electrostatic potential (ESP) computed on a CHELPG[55] grid of points at the B3LYP/cc-pVTZ level, with an ESP restraint (RESP) weight of 0.0005 [56]. Consistent with AMBER[57] and GLYCAM[19], partial charges for the intact biomolecules (lipids), to be employed in condensed-phase simulations, were subsequently derived by ESP-fitting at the HF/6-31G\* level. As earlier proposed for use with

GLYCAM[21] a RESP weight of 0.01 was employed for the biomolecular charge derivation. In order to address the issue of conformational charge dependence[22], conformationally-averaged RESP charges were computed for each lipid from sixty-three unique lipid conformations extracted from an existing equilibrated DMPC bilayer, freely available from the website of Professor D. P. Tieleman at <http://moose.bio.ucalgary.ca/>.

Two other MD-equilibrated homogeneous lipid membranes composed of 1-stearoyl-2-docosahexaenoyl-sn-glycerco-3-phosphocholine (SDPC)[58] and palmitoyl oleoyl phosphatidylethanolamine (POPE)[59] were used, along with the DMPC bilayer, to determine the average ensemble charge distribution for functional regions of the lipids. The lipids were divided into a head group region, which was composed of phosphatidycholine or ethanolamine groups and the glycerol sub-unit, and two hydrocarbon tail groups, *sn1* and *sn2*.

**Table 2.1 Atom types, bonds, angles and torsion parameters in GLYCAM06 (lipid extension)**

Atom Name	vdW radius (Å)	$\epsilon$ (kcal/mol)	Type
CK	1.908 (PARM 94)	0.086 (PARM 94)	sp <sup>2</sup> carbon for alkenes
CJ	1.908 (PARM 94)	0.086 (PARM 94)	sp <sup>2</sup> alkene carbon single-bonded to an sp <sup>2</sup> alkene carbon
CP	1.908 (PARM 94)	0.109 (PARM 94)	sp <sup>3</sup> aliphatic carbon adjacent to phosphate
HP	1.100 (PARM 94)	0.016 (PARM 94)	H bonded to C adjacent to a positively charged group
N3	1.824 (PARM 94)	0.170 (PARM 94)	Positively charged amino group nitrogen (Lys, phospholipids, etc.)
P	2.100 (PARM 94)	0.200 (PARM 94)	Phosphorous in phosphates

**Bond Terms**

Term	k	Equilibrium Value	Training Molecule or Data Source
N3-CG	355.0	1.490	Methyl ammonium
P -OS	230.0	1.610	PARM 94
P -O2	525.0	1.480	PARM 94
CP-OS	285.0	1.460	CG-OS (GLYCAM06)
CG-CP	310.0	1.520	CG-CG (GLYCAM06)
CK-CK, CJ-CJ	629.0	1.337	Ethene
OS-CK, OS-CJ	350.9	1.359	Methoxyethene
CG-CK, CG-CJ	324.0	1.514	Propene
CJ-CK	350.9	1.467	Butadiene
N3-H	434.0	1.010	PARM94
CG-HP	360.0	1.095	CG-HC (GLYCAM06)
CK-HA, CJ-HA	360.0	1.095	CG-HC (GLYCAM06)
CP-H1	340.0	1.095	CG-H1 (GLYCAM06)

**Angle Terms**

Term	k	Equilibrium Value	Training molecule or Data Source
N3-CG-CG, N3-CG-CP	67.0	111.6	Ethyl ammonium
CG-N3-CG	54.0	111.1	GLYCAM06
OS-P -OS	45.0	109.6	PARM94
P -OS-CG, P -OS-CP	50.0	119.0	Dimethyl phosphate
O2-P -OS	100.0	108.2	PARM94
O2-P -O2	140.0	119.9	PARM94
CG-CP-OS, CP-CG-OS	70.0	108.5	CG-CG-OS (GLYCAM06)
CG-OS-CJ, CG-OS-CK	38.0	107.0	Methoxyethene
OS-CJ-CJ, OS-CK-CK	59.5	119.0	Methoxyethene
P -OS-P	12.8	150.0	Dimethyl diphosphate
CJ-CG-CK, CK-CG-CK, CJ-CG-CJ	46.7	111.5	1,4-Pentadiene

CG-CJ-CG, CG-CK-CG	46.1	115.6	2-Methylpropene
CK-CJ-CJ, CK-CK-CJ	49.4	127.7	Butadiene
CG-CJ-CJ, CG-CK-CK	47.9	122.5	Propene
CG-CG-CJ, CG-CG-CK	43.0	112.0	1-Butene
OH-CG-CJ, OH-CG-CK	70.0	107.5	OH-CG-CG (GLYCAM06)
CP-CG-CG	45.0	113.5	CG-CG-CG (GLYCAM06)
N3-CG-HP	57.0	109.6	Methyl ammonium
HP-CG-HP	40.0	109.5	HC-CG-HC (GLYCAM06)
HP-CG-CG, HP-CG-CP	45.0	112.6	HC-CG-CG (GLYCAM06)
CK-CJ-HA, CJ-CK-HA	32.7	126.4	Butadiene
H -N3-CG	45.5	109.2	Methyl ammonium
HA-CJ-HA, HA-CK-HA	31.3	117.4	Ethylene
HA-CJ-CJ, HA-CK-CK	38.2	121.3	Ethylene
H1-CG-CJ, H1-CG-CK	41.4	114.0	Propene
HC-CG-CJ, HC-CG-CK	41.4	114.0	Propene
HA-CJ-CG, HA-CK-CG	50.0	114.0	Propene
H1-CP-CG, H1-CG-CP	45.0	111.0	H1-CG-CG (GLYCAM06)
H1-CP-OS	60.0	110.0	H1-CG-OS (GLYCAM06)
H -N3-H	35.0	109.5	PARM94
H1-CP-H1	45.0	109.5	H1-CG-H1 (GLYCAM06)
OS-CJ-HA, OS-CK-HA	65.1	107.5	Methoxyethene

#### Torsion Terms

Term	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	Training molecule	Average training error	Test molecule	Average test error
CG-N3-CG-CG, CG-N3-CG-CP	0.10	-0.20	0.10	Ethylmethylamine	0.363		
N3-CG-CG-CG	0.10	-0.10	0.30	N-Propylamine	0.210	Propyl Ammonium	0.344
N3-CG-CG-OS, N3-CG-CP-OS	-0.50	0.50	1.50	N-(2-Methoxyethyl)amine	0.437	2-Methoxyethyl Ammonium	0.789
N3-CG-CG-OH	-1.00	0.50	0.60	N-(2-Hydroxyethyl)amine	0.180		
N3-CG-CG-HC, N3-CG-CG-H1, N3-CG-CP-H1	0.0	0.0	0.10	Ethyl amine	0.164	Ethyl Ammonium	0.013
HP-CG-N3-CG	0.0	0.0	0.25	Dimethyl Ammonium	0.223		
H -N3-CG-HP	0.0	0.0	0.13	Methyl ammonium	0.013		
H -N3-CG-CG, H -N3-CG-CP	0.0	0.0	0.13	Ethyl ammonium	0.013		
OS-P -OS-CG, OS-P -OS-CP	0.0	0.70	0.50	Dimethyl phosphate	0.084		
O2-P -OS-CG, O2-P -OS-CP	0.10	-0.50	0.10	Dimethyl phosphate	0.140		
CG-CG-OS-P, CG-CP-OS-P	-1.20	0.05	0.10	Ethylmethylphosphate	0.348		
H1-CG-OS-P, H1-CP-OS-P			0.07	Dimethylphosphate	0.114		
CK-CG-CG-N, CJ-CG-CG-N	-0.22	0.03	0.18	N-but-3-en-1-yl acetamide	0.153		
C -CG-CG-CK, C -CG-CG-CJ	0.40	0.33	-0.42	Methyl pent-4-enoate	0.177		

OH-CG-CG-CK, OH-CG-CG-CJ	-1.06	-0.16	-0.11	3-Butenol	0.161	Pent-4-en-2-ol	0.512
CG-OS-CK-CK, CG-OS-CJ-CJ	-0.25	-3.99	-0.87	Methoxyethene	0.089	(E)-1-sec-butoxy-2-methyl but-1-ene	0.604
CG-CG-OS-CK, CG-CG-OS-CJ	0.22	-0.42	-0.01	Ethoxyethene	0.289	(E)-1-sec-butoxy-2-methyl but-1-ene	0.489
CK-CG-CK-CK, CJ-CG-CK- CK, CK-CG-CJ-CJ, CJ-CG-CJ-CJ	-0.21	-0.68	-0.04	1,4-Pentadiene	0.069	3-Methyl-1,4-heptadiene	0.145
CG-CG-CJ-CG, CG-CG-CK-CG	0.29	-0.34	-0.01	2-Methyl-1-butene	0.230	2,5-Dimethyl-2,4- heptadiene	0.640
OH-CG-CK-CK, OH-CG-CJ-CJ	-0.50	-1.00	-0.50	Prop-2-en-1-ol	0.250		
CJ-CJ-CK-CK	0.92	-1.75	0.69	Butadiene	0.132	2,5-Dimethyl-2,4- heptadiene	0.652
CG-CG-CG-CK, CG-CG-CG-CJ	0.0	0.0	0.0	1-Pentene	0.252		
CG-CG-CK-CK, CG-CG-CJ-CJ	0.0	-0.40	-0.50	1-Butene	0.093		
OS-C -CG-CG	0.0	0.0	0.20	Methyl propanoate	0.057	Methyl 2-methylpropanoate	0.140
H1-CG-CG-CK, H1-CG-CG-CJ, HC-CG-CG-CK, HC-CG-CG-CJ	0.0	0.0	0.10	1-Butene	0.036		
CG-OS-CJ-HA, CG-OS-CK-HA	0.40	0.22	0.08	Methoxyethene <sup>a</sup>	0.154		
H1-CG-OS-CJ, H1-CG-OS-CK	0.0	0.0	0.10	Methoxyethene	0.402		
CK-CG-CK-HA, CJ-CG-CK- HA,	0.05	0.03	0.31	1,4-Pentadiene <sup>b</sup>	0.500		
CK-CG-CJ-HA, CJ-CG-CJ-HA							
HC-CG-CJ-CG, HC-CG-CK-CG	0.0	0.0	0.10	2-Methyl-1-propene	0.302		
HA-CJ-CK-CK, HA-CK-CJ-CJ	0.69	0.40	0.17	Butadiene	0.183		
HA-CK-CJ-HA	0.92	-1.75	0.69	Butadiene <sup>c</sup>	0.127		
HO-OH-CG-CK, HO-OH-CG- CJ	0.38	0.58	0.20	2-Propenol	0.120		
H1-CG-CK-CK, H1-CG-CJ-CJ, HC-CG-CK-CK, HC-CG-CJ-CJ	0.10	-0.10	-0.34	Propene	0.032		
HC-CG-CK-HA, HC-CG-CJ- HA, H1-CG-CK-HA, H1-CG- CJ-HA	0.0	0.0	0.0	Propene	0.054		
OH-CG-CK-HA, OH-CG-CJ- HA	0.0	0.0	0.0	Prop-2-en-1-ol	0.198		
CG-CG-CK-HA, CG-CG-CJ-HA	0.0	0.0	0.0	1-Butene	0.200		
OS-CJ-CJ-CG, OS-CK-CK-CG	0.0	-13.00	0.0	1-Propenylmethylether			
CK-CJ-CJ-HA, CJ-CK-CK-HA	0.0	-15.00	0.0	Butadiene			
OS-CJ-CJ-HA, OS-CK-CK-HA	0.0	-20.00	0.0	Methoxyethene			
CJ-CK-CK-CG, CK-CJ-CJ-CG	0.0	-3.50	0.0	1,3-Pentadiene			
CJ-CK-CK-CJ, CK-CJ-CJ-CK	0.0	-2.00	0.0	Hexatriene			
CG-CK-CK-CG, CG-CJ-CJ-CG	0.50	-15.00	0.0	2-Butene			

HA-CK-CK-CG, HA-CJ-CJ-CG	0.0	-21.00	0.0	Propene	
HA-CK-CK-HA, HA-CJ-CJ-HA	0.0	-8.50	0.0	Propene	
HP-CG-CG-OS, HP-CG-CP-OS, H1-CG-CP-OS, H1-CP-CG-OS	0.0	0.0	0.05		HC-CG-CG-OS (GLYCAM06)
OS-CP-CG-OS	0.0	0.82	0.0		OS-CG-CG-OS (GLYCAM06)
CP-CG-OS-C	0.47	0.0	-0.04		CG-CG-OS-C (GLYCAM06)
OS-CP-CG-CG, OS-CG-CG-CP	-0.27	0.0	0.0		OS-CG-CG-CG (GLYCAM06)
CP-CG-OS-CG	0.0	0.0	0.16		CG-OS-CG-CG (GLYCAM06)
H1-CG-CP-H1	0.0	0.0	0.17		H1-CG-CG-H1 (GLYCAM06)
H1-CG-CG-CP, H1-CP-CG-CG	0.0	0.0	0.15		H1-CG-CG-CG (GLYCAM06)
HP-CG-CG-H1, HP-CG-CP-H1, HP-CG-CG-HC	0.0	0.0	0.13		HC-CG-CG-HC (GLYCAM06)
HP-CG-CG-CG	0.0	0.0	0.10		HC-CG-CG-CG (GLYCAM06)

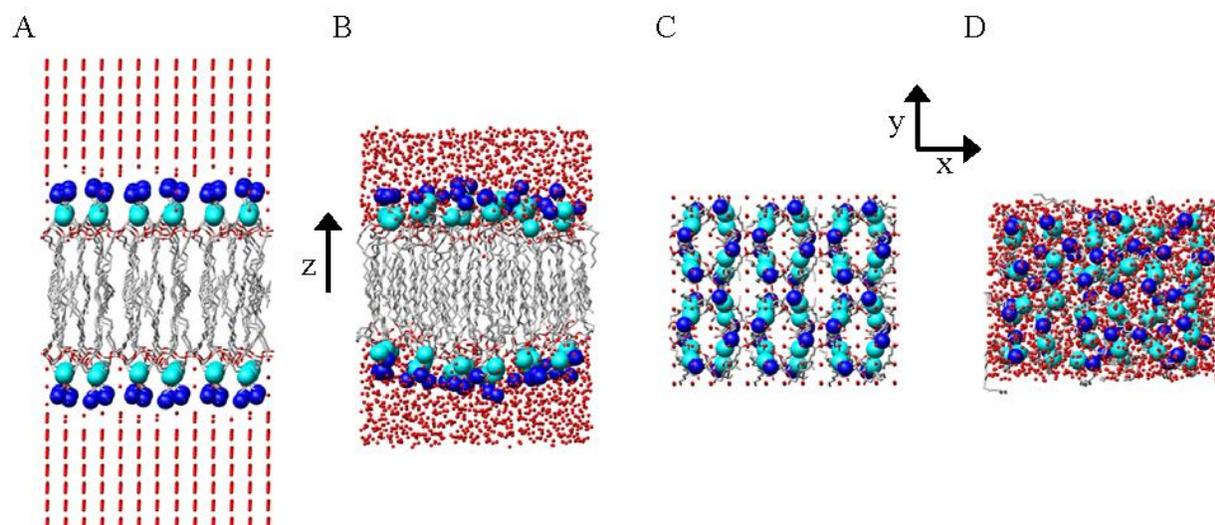
<sup>a</sup>Coupled to CG-OS-CK-CK/CG-OS-CJ-CJ

<sup>b</sup>Coupled to CK-CG-CK-CK/CK-CG-CJ-CJ/CJ-CG-CK-CK/CJ-CG-CJ-CJ

<sup>c</sup>Coupled to HA-CK-CJ-HA

### 2.3.2 Phospholipid bilayer equilibration and simulation

A lipid bilayer containing 48 DMPC molecules was constructed based on a POPC bilayer model[60], with the bilayer normal aligned along the z-axis. The starting structure had a surface area per DMPC molecule of  $78.4 \text{ \AA}^2$  and a bilayer thickness of  $41.9 \text{ \AA}$ . Using the PTRAJ program in AMBER, a  $24 \text{ \AA}$  layer of TIP3P water molecules (with a  $1.4 \text{ \AA}$  crystal spacing) was added to the upper and lower surfaces of the bilayer, resulting in the addition of 1847 waters (**Figure 2.1**). All bilayer equilibration and simulation steps were performed using AMBER 9[52]. An iterative approach was adopted to equilibrate the bilayer system and achieve properties consistent with the  $L_\alpha$ -phase of DMPC. Initially, with the lipids restrained, the water molecules were energy minimised (500 steps of steepest decent, 1500 steps of conjugate gradient) and then subjected to 10 ps of molecular dynamics in the NPT ensemble at 1 atm with anisotropic pressure scaling. The entire system was subsequently energy minimised (500 steps of steepest decent, 500 steps of conjugate gradient). Using the NVT ensemble while restraining the waters, the DMPC molecules were subjected to 5 ps of molecular dynamics at 300 K. The entire system was then minimised again for 1000 steps (500 steps of steepest decent, 500 steps of conjugate gradient). In the NPT ensemble at 1 atm with anisotropic pressure scaling, the bilayer was then restrained and only waters were equilibrated with 10 ps MD at 300 K. Finally, in the NVT ensemble, the entire system was brought to the desired temperature of 300 K over 50 ps. A production MD simulation of the bilayer-water system was then performed for 15 ns at 300 K in the NVT ensemble. A 2 fs time-step was employed throughout for integrating the equations of motion. Hydrogen-containing bonds were constrained with the SHAKE algorithm[61], and long range electrostatics were treated using the Particle Mesh Ewald method[62]. Snapshots were collected at 1 ps intervals for subsequent analysis.



**Figure 2.1** The DMPC bilayer system pre- and post-equilibration. (**A & C**) Initial model of the bilayer (pre-equilibration) and (**B & D**) after the minimisation and heating steps (the starting structure for the 15 ns simulation). To help visualize the bilayer and size of the periodic box the nitrogen (blue spheres) and phosphorous atoms (cyan spheres), and waters (small red spheres) are emphasized. Hydrogen atoms were removed for clarity.

Average bilayer thickness was computed from the distance along the z-axis (bilayer normal) between the centers of mass of the desired atoms in each leaflet, using structures taken at 100 ps intervals. The bilayer thickness ( $d_i$ ) was calculated using the center of mass of the  $N(\text{CH}_3)_3$  moiety in the head group; the hydrophobic thickness ( $d_{hc}$ ) was calculated using the carbon situated at the branch point of the hydrocarbon tails. To calculate the absolute distance between the termini of the hydrophobic groups ( $d_{\text{CH}_3}$ ), the average position of the  $\text{CH}_3$  groups from both *sn1* and *sn2* hydrocarbon tails was used.

Molecular graphics images of the bilayer were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081)[63].

## 2.4 Results and Discussion

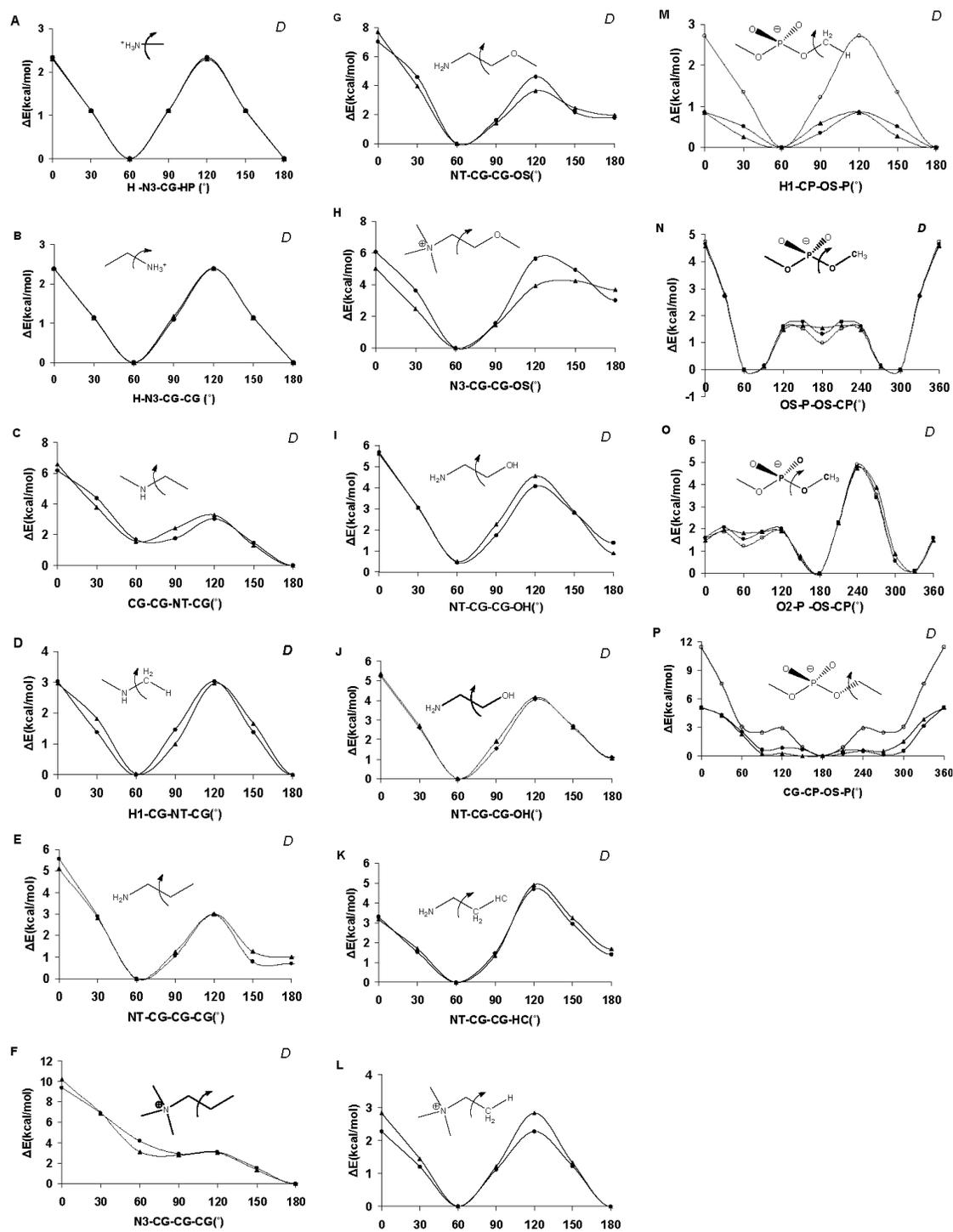
### 2.4.1 Parameters

#### 2.4.1.1 New atom types

The extension of GLYCAM06 to lipids and glycolipids required the introduction of only three lipid-specific carbon atom types. Two of the new atom types (CJ and CK) were introduced to accommodate unsaturated and polyunsaturated hydrocarbon chains and their vdW terms were transferred directly from the unsaturated carbon, CA, atom type in AMBER. Although atom types exist in AMBER to model  $sp^2$  carbon atoms, some lipids have alternating single and double bonds, which results in ambiguity regarding the placement of the double bond. To relieve the ambiguity, new atom types, CK and CJ, were added. These atom types allow the definition of a C-C single bond between two adjacent double bonded carbons (CK=CK-CJ=CJ) as in the molecule butadiene, which is a common motif in lipid tails. Notably, by dividing the rotational contribution among the constituent linkages associated with the CK-CJ bond the planarity of the  $sp^2$  center was maintained without the need to introduce out-of-plane or improper torsion terms.

The third atom type, CP, was created to describe  $sp^3$  carbon atoms attached to the oxygen atom of a phosphate group and its vdW terms were transferred directly from the saturated tetrahedral carbon, CT, atom type in AMBER. The AMBER force field (PARM94–PARM99) contains parameters for this linkage[64], employing the  $sp^3$  carbon atom type CT, however, the performance of those parameters in our test cases was not optimal. The current AMBER parameters significantly overestimated the methyl rotational barrier in methylphosphates (see **Figure 2.2M & P**). While this rotation has little significance for the overall 3D structure directly, it was important to ensure that the H-CH<sub>2</sub>-O-P rotational contribution was correctly reproduced

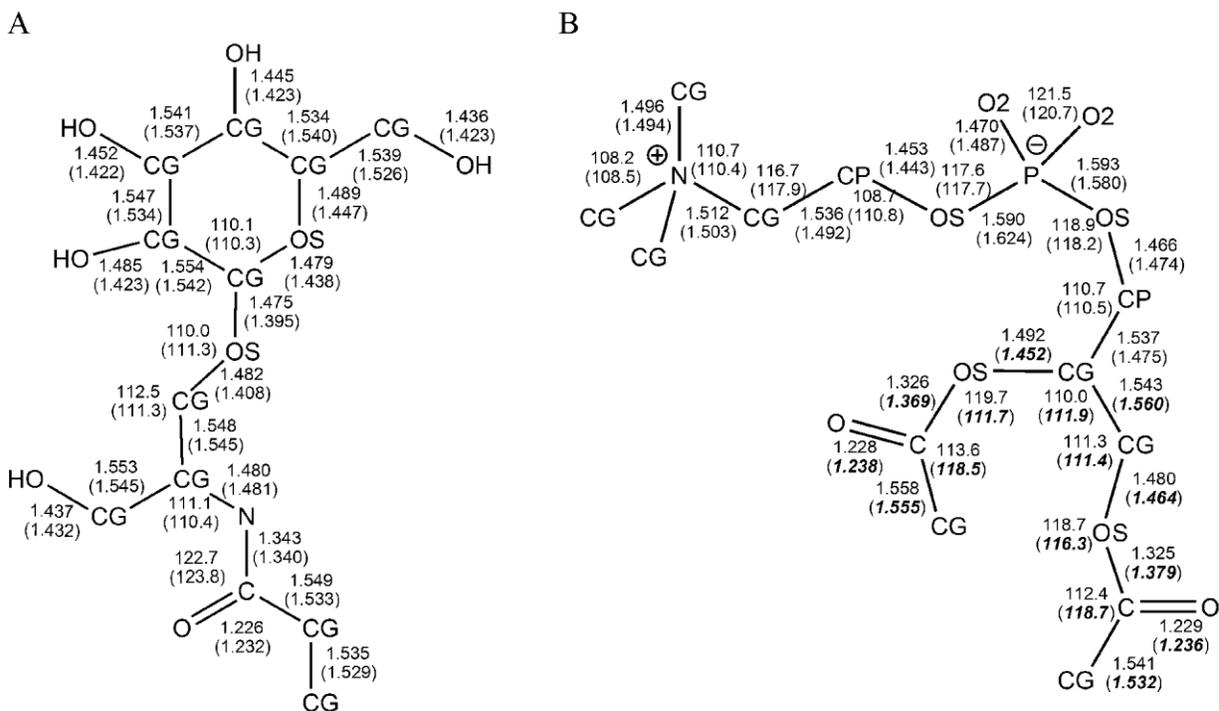
in order to employ this term in larger structures. It should be noted that unlike AMBER, or earlier versions of GLYCAM, all of the constituent torsion terms are explicitly defined in GLYCAM06. Thus, an error in the H-CH<sub>2</sub>-O-P contribution would propagate throughout the subsequent lipid components (R-CH<sub>2</sub>-O-P etc). The vdW terms for the new atom types, as well as for all other atom types employed in this extension of GLYCAM06, were transferred from the standard AMBER values and are presented in **Table 2.1** together with the valence bond lengths, angles, and associated quantum-derived force constants.



**Figure 2.2** Torsion rotation curves for phosphate and nitrogen (amine and cation) containing species. Those used in the developmental training set are indicated with a “D”, otherwise they represent test cases. Legend: QM B3LYP/6-31++G(2d,2p) data (▲), GLYCAM06 (lipid extension) (●), AMBER PARM94 (○).

#### 2.4.1.2 Valence Properties

To assess the effect of zwitterionic head group moieties, such as phosphoethanolamine, on bond lengths and angles, a truncated phospholipid was built and energy minimised. This model was selected because of its similarity to phosphatidylcholine, for which experimental crystallographic values[65] were available for comparison (**Figure 2.3**). Included in **Figure 2.3** are the pertinent valence geometries for a truncated model of the sphingolipid glucosylceramide, which has also been compared to experimental crystallographic data[66] for the intact glycolipid. The atom types employed in lipids and glycolipids are indicated in **Figure 2.3**. The average error in the bond lengths was  $0.01 \pm 0.01 \text{ \AA}$  and in the bond angles  $1.3 \pm 1.1^\circ$ . It may be observed that the values for the CG-CP and CP-OS bond lengths and CG-OS-CJ and P-OS-CP valence angles deviate somewhat from the experimental data. Alteration of equilibrium bond lengths and angles containing CP terms may be considered after a more thorough survey of relevant structures has been performed.



**Figure 2.3** GLYCAM06 energy minimized and related crystallographic values for bond lengths and angles in: **A**) glucosylceramide, experimental data from [66]; **B**) a representative phospholipid head group, experimental data from [65]. Experimentally determined values are shown in parentheses where available. Bold and italic values come from PDB id 2HG9 [67].

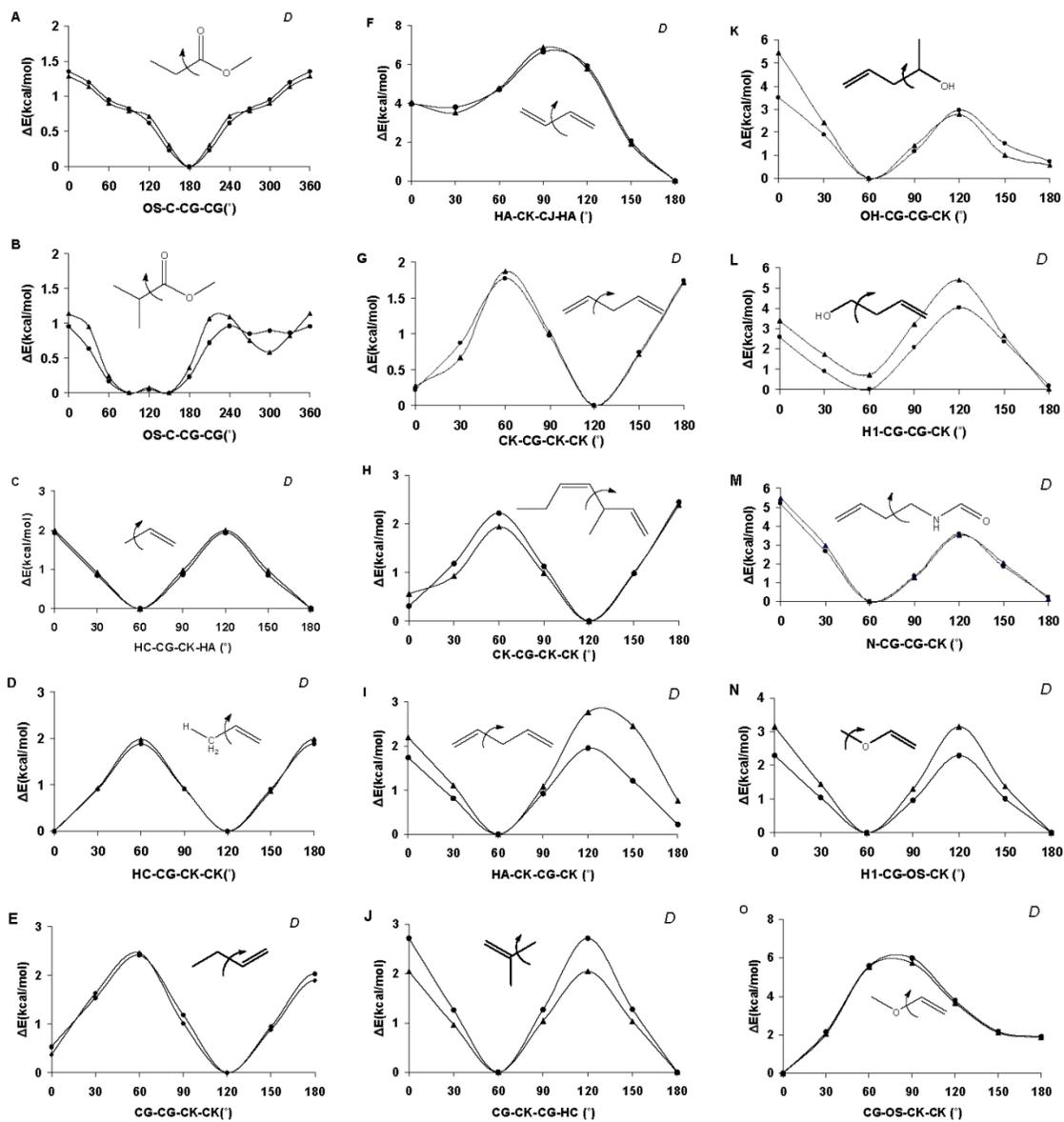
Although the CG-OS-CJ angle was correctly reproduced in the training structure methoxyethene (experimental value:  $118.3^\circ$ , GLYCAM06 value: minimized value  $118.3^\circ$ )[68], it was overestimated by  $3.7^\circ$  in the experimental structure of 1,3-bis(vinyloxy)-2,2-bis(vinyloxymethyl)propane (experimental value:  $115.8^\circ$ )[69]. This discrepancy is notable and may reflect the complex hyperconjugation present in 1,3-bis(vinyloxy)-2,2-bis(vinyloxymethyl)propane.

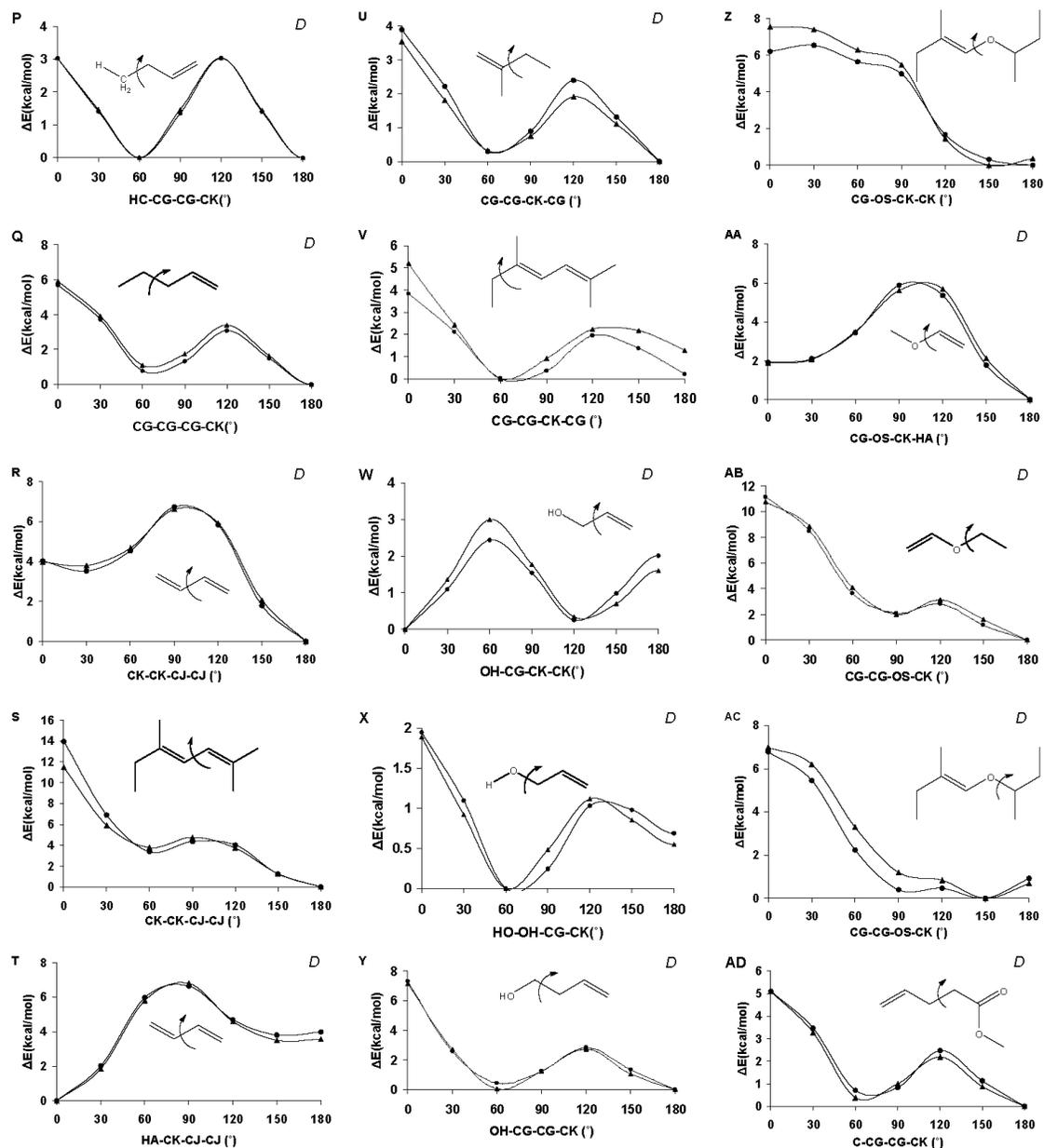
Several x-ray structures of proteins containing lipids were considered for test structures, including PDB IDs: 1LN1 [70], 1T27 [71], 2A1L [72], 2HG9 [67], 1BP1[73], and 1POB [74]. However, the B-factors for the ligands in these structures were often large (greater than  $50 \text{ \AA}^2$ )

and direct comparisons with theoretical values may suffer accordingly. For example, several ligands displayed unrealistically large C-OS bond lengths averaging 1.494 Å in all but 2HG9 (1.374 Å in 2HG9 and 1.360 Å in methoxyethene[68]). The best resolved values for the glyceryl and carbonyl portions of the head group are shown in bold and italics in **Figure 2.3** from PDB id 2HG9. B-factors in ligand PC7 ranged from 24.61 Å<sup>2</sup> to 69.48 Å<sup>2</sup>. **Figure 2.3** indicates poor agreement with 2HG9 for the angles CG-OS-C and OS-C-CG, however, the average GLYCAM06 values were found to be within the range (112.6-128.1° and 112.6-121.0° respectively) provided by the collection of experimental structures. It should be noted that no hydrogen containing valence terms were tested due to the lack of experimental structures with well-resolved hydrogens.

The torsion terms and partial charges are of profound significance to the 3D geometry and dynamics of biomolecules. The polar character of lipid head groups presents a potential for tight coupling between bond rotational properties and 1-4 electrostatic interactions. This might be expected to be a significant complication, since GLYCAM06 does not employ any 1-4 scaling to dampen such interactions. However, the quality of the fits to the QM torsion data suggests that internal electrostatic scaling is not required. It is noteworthy that a common set of torsion terms could be employed for both neutral and protonated amino groups. For example, when torsion terms were derived for neutral amine functionalities and subsequently applied with protonated ammonium species, the resulting agreement with the QM rotational energy data was reasonable (**Figure 2.2A-L**). The average error in the neutral amino training set was 0.20 kcal/mol, while that for the charged ammonium test set was only 0.57 kcal/mol. Errors in the training set were not specifically localised, but were generally distributed over the entire curve. In contrast, a large portion of the test set error was evident primarily at rotational energy barriers.

Four phosphate-containing torsion terms were examined, with an average error in the training set of 0.17 kcal/mol (**Figure 2.2M-P**).

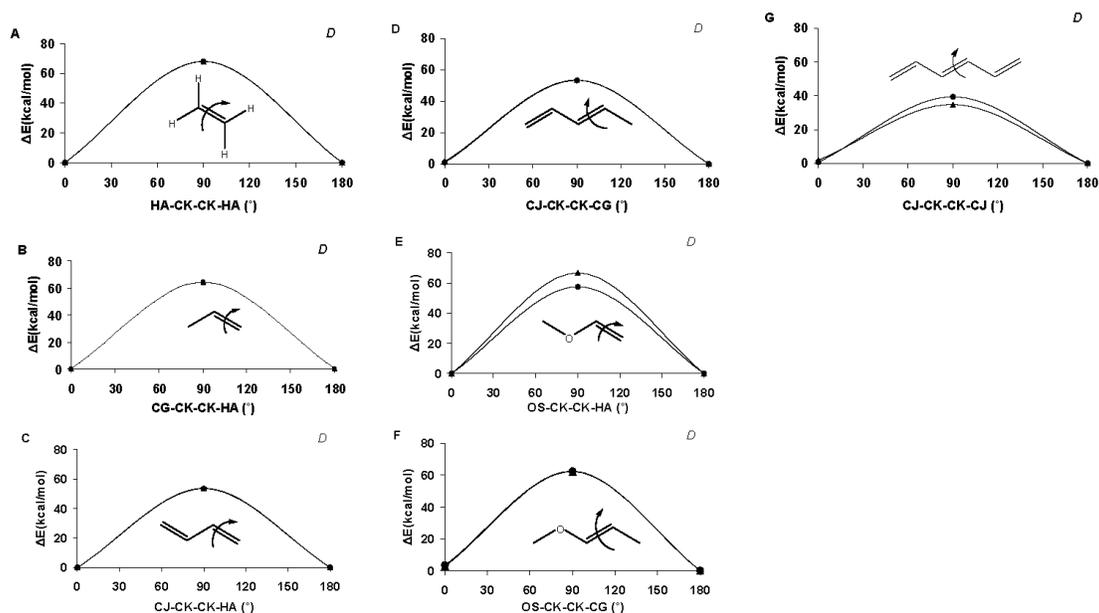




**Figure 2.4** Torsion rotation curves for lipid tail groups. Those used in the developmental training set are indicated with a “D”, otherwise they represent test cases. Legend: QM B3LYP/6-31++G(2d,2p) data (▲), GLYCAM06 (lipid extension) (●).

In order to accommodate a variety of lipid tail groups, the torsion parameterisation focused on combinations of double-bonded carbon terms with adjacent double bonds, ether, alcohol, or amide functionalities. The average error in the fit to the QM torsion curves for the tail

group training set was 0.18 kcal/mol (**Table 2.1**). Rotational energy curves for GLYCAM06 and QM data are presented in **Figure 2.4**. While the overall agreement was generally good (the average error in the test set was 0.51 kcal/mol), modeling rotational energies associated with C(sp<sup>2</sup>) atoms proved challenging. When both a hydrogen atom and an sp<sup>3</sup> carbon atom are attached to an sp<sup>2</sup> carbon, the individual contribution from the hydrogen to the rotational energy profile for the C(sp<sup>2</sup>)–C(sp<sup>3</sup>) bond is difficult to isolate. For such sequences, both the carbon (CK-CG-CK-CK) and hydrogen (HA-CK-CG-CK) torsion terms were fit concurrently, giving equal priority to adequately reproduce the C(sp<sup>2</sup>)–C(sp<sup>3</sup>) rotational energy profile, while maintaining the planarity of sp<sup>2</sup> center. This approach resulted in good agreement with the QM data, as seen in the test structure, 3-methyl-1,4-heptadiene, which displayed an average error of only 0.14 kcal/mol (**Figure 2.4H**). The C-C single bond rotation in butadiene (CK-CJ) was similarly parameterised using a simultaneous and equally weighted division between the sp<sup>2</sup> substituents (coupled fitting), and resulted in good agreement with the QM data. While some deviation was seen between the QM and GLYCAM06 rotational energies for the test structure, 2,5-dimethyl-2,4-heptadiene, the errors were primarily in the high energy eclipsed rotamers. To maintain sp<sup>2</sup> planarity and methyl rotamer preferences, parameterisation of the CG-CG-CK-CG sequence present in 2-methylbut-1-ene (**Figure 2.4U**) also required coupled fitting with HC-CG-CK-CG term. C=C bonds are common in lipid tails and apolar molecules, such as cholesterol, which is an important component of eukaryotic cell membranes. The C=C torsion parameters for substituted double bonds were all derived using coupled fitting to ensure both reasonable barrier heights and planarity of the sp<sup>2</sup> atoms, without the need for improper torsion terms. Torsion parameters for the double bonds are presented in **Table 2.1**, while the rotational energy curves are provided in **Figure 2.5**.



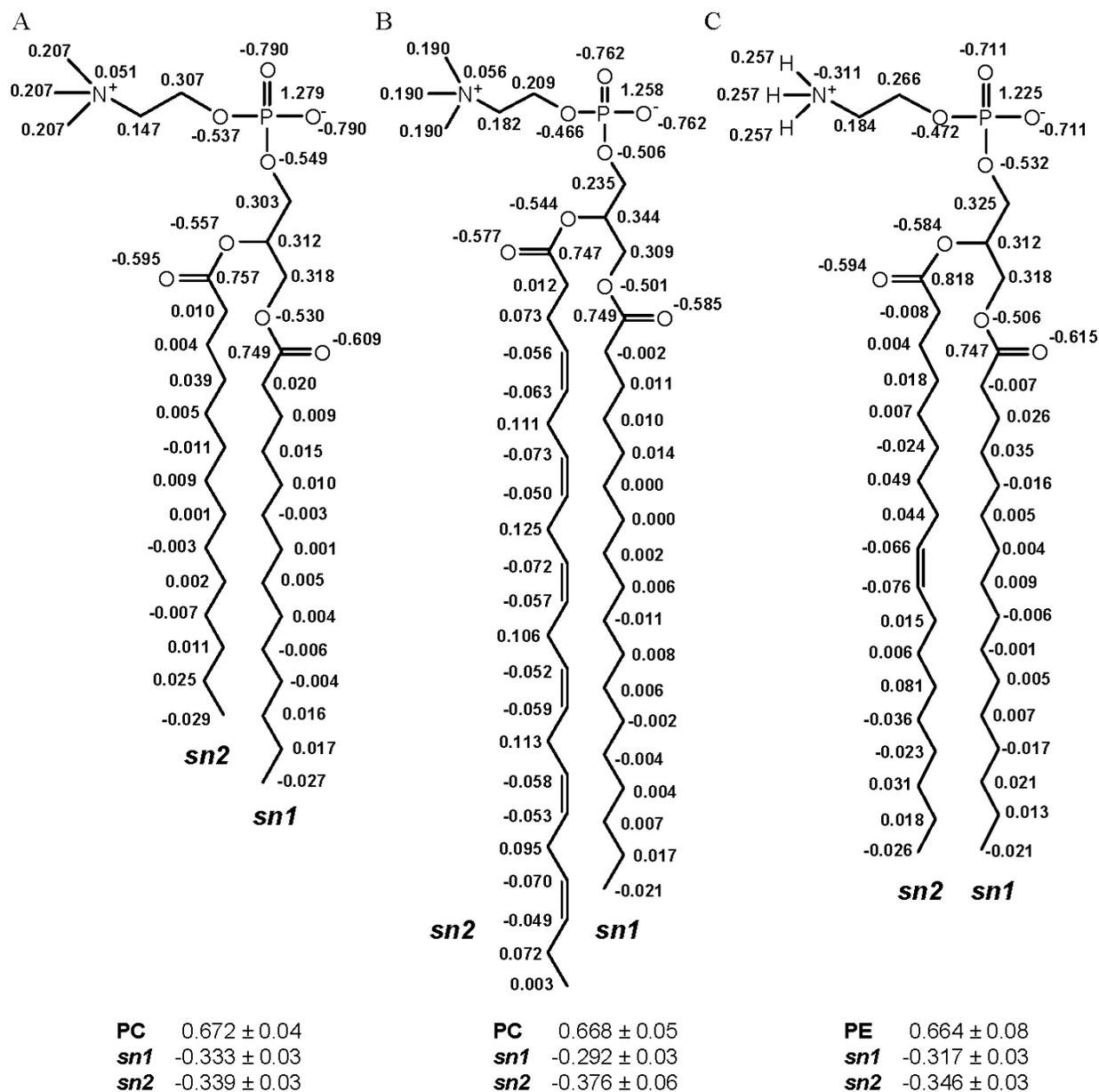
**Figure 2.5** Torsion rotation curves for C=C double bonds. Those used in the developmental training set are indicated with a “D”, otherwise they represent test cases. Legend: QM B3LYP/6-31++G(2d,2p) data (▲), GLYCAM06 (lipid extension) (●).

Additional errors in the torsions were related to the priority given to accurately fitting dependent torsion terms. For example, in the case of 3-butenol (**Figure 2.4L & X**), there are two related torsion terms: H1-CG-CG-CK and OH-CG-CG-CK. Correcting the H-containing torsion term (H1-CG-CG-CK) negatively impacted the accuracy of the heavy-atom torsion term (OH-CG-CG-CK); increasing the MM H-containing torsion barrier increased the heavy-atom torsion barrier above the QM barrier. Since accuracy of these related torsion terms was inversely related, the accuracy of the heavy-atom torsion was prioritized.

### 2.4.1.3 Partial atomic charges

Charge standardization among similar classes of lipids, i.e. phospholipids, has been examined using the ensemble average charge method[22]. The three lipid systems examined,

DMPC, SDPC, and POPE all contain similar phosphate groups and glycerol linkages to the fatty acid tails. The key differences between their head groups are in the presence of positively charged choline, in DMPC and SDPC, or of the ethanolamine found in POPE. The three structures also have different degrees of unsaturation in one of each of their tails, as seen in **Figure 2.6**. Charges for these molecules (**Figure 2.6**) were divided into the head and tail regions to determine if there was sufficient similarity to facilitate charge standardization among different functional regions of the phospholipids. The average head and tail group charges were similar, within a standard deviation, for each lipid type. The average charge on the phosphate head group was 0.668 a.u., while the charges for the *sn1* and *sn2* tail groups averaged -0.314 and -0.353 a.u., respectively. Subdividing phospholipids into functional regions and developing independent charge sets for these regions, allowed for the development of lipid bilayers from constitutive fragments with defined charges based on attachment. A similar trend was also observed for the atomic charges of the phosphate, choline, and carbonyl groups.



**Figure 2.6** Ensemble-averaged RESP charges for **A)** DMPC, **B)** SDPC, and **C)** POPE. Aliphatic protons carry zero net charge in GLYCAM06[48]. Average charges for the head (PC and PE) and tail (*sn1* and *sn2*) groups are presented with standard deviations.

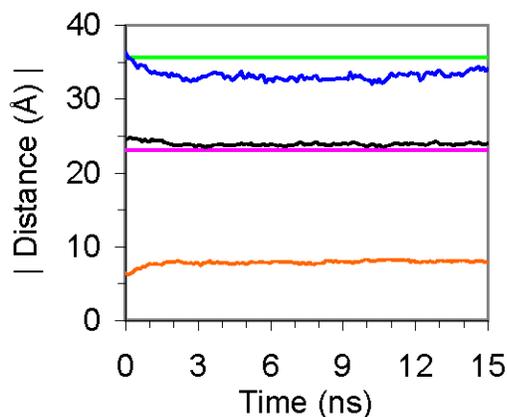
Two main atomic charge differences emerged from this study related to the head group. The difference in the choline (**Figure 2.6A & B**) and ethanolamine (**Figure 2.6C**) head group charges is apparent as the nitrogen has a moderate negative charge on the ethanolamine while the choline nitrogen has a slight positive charge. This difference was a result of the positive charges

fit to the hydrogens in the ethanolamine model, while the same positive charge was distributed over the carbon atoms in the choline models. The standard deviation on the ethanolamine nitrogen, 0.36 a.u., indicates a large degree of charge variability in contrast the nitrogen on the choline models indicates relatively consistent charges with a standard deviation of 0.05 a.u. The other significant atomic charge difference can be seen on the glyceryl carbon connecting to the oxygen atom in the phosphate (**Figure 2.6**). The average charge of this carbon atom (over the three molecules) was  $0.288 \pm 0.06$ . All other corresponding atoms within the head group were within a standard deviation of each other.

As for the lipid tails, future development will focus on the addition of fatty acid tails with alkyne, aldehyde and ketone functional groups.

#### **2.4.2 Bilayer simulations**

In the initial bilayer configuration the DMPC and water molecules were highly ordered requiring several equilibration stages, consisting of alternating rounds of energy minimisation and MD (described in 2.3.2). The iterative combination of NPT and NVT ensembles employed during the equilibration steps led to a decrease in the area per DMPC molecule and in the thickness of the membrane, from  $78.4 \text{ \AA}^2$  and  $41.9 \text{ \AA}$  in the initial structure to  $67.1 \text{ \AA}^2$  and  $36.4 \text{ \AA}$  in the post-equilibration structure (the 0 ns production MD structure). Experimental values for the area per DMPC molecule and the thickness of the membrane, measured for lamellar lattices of DMPC in excess water at 300 K, are  $61.7 \text{ \AA}^2$  and  $35.7 \text{ \AA}$ , respectively[75]. The iterative approach maintained system integrity while facilitating the equilibration of both lipid and water molecules and permitted the related adjustment in the periodic box size (**Figure 2.1**).

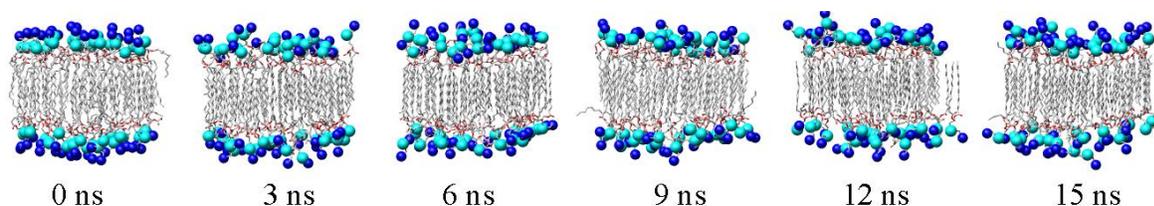


**Figure 2.7** Bilayer thickness parameters calculated from MD simulations compared to experimental spacings.  $d_l$  MD (—),  $d_l$  exp. (—),  $d_{hc}$  MD (—),  $d_{hc}$  exp. (—),  $d_{CH_3}$  MD (—)

A 15 ns NVT simulation of the bilayer was then performed, using the 0 ns equilibrated configuration as a starting structure. The overall bilayer thickness ( $d_l$ ), the hydrophobic thickness ( $d_{hc}$ ) and the distance between the tail methyl groups from the hydrocarbon tails of the opposing leaflets ( $d_{CH_3}$ ) were monitored over time to assess the stability of the membrane (**Figure 2.7**).

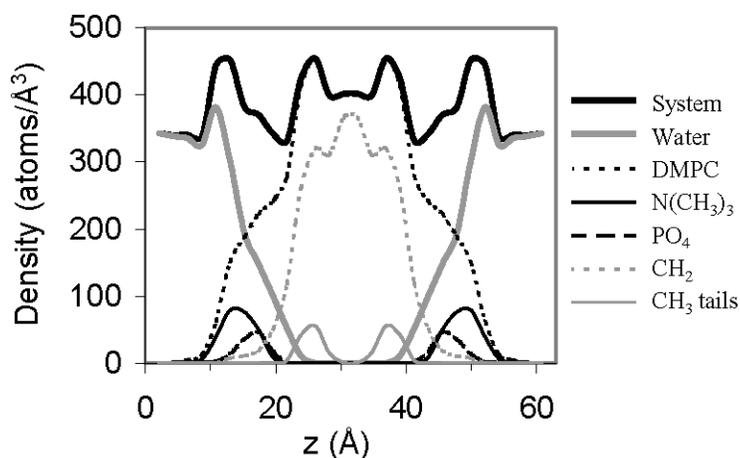
The extensive pre-equilibration procedure notwithstanding, over the first 1 ns of the 15 ns trajectory further equilibration occurred, indicated by a slight compression of  $d_l$  and expansion of  $d_{CH_3}$ . Over the 1-15 ns period, the DMPC molecules formed a fluid, but stable bilayer (**Figure 2.8**). The average thickness values over the last 14 ns of simulation were  $d_l = 33.1 \pm 0.5 \text{ \AA}$ ,  $d_{hc} = 23.9 \pm 0.2 \text{ \AA}$  and  $d_{CH_3} = 19.1 \pm 0.2 \text{ \AA}$ . These values are in good agreement with experimental parameters from lamellar DMPC lattices in excess water which give  $d_l = 35.7 \text{ \AA}$  and  $d_{hc} = 22.3 \text{ \AA}$  (error values were not reported)[75]. Although there is no experimental value for  $d_{CH_3}$ , it was calculated from the MD simulation in order to monitor fluctuations throughout the membrane. As with the measurements of  $d_l$  and  $d_{hc}$ , there was a short equilibration period during first ns of the simulation, during which  $d_{CH_3}$  expanded slightly (**Figure 2.7**). Similar to a native state simulation of a protein[76], a portion of the initial production run is considered part of the

equilibration phase. When comparing the head group atoms of the 0 ns structure to the other structures in **Figure 2.8**, the need for a longer (~1 ns) equilibration period is apparent as the head group atoms are more ordered in the 0 ns structure than the other structures. Once fully equilibrated, the membrane preserved its overall shape and structural characteristics.



**Figure 2.8** Snapshots from DMPC bilayer simulation. Coloured as in **Figure 2.1**.

To demonstrate the overall structural properties of the membrane an atom density profile of the bilayer system was calculated from the 15 ns structure (**Figure 2.9**). From the density profile, waters can be seen to penetrate the DMPC head group region ( $\text{PO}_4$  and  $\text{N}(\text{CH}_3)_3$  groups), but not the hydrophobic tail region ( $\text{CH}_2$  and  $\text{CH}_3$  groups). To conform to the experimentally determined membrane thickness, the long hydrocarbon tails from opposing leaflets overlapped, yielding a high atomic density in the central region of the bilayer. In addition to the center of the membrane, the head group region displayed a high atomic density, as it was occupied by both lipid and water molecules; a trend observed in other membrane simulations [77,78,79].

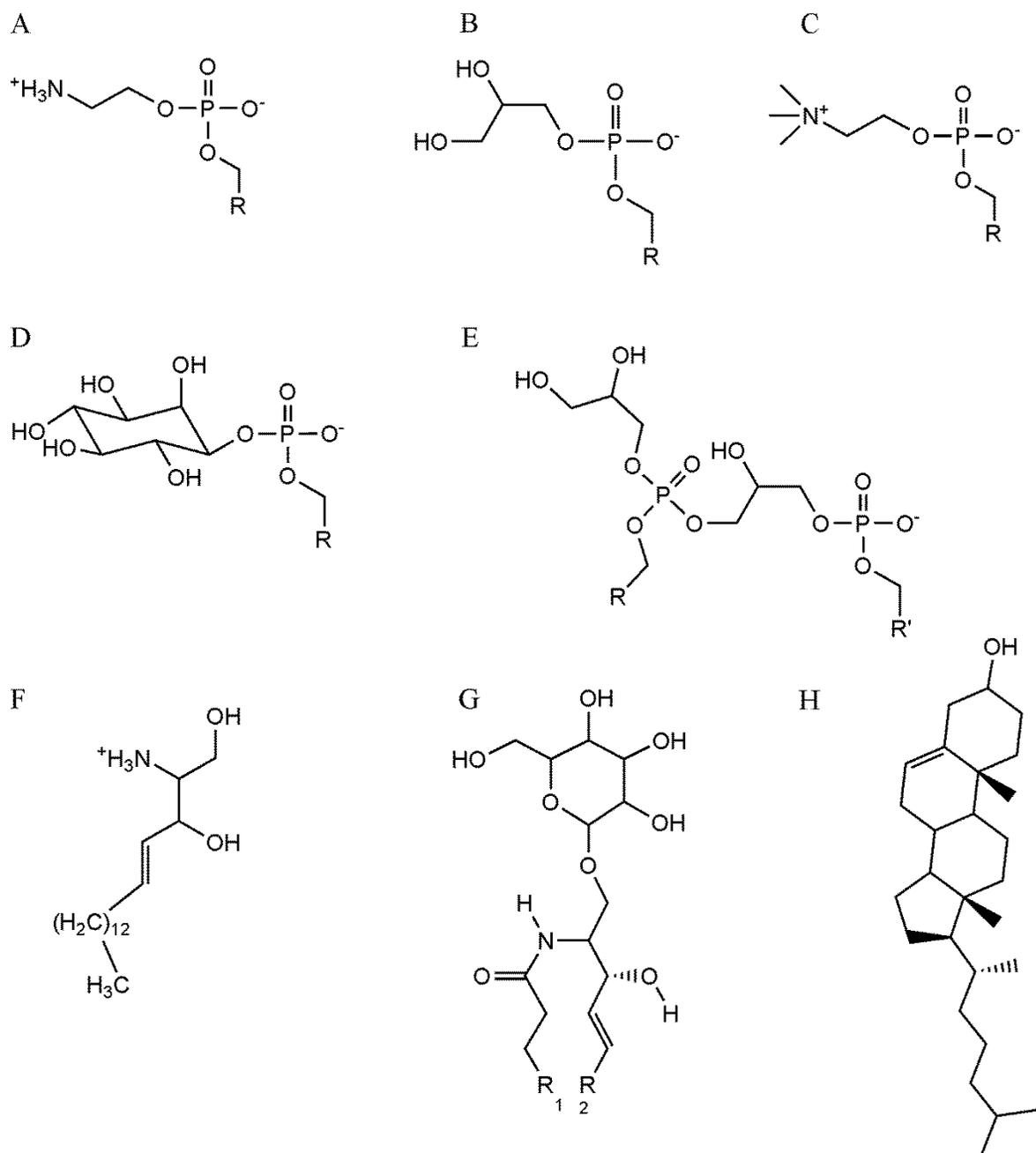


**Figure 2.9** Atom density profiles for selected groups from the 15 ns snapshot of the DMPC bilayer simulation. Values are averaged over both leaflets.

## 2.5 Conclusions

A new parameter set for lipid simulations has been developed for use with the GLYCAM06 biomolecular force field. The parameters were designed to reproduce gas-phase QM bond rotational energies and employ ESP partial charges that are consistent with the AMBER protein parameters. The development of standardized charges for phospholipid fragments (polar head groups, and *sn1* and *sn2* tail groups) will allow for faster charge assignment and the potential for interchanging fatty acid tails. GLYCAM06 parameters were developed to facilitate the combination of a broad range of lipid head and tail moieties, including carbohydrate head groups and saturated or unsaturated hydrocarbon tails. The parameters described above expand GLYCAM06 to include a large variety of lipids including triacylglycerols; sphingolipids including cerebrosides, gangliosides, and sphingomyelin; some steroids including cholesterol; and most glycerophospholipids including those incorporating cholines, ethanolamines, glycerols, inositols, and phosphatidylglycerol (**Figure 2.10**). With these parameters it is now possible to simulate heterogeneous lipid bilayers, as well as to examine the

properties of glycolipids embedded in membranes. When augmented by the AMBER parameters for proteins, it is also possible, in principle, to study protein-membrane complexes. A preliminary 15 ns NVT MD simulation of a DMPC bilayer, using the presented GLYCAM06 parameters, was in good agreement with experimental data. While achieving equilibration of lipid bilayers for MD simulations can be challenging, modification of the pressure scaling options in AMBER would facilitate this process. However, the GLYCAM06 parameters are self-contained and may be employed in a variety of software packages. The parameters and relevant structure files are available from the GLYCAM website ([www.glycam.com](http://www.glycam.com)).



**Figure 2.10** Examples of lipid structures accommodated in the extended GLYCAM06 parameters: **A)** Phosphatidylethanolamine, **B)** Phosphatidylglycerol, **C)** Phosphatidylcholine, **D)** Phosphatidylinositol, **E)** Diphosphatidylglycerol, **F)** Sphingosine, **G)** Glucosylceramide, **H)** Cholesterol.

## 2.6 Acknowledgements

The authors thank the National Institutes of Health (RR05357 and GM55230) for funding.

## 2.7 Epilogue

At the time GLYCAM06 lipids was developed, all-atom lipid simulations were just beginning to observe long-timescale (10s of ns) simulations[80]. Most simulation work to that point used united-atom (UA) or coarse-grained (CG) force fields to reduce the number of atoms in the simulation[81]. At the time, bilayer simulations required use of a surface tension factor which was also necessary for the GLYCAM06 Lipids implementation. Until recently, the AMBER simulation package did not include a surface tension restraint method[82], so simulations of lipid bilayers required use of NVT simulations to maintain the correct lipid packing density (headgroup density). NVT simulations were not used to stabilize the bilayer, as it maintained its form throughout NPT test simulations (unpublished data) but instead removed the bilayer density as a variable in the simulation. In surface tension-constrained simulations, the bilayer was allowed to expand and contract in density but a surface tension value was selected to constrain the bilayer to its average headgroup density. In this implementation, NPT simulations with surface tension restraints biased the bilayer density similar to using an NVT simulation in which the box dimensions were pre-assigned to the correct density.

GLYCAM06 Lipids was the first generalizable lipids force field in AMBER. Subsequent AMBER lipids force fields, namely Lipids11[82], were developed and identified a similar dependence on surface tension restraints to obtain the bilayer density. More recently, the

Lipids11 force field was modified (now called GAFFlipid) to remove the surface tension restraint dependence by altering the van der Waals parameters for the aliphatic tail groups[83]. A correction to the aliphatic tail group van der Waals parameter set for GLYCAM06 Lipids has been developed by fitting to solution-state properties of large aliphatic molecules (unpublished data). Outside of the AMBER force field family there have been updates to lipids force fields for CHARMM[84,85], GROMOS[86], and Martini[87,88], all of which have re-fit van der Waals terms to remove surface tension restraints.

When the work was originally completed, lipid bilayers were built ad hoc and required considerable user interaction to prepare for simulations. This meant that the DMPC bilayer, which was already publically available, was used in our test simulations for this manuscript. Since this publication, the CHARMM lipid bilayer building tool has been made publically available to generate a variety of bilayers which substantially reduces the time for setup and simulation[89].

Bilayer simulations have been observed to equilibrate over short timescales (<10 ns) while also achieving statistical convergence on measured properties in only a few nanoseconds[80]. This meant that while bilayer simulations were computationally demanding, the timescales needed were relatively short. At the time of this publication, over a week was required to simulate this 48-lipid bilayer on 16-CPU cores (15 ns). The same simulation now requires less than a day on a single GPU setup as computational resources have greatly improved with the development of more efficient code and GPU-computing[90,91,92]. This improvement in speed has resulted in more complex bilayer simulations which characterize interactions with pharmaceuticals[93] and biomolecules[94], as well as macroscopic cellular phenomena like domain formation[95].

**3. CONFORMATIONAL ANALYSIS OF GLYCOSAMINOGLYCANS BY NMR AND  
MOLECULAR DYNAMICS SIMULATIONS USING PARAMETER EXTENSIONS FOR  
THE GLYCAM FORCE FIELD<sup>2</sup>**

---

<sup>2</sup> M.B. Tessier, K. Pederson, X. Wang, A. Singh, J.H. Prestegard, A. Venot, G.-J. Boons, & R.J. Woods.  
To be submitted to *J. Am. Chem. Soc.*

### 3.1 Abstract

Glycosaminoglycan structures are a biologically important class of carbohydrates which serve roles in blood clotting, injury repair, and composition of synovial fluid, to name a few. A key component in several GAGs is the presence of heterogeneous sulfation which is known to alter biological recognition and polymer structure, especially by modifying iduronate ring conformation. Here, several heparin-like GAG disaccharides and synthetic tetrasaccharides are experimentally characterized using NMR and modeled using molecular dynamics simulations. Analysis of the NMR data revealed a sulfation-dependence to iduronate ring conformation populations while simulations show that certain sulfation positions altered the glycosidic rotamer profiles of the non-reducing glycosidic linkage to iduronate. Additionally, simulations of disaccharides with  $\Delta 4,5$ -unsaturated uronate, a non-reducing terminal cleavage product of iduronate and glucuronate, revealed that changes in the sulfation positions and levels alters the populations of the  ${}^1\text{H}_2$  and  ${}^2\text{H}_1$  ring conformations which exchange frequently on nanosecond timescales. In order to simulate these sequences, the GLYCAM06 force field was modified to include a transferable sulfate charge model,  $\Delta 4,5$ -unsaturated uronate parameters (including charges), and an independent treatment of the endo- and exo-anomeric torsion terms in iduronate.

### 3.2 Introduction

Carbohydrates play crucial roles in extracellular activity, involving immune response, regulation of biosynthetic pathways, and cell adhesion, to name a few. This makes carbohydrate structures attractive targets for development of novel diagnostic and therapeutic molecules. The polysaccharides heparin (HA) and heparan sulfate (HS) belong to a class of carbohydrates called

glycosaminoglycans (GAGs), which are composed of repeating disaccharide units of a hexosamine, and another hexose that is often an anionic residue, such as *D*-glucuronate (GlcA) or *L*-Iduronate (IdoA)[7]. A key structural and functional feature of HA and HS is the presence of sulfate groups. Both the level of sulfation, as well as sulfation patterns, are highly variable in biologically-sourced GAGs, making sequence analysis difficult. This often limits experimental characterization of GAG structures to composition-based analysis of digested fragments of native GAGs. More detailed analysis, such as by NMR or crystallography usually employs short isolated, or synthetic oligomers (GAGettes), where the sulfation patterns are well controlled. Nevertheless, sulfation patterns have been demonstrated to modulate biological function, for example in the cases of heparan sulfate in growth factor activation and cellular defense[96,97], chondroitin sulfate growth factor recognition[98], and synthetic heparin/heparan sulfate in anticoagulant activity[99].

The sulfation patterns are not just essential for altering the charge of GAGs, but have also been shown to directly impact the 3-dimensional structure of GAG fragments. Nuclear magnetic resonance spectroscopy (NMR) of GAG fragments has shown that sulfation patterns can alter the ring pucker associated with non-reducing terminal  $\Delta$ 4,5-unsaturated uronates[12,31,100] and IdoA[9,101,102]. Recently published simulations of HA/HS GAGs have shown that IdoA ring puckering (flipping from one hexopyranose ring shape to another) can have a significant impact on the 3D shape of the GAG polymer[103]. Although as noted earlier[104], not all differences in ring puckering lead to altered overall 3D shape. Whether or not sulfation significantly impacts the inter-residue linkage conformations (the glycosidic linkages), or the puckering of rings other than IdoA is uncertain. While early modeling studies of heparin disaccharides suggested that the glycosidic linkages were not significantly affected by sulfation pattern, a survey of the protein

crystallographic database (PDB) indicates that glycosidic linkages in GAGs do exhibit considerable flexibility (*vide infra*). The question as to whether GAG conformation, other than IdoA ring puckering, is affected by sulfation pattern therefore remains open.

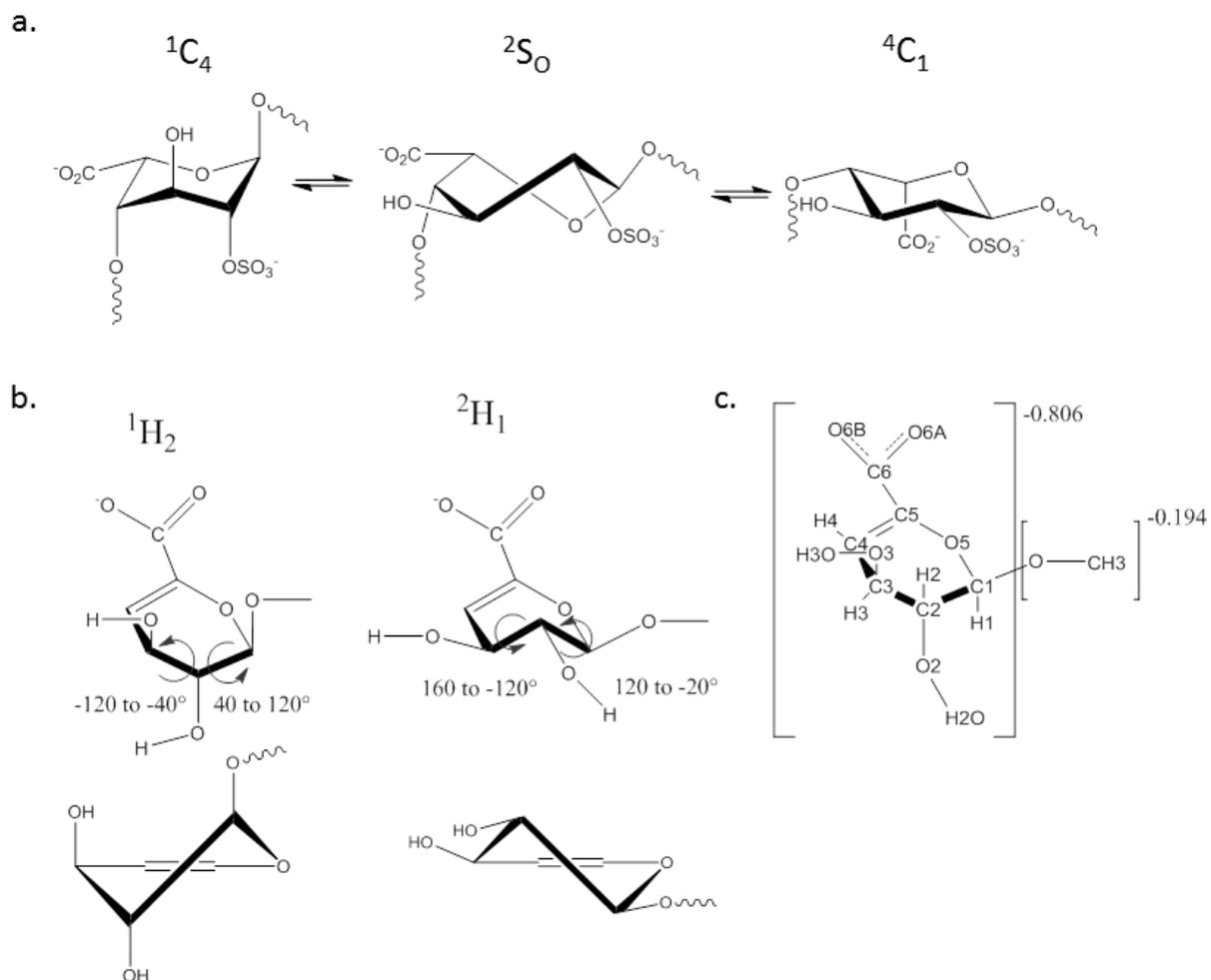
One key component to interpreting this complex data has been the development of molecular mechanics (MM) force fields to provide a basis for back-calculating NMR observables, such as scalar *J*-couplings or nuclear Overhauser effects (NOEs)[18]. Historically, existing carbohydrate MM force fields have been augmented in an *ad hoc* manner for examining sulfation patterns[18], however, generalizable sulfate models have recently been reported which support a variety of sulfation patterns and linkages[105]. Molecular simulation, employing a consistent and validated force field provides not only a basis for interpreting experimental NMR data, but also enables structure prediction of polymers and sulfation patterns that are either too large for NMR analysis or too complex for synthetic preparation.

The GLYCAM force field was specifically designed to capture both the relative energies of ring puckering states, and the frequency of IdoA transitions, which occur on the order of hundreds of nanoseconds per flip[28]. To address the ring state transitions special parameters were developed which separate the *exo*-anomeric torsion rotation from the *endo*-anomeric rotation within the pyranoside ring, as has previously been described in Sattelle et al 2012[28]. In order to obtain convergence of ring pucker populations in molecular dynamics (MD) simulation data, it has been reported that timescales as long as 6  $\mu$ s may be required; more than one order of magnitude greater than in prior simulations of similar structures[106,107,108]. One approach to overcoming such extremely long simulation times is to restrain the IdoA rings during the simulation. Multiple simulations of the GAG may then be performed with the ring held in each of the three major ring shapes,  ${}^1C_4$ ,  ${}^4C_1$ , and  ${}^2S_0$  (**Figure 3.1a**)[9,109]. While ring

puckering is also observed in  $\Delta 4,5$ -unsaturated uronates, adopting only half-chair forms  ${}^1H_2$  and  ${}^2H_1$ (**Figure 3.1b**), transition frequencies in prior simulations were two orders of magnitude faster than IdoA[30,31] making convergence times significantly shorter.

In order to address the question of the influence of sulfation pattern on GAG linkage conformation and on ring pucker preferences, we measured NMR  ${}^3J$ -couplings and nuclear Overhauser effects (NOE's) for a number of heparin sulfate component disaccharides and tetrasaccharides. Data interpretation was facilitated by deriving theoretical NMR observables from MD simulations of these GAGettes. Additionally, in this work we address two deficiencies necessary for simulating certain GAG sequences in the GLYCAM force field. The first is the creation of a generalizable *N*- and *O*-sulfate parameter set, including new valence, torsion and electrostatic charge terms that is consistent with GLYCAM partial atomic charges[22,24,25]. The second is the development of force field parameters for  $\Delta 4,5$ -unsaturated uronates that permit simulation of the non-reducing terminal residue, introduced during experimental lysis of GAGs.

The analysis presented here demonstrates that the new force field parameters (GLYCAM13) reproduce the NMR data for a number of GAG fragments, both with and without terminal  $\Delta 4,5$ -unsaturated uronates. Notably, the simulations indicate that the conformation of the glycosidic linkage between GlcN and IdoA residues changes by the introduction of *N*-sulfation to the GlcN residue.



**Figure 3.1 a.** The three major ring states of 2-*O*-sulfated L-iduronate. **b.**  $\Delta$ 4,5-Unsaturated uronate puckering states with the torsion ranges typically associated with H<sub>1</sub>-C<sub>1</sub>-C<sub>2</sub>-H<sub>2</sub> and H<sub>2</sub>-C<sub>2</sub>-C<sub>3</sub>-H<sub>3</sub> torsions. **c.** The atom naming scheme employed in the model of  $\Delta$ 4,5-unsaturated uronate as well as the net residue charges typical to GLYCAM.

### 3.3 Methods

#### 3.3.1 NMR

NMR spectroscopy was carried out on a spectrometer operating at 18.8 T, equipped with a Varian Inova console and a 5 mm cryogenically cooled probe. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was included as an internal reference in each sample. NMR samples consisted of

0.5 mg of disaccharide in 100% D<sub>2</sub>O buffer containing 20 mM sodium phosphate and 1 mM DSS, pH 6.5. The sample was shimmed to a DSS linewidth < 1 Hz.

Proton resonances were assigned using a standard COSY experiment (Varian ChemPack), processed with NMRpipe[110] and assigned in Sparky[111]. <sup>3</sup>J-coupling measurements were made from a 1D proton experiment with presaturation to suppress signal due to any residual H<sub>2</sub>O, collected with a spectral width of 9000 Hz and 32k points, processed and analyzed in MestReNova.

NOEs were measured using a standard NOESY experiment (Varian ChemPack) with a mixing time of 0.4 s, 512 increments and 9000 points, processed with NMRpipe[110]. NOE peaks were integrated in NMRViewJ[112] and the distance was calibrated using the distance from the MD simulations between either the H<sub>1</sub><sup>B</sup> and H<sub>2</sub><sup>B</sup> or H<sub>2</sub><sup>B</sup> and H<sub>3</sub><sup>B</sup> protons on the disaccharide reducing terminal residue (residue **B**), **1-5** as a standard. Tetrasaccharides, **6** and **7**, were calibrated using the distance from the MD simulations between H<sub>1</sub><sup>C</sup> and H<sub>5</sub><sup>C</sup> protons on glucuronate (residue **C**).

### 3.3.2 Valence and torsion parameter development

The GLYCAM06 force field for carbohydrates[24] and lipids[25] was adapted to include new terms describing double-bonds in unsaturated uronic acids. The only new term added in this work was the sulfur (S) atom adapted from the van der Waals parameters for the sulfate atom type found in parm99[113]. Charges were developed using methods consistent with prior GLYCAM06 development procedures, *vide infra*. All valence and torsion terms (supplementary **Table 7.1** & supplementary **Table 7.2**) were developed using the hierarchical development

procedure outlined in two prior GLYCAM06 publications[24,25] where bonds are developed first, followed by angles, and torsions. Small molecules were selected for parameter development such that each contains as few terms as possible while maintaining an electronic environment relevant to a carbohydrate[24]. Equilibrium values for valence terms were obtained from the averages of crystal structures found in the Cambridge Structural Database [53] with molecule IDs HEMKEP, KOCOJ, SRHXGU, MIZFUX, GUVFOS, GUVFEI, GUVFAE & ZULPIF (supplementary **Table 7.1**).

Quantum mechanical (QM) valence and torsion potentials were collected using B3LYP/6-31++g(2d,2p)//HF/6-31++g(2d,2p) except in the case of 2-(Methoxymethoxy) acrylate (Os-Cg-Os-Ck torsion term) where the MP2 level of theory was employed for geometry optimization. Small molecule torsion potentials were collected for the molecules found in supplementary **Table 7.2** at 30° increments with the exception of terms describing double bond rotations in which only 0, 90 and 180° increments were used to characterize the *cis/trans* relative energies and the barrier height between them. The barrier energy for *cis/trans* rotation in double-bonds was distributed equally between heavy and light atom terms, Cg-Ck-Ck-C and Ha-Ck-Ck-C, avoiding the need for improper torsions as was established in the development of lipid parameters. All torsion terms were developed without the use of a phase shift (phase shift set to 0).

### 3.3.3 Charge development

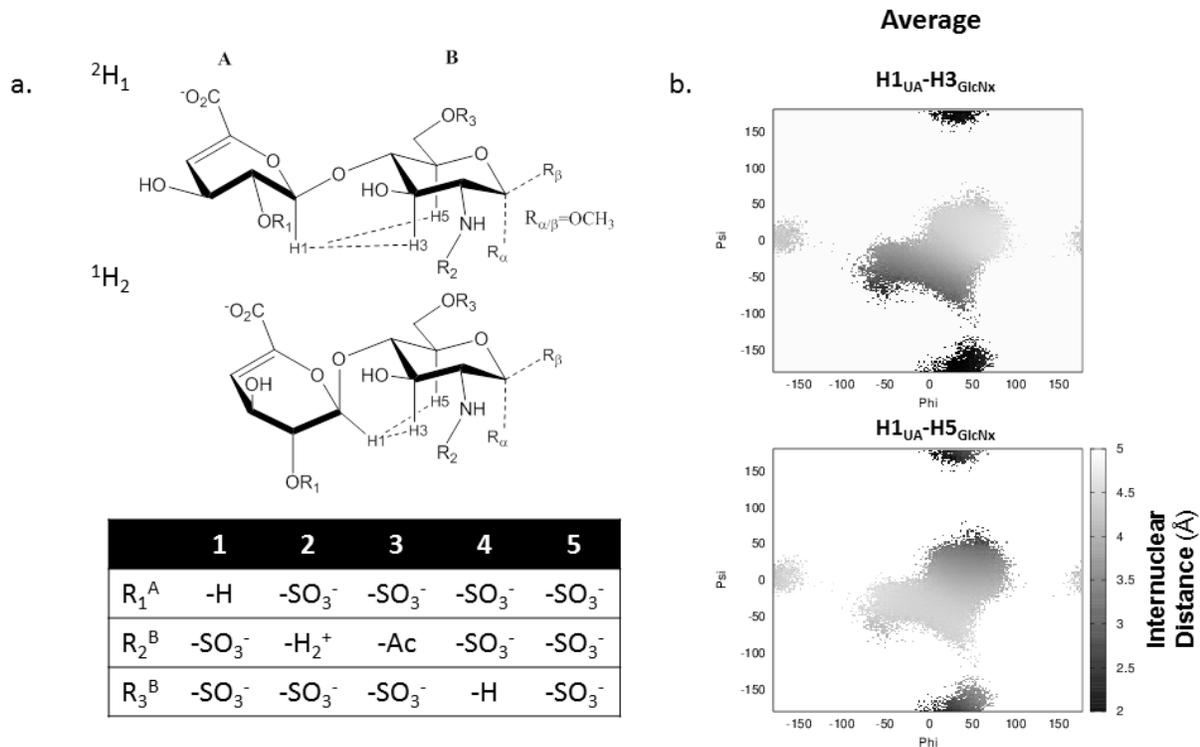
Atomic point charges were developed using the GLYCAM06 method for ensemble-averaged charges[22] using the restrained electrostatic potential (RESP) weighting method[57]

and the GLYCAM standard to maintaining net molecular integer charges (**Figure 3.1c**). The GLYCAM ensemble averaged charge method relies on two stages of charge fitting; first, a quantum mechanically optimized crystal structure is used to develop restrained electrostatic potential (RESP) charges (if a crystal structure is not available, a QM-optimized model is used). Second, an ensemble-averaged RESP charge set is developed using a fully solvated molecular dynamics (MD) simulation employing the crystallographic RESP charge set. In this simulation, exocyclic free rotamers (hydroxyls, methyls, sulfates, etc.) must be adequately sampled, typically requiring 10 to 50ns at 300K in an nPT MD simulation. One-hundred evenly-spaced frames are selected from the simulation as a representative ensemble of 3D structures. This ensemble is then subjected to QM-optimization in which the exocyclic rotamer torsions are frozen in their MD-simulation conformations. Then RESP charges are calculated for each frame and averaged together to form the ensemble averaged charge set for a particular molecule. In the case of anionic systems, HF/6-31++g(2d,2p) is used for both optimization and RESP charge calculation whereas in neutral and cationic systems HF/6-31g(d) is employed . Both Pople basis sets use a RESP weight of 0.01. In valence and torsion parameter development charges are selected from the lowest energy conformational state using HF/cc-pvtz for calculating the electrostatic potential and a RESP weight of 0.0005. The cc-pvtz basis set at this RESP weight has been shown to produce consistent results with the RESP-weighted Pople basis sets described above[22].

### 3.3.4 Molecular dynamics simulation setup

Each of the disaccharides found in **Figure 3.2a** were simulated in four distinct 100 ns nPT MD simulations at 300 K. Two pairs of simulations were started from each of the two  $\Delta$ 4,5-

unsaturated uronate half-chair conformations,  ${}^1H_2$  and  ${}^2H_1$ . The tetrasaccharide models were set into four simulation sets: three simulations in which the iduronate was restricted to the  ${}^1C_4$ ,  ${}^2S_0$  and  ${}^4C_1$  ring states using 15 kcal/mol Cartesian restraints on the iduronate ring heavy atoms; and one simulation in which no restraints were employed. All simulations used sodium ( $Na^+$ ) counter-ions to neutralize the net charge of each system and cubic boxes of explicit TIP3P waters[114] were added so that there was at least 12 Å between the glycan solute and each edge of the solvated simulation box. The GLYCAM06i[115] force field and its modified terms for iduronates, GLYCAM11[28], were used with the PMEMD CPU and GPU code[90,91,92] found in the AMBER 12 simulation package[116]. Non-bonded scaling factors were set to unity and a 10.0 Å non-bonded cut-off was employed in all steps. Initial geometries for the glycosidic linkages were obtained from the most commonly observed rotamers for  $\varphi$  (H1-C1-O4-C4) and  $\psi$  (C1-O4-C4-H4), approximately 50 and 15° respectively. The MD simulations were preceded by two minimization schemes, both with at most 10000 steps of minimization, 5000 steps of steepest descent followed by 5000 steps of conjugate gradient. One minimization was performed prior to adding counter-ions and explicit solvent by using in a Generalized-Born implicit solvent[117] (implicit solvent simulations employed an infinite non-bonded cutoff), and one minimization after each had been explicitly solvated and neutralized.



**Figure 3.2 a.** The five  $\Delta 4,5$ -unsaturated uronate for which NMR and simulation data have been collected. The H1---H3' and H1---H5' internuclear distances are shown to emphasize their distance-dependence based on the glycosidic linkage geometry. **b.** The H1---H3' (top) and H1---H5' (bottom) distance dependence is plotted relative to the phi and psi glycosidic angles. Shown in (b) is the relationship for GAGette 1 which demonstrates the relationship observed in all five  $\Delta 4,5$ -unsaturated uronates.

An important distinction in the GPU code compared to the multi-core CPU code in PMEMD is that the GPU and single-core CPU code is deterministic while the multi-core CPU code is not[90]. Taking this into consideration, all equilibration steps were performed over multiple CPU cores to introduce variations in the molecular motions before the production simulations were simulated using the GPU code. This variation could also have been introduced by varying the random seed number used in the GPU code. The simulations were equilibrated using an nPT ensemble consisting of a 50 ps heating stage raising the temperature from 0 to 300 K and an additional 100 ps in which the temperature was maintained at 300 K. All MD simulations used periodic simulations where the pressure was maintained at 1.0 atm, the external

dielectric was set to 1.0, and the system compressibility was set to that of water. The weak coupling algorithm[17] was used for all temperature controls and the SHAKE algorithm[61] was used to constrain bonds with hydrogens, allowing 2 fs timesteps to be used. Production MD frames were stored every 1 ps for further analysis.

Initially, unrestrained MD simulations of **6** and **7** were run for 1  $\mu$ s however this simulation length was found to be insufficient to obtain iduronate ring state population convergence (data not shown). Recently published work showed that without special sampling techniques the appropriate simulation lengths are likely on the 5-10  $\mu$ s timescale[103]. In lieu of simulating long timescales for sampling, and to obtain pucker-specific glycosidic linkage profiles, an approach was taken in which the iduronate was restrained in the three most populated iduronate ring states,  ${}^1C_4$ ;  ${}^2S_0$ ; and  ${}^4C_1$ . Each trajectory was run for 1  $\mu$ s to adequately sample the  $\varphi$  and  $\psi$  space for the glycosidic linkages.

### 3.3.5 NMR and MD simulation analysis

The MD trajectories for each of the disaccharides were analyzed to determine the exact populations of the  ${}^1H_2$  and  ${}^2H_1$  ring states by comparing the  $H_1-C_1-C_2-H_2$  and  $H_2-C_2-C_3-H_3$  torsion rotation rotamers. The torsion profiles of these two states allow ring-state separation around values of  $120^\circ$  for  $H_1-C_1-C_2-H_2$  and  $-120^\circ$  for  $H_2-C_2-C_3-H_3$ . Both ring states lie within  $\pm 80^\circ$  of these barriers creating an easy method for separating the two states, more specifically the  ${}^1H_2$  state is defined by  $40^\circ$  to  $120^\circ$  ( $H_1-C_1-C_2-H_2$ ) and  $-120^\circ$  to  $-40^\circ$  ( $H_2-C_2-C_3-H_3$ ) while the  ${}^2H_1$  state is defined as  $120^\circ$  to  $-20^\circ$  ( $H_1-C_1-C_2-H_2$ ) and  $160^\circ$  to  $-120^\circ$  ( $H_2-C_2-C_3-H_3$ ) (**Figure 3.1b**). The quantification of the ring puckering and the sampling of the glycosidic  $\varphi$  and  $\psi$  rotamers were

considered as diagnostic of adequate sampling times for each trajectory. Transitions between chair forms (ring flipping) was frequently observed within 100 ns for each ring.

NMR properties were calculated for comparison with the experimental properties determined in this work, particularly the proton-proton distance relationships from the nuclear Overhauser effects (NOEs) and the three-bond  $J$ -couplings ( $^3J$ -couplings). Both experimental values are comparative to the MD averages of these properties calculated over the entire trajectory. NOE's were calculated assuming a simple the proportional relationship of NOE signal to  $1/r^6$  where  $r$  is the distance between two protons.  $^3J$ -couplings were calculated using Karplus-like equation developed by Haasnoot et al.[118] using the electronegativity values indentified by Altona et al.[119] (supplementary **Equation 7.1** & supplementary

**Table 7.3).** Heteronuclear glycosidic torsions are provided however experimental data was not collected as  $C^{13}$  enriched samples were not available and these couplings were not observable at natural abundances. For each saccharide residue, except the iduronate,  $^3J$ -couplings were collected using the MD simulation trajectory. Since restraints were employed on the iduronate residues,  $^3J_{HH}$ -couplings were obtained from models of the ideal geometries for the  $^1C_4$ ,  $^2S_0$ , and  $^4C_1$  ring states. These ideal couplings were used to predict the experimental ring state populations for IdoA using an RIS least-squares fitting method outlined in [20]. Additionally, experimental ring state populations were identified which have the lowest total absolute error ( $|J_{NMR}-J_{MD}|$ ) between the computed and experimental intra-ring couplings. This analysis is less ideal as it produced an uneven distribution of errors compared to the RIS analysis, however, this type of analysis is useful in emphasizing the range of populations which fit the observed  $J$ -couplings.

## 3.4 Results

### 3.4.1 Force field development and evaluation

#### 3.4.1.1 Valence and torsion parameters

The development of a set of generalizable GAG parameters was based on the existing van der Waals parameters for the sulfate  $S$  atom type found in FF99SB[113]; all other terms were developed *de novo* for this work. The sulfate parameters and the carbon-carbon double bond terms for the  $\Delta 4,5$ -unsaturated six-membered ring represent the two major additions to the GLYCAM library from this work. A total of four bonds, ten angles, seventeen torsions, and three improper torsions were parameterized, while the rest have been adopted from

GLYCAM06[24], GLYCAM11[28], or the AMBER protein force field[113] (supplementary **Table 7.1** & supplementary **Table 7.2**). All torsions were initially fit to the quantum mechanical potentials for the relevant molecules found in supplementary **Table 7.2**. Exocyclic torsion rotations were developed using  $\Delta$ 4,5-unsaturated uronates or THP analogs (see supplementary **Table 7.2**). In the case of *N*-sulfate terms, planarity was maintained during the QM torsion rotation to reflect solution conformations of an  $sp^2$ -hybridized nitrogen. The improper torsion term used to maintain planarity in the force field was selected by determining a value with which planarity was maintained in both the gas-phase and fully solvated simulations of the  $\alpha$ - and  $\beta$ -methyl glycosides of *D*-GlcNSO<sub>3</sub><sup>-</sup>.

Despite good fitting to the QM rotational energy data, overall, preliminary simulations of  $\alpha$  and  $\beta$ -GlcNSO<sub>3</sub><sup>-</sup> revealed the barrier to rotation was insufficient to prevent rotation of the sulfonamide group. Such rotation is inconsistent with orientations observed in a survey of crystal structures (data not shown). To address this issue, the torsion barrier associated with the N-Cg bond was increased until rotation was not observed in solution MD simulations at 300K. This resulted in a degraded fit to the QM data for the S-N-Cg-Cg/H1 torsion terms (average error 2.5 kcal/mol), and indicates that when rotations involve highly ionic moieties, gas-phase QM data is limited in its applicability. Similar behavior has been reported for carboxylate-containing bond rotations[120].

A unique characteristic to the sulfate parameters is the use of a  $V_6$  term to precisely describe the H-N-S-O2 rotation. A small periodic deviation in the torsion potential of this rotation was observed after fitting to the QM potential. A small  $V_6$  term of -0.1 kcal/mol was found to correct the MM deviation, however, it represents the first time that a  $V_x$  term greater

than 3 has been introduced into the GLYCAM force fields. Simulation software which does not allow  $V_6$  terms may remove this term, with only minor deviations from the ideal behavior.

In the development of the endocyclic terms associated with the unsaturated bond, correct reproduction of certain experimentally-observed rotamer preferences required use of a higher level of QM theory (second order Moller-Plesset, MP2) during geometry optimization. This was particularly the case for rotation about the central Cg-Os bond in the Os-Cg-Os-Ck sequence, during which the adjacent (terminal) Os-Cg bond favored experimentally-inconsistent rotamers at the HF-level. This behavior was corrected by performing the geometry optimizations at the MP2 level. The need for a QM level that includes electron correlation presumably reflects the presence of hyperconjugation between the oxygen atoms (Os) and the unsaturated carbon center (Ck). Having observed a dependence of rotamer preference on level of QM theory for this term, all other terms were re-examined, and found not to show any notable dependence on the QM level.

Also during development of the Cg-Os-Ck-C term, it was noted that while it was possible to obtain good fits to the rotational energies for molecular fragments, incorporation of these terms into the cyclic uronate failed to reproduce the conformational preferences of the constrained six-membered ring. Specifically, use of the terms derived from small fragments led to a preference for the *syn* orientation, whereas only the *anti* is seen in crystallographic data. The  $V_1$  term, which is used to adjust the relative energy between 0 and 180° rotamers, was therefore removed, resulting in a larger net error when compared to QM data for the small acyclic fragments, but achieving agreement with crystallographic data.

### 3.4.1.2 Charges

Charge models for *N*- and *O*-sulfates; glucosamine; and  $\Delta$ 4,5-unsaturated uronates were developed using the standard GLYCAM ensemble-averaged charge method[22]. The charges were developed for sulfates using 4-*O*- and 6-*O*-sulfated  $\beta$ -D-GalNAc, and both anomers of *N*-sulfated  $\alpha$ - and  $\beta$ -D-glucosamine (D-GlcNS), using initial glycan geometries extracted from co-crystallized protein-sugar complexes. The computed charges for the sulfate fragments ( $\text{SO}_3^-$ ) of both *N*- and *O*-sulfates were within statistical variance, allowing the creation of an interchangeable sulfate residue. Further examination of the sulfated sugars revealed that the charges of the remaining atoms were similar to those of their non-sulfated counterparts[24]. The only significant deviation between the non-sulfated and sulfated sugar charges was found at the linking atom (O or N) associated with the sulfate residue. The adjustment of the charge on this atom was sufficient to obtain a net integer charge on each sulfated sugar. The charges are presented supplementary **Table 7.4**.

Charges for protonated  $\alpha$ - and  $\beta$ -D-glucosamine ( $\text{GlcNH}_3^+$ ) were similarly developed, and found to significantly vary from the GLYCAM charges for  $\alpha$ - and  $\beta$ -D-Glc and GlcNAc, particularly for the ring carbon atoms (supplementary **Table 7.4**). This variation suggests that such analogs will require unique charge sets for each monosaccharide. This is not surprising, as the charge site is directly adjacent to the sugar ring.

Charges for the  $\Delta$ 4,5-unsaturated uronate monomers were obtained from by averaging the charges for each of the low-energy half-chair states,  ${}^1\text{H}_2$  and  ${}^2\text{H}_1$ , (**Figure 3.2a** and supplementary **Table 7.4**).

### 3.4.1.3 Optimization of $\Delta 4,5$ -unsaturated uronate ring state populations

In order to determine the experimental populations of the two half-chair states,  ${}^1H_2$  and  ${}^2H_1$ , in  $\Delta 4,5$ -unsaturated uronates ( $\Delta 4,5uu$ ), scalar NMR  ${}^3J$ -couplings were measured for five disaccharides terminating in  $\Delta 4,5uu$  residues. The  ${}^3J_{HH}$ -couplings relevant to the ring conformation (associated with the  $H_1-C_1-C_2-H_2$ ,  $H_2-C_2-C_3-H_3$ , and  $H_3-C_3-C_4-H_4$  torsion angles) were decomposed into populations for the half-chair states, by performing a rotational isomeric state (RIS) analysis[121], in which the total observed  $J$ -value is assumed to arise from a linear combination of the population-weighted  $J$ -values computed[118] for each individual state (see **Table 3.1**).

**Table 3.1** Experimental observables for residue **A** of the  $\Delta 4,5uu$  disaccharides **1-5**.

Residue <b>A</b>	<b>1</b>		<b>2<sub><math>\alpha,\beta</math></sub></b>		<b>3<sub><math>\alpha,\beta</math></sub></b>		<b>4</b>		<b>5</b>	
	<i>Exp</i>	<i>MD</i>	<i>Exp</i>	<i>MD</i>	<i>Exp</i>	<i>MD</i>	<i>Exp</i>	<i>MD</i>	<i>Exp</i>	<i>MD</i>
<i>J</i> -coupling* (Hz)										
H <sub>1</sub> -C <sub>1</sub> -C <sub>2</sub> -H <sub>2</sub>	5.5	4.6	3.4	3.4	3.4	2.7	3.0	2.7	3.7	2.9
H <sub>2</sub> -C <sub>2</sub> -C <sub>3</sub> -H <sub>3</sub>	4.9	5.0	2.8	3.4	-	2.5	2.5	2.5	2.8	2.8
H <sub>3</sub> -C <sub>3</sub> -C <sub>4</sub> -H <sub>4</sub>	3.8	3.4	4.4	4.4	4.7	5.0	4.7	5.0	4.4	4.8
Pop. ( <sup>1</sup> H <sub>2</sub> : <sup>2</sup> H <sub>1</sub> )	40:60	42:58	69:31	67:33	70:30	80:20	76:24	80:20	67:33	76:24
<i>NOE-derived distances (Å)</i>										
H <sub>1</sub> <sup>A</sup> -H <sub>3</sub> <sup>B</sup>	3.0	4.2	3.1	4.4	-	3.5	2.7	3.3	2.4	2.9
H <sub>1</sub> <sup>A</sup> -H <sub>5</sub> <sup>B</sup>	-	3.9	2.8	3.9	-	3.5	3.1	3.6	-	3.3

\*Average <sup>1</sup>H<sub>2</sub> couplings for H<sub>1</sub>-C<sub>1</sub>-C<sub>2</sub>-H<sub>2</sub>, H<sub>2</sub>-C<sub>2</sub>-C<sub>3</sub>-H<sub>3</sub>, and H<sub>3</sub>-C<sub>3</sub>-C<sub>4</sub>-H<sub>4</sub> are 1.7, 1.2, and 5.8 Hz. Average <sup>2</sup>H<sub>1</sub> couplings are 6.7, 7.8, and 1.6 Hz

In the case of disaccharides **1** and **2**, the NMR-derived populations, 40:60 and 69:31 (<sup>1</sup>H<sub>2</sub>:<sup>2</sup>H<sub>1</sub>), respectively, were in poor agreement with the independently-derived values from the MD simulation, namely 70:30 and 35:65, respectively. Disaccharides **1** and **2** were therefore selected as reference sets for the development of new force field parameters associated with the ring torsion terms in  $\Delta 4,5uu$  residues. As the partial atomic charges in the  $\Delta 4,5uu$  residue had been derived under the assumption of an equal population of half-chair states, this appeared to be a potential source of error. However, data from MD simulations, in which the contribution of the partial charges from each half-chair was varied (from 0% <sup>1</sup>H<sub>2</sub> to 100% <sup>1</sup>H<sub>2</sub>), indicated that the half-chair populations were relatively insensitive to the atomic charges (supplementary **Figure 7.1**).

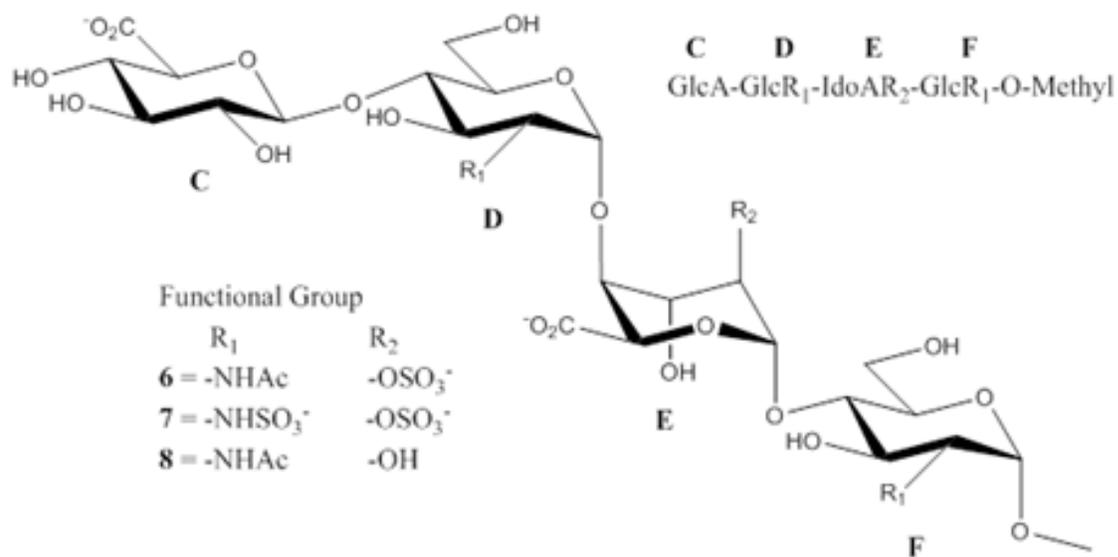
Subsequently, an examination of ring torsion terms was undertaken, as an alternative approach to altering the ring pucker populations. Only one torsion term (Oh-Cg-Cg-Ck, **1**; Os-Cg-Cg-Ck, **2**) differentiates the ring systems in **1** and **2**, suggesting that it could be employed to tune the relative populations of each half-chair in these reference systems. These torsion terms

were iteratively adjusted so as to obtain optimal agreement with the NMR populations. MD simulations with the optimized torsion terms, not surprisingly, yielded average population ratios for **1** and **2** of 42:58 and 67:33 ( $^1\text{H}_2:^2\text{H}_1$ ), respectively, that were in good agreement with experimental values. Employing these optimized parameters, independent MD simulations of the remaining three disaccharides (**3-5**) were performed.

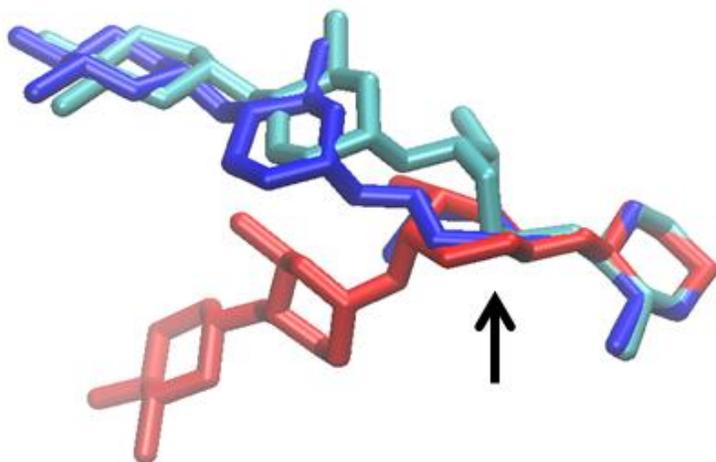
### 3.4.2 GAG structural analysis

#### 3.4.2.1 GAG tetrasaccharides

The conformational properties of two synthetic GAG tetrasaccharides, differing only in the N-substituent (an acetyl group in **6** and a sulfate in **7**, **Figure 3.3**) were characterized by NMR. Homonuclear  $^3J_{\text{HH}}$ -couplings were used to determine the ring pucker states and populations, while NOEs were collected to further characterize the global 3D shape of these GAGettes. 3D models, and MD simulations, of oligosaccharides **6** and **7** were developed to determine the ring pucker populations. These static models show that altering the IdoA ring shape will have change the 3D geometry of the polymer (**Figure 3.4**).



**Figure 3.3** The core structure of the GAG tetrasaccharides (**6-8**) in sequence and chemical structure forms. Labels **C-F** are used to identify the monosaccharide residue components.



**Figure 3.4** Shown are the three dominant L-IdoA ring pucker shapes, indicated by a black arrow, and the impact they have on GAG 3D structure. The  ${}^1C_4$  shape is indicated in cyan,  ${}^4C_1$  in blue and  ${}^2S_0$  in red when aligned to the reducing terminal sugar (far right). These models were obtained from MD simulations in which glycosidic linkages are near the dominant conformation and each linkage is within  $\pm 5^\circ$  of each other.

The  $J$ -couplings for GlcA (residue **C**) and GlcNx ( $x = \text{Ac}$  or  $\text{SO}_3^-$ , residues **D** or **F**, respectively) are presented in **Table 3.2**. All experimental  ${}^3J_{\text{HH}}$ -couplings for residues **C**, **D** and **F** were within 1 Hz of MD simulation values, except for the  $\text{H}_1$ - $\text{H}_2$  coupling in **C**, for which simulations predict a coupling that is 2 Hz larger than experiment. However, the MD simulations for **C** only sampled the  ${}^4C_1$  ring conformation. In the  ${}^4C_1$  conformation, protons  $\text{H}_1$  and  $\text{H}_2$  in the  $\alpha$ -GlcA ring are *anti*- to each other, leading to a large  $J$ -coupling, whereas in the  ${}^1C_4$  conformation they would be *gauche*, leading to a minimal coupling. Thus the difference between the NMR and MD values for this  $J$ -coupling suggests that the non-reducing terminal residue adopts multiple ring conformations. Computed intra-residue signal-averaged distances (described in the 3.3.5) correlate well with distances derived from experimental NOEs (supplementary **Table 7.7**); thus it is also possible that the 2 Hz coupling difference may be also

be the result of approximation errors in the Karplus equation. Otherwise, MD and NMR couplings for **6** and **7** are indicative of a  ${}^4C_1$  chair conformation for residues **C**, **D** and **F**.

**Table 3.2** Dependence of NMR and MD tetrasaccharide  ${}^3J$ -couplings on IdoA ring states.

	<b>6</b>				<b>7</b>			
	Theoretical		Observed		Theoretical		Observed	
Residue: ${}^3J$ -coupling	${}^1C_4$	${}^2S_0$	${}^4C_1$	NMR	${}^1C_4$	${}^2S_0$	${}^4C_1$	NMR
<b>C</b> : H <sub>1</sub> -H <sub>2</sub>	9.8	9.8	9.8	7.9	9.9	9.8	9.8	7.9
<b>D</b> : H <sub>1</sub> -H <sub>2</sub>	3.4	3.6	3.5	3.6	3.3	3.3	3.5	3.6
<b>D</b> : H <sub>2</sub> -H <sub>3</sub>	10.1	10.0	10.0	10.6	10.1	10.1	10.0	10.2
<b>E</b> : H <sub>1</sub> -H <sub>2</sub>	1.7*	8.1*	10.1*	0.9	1.7*	8.1*	10.1*	3.2
<b>E</b> : H <sub>2</sub> -H <sub>3</sub>	1.7*	10.1*	10.1*	-	1.7*	10.1*	10.1*	5.9
<b>E</b> : H <sub>3</sub> -H <sub>4</sub>	2.0*	7.5*	10.1*	-	2.0*	7.5*	10.1*	3.7
<b>E</b> : H <sub>4</sub> -H <sub>5</sub>	3.3*	5.2*	3.4*	2.4	3.3*	5.2*	3.4*	2.7
<b>F</b> : H <sub>1</sub> -H <sub>2</sub>	3.5	3.5	3.5	3.6	3.5	3.5	3.5	3.5
<b>F</b> : H <sub>2</sub> -H <sub>3</sub>	10.0	10.0	10.0	-	10.0	10.0	10.0	10.2
<b>F</b> : H <sub>4</sub> -H <sub>5</sub>	10.0	10.0	10.0	-	10.0	10.0	10.0	9.5
<b>F</b> : H <sub>5</sub> -H <sub>6</sub>	2.4	2.2	2.3	-	2.3	2.1	2.3	4.6
<b>F</b> : H <sub>5</sub> -H <sub>6</sub> '	1.8	1.6	1.6	-	1.8	1.6	1.6	2.4
<b>C-D</b> : $\phi$ H <sub>1</sub> -C <sub>4</sub>	3.1	3.1	3.2	-	3.0	3.1	3.1	-
<b>C-D</b> : $\psi$ C <sub>1</sub> -H <sub>4</sub>	5.1	5.1	5.3	-	5.1	5.1	5.1	-
<b>D-E</b> : $\phi$ H <sub>1</sub> -C <sub>4</sub>	3.1	3.3	3.2	-	3.8	4.1	3.5	-
<b>D-E</b> : $\psi$ C <sub>1</sub> -H <sub>4</sub>	4.0	3.5	4.7	-	5.0	4.5	4.9	-
<b>E-F</b> : $\phi$ H <sub>1</sub> -C <sub>4</sub>	3.5	3.3	3.4	-	3.3	3.4	3.4	-
<b>E-F</b> : $\psi$ C <sub>1</sub> -H <sub>4</sub>	4.7	5.1	5.0	-	5.0	4.9	5.0	-

\*Assigned from ideal geometries of static IdoA rings due to ring restraints in the MD simulations.

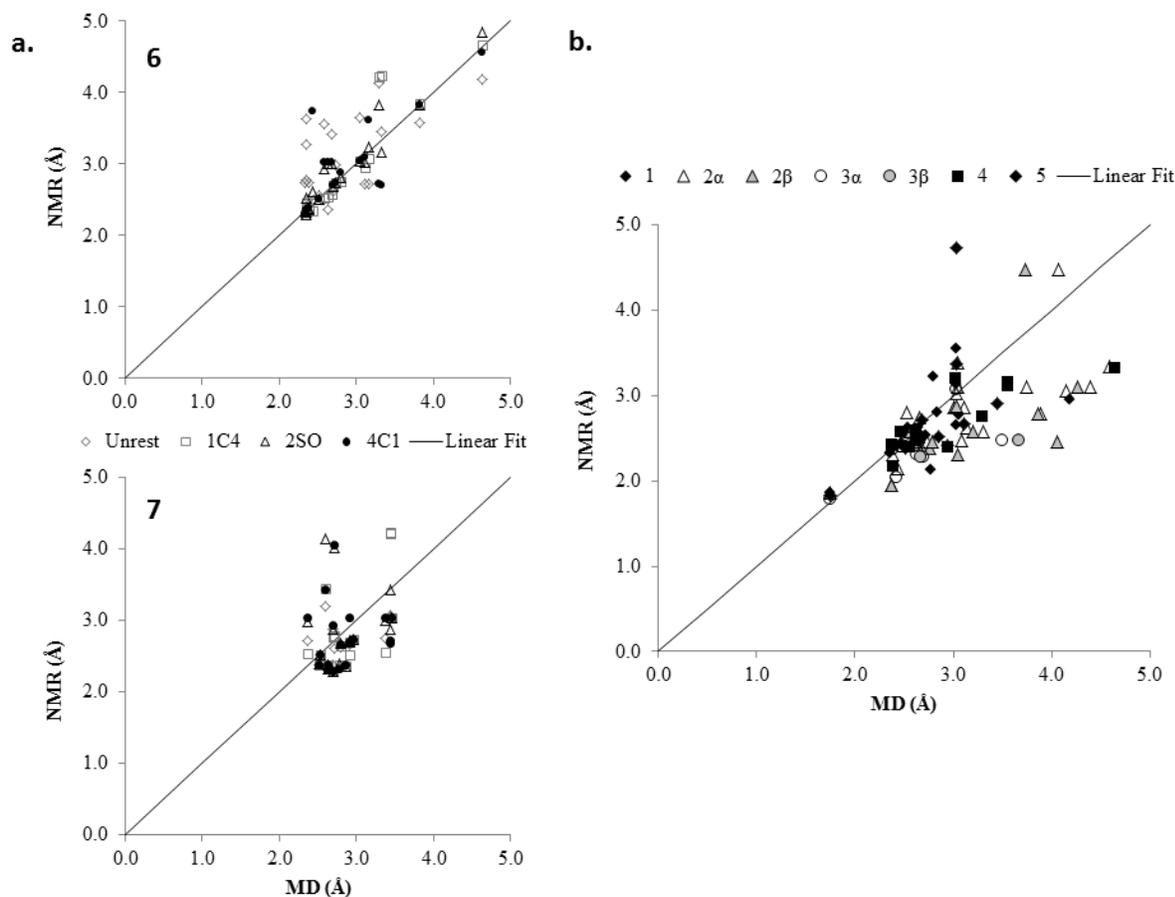
Characterization of the IdoA (residue **E**) ring structure required a best-fit analysis of contributions from multiple ring forms to the experimental coupling data. In both **6** and **7**, the  $^3J_{\text{H4H5}}$  coupling is insensitive to the ring conformation, and was not included in the fitting analysis (**Table 3.2**).

Three experimental  $^3J_{\text{HH}}$ -couplings were available to characterize the ring structure of **7**. Least-squares fitting of the couplings revealed the most probable population distribution is 61:39:0 ( $^1\text{C}_4$ : $^2\text{S}_0$ : $^4\text{C}_1$ ) while the population with lowest absolute net error was 69:31:0 suggesting a population range between 60-70%  $^1\text{C}_4$ , 30-40%  $^2\text{S}_0$  and negligible  $^4\text{C}_1$ . The absence of the  $^4\text{C}_1$  state is supported by NMR data for similar GAGs[9,104,109,122,123], which indicate this state to be the least populated of the three, if present at all. A search of the Protein Databank (PDB)[124] revealed the  $^1\text{C}_4$  (73%) and  $^2\text{S}_0$  (24%) states to be the dominant forms of IdoA.

For **6**, only the  $^3J_{\text{H1H2}}$  coupling was experimentally observed (0.9 Hz); the small value of this coupling indicates that the  $^1\text{C}_4$  state, calculated at 1.7 Hz, dominates the IdoA geometry since both the  $^2\text{S}_0$  and  $^4\text{C}_1$  couplings are large (>8 Hz).

The global 3D structure of a polysaccharide may be characterized by the glycosidic torsion angles, which may be derived from *trans*-glycosidic heteronuclear  $^3J_{\text{CH}}$ -couplings. Although these couplings were not accessible in the present study, theoretical values derived from MD simulation are included in **Table 3.2**, for reference. The  $^3J_{\text{CH}}$ -couplings for **6** and **7** suggest that the glycosidic linkage of these GAGettes vary by only 1-1.5 Hz with the IdoA ring conformation. This narrow range of coupling values suggests that sulfation has a marginal effect on the glycosidic linkage conformation.

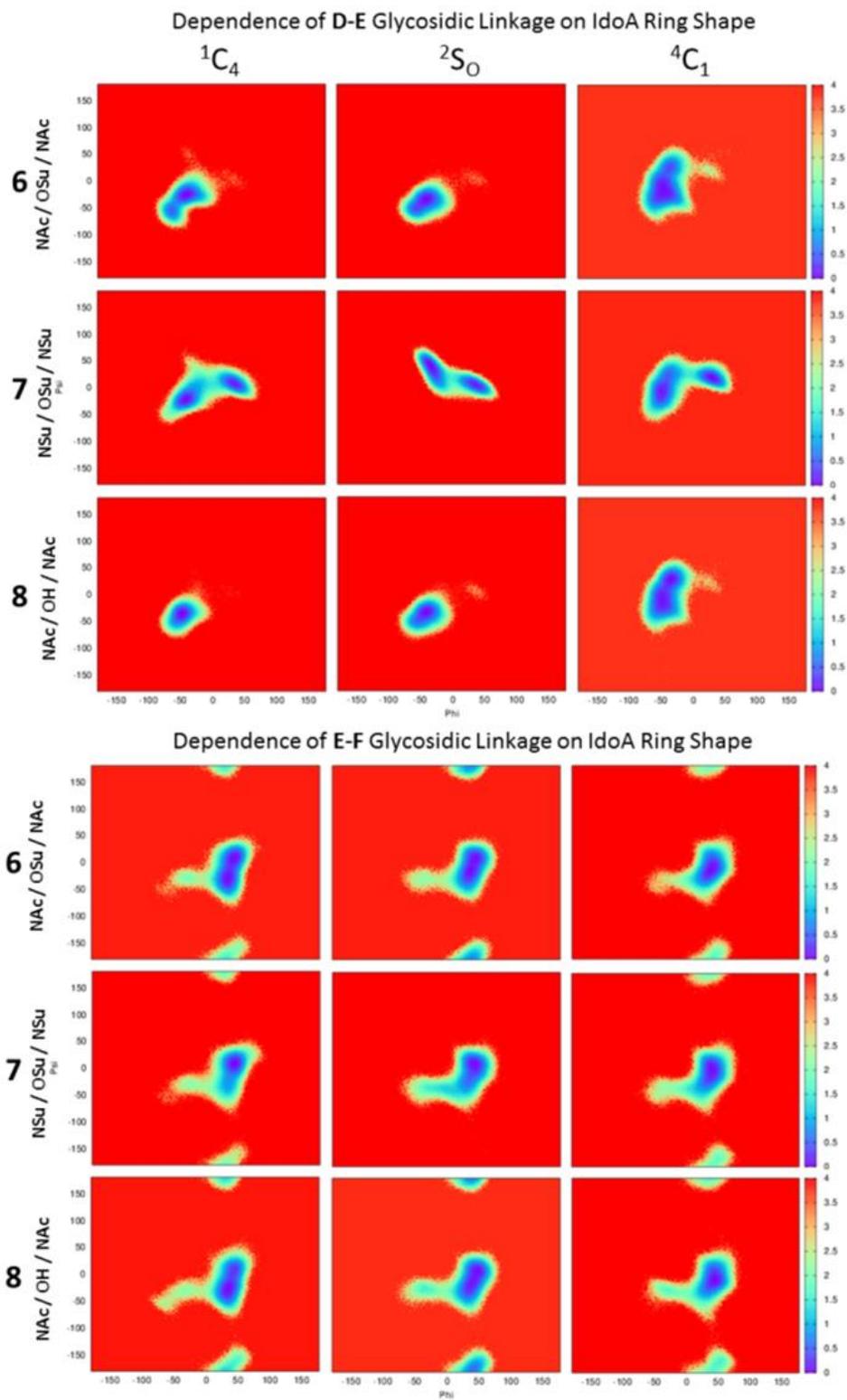
In lieu of  $^3J_{\text{CH}}$ -couplings, inter-residue NOE data were collected and compared to simulation data. The majority of the predicted inter-proton distances agree to within 0.5 Å of experimentally-derived distances, indicating that **6** exists predominantly in the  $^1\text{C}_4$  conformation, and **7** as mixture of  $^1\text{C}_4$  and  $^2\text{S}_0$  conformations (**Figure 3.5a**). A disagreement of just over 1.0 Å between experimental and computational NOEs was noted between protons  $\text{H}_1^{\text{E}}-\text{H}_3^{\text{F}}$  of **7**. The NOE between these protons is strong, equivalent to a distance of approximately 2.6 Å, however, these protons were never closer than 3.4 Å in the MD simulations of the three ring states. Interestingly, the distance between the neighboring *trans*-glycosidic proton pair  $\text{H}_1^{\text{E}}-\text{H}_4^{\text{F}}$  (2.6 Å), agrees well with MD predictions (2.3 Å). Without additional experimental data it is not possible to determine the cause of the disagreement, however, it is notable that there is a slight structural difference in the aglycone groups between the experimental compound (n-pentylamine) and computational models (methyl).



**Figure 3.5** Correlation between experimental and computational NOE's for the tetrasaccharides, molecules **6** and **7** (a). Correlation between experimental and computational NOE values for each of the disaccharides, molecules **1-5** (b). Both epimers for disaccharides **2** and **3** and shown as white and gray filled triangles and circles, respectively. Each of the restrained simulations is indicated for the tetrasaccharides along with the nOe's for the iduronate ring state mixtures for **7**, 60:40:0 and 70:30:0. Iduronate (residue **E**) NOE's are not included as these were not dynamic distances in the simulation. A complete listing of the experimental and computational NOE's is available in supplementary **Table 7.6** & supplementary **Table 7.7**.

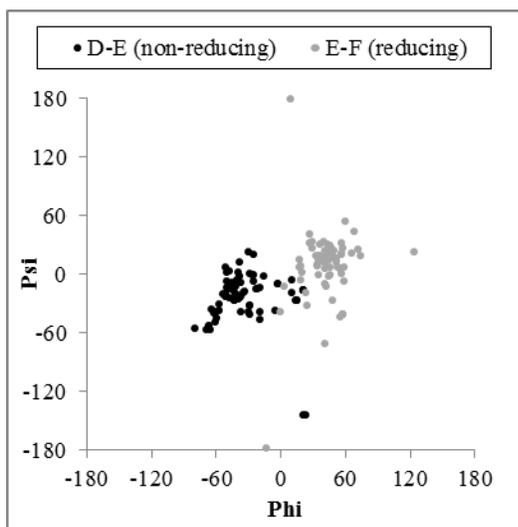
To further understand the role of IdoA shape on altering the glycosidic linkage, MD simulations were performed on each of the three ring states ( ${}^1C_4$ ,  ${}^2S_0$ , and  ${}^4C_1$ ) for **6** and **7**, as well as their de-sulfated GAGette (**8**). The glycosidic linkages between the non-reducing GlcNx (**D**) and IdoA (**E**) showed the most substantial dependence on ring shape and sulfation pattern (top half of **Figure 3.6**). For the non-sulfated tetrasaccharide (**8**) the  ${}^1C_4$  and  ${}^2S_0$  IdoA structures yielded similar glycosidic linkage profiles for the **D-E** linkage (**Figure 3.6**), with single a

dominant geometry around  $\varphi \approx -50^\circ$ ,  $\psi \approx -50^\circ$ . The introduction of a 2-*O*-sulfate moiety into the  ${}^1C_4$  conformation of IdoA (**6**) resulted in the creation of a new state characterized by  $\varphi \approx -40^\circ$ ,  $\psi \approx -30^\circ$ , but had no effect when IdoA was in the  ${}^2S_0$  conformation. *N*-sulfation of **6** (**7**) resulted in the formation of two stable glycosidic linkage rotamers for each ring conformation of IdoA. In the case of the  ${}^1C_4$  and  ${}^4C_1$  ring forms, the dominant conformation was the same as seen in both the non-sulfated and 2-*O*-sulfated systems, with a new minor state defined by a positive value of  $\varphi \approx 40^\circ$ . In the case of the  ${}^2S_0$  ring form, *N*-sulfation further perturbed the linkage conformation by altering the dominant conformation to a state characterized by a positive value of  $\psi \approx 40^\circ$ . Thus, for these **D-E** linkages, only *N*-sulfation appears to markedly perturb the rotameric preferences of the glycosidic linkage.



**Figure 3.6** The Boltzmann relative energy heat maps for the glycosidic linkages ( $\phi, \psi$ ) between residues D-E (top) and E-F (bottom) for polysaccharides 6-8.

The non-reducing (**C-D**) and reducing (**E-F**) terminal glycosidic linkages are largely insensitive to either sulfation or IdoA ring conformation (**Figure 3.6** and supplementary **Figure 7.2**). Each of these linkages populate the same dominant rotamer ( $\phi \approx 45^\circ$ ,  $\psi \approx 0^\circ$ ) and two minor rotamers ( $-45^\circ$ ,  $-15^\circ$ ) and ( $45^\circ$ ,  $180^\circ$ ).



**Figure 3.7** The distribution of  $\phi$  and  $\psi$  angles from GAG crystal structures with analogous **D-E-F** sequences to **6-8**.

The range of MD conformations presented in **Figure 3.6** may be compared to experimental crystallographic data for GAG sequences with non-terminal IdoA (**Figure 3.7**). It should be noted that there are no free GAG sequences, or protein co-complexes, containing GAGs identical to **6**, **7**, or **8** in the protein database. However, each of the structures reported in **Figure 3.7** contains at least the core GlcNS- $\alpha$ -1,4-IdoA- $\alpha$ -1,4-GlcNS- $\alpha$  sequence with varying degrees of *O*-sulfation. All but one of the seventy-four linkages examined have 2-*O* sulfated IdoA making them most similar to **7**. For both the analogous non-reducing and reducing

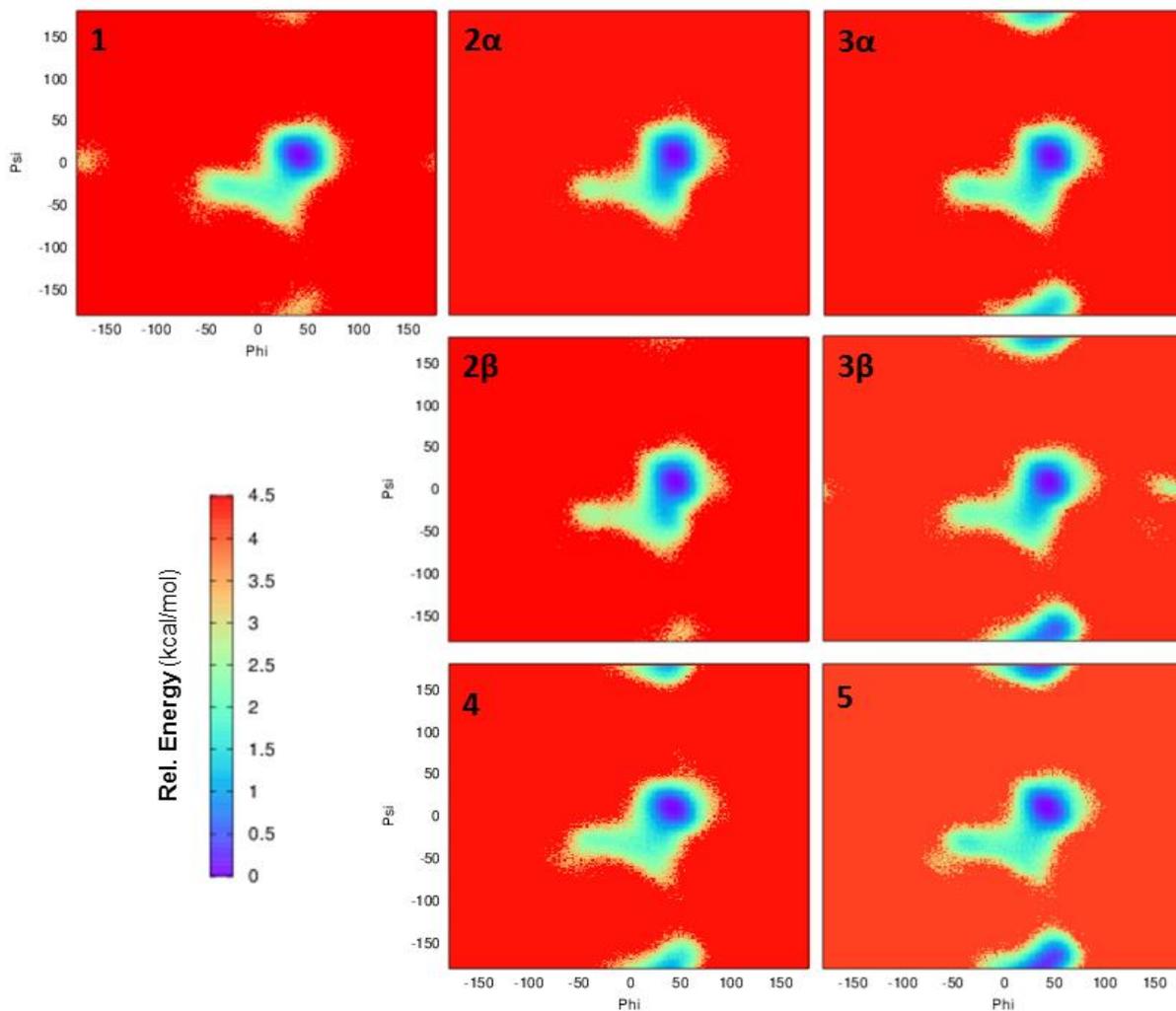
terminal linkages, the dominant rotamers predicted by MD for **7** were found to be the most commonly observed in the crystals data.

### 3.4.2.2 $\Delta$ 4,5-unsaturated uronate disaccharides

Five disaccharide GAGettes, molecules **1-5** (**Figure 3.2a**), containing a non-reducing terminal  $\Delta$ 4,5-unsaturated uronate (residue **A**) and a reducing terminal GlcNx ( $x=Ac, SO_3^-,$  or  $H_3^+$ ; residue **B**) were analyzed using NMR and MD simulations to determine ring conformations and glycosidic linkage geometry profiles. Ring state populations for **A** from MD simulations agree well with those derived from NMR  $J$ -couplings (**Table 3.1**). Residue **A** shows a preference for the  ${}^1H_2$  state in all but the non-sulfated  $\Delta$ 4,5 $uu$  disaccharide, molecule **4**. Meanwhile, conformation analysis of  ${}^3J$ -couplings for all **B** residues (data not shown) was consistent with only the  ${}^4C_1$  conformation. It should be noted that NMR data for molecules **2** and **3** were collected as a mixture of both anomers of GlcNx. The MD  ${}^3J$ -couplings for **2** and **3** showed that the ring state populations of **A** and **B** are unaffected by the orientation of the reducing terminal aglycone.

Ring structures were also probed using intra-residue NOE calculations. Overall, intra-residue NOE-derived distances correlate well with experiment in models **1, 3, 4** and **5** however an overall agreement error of 0.5 Å was noted for **2**. This is principally due to the the  $H_1^B-H_4^B$  interaction predicted by NMR as 3.1 and 2.5 Å ( $\alpha$  and  $\beta$ , respectively) and as 4.2 and 4.1 Å in the MD simulation. The distance for this interaction has a narrow range of fluctuation in all the MD simulations of GlcNx as the structure remains in the  ${}^4C_1$  conformation. The overall agreement of

the NOE data with computational predictions supports the ring structure predictions made using the J-couplings.



**Figure 3.8** MD solution relative free energies for glycosidic rotamers based on Boltzmann weighted populations at 300K. Most regions are not sampled (red colors) and have  $>4$  kcal/mol relative energy. All models show a similar global minimum around  $50^\circ$  and  $0^\circ$  for  $\phi$  and  $\psi$ , respectively.

The 3D structure of the disaccharides also depends upon the conformations adopted by the glycosidic linkage. The heat maps in **Figure 3.8** show the global minimum from the MD simulations around  $\varphi \approx 50$ ,  $\psi \approx 0$  for all the linkages. Sugars **3**, **4** and **5** show a second stable state near  $\varphi \approx 50$ ,  $\psi \approx 180$ , however, this state is infrequently populated in **1** and **2**. NMR characterization of the glycosidic linkage was limited to analysis of NOE distances measured for several inter-residue proton-proton interactions. The  $H_1^A-H_3^B$  and  $H_1^A-H_5^B$  NOE's were used to determine the approximate orientation of the glycosidic linkage (**Figure 3.2a**) as is shown by computing these two proton-proton distances relative to  $\varphi$  and  $\psi$  torsion angles. Comparison of the  $H_1^A-H_3^B$  and  $H_1^A-H_5^B$  NOE's (**Table 3.1**) shows reasonable agreement ( $\pm 0.5 \text{ \AA}$ ) with experiment in **4** and **5** yet in the MD simulations of **1** and **2** this distance is longer by more than  $1.0 \text{ \AA}$ . This disagreement may partly arise from under sampling the  $180^\circ$   $\psi$ -state which has shorter distances for both proton pairings as can be seen on a plot of the relationship between the NOE distance and  $\varphi$ ,  $\psi$  (**Figure 3.2b**). If this state was sampled in **1** and **2** as frequently as in the other disaccharides then these NOE's would be more consistent with experimental observations. Additionally, integration inaccuracies or fast molecular tumbling times also increase experimental disagreement. Despite these two cases, cross-residue signal-averaged distances agree reasonably well with experiment, having an overall error of  $< 0.5 \text{ \AA}$  in all but disaccharide **2** (**Figure 3.5** & supplementary **Table 7.5** & supplementary **Table 7.6**).

### 3.5 Discussion

Sulfation patterns in GAGs have been previously shown to alter the ring conformations of IdoA and  $\Delta 4,5$ -unsaturated uronates, a trend also noted in analysis of NMR observables for the GAGettes described here. Yet the effects of sulfation, particularly N-sulfation, have been

shown in this work to also impact the glycosidic rotamer profiles. The challenge is now to experimentally characterize these states as the predicted  $^3J_{\text{CH}}$ -couplings for the  $\psi$  glycosidic torsion angle varies by only a few Hertz and the NOE intensities are not sensitive enough to distinguish the changes. A survey of the PDB for analogous IdoA-containing sequences does show a similar sampling to the glycosidic patterns observed in the simulation work. Still, the effects of sulfation do not always substantially alter the glycosidic rotamer profiles adjacent to IdoA as was observed in the transition from a non-sulfated sequence to a 2-*O*-sulfated IdoA.

The most notable effects of sulfation for these GAGs was on ring geometries for the IdoA and  $\Delta$ 4,5-unsaturated uronate sugars. Decoupling of the NMR data showed that the tetrasaccharide with only 2-*O*-sulfated IdoA (**6**) exclusively favored the  $^1C_4$  conformation while adding *N*-sulfating to adjacent glucosamines introduces a substantial  $^2S_0$  population. While the ring flip dynamics were not captured by this work, long timescale simulations of IdoA using GLYCAM have previously shown experimentally-consistent ring populations[28]. This work does show the ring shape dynamics of  $\Delta$ 4,5-unsaturated uronates on timescales which allow direct parameterization of the ring populations. Analysis of NMR  $J_{\text{HH}}$ -couplings shows that conformation populations of the  $\Delta$ 4,5-uu ring is largely insensitive to the adjacent sulfation patterns and the *N*-substituent; only 2-*O*-sulfation (**1**) of  $\Delta$ 4,5-uu altered the favored geometry. These results were mirrored by MD simulations of all five disaccharides.

The new parameter set for GAGs containing IdoA, sulfate, and  $\Delta$ 4,5-unsaturated uronate parameters have been developed to augment existing GLYCAM06 parameters. The identification of a transferable sulfate model and singular charge model for  $\Delta$ 4,5-unsaturated uronate avoids the need for developing ensemble-averaged charges for multiple attachment points and ring conformations, respectively. This allows easy integration into the automated

carbohydrate builders like the GLYCAM webtool[32] making GAG model building for simulations substantially more efficient. Additionally, the inclusion of parameters for unsaturated carbohydrate analogs like  $\Delta$ 4,5-unsaturated uronates allows a wide variety of carbohydrate transition state mimics to be modeled using GLYCAM. These new unsaturated parameters are capable of modeling other unsaturated sugars including the sialic acid transition state analogs which are often co-crystallized in sialidases. These parameters and related structure files are available for download from the GLYCAM website ([www.glycam.org](http://www.glycam.org)) and are included in the new GLYCAM webtool which includes a GAG sequence builder[32].

### **3.6 Acknowledgements**

The authors thank the Resource for Integrated Glycotechnology at the University of Georgia which is supported by grants from the National Center for Research Resources (5 P41 RR005351) and the National Institute of General Medical Sciences (8 P41 GM103390) from the National Institutes of Health.

**4. COMPUTATIONAL SCREENING OF THE HUMAN TF-GLYCOME PROVIDES A  
STRUCTURAL DEFINITION FOR THE SPECIFICITY OF THE ANTI-TUMOR  
ANTIBODY JAA-F11<sup>3</sup>**

---

<sup>3</sup> Matthew B Tessier, Oliver C Grant, Jamie Heimburg-Molinaro, David Smith, Snehal Jadey, Andrew M. Gulick, John Glushka, Susan L. Deutscher, Kate Rittenhouse-Olson & Robert J. Woods. 2013. *PLoS One* 8(1):e54874. Reprinted here with permission of the publisher.

## 4.1 Abstract

Recombinant antibodies are of profound clinical significance, yet, anti-carbohydrate antibodies are prone to undesirable cross-reactivity with structurally related-glycans. Here we introduce a new technology called Computational Carbohydrate Grafting (CCG), which enables a virtual library of glycans to be assessed for protein binding specificity, and employ it to define the scope and structural origin of the binding specificity of antibody JAA-F11 for glycans containing the Thomsen-Friedenreich (TF) human tumor antigen. A virtual library of the entire human glycome (GLibrary-3D) was constructed, from which 1,182 TF-containing human glycans were identified and assessed for their ability to fit into the antibody combining site. The glycans were categorized into putative binders, or non-binders, on the basis of steric clashes with the antibody surface. The analysis employed a structure of the immune complex, generated by docking the TF-disaccharide (Gal $\beta$ 1-3GalNAc $\alpha$ ) into a crystal structure of the JAA-F11 antigen binding fragment, which was shown to be consistent with saturation transfer difference (STD) NMR data. The specificities predicted by CCG were fully consistent with data from experimental glycan array screening, and confirmed that the antibody is selective for the TF-antigen and certain extended core-2 type mucins. Additionally, the CCG analysis identified a limited number of related putative binding motifs, and provided a structural basis for interpreting the specificity. CCG can be utilized to facilitate clinical applications through the determination of the three-dimensional interaction of glycans with proteins, thus augmenting drug and vaccine development techniques that seek to optimize the specificity and affinity of neutralizing proteins, which target glycans associated with diseases including cancer and HIV.

## 4.2 Introduction

Aberrant glycosylation is a hallmark of many diseases, including cancer[125], and can therefore provide a basis for disease diagnosis and staging, and may potentially be exploited for therapeutic intervention[126]. An established carbohydrate-based cancer marker is the Thomsen-Friedenreich (TF) antigen ( $\text{Gal}\beta 1\text{-3GalNAc}\alpha$ ), which is typically found O-linked to serine or threonine residues. The TF antigen (also known as T antigen) has been associated with several human carcinomas, including those found in the pancreas, colon, and breast, and on this basis has been referred to as a pan-carcinoma marker[127,128,129,130]. TF antigen is concealed from the immune system in normal adult tissues as a result of extension with larger glycan chains[129,131]. In cancer, the cellular glycosylation machinery may be disrupted, leading to truncation of these chains and exposure of the TF antigen[132]. Here, we examine the specificity of a potentially diagnostic and therapeutic monoclonal antibody (mAb JAA-F11)[133] that was raised against the TF antigen[134]. JAA-F11 preferentially binds to tumor tissue over normal[34], and in vivo it enhances survival and decreases metastasis in the mouse 4T1 metastatic model[133], indicating a potential for this mAb to be used, after humanization, in cancer patient therapy.

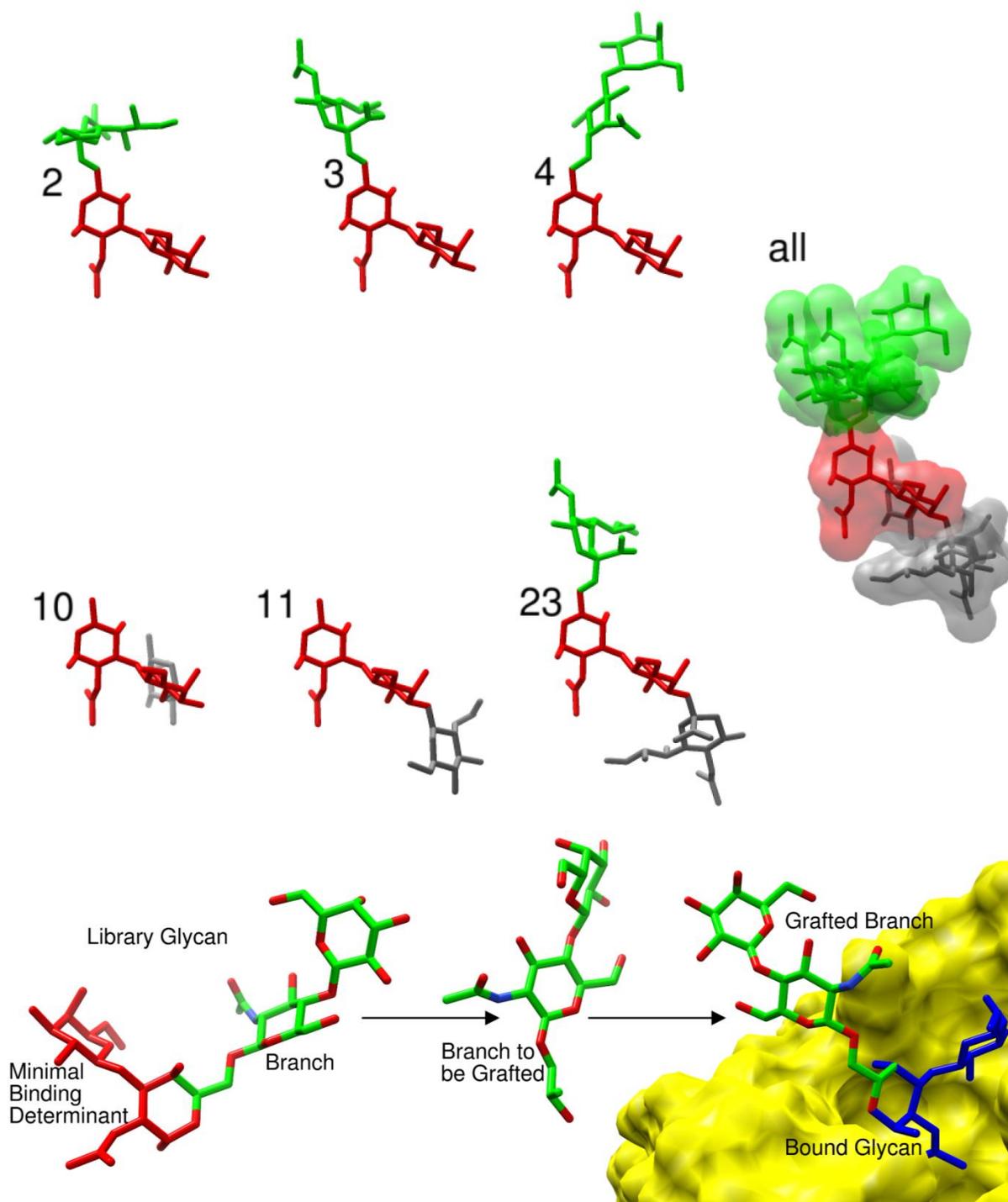
The potential also exists for anti-carbohydrate mAbs, such as JAA-F11, to be used as diagnostic agents, however, the diversity of glycans present in eukaryotic organisms leads to the possibility for cross-reactivity among structurally similar carbohydrates, which may nevertheless have unrelated biological roles. Thus, it is particularly critical to determine the specificity of any reagent proposed for use in glycan-based disease-marker detection[4,15,135]. Over the past decade, glycan microarray screening has gained wide-spread popularity as a technique for assessing carbohydrate-binding specificity. The largest glycan microarrays currently contain on

the order of 600 members[5], enabling rapid assessment of binding specificity, and requiring far less protein and carbohydrate for the analysis than would be necessary for more detailed affinity measurements. Despite these advances the human glycome is far more diverse than even the largest experimental glycan array[15]. Thus, experimental screening of the entire human glycome is not yet feasible, leaving the potential for cross-reactive binding to go undetected. Moreover, although glycan array screening can provide specificity data for many glycans simultaneously, the data are difficult to relate directly to binding affinities, and do not provide insight into the structural mechanisms of a binding interaction.

Structural information is traditionally provided by experimental methods such as X-ray crystallography and NMR spectroscopy. Despite the importance of 3D structural data in defining structure-function relationships, neither NMR spectroscopy nor protein crystallography can be considered high throughput methods in this role. Additionally, both techniques face significant challenges when applied to the characterization of certain classes of protein-ligand complexes, such as those formed between antibodies and large glycans[136,137]. Issues such as glycan flexibility, structural heterogeneity, and challenges in the synthesis, isolation or crystallization of complex glycans contribute to difficulties in such studies. In addition, to enhance the likelihood of crystallization, and to facilitate NMR data interpretation, such studies typically employ only di- or tri-saccharide fragments, rather than the whole, intact glycan.

Here we present a new technology, Computational Carbohydrate Grafting (CCG) that is complementary to glycan array screening, NMR spectroscopy and crystallography. CCG leverages available 3D structural data for carbohydrate-protein complexes, with virtual glycan library screening to generate 3D models of glycan-protein complexes. We use CCG to predict the binding specificity of JAA-F11, and demonstrate that the theoretical predictions are fully

consistent with experimental specificity data for the same antibody generated by screening against an experimental glycan array. In contrast to traditional virtual screening, which would attempt to dock the entire glycan into the binding site, CCG splices the glycan branches onto the appropriate positions in a fragment (TF antigen in this case, **Figure 4.1**) of the glycan present in a protein-carbohydrate complex. The 3D orientations of the grafted branches relative to the fragment (or minimal binding determinant) are generated on the basis of established carbohydrate conformational preferences[138]. Quantification of any steric overlaps between the grafted glycan and the protein surface enables discrimination between potential binding partners and non-binders. It should be noted that the CCG analysis does not rank the ligands in terms of theoretical affinities. Nevertheless, the power of the CCG method is the rapid identification of a subset of putative binders, which can subsequently be examined in more detail either experimentally or theoretically. By grafting the virtual glycan structures onto a bound carbohydrate motif, a level of speed and accuracy in the prediction of the 3D structures is achieved that would otherwise be impossible using either traditional virtual screening or experimental techniques alone. In addition, CCG facilitates the screening of vast libraries of glycans that can encompass the entire known human glycome, as well as synthetic or hypothetical structures, extending the CCG screening capability far beyond the scope of current experimental glycan microarrays.



**Figure 4.1** An illustration of Computational Carbohydrate Grafting (CCG) method applied to predict binding conformations of TF-containing glycans binding to the JAA-F11 antibody. **Upper.** Examples of glycans that bind to JAA-F11: Neu5Ac $\beta$ 2-6(Gal $\beta$ 1-3)GalNAc $\alpha$  (**2**); Neu5Ac $\alpha$ 2-6(Gal $\beta$ 1-3)GalNAc $\alpha$  (**3**); Gal $\beta$ 1-4GlcNAc $\beta$ 1-6(Gal $\beta$ 1-3)GalNAc $\alpha$  (**4**), as well as non-binding sequences (**middle**): Fuca1-2Gal $\beta$ 1-3GalNAc $\alpha$  (**10**); GlcNAc $\beta$ 1-3Gal $\beta$ 1-3GalNAc $\alpha$  (**11**); and Neu5Ac $\alpha$ 2-6(Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3)GalNAc $\alpha$  (**23**), showing the minimal binding

determinant in red, the tolerated glycan branches in green, and the disallowed branches in grey. Also presented are the combined solvent-accessible surfaces from a superimposition of the sequences based on aligning the minimal determinant. **Lower.** In the grafting process branches from TF-containing glycans in the library are excised and spliced onto the bound minimal determinant. The grafted branches are then assessed for steric clashes with the antibody surface. This process is illustrated for the grafting of the glycan branch Gal $\beta$ 1-4GlcNAc $\beta$ 1-6 (green carbon frame) from **4** onto the TF antigen in the JAA-F11 binding site (yellow solvent-accessible surface). Figures generated with Chimera[63].

A CCG analysis of JAA-F11 was performed against a virtual array of 1,182 TF-containing human glycans, extracted from a library of glycan 3D structures (GLibrary-3D), comprising the known human glycome as present in the GlycomeDB database ([www.glycome-db.org](http://www.glycome-db.org))[139]. The virtual screening employed a crystal structure of the Fab (**Table 4.1**); however, co-crystals of the Fab – TF antigen complex proved to be elusive, thus computational docking of the minimal motif (TF disaccharide) was performed. Experimental support for the predicted orientation of the TF disaccharide in the binding site was provided by data from saturation-transfer difference (STD) NMR experiments on the mAb-TF complex (see supporting information: **7.2.1 Supplementary Results, Figure 7.3 & Figure 7.4**). The docked orientation of the minimal determinant was consistent with the STD NMR data, and subsequent CCG analysis employing this complex fully explained the observed binding specificity of this mAb from experimental glycan array screening. In addition, the CCG screening led to the prediction of a small number of putative binders that are not present on the current experimental arrays. Taken together, the CCG analysis and experimental glycan array data confirm the specificity of JAA-F11 for TF antigen and certain extended core-2 type mucins.

**Table 4.1** Data collection and refinement statistics for Fab JAA-F11

<b>Data collection</b>	
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	94.2, 94.2, 95.0
$\alpha$ , $\beta$ , $\gamma$ (°)	90°, 90°, 90°
Resolution (Å)	30.0–2.1
<i>R</i> <sub>merge</sub> (%)	9.6 (43.3) <sup>a</sup>
$\langle I/\sigma(I) \rangle$	9.6 (2.0)
Completeness (%)	99.0 (98.2)
Redundancy	10.5 (9.5)
<b>Refinement</b>	
Resolution (Å)	30.0 – 2.1
Number of reflections	24063
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	18.5 / 26.2 (19.9 / 29.9)
Number of atoms	
Protein	3342
Ligands	20
Water	176
Average <i>B</i> -factors	
Protein (Å <sup>2</sup> )	21.0
Ligands (Å <sup>2</sup> )	54.8
Water (Å <sup>2</sup> )	24.3
RMS deviations	
Bond lengths (Å)	0.02
Bond angles (°)	1.9

<sup>a</sup>Values in parentheses are for highest-resolution shell (2.2–2.1 Å)

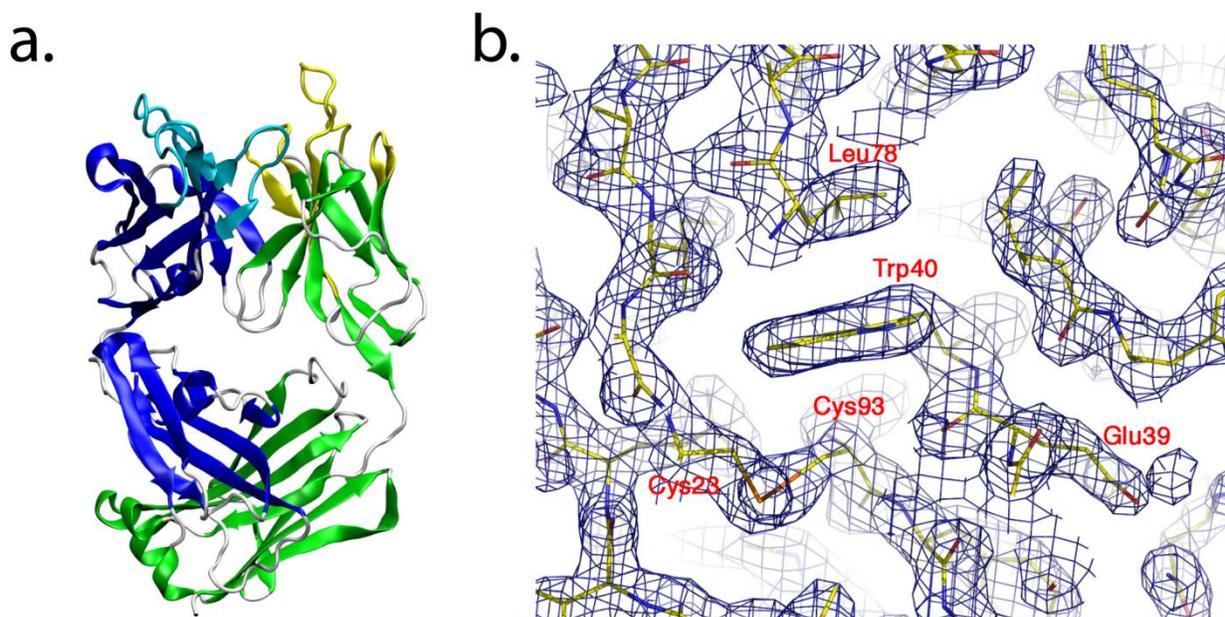
## 4.3 Results

### 4.3.1 Crystal structure of Fab JAA-F11

The unliganded crystal structure of Fab JAA-F11 was determined at 2.1 Å (see **Table 4.1**), and consists of Leu<sup>L1</sup>-Asn<sup>L217</sup> of the variable light chain (V<sub>L</sub>) and Ala<sup>H1</sup>-Arg<sup>H218</sup> of the variable heavy chain (V<sub>H</sub>), as well as 176 water molecules.

A ribbon diagram of the Fab shows that it displays the typical overall fold of a Fab fragment (**Figure 4.2a**). The quality of the electron density for a region of the light chain is also provided (**Figure 4.2b**). Complementarity determining regions (CDRs) in the Fab were assigned to canonical structure class 1 for loops L2 (Lys<sup>L55</sup> to Ser<sup>L61</sup>), L3 (Phe<sup>L94</sup> to Thr<sup>L102</sup>), and H1

(Thr<sup>H31</sup> to His<sup>H35</sup>), while loops L1 (Arg<sup>L24</sup> to Glu<sup>L39</sup>) and H2 (Phe<sup>H50</sup> to Asp<sup>H65</sup>) belong to canonical structures 4 and 2, respectively. CDR H3 consisted of residues Ser<sup>H99</sup> to Phe<sup>H107</sup> and could not be assigned a canonical conformation. Together these CDRs form a canyon shaped [140] binding pocket of 712 Å<sup>3</sup> volume.

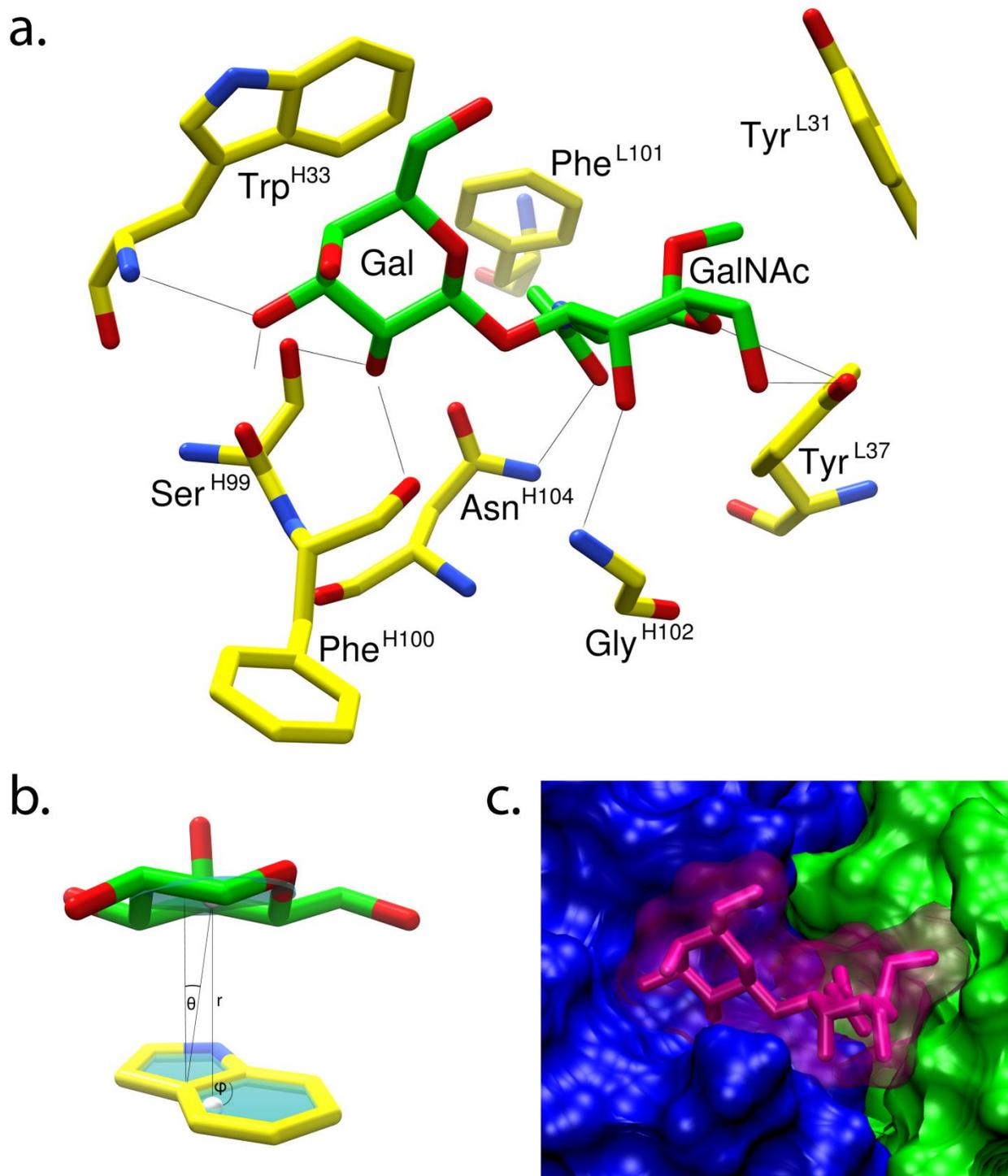


**Figure 4.2** Illustrations of the 3D structure of the JAA-F11 antibody as determined by X-ray crystallography. **a.** The overall JAA-F11 Fab fold in ribbons (heavy chain residues in blue, light chain in green) indicating the CDRs (heavy chain in light blue, light chain in yellow). **b.** Representative electron density of the final map for a region of the light chain; the map is calculated with coefficients of the form  $2F_o - F_c$ , and contoured at 1.2 Å. The conserved disulfide bond between Cys<sup>L23</sup> and Cys<sup>L93</sup> is shown.

#### 4.3.2 3D Model of the Minimal Determinant – Fab complex

Computational docking of the TF disaccharide was performed with AutoDock 3.05[141] to identify possible poses of the antigen in the Fab JAA-F11 binding site. The crystal structure of the Fab and a 3D structure for the TF antigen, obtained from GLYCAM-Web ([www.glycam.org](http://www.glycam.org)) [32], were employed in the docking. During docking, the  $\phi$  and  $\psi$  torsion angles of the glycosidic

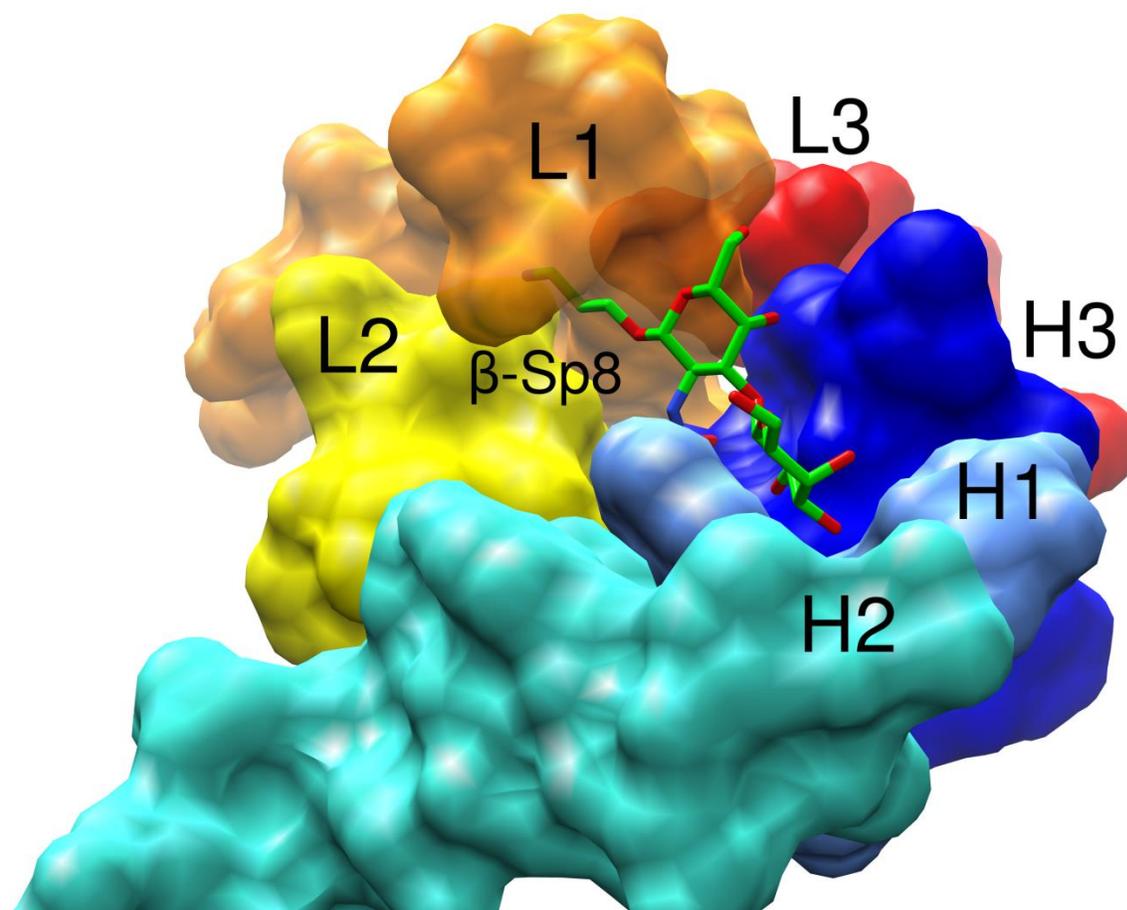
linkages were maintained in the low energy conformation generated by energy minimization with the GLYCAM06 force field[24], while all hydroxyl and C<sub>5</sub>-C<sub>6</sub> bonds were allowed to rotate freely. Fifty poses were obtained (**Table 7.8**) and ranked in terms of predicted interaction energy and pose clustering. Clustering of the docking results was performed, based on placement and orientation of each pose relative to the protein.



**Figure 4.3** The binding interactions predicted from the docked model of the TF-disaccharide bound to JAA-F11. **a.** Validated model of the bound minimal determinant (green carbon frame) in the mAb binding site, hydrogen atoms removed for clarity. Protein residues (yellow carbon frame) involved in hydrogen bonds (black lines) or hydrophobic interactions with the TF antigen are shown. **b.** Depiction of the stacking interaction between the Gal and the Trp<sup>H33</sup> ( $r = 4.0 \text{ \AA}$ ,  $\theta$

= 12.1°,  $\phi$  = 90.3°); the geometry of this interaction is comparable to literature values[142] (lower left). **c.** Solvent accessible surface of the bound minimal determinant (magenta) and the Fab (showing V<sub>H</sub> and V<sub>L</sub> regions in blue and green, respectively).

The most highly populated cluster (see **Table 7.8**) also exhibited the lowest average energy of all clusters and contained the lowest energy pose (pose 1, **Figure 4.3**). In this optimal pose, four hydrogen bonds were formed between the protein and oxygen atoms in the terminal Gal residue; Ser<sup>H99</sup> O<sub>γ</sub> to O-2, Phe<sup>H100</sup> O to O-2, Trp<sup>H33</sup> N to O-3 and Ser<sup>H99</sup> O to O-3. Four additional hydrogen bonds were formed with the GalNAc residue; Asn<sup>H104</sup> N<sub>δ</sub> to the carbonyl oxygen atom of the N-acetyl group, Gly<sup>H102</sup> N to O-4, and Tyr<sup>L37</sup> OH to O-5 and O-6 (**Figure 4.3a**). These data are consistent with earlier conclusions that neither of the primary hydroxyl groups was involved in binding[134]. The presence of an aromatic stacking interaction between the Gal residue and Trp<sup>H33</sup> was also observed, which is a common feature in carbohydrate-protein complexes (**Figure 4.3b**)[143]. Based on this structure for the immune complex, the important observation that this mAb binds α-linked TF antigen, but not β-TF, yielding tumor specificity, could be rationalized on the basis that a β-linkage at the reducing terminus would result in overlaps between the aglycon and several residues in CDR L1, including, Tyr<sup>L31</sup>, Ser<sup>L32</sup>, and Asn<sup>L33</sup> (**Figure 4.4**).

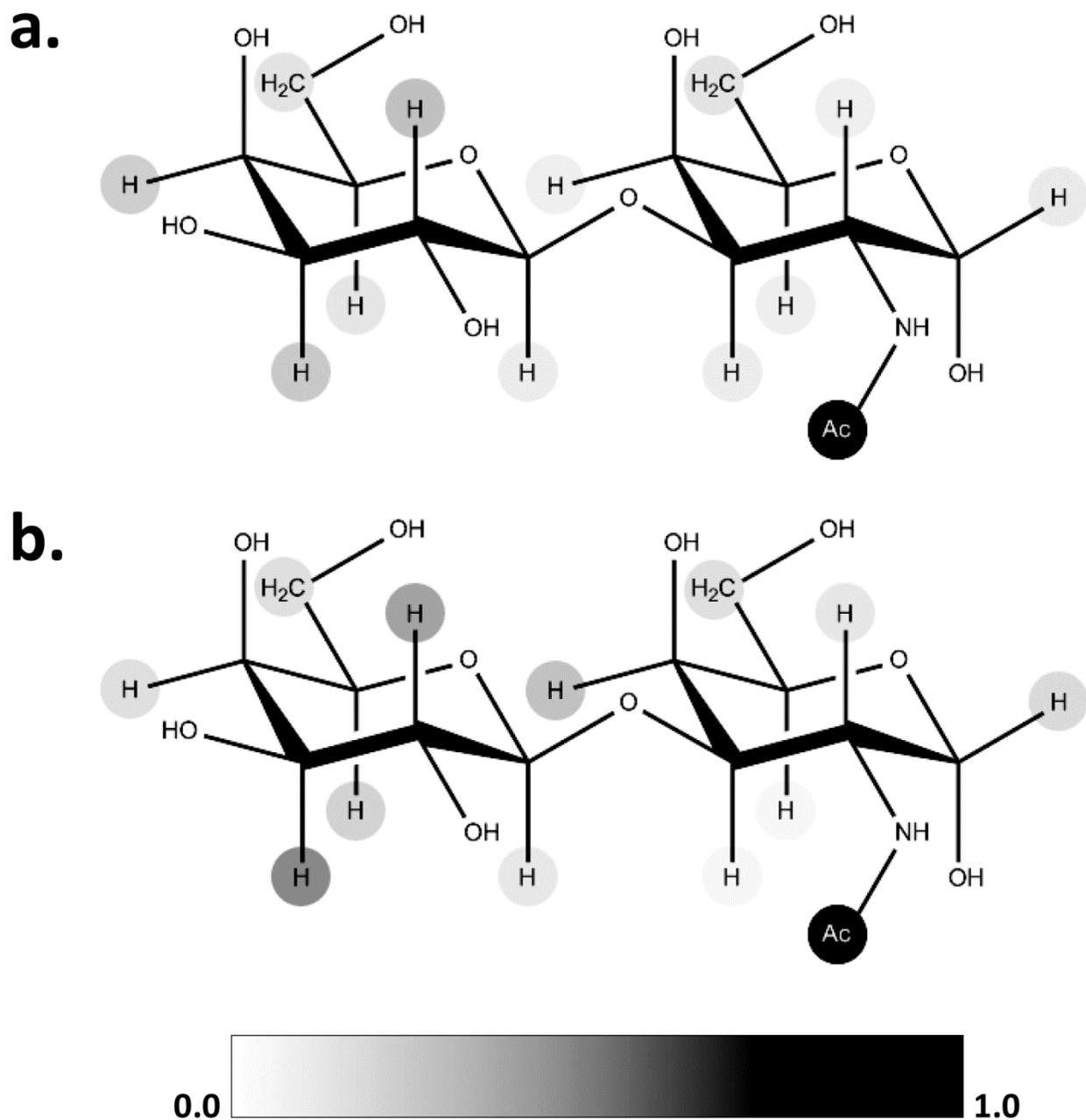


**Figure 4.4** Solvent accessible surface of the CDRs illustrating the predicted overlap between a  $\beta$ -linked aglycon in the TF disaccharide and CDR L1, responsible for ensuring the  $\alpha$ -TF specificity of JAA-F11.

#### 4.3.3 Antigen pose validation through STD-NMR

Pose 1 from the docking was both top ranked in terms of cluster population, and lowest energy, and explained the specificity of this mAb for  $\alpha$ -linked TF antigen. To further support the theoretical docked model, experimental confirmation was sought from saturation transfer difference (STD) NMR experiments. As the name implies, STD experiments detect the difference in nuclear Overhauser enhancement (nOe) magnetization transfer from the irradiated protein to the bound and free states of a ligand. The relative enhancements of the proton signals

in the bound state of the ligand are proportional to the proximity of those protons to protons in the protein. Thus, STD data provide important insight into the bound orientation of the ligand. This information permits a direct comparison between the experimental STD enhancements, and those derived from the theoretical orientation produced by docking the ligand in the Fab – TF antigen complex. STD enhancements were computed from the theoretical complex following an adaptation of the isolated spin-pair approximation (ISPA)[144], frequently employed for estimating nOe values. In ISPA, the assumption is made that the nOe intensities are proportional to  $R_{ij}^{-6}$ , where  $R_{ij}$  is the inter-proton distance between spins  $i$  and  $j$ . Here, STD intensities were derived for each proton in the ligand based on the sum of the  $R_{ij}^{-6}$  values between each proton in the Fab fragment ( $i$ ) and those in the ligand ( $j$ ). In the case of methyl groups, contributions from each proton were computed in the sum of  $R_{ij}^{-6}$  values. The agreement between the predicted and experimental STDs is illustrated in **Figure 4.5**, which shows that the strongest interactions with the protein surface involve the N-Acetyl methyl group (GalNAc) followed by H-2, H-3 and H-4 (Gal). Both the STD data and the theoretical structure indicate a key role for the N-acetyl moiety in the binding of TF antigen to JAA-F11. Each of the other poses obtained by docking was subjected to a similar analysis, however, only pose 1 resulted in satisfactory agreement between the theoretical and experimental STD data (**Figure 7.3**). The theoretical model based on pose 1 was therefore adopted as a basis for the subsequent CCG specificity analysis.



**Figure 4.5** Graphical representation of the normalized NMR-STD data for TF antigen in complex with JAA-F11, experimental (**a**) and theoretical (**b**).

#### 4.3.4 CCG specificity predictions and comparison to experimental specificity data

The number of glycan sequences included in the CCG specificity analysis is not limited by physical constraints, and the method can therefore be extended far beyond the capacity of current experimental arrays. Even the largest glycan arrays to date likely represent only 10 – 15 percent of the known human glycome[15], although even with a relatively small coverage it is possible to include a representative diversity of glycans[5]. The present virtual library of 3D glycan structures (GLibrary-3D) contains over 7,000 glycans, represented by more than 200,000 unique 3D rotamers, spanning the known human glycome.

It is important to note that because of the specificity of glycosyltransferases, not all permutations of glycosidic linkages are biologically relevant in all contexts. For example, the Neu5Ac $\alpha$ (2-6)Gal linkage is not possible when the Gal residue is present in mucins containing the core-1 (TF) disaccharide, but is commonplace in complex N-linked glycans. Biological relevance was assessed after CCG analysis for the glycans predicted to be binders. Any non-natural sequences were discarded from further consideration.

To ensure that CCG identified only true non-binders (i.e. glycans whose lack of binding arose solely from steric overlaps with the mAb), the analysis was performed without explicit consideration of the spacer type used on the experimental array[145]. In the case of glycans containing flexible linkages, each stable rotamer was generated and analyzed for steric clashes with the protein surface. Overlaps were quantified in terms of the area of the van der Waals overlap between the atoms in the glycan and the protein. In all cases, methyl glycosides were employed and confirmed the exclusive tolerance for the  $\alpha$ -configuration at the reducing terminus of the TF antigen; all  $\beta$ -anomers led to steric overlaps between the methyl and the mAb surface. The virtual screening indicated that in addition to the TF antigen, the only other glycans that did

not make significant steric clashes with the mAb surface were those that contained branches emanating from the O-6 positions of either the TF Gal or GalNAc residues (**Table 4.2**).

**Table 4.2** Human glycans predicted to fit<sup>a</sup> into the binding site of JAA-F11, in addition to those present on the CFG v4.0 array

GlycomeDB ID	Glycan Sequence
10743, 18135, 3618, 32649, 32532	Galβ1-4GlcNAcβ1(-3Galβ1-4GlcNAcβ1) <sub>1-4</sub> -6( <b>Galβ1-3</b> )GalNAc
32608	Neu5Acα2(-3Galβ1-4(Fuca1-3)GlcNAcβ1) <sub>3</sub> -6( <b>Galβ1-3</b> )GalNAc
10752, 22152	Galβ1-4(Fuca1-3)GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAc
3184, 10753	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAc
985	Galβ1-3GlcNAcβ1-3Galβ1-4GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAc
1271, 13480, 10751, 21997	Fuca1-2Galβ1-4GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAc

<sup>a</sup>CCG analysis indicated no clashes between these glycans and the JAA-F11 surface

In terms of the application of JAA-F11 as a therapeutic or diagnostic reagent, despite the large number of human glycans present in GLibrary-3D, only core 2 mucins with sialic acid, or polylactosamine extensions, at the 6-position in the core GlcNAc residue were predicted to bind, in addition to those present in the glycan array (**Table 4.3**). Thus, CCG screening of the virtual library predicted that JAA-F11 would be specific for TF antigen and a very limited subset of TF-containing human glycans.

**Table 4.3** Comparison of theoretical and experimental specificity data for mAb JAA-F11 with glycans containing the TF motif present in the CFG v4.0 glycan array

ID	Glycan Sequence	Theoretical Clash Score <sup>b</sup>	Experimental RFU <sup>a</sup>		
			Sp8 <sup>c</sup>	Sp14	Sp0 <sup>d</sup>
1	Galβ1-3GalNAcα	0	98	0	--- <sup>d</sup>
2	Neu5Acβ2-6( <b>Galβ1-3</b> )GalNAcα	0	81	---	---

3	Neu5Ac $\alpha$ 2-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	0.1	78	0	---
4	Gal $\beta$ 1-4GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	0	52	0	---
5	GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	0	51	0	---
6	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	0	---	0	---
7	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	0	---	0	---
8	(3S) <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	2.8	0	---	---
9	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	7.6	0	0	---
10	GlcNAc $\beta$ 1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	11.6	0	---	---
11	GlcNAc $\beta$ 1-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	12.4	0	---	---
12	6S(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	17.3	0	---	---
13	Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	17.4	0	0	---
14	Neu5Ac $\alpha$ 2-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	17.6	0	0	---
15	KDN $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	13.8	---	0	---
16	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	17.4	---	0	---
17 <sup>e</sup>	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4(Fuca1-3)GlcNAc $\beta$ 1-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	17.5	---	0	---
18 <sup>e</sup>	GlcNAc $\alpha$ 1-4 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	6.1	---	0	---
19 <sup>e</sup>	<b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc	5.1	---	---	0
20 <sup>e</sup>	<b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4Glc	5.3	---	---	0
21	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc $\beta$	12.6	---	---	0
22	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4Glc $\beta$	12.8	---	---	0
23	GalNAc $\alpha$ 1-3(Fuca1-2) <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc $\beta$	26.3	---	---	0

<sup>a</sup>Normalized RFUs averaged over all protein concentrations (0.1, 5 and 200  $\mu$ g/mL) and over multiple values for the same glycan, when present on the CFG array, see Methods. <sup>b</sup>Relative van der Waals overlap, see Methods. <sup>c</sup>Spacers, Sp0:  $-(\text{CH}_2)_2\text{NH}_2$ ; Sp8:  $-(\text{CH}_2)_3\text{NH}_2$ ; Sp14: -threonine. <sup>d</sup>Not present on glycan array. <sup>e</sup>Reducing anomeric configuration undefined on the CFG array,  $\alpha$ -configuration assumed for the CCG analysis.

To provide experimental confirmation of the specificity of JAA-F11, the mAb was screened against the Consortium for Functional Glycomics (CFG) printed glycan array (v4.0) at three protein concentrations (0.1, 5.0 and 200  $\mu$ g/mL, **Table 4.3**, and **Table 7.9**). As expected, the mAb displayed selectivity for the TF disaccharide at all concentrations, however, it also reacted with four other glycans that contain the TF disaccharide in their sequence. At the lowest antibody concentration only three strong binders were observed. Only one non-TF glycan (Gal $\alpha$ 1-3GalNAc $\alpha$ ) bound well, but only at the highest concentration of the antibody, indicating that the interaction is likely to be non-specific[4]. Thus, while the TF disaccharide is the minimal binding and preferred determinant for this mAb, it is not the exclusive ligand.

It is important to note that little is known about the cell-surface densities, abundances, or tissue distributions, for any of the putative binders other than the TF disaccharide. And only in the case of the TF disaccharide has the alteration of these properties in disease states, such as cancer, been examined. Thus, not all binders are necessarily biologically significant for this mAb. This fact may explain why JAA-F11 has a demonstrated specificity for tumor tissue over normal in mice[34].

#### 4.3.5 Effect of Glycan Spacers

The nature of the chemical conjugation of glycans to an array surface can lead to false negative binding. This is presumably because the spacer moiety alters either the presentation or accessibility of the glycan, or is itself in overlap with the protein surface[146]. By including multiple replicates of a glycan in the array, each conjugated through different linkers, it is possible to readily identify such spacer effects. Indeed, the data in **Table 4.3** demonstrate that conjugation of any of the high affinity binders via linker Sp14 completely abrogates binding (glycans **1, 3 - 5**). Thus, it was not possible to determine on the basis of the experimental array data alone, whether a glycan that was conjugated solely via Sp14 (**6, 7, 15 - 18**) might in fact be a binder if it were conjugated through a non-interfering linker. Similarly, none of the glycans linked through Sp0 (**19 - 23**) bound to the mAb. Unlike the case of Sp14, there is no evidence from the experimental data to confirm whether or not Sp0 is itself interfering with binding. However, given the similarity of spacers Sp8 ( $-(\text{CH}_2)_3\text{NH}_2$ ) and Sp0, ( $-(\text{CH}_2)_2\text{NH}_2$ ), it is reasonable to infer that Sp0 would likely be non-interfering.

The CCG analysis provides a complementary structure-based approach for identifying linkage effects, correctly predicting each of the five glycans (**1 – 5**) confirmed to bind by glycan array screening, as well as all of the true non-binders (**8 – 23**). Glycans **6** and **7** (Sp14) were predicted to be binding partners for JAA-F11, which may not be unexpected, given their structural similarity to other binding glycans (**4** and **5**). What is notable in the case of **6** and **7**, however, is that the theoretical analysis independently identified a potential linker effect for Sp14. The present analysis clearly indicates a role for CCG in identifying putative false negative binding, and additionally draws attention to the need for printing experimental arrays with as many glycan – linker permutations as possible.

#### **4.4 Discussion**

Carbohydrate-specific antibodies have a key role as diagnostic and therapeutic agents[147,148], yet these interactions are some of the most challenging to characterize using traditional structural biology methods. In addition, glycan chemical synthesis remains a laborious undertaking. The CCG method requires minimal experimental 3D structural data, and can be used to both guide the prioritization of chemical synthetic efforts and to provide important insight into the structural basis for biological recognition and specificity. Once the 3D structure of the bound minimal determinant is established, through either theoretical or experimental methods, or a combination of both as employed here, CCG can be used to screen an effectively unlimited range of glycans that contain this minimal determinant. Experimental microarray data can subsequently be employed to provide cross-validation for a sub-set of the CCG predictions. While crystallographic analyses of such complexes may remain challenging, the predicted structures can potentially be corroborated using a number of experimental approaches, including,

STD NMR, site-directed mutagenesis, as well as additional binding assays for predicted ligands or ligand analogs.

Experimentally-consistent structures for Fab – glycan complexes not only provide a basis for defining and predicting specificity, they also facilitate structure-based strategies for the directed evolution of antibodies with either varied specificity or affinity. In the case of JAA-F11, the 3D model can be further employed to guide the rational design of peptide mimotopes of the carbohydrate, for use as immunogens[149,150,151,152], as well as to guide the selection of key residues to be included during antibody humanization. Humanization of a mouse mAb by insertion of the CDRs into a human antibody scaffold, can lead to changes in the 3D conformation, particularly in the interface between the variable light and heavy chains[153], with a corresponding loss of affinity or specificity. As the framework regions that support and give structure to the binding site are altered during the humanization process, the resulting Ab may have a reduced affinity or specificity for the target due to conformational changes in the binding site. CCG provides an opportunity to predict the effects that structural changes or point mutations have on antigen specificity and provides a structural basis from which to assess those CDRs and antigen contacts that must be maintained during antibody humanization (**Figure 4.3**).

Several caveats to the applicability of the CCG method require consideration. The first relates to the requirement for a 3D structure of the minimal glycan determinant in the binding site. Inaccuracies in this complex will degrade the reliability of the theoretical specificity predictions. In the case of JAA-F11, automated docking was able to generate a 3D model for the immune complex that was consistent with data from STD-NMR experiments performed on the same system. In addition, specificity data from experimental glycan array screening may be employed as filters to eliminate experimentally-inconsistent poses predicted by automated

docking (**Table 7.10**). This is a novel and enabling combination of two complementary high-throughput methods. Secondly, the scope and utility of the virtual glycan library is directly dependent on the content of experimental glycomic databases. The predictive power of virtual screening will improve as the number of experimentally confirmed glycans and their detailed characterization and annotation increases. Thirdly, CCG is a high-throughput screening method that currently treats the ligand and receptor as rigid, and thus ignores the potential for induced fit to enable binding. Fourthly, the method predicts only the potential for glycans to bind to a particular receptor; it does not take into account the relative affinity of the interaction, and, just as in the case of experimental array screening, neither the natural abundance nor the cellular localization of the glycans are considered. The determination of these properties will be essential for placing binding specificity data into biologically relevant contexts. Finally, unlike experimental glycan array screening, the CCG approach requires that the minimal binding determinant be known.

Over the past decade, data from glycan array screening initiatives, such as from the CFG, have confirmed the importance of glycans and glycan binding proteins in infectious diseases, cancer metastasis and progression, immune cell interaction, and congenital diseases, such as muscular dystrophy. Concurrently, glycomics studies employing mass spectrometry have greatly increased our knowledge of the diversity of the human glycome and its alteration in disease states. CCG provides a new tool to aid in the efficient translation of this information into the practical development of therapeutics and diagnostics.

## 4.5 Materials and Methods

### 4.5.1 Crystallization of JAA-F11 Fab and Structure Determination

The JAA-F11 IgG<sub>3</sub> antibody was purified by triple ammonium sulfate precipitation, followed by Cibacron Blue 3GA agarose gel and hydroxyapatite column chromatography. Digestion and purification of the Fab utilized papain digestion followed by separation on a Protein A column. Purified JAA-F11 Fab fragment[134] was subjected to crystallization screening using the high-throughput crystallization screening facility at the Hauptman-Woodward Institute[154]. Multiple conditions were tested for optimization with hanging drop vapor diffusion. Final crystals were grown by vapor diffusion using a precipitant containing 26% PEG 5000, 50 mM lithium bromide, and 50 mM sodium citrate (pH 4.0). Crystals appeared in 3-4 days. Crystals were cryoprotected by transferring the crystal to solutions of mother liquor containing incrementally higher solutions of ethylene glycol (8, 16, and 24 %) and flash cooled in liquid nitrogen. Data were collected at the A1 beamline of the Cornell High Energy Synchrotron Source and scaled with HKL2000[155]; the intensities were converted to structure factors using the TRUNCATE program of the CCP4 suite[156]. The structure was solved by molecular replacement with MOLREP[157]. Multiple Fab antibody fragments were tested as search models. A satisfactory solution was found using a search model consisting of the heavy and light chains from 1CLZ, the crystal structure of an antitumor directed antibody that recognizes the Lewis Y tetrasaccharide[158]. The molecular replacement solution was refined through iterative manual model building with COOT[159] and maximum likelihood refinement with REFMAC5[160]. Diffraction and refinement statistics are presented in **Table 4.1**. The binding pocket volume was determined with the fpocket software package[161]. The final structure factors and coordinates are deposited with the Protein Database (PDB ID 3GNM).

### 4.5.2 Glycan Array Screening

A sample of antibody JAA-F11, isolated by ammonium sulfate precipitation, was submitted to the Consortium for Functional Glycomics (CFG) for screening at concentrations of 0.1, 5.0, 5.0 and 200  $\mu\text{g}/\text{mL}$  on version 4.0 of the glycan array. Fluorescence was obtained by detecting the antibody with an Alexa Fluor-488 labeled anti-mouse IgG (Invitrogen) at 5.0  $\mu\text{g}/\text{ml}$ . Data collection and interpretation methods have been reported[5] and are available on the CFG website ([www.functionalglycomics.org](http://www.functionalglycomics.org)).

### 4.5.3 Saturation Transfer Difference NMR

A sample of JAA-F11 antibody (6.7  $\mu\text{M}$ , two binding sites per dimer) and TF disaccharide from Carbosynth, (1.9 mM, approximately equal amounts of  $\alpha$  and  $\beta$  anomers at the reducing terminus) was prepared in buffered  $\text{D}_2\text{O}$ , giving an approximate molar ratio of 100:1 for each anomer per binding site. The JAA-F11 antibody was obtained from the Rittenhouse-Olson lab and the free TF-disaccharide was purchased from Carbosynth. STD data were collected on an 800 MHz Varian (Agilent) Inova spectrometer at 25  $^\circ\text{C}$  using the double pulsed field gradient spin echo[162] method for water suppression. Antibody protons were selectively irradiated using a train of 50 ms Gaussian pulses at 0.5 ppm and a difference spectra produced by subtracting a reference spectrum irradiated at 25 ppm[163]. A total of 2048 scans were acquired for each interleaved spectrum. Different irradiation times were obtained at 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, and 8.0 seconds to obtain a build-up curve (**Figure 7.4**). Data were processed and integrated using Mnova software[164]. The 1.0 s irradiation time was used to compare to the docking model results.

#### 4.5.4 Virtual Docking

Docking was performed with AutoDock 3.05[141]. Waters of crystallization were removed from the Fab structure of JAA-F11 before assigning partial charges to the protein (Kollman)[165] and to the TF disaccharide (Gasteiger)[166]. Initial 3D models for the TF disaccharide were generated using the carbohydrate 3D structure generation tools at GLYCAM-Web ([www.glycam.org](http://www.glycam.org))[32] which energy minimizes the glycan in implicit solvent using the GLYCAM06 force field[24]. The final structure can be characterized by the values for the glycosidic linkage:  $\varphi$  ( $H_1-C_1-O_3-C_3$ ) =  $61.5^\circ$  and  $\psi$  ( $C_1-O_3-C_3-H_3$ ) =  $6.2^\circ$ , which are consistent with the dominant solution NMR conformation[167]. Torsion rotation within the epitope was limited to the exocyclic free rotors (H-O and  $C_5-C_6$  bonds); all other torsions were frozen at the values of the initial 3D model. The docking region was defined so as to include all the hypervariable loops, by centering a grid box (33 Å per side) on the sidechain nitrogen of Asp<sup>H100</sup>. A cubic grid spacing of 0.375 Å was employed. Docking was performed using the Lamarckian Genetic Algorithm[168] with a population size of 150 and 2.5 million energy evaluations in each of 50 docking runs and a clustering cutoff of 2.0 Å was employed.

#### 4.5.5 Creation of GLibrary-3D

Glycans to be included in the virtual glycan library (GLibrary-3D) were each selected from GlycomeDB ([www.glycome-db.org](http://www.glycome-db.org))[139], which is an online database that currently contains 3,570 N- and O-linked glycan sequences found in humans, of which 3,086 contained sufficient information to be converted into 3D structures. Because most of the structures contained in GlycomeDB were determined using mass spectrometric techniques, not all reported sequences included sufficient information to uniquely define the glycan. For example, many

sequences do not include information regarding inter-residue linkage positions, and these were generally excluded from the virtual library. However, in the case of certain human glycan sequences, which display only a limited number of linkage possibilities, such as the disaccharides Neu5Ac $\alpha$ (2-3/6)Gal or Gal $\beta$ (1-3/4)GlcNAc, each linkage permutation was constructed. Additionally, on the basis of known glycan structures, all ring types were assumed to be pyranose. These assumptions resulted in a total of 7,127 unique putative human glycan sequences. For glycans containing 1-6 or 2-6 linkages, each stable rotamer of the  $\omega$ -angle ( $\pm 60^\circ$ ,  $180^\circ$ ) was generated. Additional rotamers were built for the  $\phi$ -angles in 2-3 linkages ( $-60^\circ$  and  $180^\circ$ )[169], leading to a library of 207,693 glycan 3D structures (GLibrary-3D). Glycan sequences were converted to 3D structures using an automated version of the Carbohydrate Model Building Tool of GLYCAM-Web ([www.glycam.org](http://www.glycam.org))[32].

#### 4.5.6 Computational Carbohydrate Grafting (CCG)

All putative human glycans containing the TF disaccharide (1,182 glycans) were extracted from GLibrary-3D. The glycan branches to be grafted onto the minimal determinant were then translated and rotated as required in order to ensure correct relative alignment of the branch with respect to the minimal determinant, as defined on the basis of glycosidic bond lengths and angles (**Figure 4.1**). The  $\phi$ - and  $\psi$ -glycosidic torsion angles of the newly-formed linkage, were assigned on the basis of known carbohydrate conformational properties[138]. In the case of glycans that contained torsion angles known to populate more than one stable rotamer, such as 1-6 or 2-6 linkages, each rotamer was generated and treated as an independent molecule in the grafting process leading to a total of 3,109 rotamers. Once assembled, the intact glycan was energy minimized with the GLYCAM force field as described above.

Steric overlaps between the grafted branches and the protein surface were determined from the area of overlap of the atomic van der Waals surfaces[170]. As even small van der Waals overlaps can lead to very high repulsion energies, in a rigid molecular alignment, clash scores were not determined energetically. To assess the significance of the clash, the van der Waals overlap values were reported relative to the surface area of a single carbon atom ( $36.3 \text{ \AA}^2$ )[170]. Only glycans with relative overlaps (clash scores) of less than 1.0, corresponding to a single occluded carbon atom, were considered as satisfying the no-overlap criterion.

#### **4.6 Acknowledgements**

The authors would like to thank Kausar N. Samli and Mari DeMarco for preparing samples for NMR analysis.

#### **4.7 Epilogue**

A relatively new glycan microarray technology has been developed in which cellular glycans are cleaved directly from the source, separated by chromatography and printed on microarrays[171]. This type of glycomics screening does not require pre-sequencing of the glycans but instead relies on characterizing glycan sequences after binding has been observed. This exciting new technology could be easily integrated into the CCG methodology described in the published manuscript as it is simply another tool to define specificity. While this shotgun approach is useful in identifying cellular glycans which bind, the technology would be particularly more beneficial for CCG analysis if the entire printed cellular glycome were characterized. This would allow CCG analysis to characterize structural characteristics related to

both binding and non-binding glycans such as spacer type and substituents off the minimal determinant. This information is what makes synthetic glycan microarray data so useful in determining the specificity as it provides information on what cannot be tolerated by the binding pocket.

## 5. CONCLUSIONS AND FUTURE DIRECTIONS

Modeling carbohydrate structures requires an understanding of its dynamics and flexibility. Molecular dynamics simulations have proven a useful tool in characterizing this flexibility and relating it to properties observed in a variety of experiments. In order to perform these simulations, the GLYCAM force field has been developed for glycans and glycoconjugates and serves as one of the most utilized carbohydrate force fields available[18]. This work has shown the development and application of this force field in MD simulations of a lipid bilayer and glycosaminoglycans. The MD simulations of the GAG structures were able to expand on limited experimental data to predict experimentally-consistent ring shape properties and glycosidic linkage profiles which define the 3D shape of a glycan. The next step beyond pyranoside ring dynamics is the characterization of furanoside ring properties. Unique to furanosides is their role in the structure of DNA/RNA. The accurate modeling of furanose dynamics would allow the GLYCAM force field to expand into the field of nucleic acid simulations.

The extensive experimental validation of the GLYCAM force field in predicting glycan structure made it ideal for generating static 3D models of glycan structures for non-dynamic computational methods. These static models have been employed in a novel method called Computational Carbohydrate Grafting which uses these models to identify specificity constraints within the binding pocket of a protein. This technology was used to characterize the 3D features of a mouse antibody that defines its specificity to certain TF-containing saccharides. Knowledge of these features allows for better design of antigenic mimics to create a vaccine specific to these

sequences and identifies the amino acids which must be conserved when humanizing this antibody. This method is already being used to characterize the specificity restraints for a number of carbohydrate-protein complexes found in the protein databank for which glycan array specificity screenings have already been performed. Ideally, this method will be integrated into a website service that incorporates it into an automated docking scheme to identify the best docking poses based on agreement with experimental specificity screening.

Improvements in carbohydrate force field design and molecular classes will continue to depend on a symbiotic relationship with NMR technology as has been the case for the last few decades[18]. The long-term benefits to this relationship have given carbohydrate force fields reliability that makes them an appropriate supplement to sparse or non-existent experimental data. This reliability, coupled with advancements in computational speed and structure builders means that simulations of many glycan structures could be automatically simulated and data collected for bioinformatics purposes including automated docking methods. Using these automated *in silico* techniques would then reduce costs for designing synthetic glycomimetics by focusing on high affinity and specificity interactions.

## 6. REFERENCES

1. Dennis JW, Granovsky M, Warren CE (1999) Glycoprotein glycosylation and cancer progression. *Biochimica et Biophysica Acta* 1473: 21 - 34.
2. Rademacher TW, Parekh RB, Dwek RA (1988) Glycobiology. *Ann Rev Biochem* 57: 785-838.
3. Varki A, Sharon N (2009) Historical Background and Overview. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. *Essentials of Glycobiology*, 2nd Edition. 2 ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
4. Oyelaran O, Gildersleeve JC (2009) Glycan Arrays: Recent Advances and Future Challenges. *Curr Op Chem Bio* 13: 406-413.
5. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, et al. (2004) Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins. *Proc Natl Acad Sci USA* 101: 17033-17038.
6. Fukui S, Feizi T, Galustian C, Lawson AM, Chai W (2002) Oligosaccharide Microarrays for High- Throughput Detection and Specificity Assignments of Carbohydrate- Protein Interactions. *Nat Bio Tech* 20: 1011-1017.
7. Bertozzi CRR, D. (2009) Structural Basis of Glycan Diversity. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. *Essentials of Glycobiology*, 2nd Edition. 2 ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
8. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, et al. (2009) Symbol nomenclature for glycan representation. *Proteomics* 9: 5398-5399.

9. Ferro DR, Provasoli A, Ragazzi M, Casu B, Torri G, et al. (1990) Conformer Populations of L-Iduronic Acid Residues in Glycosaminoglycan Sequences. *Carbohydr Res* 195: 157-167.
10. Guo Y, Conrad HE (1989) The disaccharide composition of heparins and heparan sulfates. *Analytical Biochemistry* 176: 96-104.
11. Mann M, Islam T, Dyason J, Florio P, Trower C, et al. (2006) Unsaturated N-acetyl- D-glucosaminuronic acid glycosides as inhibitors of influenza virus sialidase. *Glycoconjugate Journal* 23: 127-133.
12. Bazin HG, Capila I, Linhardt RJ (1998) Conformational study of synthetic  $\Delta$ 4-uronate monosaccharides and glycosaminoglycan-derived disaccharides. *Carbohydrate Research* 309: 135-144.
13. Varki A, Esko JD, Colley KJ (2009) Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. *Essentials of Glycobiology*, 2nd Edition. 2 ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
14. DeMarco ML, Woods RJ (2009) Atomic-Resolution Conformational Analysis of a Ganglioside in a Lipid Bilayer and its Implications for Ganglioside-Protein Interactions at the Plasma Membrane Surface. *Glycobiology* 19: 344–355.
15. Cummings RD (2009) The Repertoire of Glycan Determinants in the Human Glycome. *Mol BioSyst* 5: 1087-1104.
16. Woods RJ, Tessier MB (2010) Computational Glycoscience: Characterizing the Spatial and Temporal Properties of Glycans and Glycan—Protein Complexes. *Curr Opin Struct Biol* 20: 575-583.

17. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* 81: 3684-3690.
18. Foley BL, Tessier MB, Woods RJ (2011) Carbohydrate Force Fields. *WIREs Computational Molecular Science*: 1-69.
19. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B (1995) Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM\_93 Parameter Development. *J Phys Chem* 99: 3832-3846.
20. Kirschner KN, Woods RJ (2001) Solvent Interactions Determine Carbohydrate Conformation. *Proc Natl Acad Sci USA* 98: 10541-10545.
21. Woods RJ, Chappelle R (2000) Restrained Electrostatic Potential Atomic Charges for Condensed Phase Simulations of Carbohydrates. *J Mol Struct (Theochem)* 527: 149-156.
22. Basma M, Sundara S, Calgan D, Vernali T, Woods RJ (2001) Solvated Ensemble Averaging in the Calculation of Partial Atomic Charges. *J Comput Chem* 22: 1125-1137.
23. Tschampel SM, Kennerty MR, Woods RJ (2007) TIP5P-Consistent Treatment of Electrostatics for Biomolecular Simulations. *J Chem Theory Comput* 3: 1721-1733.
24. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, et al. (2008) GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates. J Comput Chem* 29: 622–655.
25. Tessier MB, DeMarco ML, Yongye AB, Woods RJ (2008) Extension of the GLYCAM06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Molecular Simulation* 34: 349-364.

26. Tessier MB, Grant OC, Heimbürg-Molinaro J, Smith D, Jadey S, et al. (2013) Computational Screening of the Human TF-Glycome Provides a Structural Definition for the Specificity of Anti-Tumor Antibody JAA-F11. PLoS ONE 8: e54874.
27. Gandhi NS, Mancera RL (2010) Can Current Force Fields Reproduce Ring Puckering in 2-O-Sulfo- $\alpha$ -L-Iduronic Acid? A Molecular Dynamics Simulation Study. Carbohydr Res 345: 689-695.
28. Sattelle BM, Bose-Basu B, Tessier M, Woods RJ, Serianni AS, et al. (2012) Dependence of Pyranose Ring Puckering on Anomeric Configuration: Methyl Idopyranosides. The Journal of Physical Chemistry B 116: 6380-6386.
29. Tessier MB, Pederson K, Singh A, Prestegard JH, Woods RJ (2013) in preparation.
30. Adeyeye J, Azurmendi HF, Stroop CJM, Sozhamannan S, Williams LJ, et al. Conformation of the Hexasaccharide Repeating Subunit from the *Vibrio Cholerae* O139 Capsular Polysaccharide. abstract.
31. Jin L, Hricovíni M, Deakin JA, Lyon M, Uhrín D (2009) Residual Dipolar Coupling Investigation of a Heparin Tetrasaccharide Confirms the Limited Effect of Flexibility of the Iduronic Acid on the Molecular Shape of Heparin. Glycobiology 19: 1185-1196.
32. WoodsGroup (2005 - 2012) GLYCAM Web. In: Woods RJ, editor. Athens, GA: Complex Carbohydrate Research Center, The University of Georgia.
33. Ficko-Blean E, Stuart CP, Suits MD, Cid M, Tessier M, et al. (2012) Carbohydrate Recognition by an Architecturally Complex  $\alpha$ -N-Acetylglucosaminidase from *Clostridium perfringens*. PLoS One 7: e33524.

34. Rittenhouse-Diakun K, Xia Z, Pickhardt D, Morey S, Baek M, et al. (1998) Development and Characterization of Monoclonal Antibody to T-antigen: (gal beta1-3GalNAc-alpha-O). *Hybridoma* 17: 165-173.
35. Spector AA, Yorek MA (1985) Membrane Lipid-Composition and Cellular Function. *Journal of Lipid Research* 26: 1015-1035.
36. Hakomori S (2003) Structure, organization, and function of glycosphingolipids in membrane. *Current Opinion in Hematology* 10: 16-24.
37. Evans SV, Mackenzie CR (1999) Characterization of Protein-Glycolipid Recognition at the Membrane Bilayer. *J Mol Recognit* 12: 155-168.
38. Nagle JF, Tristram-Nagle S (2000) Structure of Lipid Bilayers. *Biochim Biophys Acta* 1469: 159-195.
39. Hare BJ, Howard KP, Prestegard JH (1993) Torsion Angle Analysis of Glycolipid Order at Membrane Surfaces. *Biophys J* 64: 392-398.
40. Kishore AI, Prestegard JH (2003) Molecular Orientation and Conformation of Phosphatidylinositides in Membrane Mimetics Using Variable Angle Sample Spinning (VASS) NMR. *Biophys J* 85: 3848-3857.
41. Tieleman DP, Marrink SJ, Berendsen HJ (1997) A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochim Biophys Acta* 1331: 235-270.
42. Lague P, Roux B, Pastor RW (2005) Molecular Dynamics Simulations of the Influenza Hemagglutinin Fusion Peptide in Micelles and Bilayers: Conformational Analysis of Peptide and Lipids. *Journal of Molecular Biology* 354: 1129-1141.

43. Noskov SY, Roux B (2007) Importance of Hydration and Dynamics on the Selectivity of the KcsA and NaK Channels. pp. jgp.200609633.
44. La Rocca P, Biggin PC, Tieleman DP, Sansom MSP (1999) Simulation studies of the interaction of antimicrobial peptides and lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1462: 185-200.
45. de Groot BL, Grubmuller H (2001) Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. pp. 2353-2357.
46. Zhu F, Tajkhorshid E, Schulten K (2001) Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Letters* 504: 212-218.
47. Edidin M (2003) The state of lipid rafts: from model membranes to cells. *Annu Rev Biophys Biomol Struct* 32: 257-283.
48. Kirschner KN, Yongye AB, Tschampel SM, Daniels CR, Foley BL, et al. (2007) GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates*. *J Comput Chem* In Press.
49. Kirschner KN, Woods RJ (2001) Quantum Mechanical Study of the Nonbonded Forces in Water-Methanol Complexes. *J Phys Chem A* 105: 4150-4155.
50. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. (1998) *Gaussian 98*. Pittsburg: Gaussian, Inc.
51. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, et al. (2004) *AMBER 8*. San Francisco: University of California.
52. Case DA, Cheatham III TE, Darden TA, Gohlke H, Luo R, et al. (2005) The AMBER Biomolecular Simulation Programs. *J Comput Chem* 26: 1668-1688.

53. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B*58: 380-388.
54. Bruno IJ, Cole JC, Edgington PR, Kessler M, Macrae CF, et al. (2002) New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr B*58: 389-397.
55. Breneman CM, Wiberg KB (1990) Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J Comput Chem* 11: 361-373.
56. Cieplak P, Caldwell JW, Kollman PA (2001) Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases. *J Comput Chem* 22: 1048-1057.
57. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J Am Chem Soc* 115: 9620-9631.
58. Feller SE, Gawrisch K, Mackerell ADJ (2002) Polyunsaturated Fatty Acids in Lipid Bilayers: Intrinsic and Environmental Contributions to Their Unique Physical Properties. *J Am Chem Soc* 124: 318-326.
59. Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Molec Graphics* 14: 33-38.
60. Heller H, Schaefer M, Schulten K (1993) Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *J Comput Chem* 14: 8343-8360.

61. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J Comput Phys* 23: 327-341.
62. Darden TA, York D, Pedersen L (1993) Particle Mesh Ewald: An  $N \log(N)$  Method for Ewald Sums in Large Systems. *J Chem Phys* 98: 10089-10092.
63. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J Comp Chem* 25: 1605-1612.
64. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J Comput Chem* 7: 230-252.
65. Pascher I, Sundell S, Eibl H, Harlos K (1986) The single crystal structure of octadecyl-2-methyl-glycero-phosphocholine monohydrate. A multilamellar structure with interdigitating head groups and hydrocarbon chains. *Chemistry and Physics of Lipids* 39: 53-64.
66. Gadola SD, Zaccai NR, Harlos K, Shepherd D, Castro-Palomino JC, et al. (2002) Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. *Nat Immunol* 3: 721-726.
67. Roszak AW, Gardiner AT, Isaacs NW, Cogdell RJ (2007) Brominated lipids identify lipid binding sites on the surface of the reaction center from *Rhodobacter sphaeroides*. *Biochemistry* 46: 2909-2916.
68. Samdal S, Seip HM (1975) Potential for Rotation About C(sp<sup>\*</sup>)-O and C(sp<sup>\*</sup>)-S Bonds: Electron Diffraction Results for CH<sub>2</sub>=CH-OCH<sub>3</sub> and CH<sub>2</sub>=CH-SCH<sub>3</sub> and AB Initio Calculations ON CH<sub>2</sub>=CH-OH and CH<sub>2</sub>XH-SH. *J Molec Struct* 28: 193-203.

69. Irngartinger H, Herpich R, Huber-Patz U, Trofimov BA, Malysheva SF (2003) CSD id IPUMAH. Heidelberg, Germany: Organisch-Chemisches Institut der Universitat.
70. Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, et al. (2002) Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol* 9: 507-511.
71. Yoder MD, Thomas LM, Tremblay JM, Oliver RL, Yarbrough LR, et al. (2001) Structure of a multifunctional protein. Mammalian phosphatidylinositol transfer protein complexed with phosphatidylcholine. *J Biol Chem* 276: 9246-9252.
72. Vordtriede PB, Doan CN, Tremblay JM, Helmkamp GM, Yoder MD (2005) Structure of PITPbeta in complex with phosphatidylcholine: Comparison of structure and lipid transfer to other PITP isoforms. *Biochemistry* 44: 14760-14771.
73. Beamer LJ, Carroll SF, Eisenberg D (1997) Crystal Structure of Human BPI and Two Bound Phospholipids at 2.4 Angstrom Resolution. *Science* 276: 1862-1864.
74. White SP, Scott DL, Otwinowski Z, Gelb MH, Sigler PB (1990) Crystal structure of cobra-venom phospholipase A2 in a complex with a transition-state analogue. *Science* 250: 1560-1563.
75. Rand RP, Parsegian VA (1989) Hydration forces between phospholipid bilayers. *Biochimica et Biophysica Acta* 988: 351-376.
76. DeMarco ML, Alonso DOV, Daggett V (2004) Diffusing and Colliding: The Atomic Level Folding/Unfolding Pathway of a Small Helical Protein. *Journal of Molecular Biology* 341: 1109-1124.

77. Tieleman DP, Berendsen HJC (1996) Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *The Journal of Chemical Physics* 105: 4871-4880.
78. Bassolino-Klimas D, Alper HE, Stouch TR (1995) Mechanism of Solute Diffusion through Lipid Bilayer Membranes by Molecular Dynamics Simulation. *J Am Chem Soc* 117: 4118-4129.
79. Marrink SJ, Sok RM, Berendsen HJC (1996) Free volume properties of a simulated lipid membrane. *The Journal of Chemical Physics* 104: 9090-9099.
80. Feller SE (2000) Molecular Dynamics Simulations of Lipid Bilayers. *Current Opinion in Colloid and Interface Science* 00: 00-00.
81. Scott HJ (2002) Modeling the Lipid Component of Membranes. *Current Opinion in Structural Biology* 12: 495-502.
82. Skjerveik ÅA, Madej BD, Walker RC, Teigen K (2012) LIPID11: A modular framework for lipid simulations using amber. *The Journal of Physical Chemistry B* 116: 11124-11136.
83. Dickson CJ, Rosso L, Betz RM, Walker RC, Gould IR (2012) GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* 8: 9617-9627.
84. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* 114: 7830-7843.
85. Pastor R, MacKerell Jr A (2011) Development of the CHARMM force field for lipids. *The journal of physical chemistry letters* 2: 1526-1532.

86. Poger D, Van Gunsteren WF, Mark AE (2010) A new force field for simulating phosphatidylcholine bilayers. *Journal of Computational Chemistry* 31: 1117-1125.
87. López CsA, Sovova Z, van Eerden FJ, de Vries AH, Marrink SJ (2013) Martini force field parameters for glycolipids. *Journal of Chemical Theory and Computation* 9: 1694-1708.
88. Periole X, Marrink S-J (2013) The martini coarse-grained force field. *Biomolecular Simulations: Springer*. pp. 533-565.
89. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal* 97: 50-58.
90. Goetz AW, Williamson MJ, Xu D, Poole D, Le Grand S, et al. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs - Part I: Generalized Born. *J Chem Theory Comput*.
91. Salomon-Ferrer R, Goetz AW, Poole D, Le Grand S, Walker R (2013) Routine microsecond molecular dynamics simulations with AMBER - Part II: Particle Mesh Ewald. *J Chem Theory Comput*: in prep.
92. Le Grand S, Götz AW, Walker RC (2013) SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications* 184: 374-380.
93. Kopeć W, Telenius J, Khandelia H (2013) Molecular dynamics simulations of the interactions of medicinal plant extracts and drugs with lipid bilayer membranes. *FEBS Journal*.
94. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology* 19: 120-127.

95. Bennett W, Tieleman DP (2013) Computer Simulations of Lipid Membrane Domains. *Biochimica et Biophysica Acta (BBA)-Biomembranes*.
96. Kreuger J, Spillmann D, Li J-p, Lindahl U (2006) Interactions between heparan sulfate and proteins: the concept of specificity. *The Journal of cell biology* 174: 323-327.
97. Bernfield M, Götte M, Park PW, Reizes O, Fitzgerald ML, et al. (1999) Functions of cell surface heparan sulfate proteoglycans. *Annual review of biochemistry* 68: 729-777.
98. Gama CI, Tully SE, Sotogaku N, Clark PM, Rawat M, et al. (2006) Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nature chemical biology* 2: 467-473.
99. van Boeckel CAA, Petitou M (1993) The Unique Antithrombin III Binding Domain of Heparin: A Lead to New Synthetic Antithrombotics. *Angew Chem Int Ed Engl* 32: 1671-1690.
100. Mikhailov D, Linhardt RJ, Mayo KH (1997) NMR Solution Conformation of Heparin-Derived Hexasaccharide. *Biochem J* 328: 51-61.
101. Blanchard V, Chevalier F, Imberty A, Leeftang BR, Basappa, et al. (2007) Conformational Studies on Five Octasaccharides Isolated from Chondroitin Sulfate Using NMR Spectroscopy and Molecular Modeling. *Biochemistry* 46: 1167-1175.
102. Sattelle BM, Shakeri J, Roberts IS, Almond A (2010) A 3D-structural model of unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone conformation. *Carbohydr Res* 345: 291-302.
103. Sattelle BM, Shakeri J, Almond A (2013) Does Microsecond Sugar Ring Flexing Encode 3D-Shape and Bioactivity in the Heparanome? *Biomacromolecules* 14: 1149-1159.

104. Mulloy B, Forster MJ, Jones C, Drake AF, Johnson EA, et al. (1994) The Effect of Variation of Substitution on the Solution Conformation of Heparin - a Spectroscopic and Molecular Modeling Study. *Carbohydr Res* 255: 1-26.
105. Mallajosyula SS, Guvench O, Hatcher E, MacKerell AD (2011) CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J Chem Theory Comput* 8: 759-776.
106. Almond A, Sheehan JK (2000) Glycosaminoglycan Conformation: Do Aqueous Molecular Dynamics Simulations Agree with X-ray Fiber Diffraction? *Glycobiology* 10: 329-338.
107. Becker CF, Guimarães JA, Verli H (2005) Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution. *Carbohydrate Research* 340: 1499-1507.
108. Pol-Fachin L, Verli H (2008) Depiction of the forces participating in the 2-O-sulfo- $\alpha$ -L-iduronic acid conformational preference in heparin sequences in aqueous solutions. *Carbohydrate Research* 343: 1435-1445.
109. Ferro DR, Provasoli A, Ragazzi M, Torri G, Casu B, et al. (1986) Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligo-saccharides: NMR and force-field studies. *Journal of the American Chemical Society* 108: 6773-6778.
110. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 6: 277-293.
111. Goddard T, Kneller D (2004) SPARKY 3. University of California, San Francisco.

112. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Protein NMR Techniques*: Springer. pp. 313-352.
113. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics* 78: 1950-1958.
114. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Phys Chem* 79: 926-935.
115. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, et al. (2007) GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates. Journal of Computational Chemistry* 29: 622–655.
116. Case D, Darden T, Cheatham III T, Simmerling C, Wang J, et al. AMBER 12; University of California: CA, 2012. There is no corresponding record for this reference.
117. Onufriev A, Bashford D, Case DA (2004) Exploring Protein Native States and Large-scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct Funct Bioinformatics* 55: 383-394.
118. Haasnoot CAG, De Leeuw M, Altona C (1980) The Relationship between Proton-Proton NMR Coupling Constants and Substituent Electronegativities - I. An Empirical Generalization of the Karplus Equation. *Tetrahedron* 36: 2783-2792.
119. Altona C, Francke R, de Haan R, Ippel JH, Daalmans GJ, et al. (1994) Empirical group electronegativities for vicinal NMR proton–proton couplings along a C–C bond: Solvent effects and reparameterization of the Haasnoot equation. *Magnetic resonance in chemistry* 32: 670-678.

120. Yongye AB, Gonzales Outeriño J, Glushka J, Schultheis V, Woods RJ (2008) The Conformational Properties of Methyl  $\alpha$ -(2,8)-di/trisialosides and Their *N*-acyl Analogs: Implications for Anti-*Neisseria meningitidis* B Vaccine Design. *Biochemistry* 47: 12493–12514.
121. Gonzalez-Outeiriño J, Kirschner KN, Thobhani S, Woods RJ (2006) Reconciling Solvent Effects on Rotamer Populations in Carbohydrates: a Joint MD and NMR Analysis. *Can J Chem* 84: 569-579.
122. Gatti G, Casu B, Hamer G, Perlin A (1979) Studies on the conformation of heparin by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. *Macromolecules* 12: 1001-1007.
123. Mulloy B, Forster M, Jones C, Davies D (1993) Nmr and molecular-modelling studies of the solution conformation of heparin. *Biochemical Journal* 293: 849.
124. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, et al. (2002) The Protein Data Bank. *Acta Crystallogr D* 58: 899-907.
125. Orntoft TF, Vestergaard EM (1999) Clinical Aspects of Altered Glycosylation of Glycoproteins in Cancer. *Electrophoresis* 20: 362-371.
126. Kosma P, Müller-Loennies S (2011) Anticarbhydrate Antibodies: From Molecular Basis to Clinical Application: Springer.
127. Xu YF, Sette A, Sidney J, Gendler SJ, Franco A (2005) Tumor-associated carbohydrate antigens: A possible avenue for cancer prevention. *Immunol Cell Bio* 83: 440-448.
128. Irazoqui FJ, Sendra VG, Lardone RD, Nores GA (2005) Immune Response to Thomsen-Friedenreich Disaccharide and Glycan Engineering. *Immunol Cell Biol* 83: 405-412.
129. Yu L-G (2007) The Oncofetal Thomsen–Friedenreich Carbohydrate Antigen in Cancer Progression. *Glycoconj J* 24: 411-420.

130. Almogren A, Abdullah J, Ghapure K, Ferguson K, Glinsky VV, et al. (2012) Anti-Thomsen-Friedenreich-Ag (anti-TF-Ag) Potential for Cancer Therapy. *Front Biosci S4*: 840-863.
131. Ju T, Otto VI, Cummings RD (2011) The Tn antigen-structural simplicity and biological complexity. *Angew Chem Int Ed* 50: 1770-1791.
132. Henderson G, Ulsemer P, Schöber U, Löffler A, Alpert C-A, et al. (2011) Occurrence of the Human Tumor-Specific Antigen Structure Gal $\beta$ 1-3GalNAc $\alpha$ - (Thomsen-Friedenreich) and Related Structures on Gut Bacteria: Prevalence, Immunochemical Analysis and Structural Confirmation. *Glycobiol* 21: 1277-1289.
133. Heimburg J, Yany J, Morey S, Glinskii OV, Huxley VH, et al. (2006) Inhibition of Spontaneous Breast Cancer Metastasis by Anti-Thomsen-Friedenreich Antigen Monoclonal Antibody JAA-F111. *Neoplasia* 8: 939-948.
134. Chaturvedi R, Heimburg J, Yan J, Koury S, Sajjad M, et al. (2008) Tumor Immunolocalization Using <sup>124</sup>I-Iodine-Labeled JAA-F11 Antibody to Thomsen-Friedenreich Alpha-Linked Antigen. *Appl Radiat Isotopes* 66: 278-287.
135. Manimala JC, Roach TA, Li Z, Gildersleeve JC (2007) High-throughput Carbohydrate Microarray Profiling of 27 Antibodies Demonstrates Widespread Specificity Problems. *Glycobiol* 17: 17C-23C.
136. DeMarco ML, Woods RJ (2008) Structural Glycobiology: A Game of Snakes and Ladders. *Glycobiol* 18: 426-440.
137. Homans SW (1993) Conformation and Dynamics of Oligosaccharides in Solution. *Glycobiol* 3: 551-555.

138. Rao VSR, Qasba PK, Balaji PV, Chandrasekaran R (1998) Conformation of Carbohydrates. Amsterdam: Harwood Academic.
139. Ranzinger R, Herget S, Wetter T, von der Lieth C-W (2008) GlycomeDB - Integration of Open-access Carbohydrate Structure Databases. BMC Bioinf 9: 384.
140. Lee M, Lloyd P, Zhang X, Schallhorn JM, Sugimoto K, et al. (2006) Shapes of Antibody Binding Sites: Qualitative and Quantitative Analyses Based on a Geomorphic Classification Scheme. J Org Chem 71: 5082-5092.
141. Morris GM, Goodsell DS, Huey R, Hart WE, Halliday S, et al. (1999) Autodock. 3.0.5 ed. La Jolla: Scripps Research Institute.
142. Sujatha MS, Balaji PV (2004) Identification of Common Structural Features of Binding Sites in Galactose-Specific Proteins. Proteins: Structure, Function, and Bioinformatics 55: 44-65.
143. Lütteke T, Frank M, von der Lieth C-W (2005) Carbohydrate Structure Suite (CSS): Analysis of Carbohydrate 3D Structures Derived from the PDB. Nucl Acids Res 33: D242-D246.
144. Gronenborn AM, Clore GM (1985) Investigation of the Solution Structures of Short Nucleic Acid Fragments by Means of Nuclear Overhauser Enhancement Measurements. Prog Nuc Magn Res Spect 17: 1-32.
145. Taylor ME, Drickamer K (2009) Structural Insights into what Glycan Arrays tell us About how Glycan-binding Proteins Interact with their Ligands. Glycobio 19: 1155–1162.
146. Lewallen DM, Siler D, Iyer SS (2009) Factors Affecting Protein–Glycan Specificity: Effect of Spacers and Incubation Time. ChemBioChem 10: 1486-1489.

147. Yarema KJ, Bertozzi CR (1998) Chemical Approaches to Glycobiology and Emerging Carbohydrate-based Therapeutic Agents. *Curr Op Chem Bio* 2: 49-61.
148. Oyelaran O, McShane LM, Dodd L, Gildersleeve JC (2009) Profiling Human Serum Antibodies with a Carbohydrate Antigen Microarray. *J Proteome Res* 8: 4301-4310.
149. Beenhouwer DO, May RJ, Valadon P, Scharff MD (2002) High Affinity Mimotope of the Polysaccharide Capsule of *Cryptococcus neoformans* Identified from an Evolutionary Phage Peptide Library<sup>1</sup>. *J Immunol* 169: 6992-6999.
150. Bhunia A, Vivekanandan S, Eckert T, Burg-Roderfeld M, Wechselberger R, et al. (2009) Why Structurally Different Cyclic Peptides Can Be Glycomimetics of the HNK-1 Carbohydrate Antigen. *J Am Chem Soc* 132: 96-105.
151. Johnson MA, Pinto BM (2008) Structural and Functional Studies of Peptide–Carbohydrate Mimicry. *Top Curr Chem* 273: 55-116.
152. Agostino M, Sandrin MS, Thompson PE, Ramsland PA, Yuriev E (2011) Peptide Inhibitors of Xenoreactive Antibodies Mimic the Interaction Profile of the Native Carbohydrate Antigens. *Pep Sci* 96: 193-206.
153. Nakanishi T, Tsumoto K, Yokota A, Kondo H, Kumagai I (2008) Critical Contribution of VH–VL Interaction to Reshaping of an Antibody: The Case of Humanization of Anti-Lysozyme Antibody, HyHEL-10. *Prot Sci* 17: 261-270.
154. Luft JR, Collins RJ, Fehrman NA, Lauricella AM, Veatch CK, et al. (2003) A Deliberate Approach to Screening for Initial Crystallization Conditions of Biological Macromolecules. *J Struct Bio* 142: 170-179.
155. Otwinowski Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Meth Enzym* 276: 307-326.

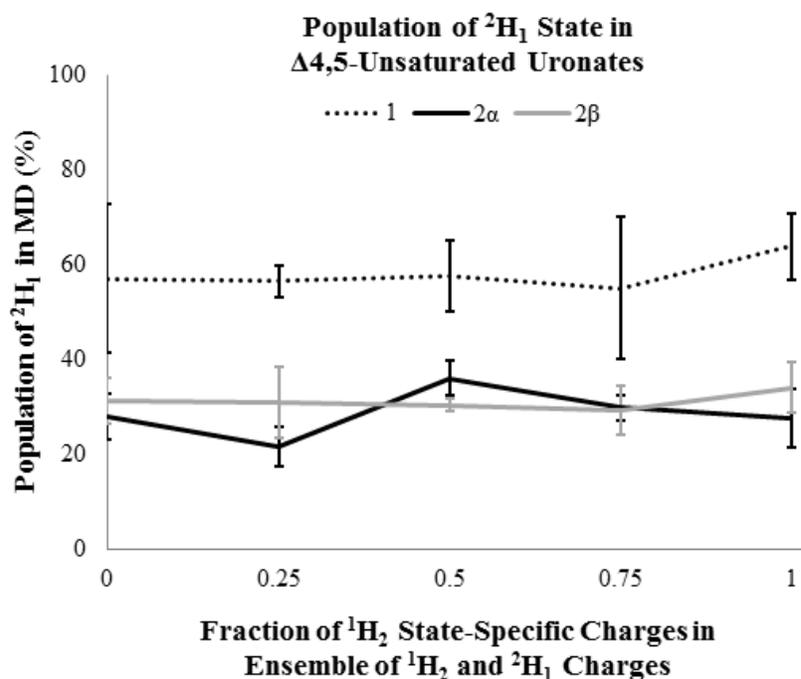
156. Collaborative Computational Project N (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr D* 50: 760-763.
157. Vagin A, Teplyakov A (2010) Molecular Replacement with MOLREP. *Acta Cryst D* D66: 22-25.
158. Jeffrey PD, Bajorath J, Chang CYY, Yelton D, Hellstrom I, et al. (1995) The X-ray Structure of an Anti-Tumour Antibody in Complex With Antigen. *Nature Struct Biol* 2: 466-471.
159. Emsley P, Cowtan K (2004) *Coot*: Model-Building Tools for Molecular Graphics. *Acta Cryst D* D60: 2126-2132.
160. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst D* D53: 240-255.
161. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: An Open Source Platform For Ligand Pocket Detection. *BMC Bioinform* 10: 168.
162. Hwang T-L, Shaka AJ (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Waveforms and Pulsed Field Gradients. *J Mag Res A* 112: 275-279.
163. Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* 123: 6108-6117.
164. (2011) MestRe Nova. 7.0.1 ed. Santiago de Compostela, Spain: Mestrelab, Inc. pp. Mnova software package.
165. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, et al. (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J Am Chem Soc* 106: 765-784.

166. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity - A rapid access to atomic charges. *Tetrahedron* 36: 3219 - 3228.
167. Bukowski R, Morris LM, Woods RJ, Weimar T (2001) Synthesis and Conformational Analysis of the T-Antigen Disaccharide ( $\beta$ -D-Gal-(1 $\rightarrow$ 3)- $\alpha$ -D-GalNAc-O-Me). *Eur J Org Chem* 2001: 2697-2705.
168. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J Comp Chem* 19: 1639-1662.
169. Poppe L, Stuike-Prill R, Meyer B, van Halbeek H (1992) The Solution Conformation of Sialyl-et(2 6)-lactose studied by Modern NMR Techniques and Monte Carlo Simulations. *JBiomol* 2: 109-136.
170. Bondi A (1964) Van der Waals Volumes and Radii. *J Phys Chem* 68: 441-445.
171. Song X, Lasanajak Y, Xia B, Heimburg-Molinaro J, Rhea JM, et al. (2010) Shotgun glycomics: a microarray strategy for functional glycomics. *Nature methods* 8: 85-90.
172. Case D, Darden T, Cheatham Iii T, Simmerling C, Wang J, et al. (2010) Amber Tools 1.5. AMBER 11, University of California, San Francisco, CA.

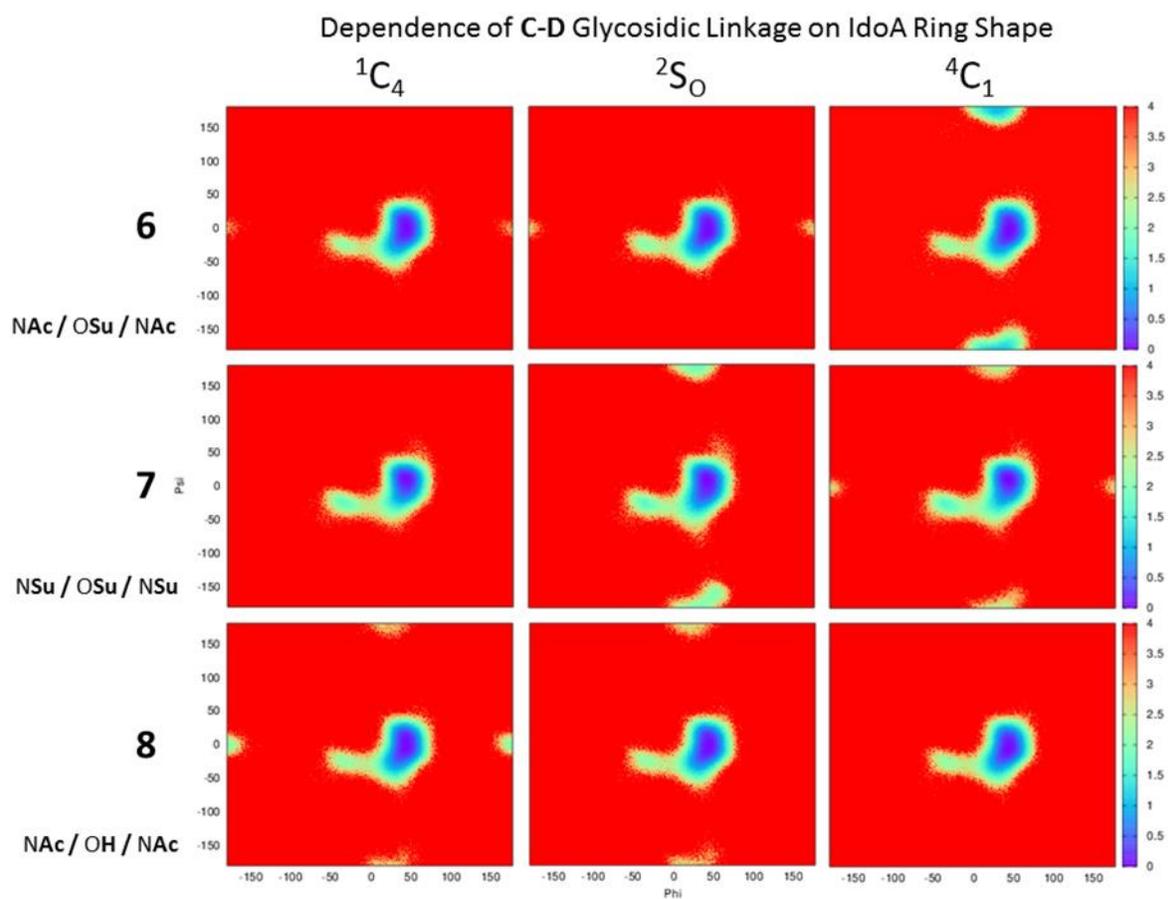
## 7. APPENDIX

### 7.1 Conformational analysis of glycosaminoglycans by NMR and molecular dynamics simulations using parameter extensions for the GLYCAM force field

#### 7.1.1 Supplementary Figures



**Figure 7.1** Population of ring states as a function of the fractional contribution of the charge model going from 0%  ${}^1\text{H}_2$  (100%  ${}^2\text{H}_1$ ) to 100%  ${}^1\text{H}_2$  (0%  ${}^2\text{H}_1$ ). Shown are disaccharides **1** and **2** with error bars representing the standard deviation of the population based on the four simulations at each charge fraction set. MD simulation populations are obtained directly from trajectory analysis, not by the computed J-couplings.



**Figure 7.2** The Boltzmann relative energy heat maps for the glycosidic linkage ( $\varphi, \psi$ ) between residues C-D for polysaccharides **6-8**.

### 7.1.2 Supplementary Equations

#### Equation 7.1

$${}^3J_{HH} = 13.24 \cos^2 \theta - 0.91 \cos \theta + \sum \Delta\chi_i (0.53 - 2.41 \cos^2(\xi_i \theta + 0.27|\Delta\chi_i|))$$

### 7.1.3 Supplementary Tables

**Table 7.1** Bond & Angle terms for GLYCAM13

Bond	Target Minimum	Fit Value	$k_r$ (kcal/mol)
C-Ck	1.494	1.466	214.0
S-N	1.638	1.675	238.0
S-Os	*	1.589	206.0
S-O2	*	1.440	620.0
Cg-Oa**		1.460	285.0
Cg-Oe**		1.460	285.0

Angle	Fit Target (°)	Parameter (°)	$k_\theta$ (kcal/mol)
C-Ck-Ck	124.3	126.0	48.0
O2-C-Ck	117.9	113.4	49.5
O2-S-N	106.4	108.0	84.0
O2-S-O2	*	113.9	123.0
O2-S-Os	*	106.9	104.0
OS-Ck-C	111.9	109.5	73.0
H1-Cg-N3†		109.5	50.0
H-N-S	114.0‡	121.2	37.5
S-N-Cg	118.4	110.0	31.0
S-Os-Cg	*	118.9	50.0
C-Cg-Oa**		112.4	63.0
Cg-Cg-Oa**		108.5	70.0
Cg-Cg-Oe**		108.5	70.0
Cg-Oa-Cg**		111.6	50.0
Cg-Oe-Cg**		111.6	50.0
Oh-Cg-Oa**		112.0	110.0
Os-Cg-Oa**		112.0	100.0
Os-Cg-Oe**		112.0	100.0
Oa-Cg-Oe**		112.0	100.0
H1-Cg-Oa**		110.0	60.0
H1-Cg-Oe**		110.0	60.0
H2-Cg-Oa**		110.0	60.0
H2-Cg-Oe**		110.0	60.0

\*Equilibrium value from Theochem 395/396 (1997) pp 107-122, \*\*Copied from GLYCAM06[24] †Copied from parm99 HP-CT-N3 [113] ‡From HF/6-31++g(2d,2p) ethylsulfamate.

**Table 7.2** Torsion and improper torsion terms for GLYCAM06 & GLYCAM11

Torsion	Error	$V_1$	$V_2$	$V_3$	Molecule
Cg-Os-Ck-C	1.0		-1.40	0.67	2-Methoxyacrylate
Cg-Ck-Ck-C	-		-11.00		But-2-enoate
Oh-Cg-Cg-Ck	0.4	-1.06	-0.60	-0.11	3-Butenol*
Os-Cg-Os-Ck	0.8	0.75	0.50		Molecule 4
Os-Cg-Cg-Ck	0.3 (0.4)	-0.27	0.40		2-(Methoxymethoxy) acrylate
O2-C-Ck-Ck	0.5		-0.80		4-Methoxybut-1-ene*
O2-C-Ck-Os	0.5		-0.80		Molecule 3
O2-S-N-Cg	0.5			0.11	(2R,3S,4S)-3,4-dihydroxy-2-methoxy-3,4-dihydro-2H-pyran-6-carboxylate**
					(2R,3S,4S)-3,4-dihydroxy-2-methoxy-3,4-dihydro-2H-pyran-6-carboxylate**
					Methylsulfamate

Ha-Ck-Ck-C	-		-11.00		Acrylate
H1-Cg-N-S	-	2.00			((2S,3R)-2-methoxytetrahydro-2H-pyran-3-yl)sulfamate†
H1-Cg-N3-H	-			0.16	copy of X-CT-N3-X [113]
H2-Cg-Os-Ck	-	1.00	0.50	0.10	2-(1-Methoxyethoxy) acrylate
H2-Cg-Cg-N3	-			0.10	2-(Methoxymethoxy) acrylate
S-N-Cg-Cg	2.5	0.50	0.50		Copy of H2-CT-CT-N (parm99)[113]
Oa-Cg-Cg-Cg‡	-	0.19	-0.11	0.14	((2S,3R)-2-methoxytetrahydro-2H-pyran-3-yl)sulfamate†
Oe-Cg-Cg-Cg	-	-0.27			1,1-dimethoxypropane[28]
Oa-Cg-Cg-Oe	-		0.40		Copy of Os-Cg-Cg-Cg (GLYCAM06)[24]
Oe-Cg-Cg-Oe	-		0.40		Copy of Os-Cg-Cg-Os (GLYCAM06g)(unpublished)
Oa-Cg-Cg-Os	-		0.40		Copy of Os-Cg-Cg-Os (GLYCAM06g)(unpublished)
Oe-Cg-Cg-Os	-		0.40		Copy of Os-Cg-Cg-Os (GLYCAM06g)(unpublished)
Oa-Cg-Cg-Oh	-	-1.10	0.25		Copy of Os-Cg-Cg-Oh (GLYCAM06)[24]
Oe-Cg-Cg-Oh	-	-1.10	0.25		Copy of Os-Cg-Cg-Oh (GLYCAM06)[24]
Oa-Cg-Oe-Cg	-	1.08	1.38	0.96	Copy of Os-Cg-Os-Cg (GLYCAM06e)(unpublished)
Oa-Cg-Os-Cg	-	1.08	1.38	0.96	Copy of Os-Cg-Os-Cg (GLYCAM06e)(unpublished)
Os-Cg-Oa-Cg	-	1.08	1.38	0.96	Copy of Os-Cg-Os-Cg (GLYCAM06e)(unpublished)
Oe-Cg-Oa-Cg‡	-	-0.25	0.76	1.20	Dimethoxymethane[28]
O2-C-Cg-Oa	-	0.04	-1.45	0.04	Copy of O2-C-Cg-Os (GLYCAM06e)(unpublished)
C-Cg-Oa-Cg	-	-0.60	0.45	0.32	Copy of C-Cg-Os-Cg (GLYCAM06e)(unpublished)
Cg-Oa-Cg-Cg‡	-	-0.70	-0.30	-0.33	2,2-dimethoxypropane[28]
Cg-Oe-Cg-Cg	-			0.16	Copy of Cg-Os-Cg-Cg (GLYCAM06)[24]
N-Cg-Cg-Oa	-	-1.30			Copy of N-Cg-Cg-Os (GLYCAM06)[24]
N-Cg-Cg-Oe	-	-1.30			Copy of N-Cg-Cg-Os (GLYCAM06)[24]
H1-Cg-Cg-Oa	-			0.05	Copy of N-Cg-Cg-Os (GLYCAM06)[24]
H1-Cg-Cg-Oe	-			0.05	Copy of H1-Cg-Cg-Os (GLYCAM06)[24]
H1-Cg-Oa-Cg	-			0.27	Copy of H1-Cg-Os-Cg (GLYCAM06)[24]
H1-Cg-Oe-Cg	-			0.27	Copy of H1-Cg-Os-Cg (GLYCAM06)[24]
H2-Cg-Cg-Oe	-			0.05	Copy of H2-Cg-Cg-Os (GLYCAM06)[24]
H2-Cg-Oa-Cg‡	-	-1.20	0.10	0.02	1,1-dimethoxyethane[28]
H2-Cg-Oe-Cg	-		0.60	0.10	Copy of H2-Cg-Os-Cg (GLYCAM06)[24]

V6

H-N-S-O2			-0.10		Methylsulfamate, Ethylsulfamate
Improper Torsion	V <sub>2</sub>				
Ck-Ha-Ck-Cg	1.0				
Os-C-Ck-Ck	1.0				
Cg-S-N-H	5.5				

\* Used for an initial fit, \*\* Fitting equally split between O2-C-Ck-Ck & O2-C-Ck-Os, †Fitting to both 2S and 2R epimers, ‡Published in Sattelle et al.[28]

**Table 7.3**  $\Delta\chi_i$  Substituents by torsion for  $\Delta 4,5$ -unsaturated uronates for **Equation 7.1**

Substituent*	$\xi_i$	H1-C1-C2-H2	H2-C2-C3-H3	H3-C3-C4-H4
S <sub>1</sub>	1	1.25 O2 (-OH, -OSO <sub>3</sub> <sup>-</sup> )	1.25 O3 (-OH)	1.25 O3 (-OH, -OSO <sub>3</sub> <sup>-</sup> )
S <sub>2</sub>	-1	0.35 C3 (-C)	0.35 C4 (-C)	0.35 C2 (-C)
S <sub>3</sub>	1	1.26 O1 (-OR)	0.35 C1 (-C)	0.35 C4 (-C)
S <sub>4</sub>	-1	1.26 O5 (-OR)	1.25 O2 (-OH)	0.35 C4 (-C)

\*Substituent numbers defined in Haasnoot et al.[118]

**Table 7.4** Ensemble averaged charges for new residues.Transferable *N*- & *O*- Sulfate

Atom	Type*	Charge (e <sup>-</sup> )
S1	S	1.245
O[1,2,3]	O2	-0.694
N/O <sub>linking</sub>	N/OS	q <sub>linking</sub> + 0.031

 $\Delta 4,5$ -unsaturated uronate

Atom	Type*	Charge (e <sup>-</sup> )
C1	Cg	0.241
C2	Cg	0.379
O2	Oh	-0.718
H2O	Ho	0.392
C3	Cg	0.301
O3	Oh	-0.718
H3O	Ho	0.407
C4	CK**	-0.110
C5	CK**	-0.006
O5	Oa	-0.250
C6	C	0.936
O6[A,B]	O2	-0.830

 $\alpha/\beta$ -D-Glucosamine

Atom	Type*	Charge	[ $\alpha$ -GlcNH <sub>3</sub> <sup>+</sup> ]-[ $\alpha$ -X]		Charge	[ $\beta$ -GlcNH <sub>3</sub> <sup>+</sup> ]-[ $\beta$ -X]			
		(e <sup>-</sup> )	X=	GlcNAc	Glc	(e <sup>-</sup> )	X=	GlcNAc	Glc

C1	Cg	0.529 (0.031)	0.061	0.020	0.458 (0.037)	0.171	0.074
C2	Cg	0.180 (0.051)	-0.065	-0.066	0.267 (0.054)	-0.213	-0.043
N2	N3	-0.085 (0.051)			-0.140 (0.085)		
H[1,2,3]N	H	0.231 ( $<0.001$ )			0.247 (0.023)		
C3	Cg	0.212 (0.081)	0.047	-0.074	0.123 (0.067)	-0.057	-0.161
O3	Oh	-0.641 (0.024)	-0.011	0.058	-0.641 (0.020)	0.040	0.068
H3O	Ho	0.436 ( $<0.001$ )	0.023	0.009	0.448 ( $<0.001$ )	0.025	0.016
C4	Cg	0.257 (0.094)	-0.065	0.003	0.326 (0.101)	0.024	0.050
O4	Oh	-0.696 (0.029)	0.037	0.014	-0.663 (0.027)	0.053	0.051
H4O	Ho	0.457 (0.018)	0.008	0.021	0.439 ( $<0.001$ )	0.003	-0.001
C5	Cg	0.345 (0.089)	0.099	0.062	0.151 (0.098)	-0.057	-0.074
O5	Oa	-0.547 (0.055)	0.021	0.027	-0.407 (0.062)	0.026	0.064
C6	Cg	0.294 (0.048)	-0.034	0.018	0.330 (0.048)	0.041	0.048
O6	Oh	-0.671 (0.050)	0.017	0.011	-0.674 (0.040)	0.015	0.014
H6O	Ho	0.431 (0.035)	0.010	0.013	0.436 (0.028)	0.012	0.012

\*Under the new GLYCAM atom-type convention, the first letter is capitalized and the second letter is lower-case. This allows more atom types to be used and modified without conflicting with other AMBER force fields[172]. \*\* Note that double bond atom-type pairs Ck=Ck are interchangeable with Cj=Cj. The redundancy is to allow for a special set of parameters to describe a single bond between two double-bonds where this single bond would be defined by Cj-Ck.

**Table 7.5** Absolute average errors between NMR and MD simulations for GAGs 1-7

	Average NOE Errors (Å)				
	1	2 <sub><math>\alpha,\beta</math></sub>	3 <sub><math>\alpha,\beta</math></sub>	4	5
Intra-ring	0.17	0.50	0.20	0.30	0.30

<b>A</b> ( <i>d4,5uu</i> )	0.25	0.44	0.23	0.14	0.05
<b>B</b> (GlcNX)	0.15	0.52	0.20	0.35	0.36
Inter-ring	0.38	0.59	0.46	0.40	0.38
Overall	0.23	0.51	0.24	0.34	0.32

	<b>6</b>			<b>7</b>				
	<sup>1</sup> C <sub>4</sub>	<sup>2</sup> S <sub>0</sub>	<sup>4</sup> C <sub>1</sub>	<sup>1</sup> C <sub>4</sub>	<sup>2</sup> S <sub>0</sub>	<sup>4</sup> C <sub>1</sub>	60:40:0	70:30:0
Intra-ring (C, D, F)	0.25	0.25	0.25	0.21	0.21	0.21	0.21	0.21
Inter-ring	0.37	0.40	0.59	0.34	0.62	0.53	0.31	0.42
Overall	0.31	0.33	0.43	0.28	0.43	0.39	0.26	0.33

**Table 7.6** NOE values for disaccharides **1-5** from NMR and MD simulations measured in Å.

<b>A</b> ( <i>d45uu</i> )	<b>A</b> ( <i>d45uu</i> )	<i>1<sub>NMR</sub></i>	<i>1<sub>MD</sub></i>	<i>2<sub>NMR</sub></i> *	<i>2<sub>MD,α</sub></i>	<i>2<sub>MD,β</sub></i>	<i>3<sub>NMR</sub></i> *	<i>3<sub>MD,α</sub></i>	<i>3<sub>MD,β</sub></i>	<i>4<sub>NMR</sub></i>	<i>4<sub>MD</sub></i>	<i>5<sub>NMR</sub></i>	<i>5<sub>MD</sub></i>
H <sub>1</sub>	H <sub>2</sub>	2.1	2.8	2.4	2.7	2.8	2.4	2.6	2.6	2.5	2.6	2.6	2.6
H <sub>1</sub>	H <sub>3</sub>	2.8	3.1	2.6	3.3	3.2	2.5	3.5	3.7	3.1	3.6		
H <sub>2</sub>	H <sub>3</sub>	3.2	2.8	2.5	2.7	2.8	2.3	2.7	2.7	2.5	2.7	2.7	2.7
H <sub>2</sub>	H <sub>4</sub>		3.9	4.5	4.1	3.7		4.1	4.2		4.2		
H <sub>3</sub>	H <sub>4</sub>	2.6	2.5		2.5	2.6	2.4	2.5	2.5	2.6	2.5	2.6	2.5

<b>A</b> ( <i>d45uu</i> )	<b>B</b> (GlcNx)	<i>1<sub>NMR</sub></i>	<i>1<sub>MD</sub></i>	<i>2<sub>NMR</sub></i> *	<i>2<sub>MD,α</sub></i>	<i>2<sub>MD,β</sub></i>	<i>3<sub>NMR</sub></i> *	<i>3<sub>MD,α</sub></i>	<i>3<sub>MD,β</sub></i>	<i>4<sub>NMR</sub></i>	<i>4<sub>MD</sub></i>	<i>5<sub>NMR</sub></i>	<i>5<sub>MD</sub></i>
H <sub>1</sub>	H <sub>3</sub>	3.0	4.2	3.1	4.4	4.3		3.5	3.4	2.7	3.3	2.4	2.9
H <sub>1</sub>	H <sub>4</sub>	2.3	2.4	1.9	2.4	2.4	2.0	2.4	2.6	2.2	2.4	2.5	2.6
H <sub>1</sub>	H <sub>5</sub>		3.9	2.8	3.9	3.9		3.6	3.5	3.1	3.6		
H <sub>1</sub>	H <sub>6</sub>	2.5	2.7	2.6	2.6	2.6		3.1	3.0	2.4	2.9	2.9	3.4
H <sub>1</sub>	H <sub>6'</sub>	2.8	2.8	2.9	3.1	3.0		2.6	3.0	2.4	2.6	2.5	2.8

<b>B</b> (GlcNx)	<b>B</b> (GlcNx)	<i>1<sub>NMR</sub></i>	<i>1<sub>MD</sub></i>	<i>2<sub>NMR</sub></i> *	<i>2<sub>MD,α</sub></i>	<i>2<sub>MD,β</sub></i>	<i>3<sub>NMR</sub></i> *	<i>3<sub>MD,α</sub></i>	<i>3<sub>MD,β</sub></i>	<i>4<sub>NMR</sub></i>	<i>4<sub>MD</sub></i>	<i>5<sub>NMR</sub></i>	<i>5<sub>MD</sub></i>
H <sub>1</sub>	H <sub>2</sub>	2.4	2.4	2.3	2.4	3.0	2.4	2.4	3.0	2.4	2.4	2.4	2.4
H <sub>1</sub>	H <sub>4</sub>		4.1	3.1	4.2	4.1		4.1	4.0		4.1		
H <sub>1</sub>	H <sub>5</sub>		3.7	3.1	3.8	2.5		3.8	2.5		3.7		
H <sub>2</sub>	H <sub>3</sub>	3.5	3.0	3.0	3.0	3.0		3.0	3.0	3.2	3.0	4.7	3.0
H <sub>2</sub>	H <sub>4</sub>	2.6	2.6	2.6	2.6	2.7		2.6	2.6	2.6	2.6	2.6	2.6
H <sub>3</sub>	H <sub>4</sub>	2.7	3.0	3.4	3.0	3.0		3.0	3.0		3.0		
H <sub>3</sub>	H <sub>5</sub>	2.5	2.6	2.4	2.6	2.6	2.3	2.6	2.6	2.5	2.6	2.6	2.7
H <sub>3</sub>	H <sub>6'</sub>		4.6	3.3	4.6	4.6		4.6	4.6	3.3	4.6		
H <sub>4</sub>	H <sub>5</sub>	3.1	3.0	3.1	3.0	3.0	3.1	3.0	3.0		3.0	3.4	3.0
H <sub>4</sub>	H <sub>6</sub>		2.9	2.6	3.1	3.1		3.1	3.0	2.7	3.1	2.7	3.1
H <sub>4</sub>	H <sub>6'</sub>		3.1	2.5	3.1	3.0		3.0	3.0				
H <sub>5</sub>	H <sub>6</sub>	2.4	2.5	2.8	2.5	2.5	2.5	2.5	2.5	2.2	2.5	2.4	2.5

H <sub>5</sub>	H <sub>6</sub>	2.5	2.6	2.1	2.4	2.5	2.4	2.5				
H <sub>6</sub>	H <sub>6</sub>	1.9	1.8	1.9	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8

\*NMR experimental NOEs represent the average from both anomers.

**Table 7.7** NOE values for tetrasaccharides **6** & **7** from NMR and MD simulations (<sup>1</sup>C<sub>4</sub>, <sup>2</sup>S<sub>0</sub>, <sup>4</sup>C<sub>1</sub>) measured in Å.

		<b>6</b> (x = Ac)				<b>7</b> (x = SO <sub>3</sub> <sup>-</sup> )			
<b>C</b> (GlcA)	<b>C</b> (GlcA)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H1	H2	3.6	3.0	3.0	3.0	3.5	3.0	3.0	3.0
H1	H3	3.0	2.7	2.7	2.7	3.0	2.7	2.7	2.7
H1	H5	2.6	2.5	2.5	2.5	2.5	2.5	2.5	2.5
H3	H5	2.6	2.7	2.7	2.7	2.9	2.7	2.7	2.7

<b>D</b> (GlcNx)	<b>C</b> (GlcA)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H4	H1	2.8	2.3	2.3	2.4	2.9	2.4	2.4	2.4
H6'	H1	2.8	2.8	2.8	2.9	2.8	2.6	2.7	2.7

<b>D</b> (GlcNx)	<b>D</b> (GlcNx)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H1	H2	2.5	2.4	2.4	2.4	2.5	2.4	2.4	2.4

<b>E</b> (IdoA)	<b>D</b> (GlcNx)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H2	H1	4.2	4.7	4.8	4.6				
H3	H1	2.5	2.4	2.6	3.7	2.7	2.8	4.0	4.1
H4	H1	3.3	2.5	2.5	2.3	2.7	2.3	2.3	2.3
H5	H4	2.7	3.1	3.2	3.6				

<b>E</b> (IdoA)	<b>E</b> (IdoA)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H1	H2	3.4	2.6	3.0	3.0	3.4	2.6	3.0	3.0
H1	H3	4.1	4.2	3.8	2.7	3.4	4.2	2.9	2.7
H2	H3	2.4	2.5	3.0	3.0	2.4	2.5	3.0	3.0
H2	H4	3.5	4.2	3.2	2.7	3.4	4.2	3.4	2.7
H3	H4	3.6	2.5	2.9	3.0	2.9	2.5	2.7	3.0
H4	H5	3.6	2.4	2.3	2.3	2.8	2.4	2.4	2.3

<b>F</b> (GlcNx)	<b>E</b> (IdoA)	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>	<b>NMR</b>	<sup>1</sup> <b>C<sub>4</sub></b>	<sup>2</sup> <b>S<sub>0</sub></b>	<sup>4</sup> <b>C<sub>1</sub></b>
H3	H1					2.6	3.5	4.1	3.4
H4	H1	2.7	2.3	2.4	2.3	2.6	2.3	2.3	2.3
H6'	H1	2.7	3.0	3.0	3.1	2.7	2.8	2.9	2.9

<b>F</b> (GlcNx)	<b>F</b> (GlcNx)	<b>NMR</b>	<b><sup>1</sup>C<sub>4</sub></b>	<b><sup>2</sup>S<sub>O</sub></b>	<b><sup>4</sup>C<sub>1</sub></b>	<b>NMR</b>	<b><sup>1</sup>C<sub>4</sub></b>	<b><sup>2</sup>S<sub>O</sub></b>	<b><sup>4</sup>C<sub>1</sub></b>
H1	H2	2.7	2.4	2.4	2.4	2.6	2.4	2.4	2.4
H1	H3	3.6	3.8	3.8	3.8				

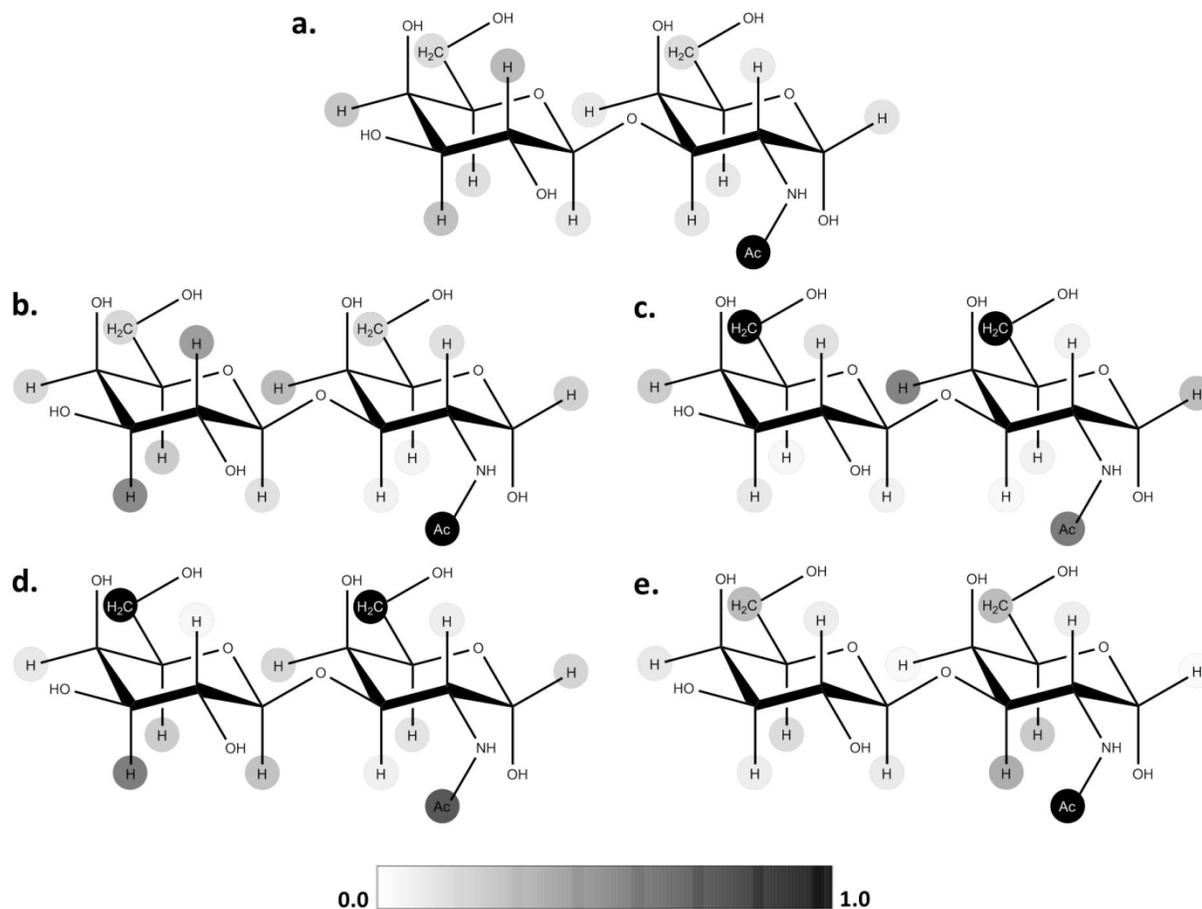
## 7.2 Computational screening of the human TF-glycome provides a structural definition for the specificity of anti-tumor antibody JAA-F11

### 7.2.1 Supplementary Results

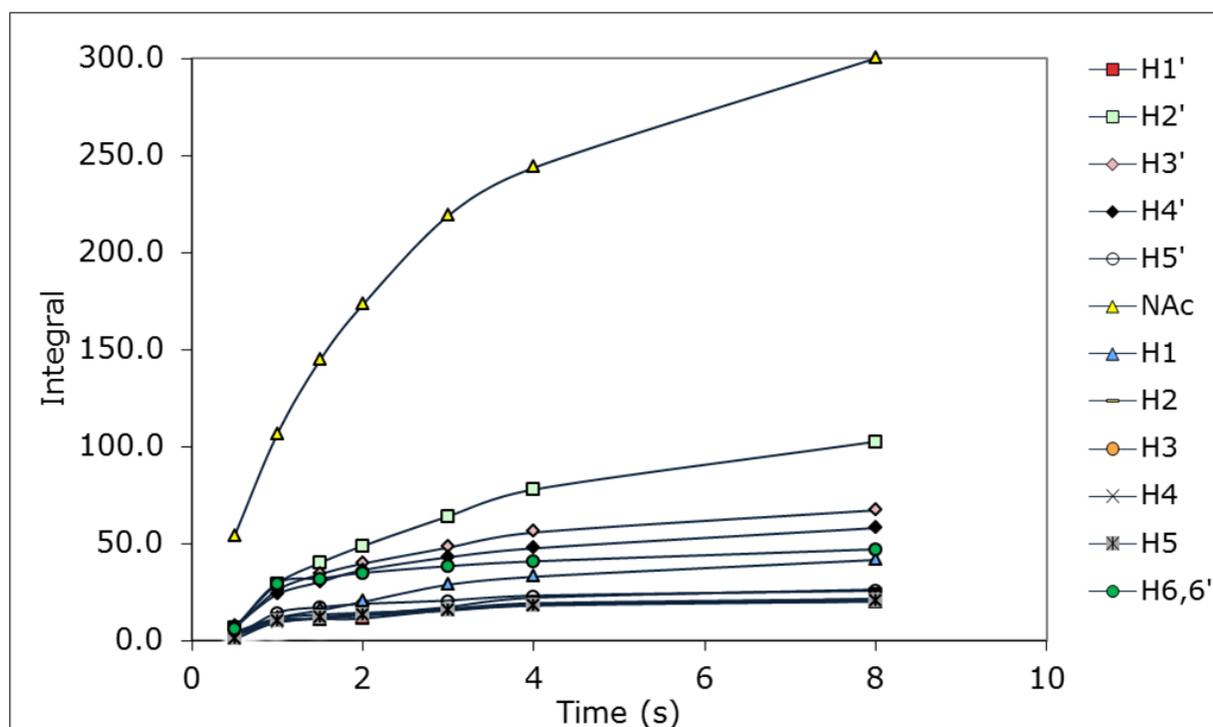
#### *Corrections or Clarifications Associated with CFG v4.0 Glycan Array Annotations*

- 1) When an anomeric center is not specified at the reducing end, the anomeric configuration is either not known or is present as a mixture.
- 2) The sequence for ligand **158** (CFG v4.0 ID) is Gal $\beta$ 1-3(Gal $\beta$ 1-4GlcNAc $\beta$ 1-6)GalNAc-**Sp14**. The anomeric center is undetermined, or is a mixture, the spacer is number 14.
- 3) The sequence for ligand **159** is Gal $\beta$ 1-3(Gal $\beta$ 1-4GlcNAc $\beta$ 1-6)GalNAc $\alpha$ -**Sp8**. The anomeric center is  $\alpha$ , the spacer is 8.
- 4) Ligands **157** and **159** are identical, Gal $\beta$ 1-3(Gal $\beta$ 1-4GlcNAc $\beta$ 1-6)GalNAc $\alpha$ -Sp8.
- 5) Ligands **125** and **182** are identical, Gal $\beta$ 1-3(GlcNAc $\beta$ 1-6)GalNAc $\alpha$ -Sp8.

## 7.2.2 Supplementary Figures



**Figure 7.3** Normalized experimental STD intensities for the TF disaccharide bound to JAA-F11 Fab (a). Theoretical STD intensities for the top four poses from docking (b-e); pose 1(b) shows the highest consistency with experimental STD data.



**Figure 7.4** STD-NMR integral build up curves for a 200:1 mixture of Gal $\beta$ 1-3GalNAc:JAA-F11 antibody, normalized to the intensity from the GalNAc N-Acetyl methyl group protons. The experiment employed free disaccharide, and only the binding of the TF- $\alpha$  disaccharide was observed.

### 7.2.3 Supplementary Tables

**Table 7.8** Docking results summary for the four pose clusters identified by docking the TF-disaccharide (Gal $\beta$ 1-3GalNAc $\alpha$ -OMe) to the FAB fragment of JAA-F11. The lowest energy conformer from each cluster was used as a representative pose (Poses 1 – 4) in the CCG analysis.

Cluster Rank <sup>a</sup>	1	2	3	4
Population (%)	64	10	6	12
Average docked energy <sup>b</sup>	-8.8	-8.0	-7.3	-6.7

<sup>a</sup>Based on a 2.0 Å cutoff. <sup>b</sup>kcal mol<sup>-1</sup>.

**Table 7.9** Experimental values for the glycan array screening of JAA-F11. Shown are the glycans containing the minimal binding determinant, Gal $\beta$ 1-3GalNAc $\alpha$ , and sequences that characterize the specificity of the mAb.

	Glycan <sup>a</sup>	Sp <sup>b</sup>	Mean Relative Fluorescence units (RFUs)		
			0.1 $\mu$ g/mL	5 $\mu$ g/mL	200 $\mu$ g/mL
1	<b>Gal<math>\beta</math>1-3GalNAc<math>\alpha</math></b>	8	16223	39364	58481
		14	8	30	14
		16	15	28	15565
2	Neu5Ac $\beta$ 2-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8	12388	33481	52238
3	Neu5Ac $\alpha$ 2-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8	13105	32998	45410
		14	7	11	6
4	Gal $\beta$ 1-4GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8 <sup>c</sup>	2023	18060	62789
		14 <sup>d</sup>	-2	11	1
5	GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8	2357	16637	61857
		14 <sup>d</sup>	-1	12	8
6	Gal $\alpha$ 1-3GalNAc $\alpha$	8	17	256	45611
		16	2	8	8
7	Gal $\beta$	8	12	2	45
10	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	8	-5	17	94
		14	4	31	24
11	GlcNAc $\beta$ 1-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	8	17	29	53
12	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4(Fuca1-3)GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	14	31	48	42
13	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-6( <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	14	-1	21	21
14	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4Glc $\beta$	0	5	23	20
15	GlcNAc $\alpha$ 1-4 <b>Gal<math>\beta</math>1-3</b> GalNAc	14 <sup>d</sup>	10	48	18
16	GlcNAc $\beta$ 1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	8	5	23	17
17	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4(Fuca1-3)GlcNAc $\beta$ 1-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc	14 <sup>d</sup>	38	35	14
18	GalNAc $\alpha$ 1-3(Fuca1-2) <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc $\beta$	0	1	8	14
19	<b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc	0 <sup>d</sup>	8	10	14
20	<b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4Glc	0 <sup>d</sup>	13	19	13
21	(3S) <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	8	-2	6	12
22	Fuca1-2 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$ 1-3(Fuca1-2)Gal $\beta$ 1-4GlcNAc $\beta$	0	1	17	10
23	Neu5Ac $\alpha$ 2-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8	2	23	8
		14	19	22	12
24	6S(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	8	8	19	7
25	Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	8	5	14	5
		14	4	7	6
26	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-6(Neu5Ac $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> )GalNAc $\alpha$	14	19	11	2
27	KDN $\alpha$ 2-3 <b>Gal<math>\beta</math>1-3</b> GalNAc $\alpha$	14	-8	6	0

<sup>a</sup>Binding is considered to be present if the mean relative fluorescence signal is above at least 5% of the maximum signal in the sample at 200 ug/mL. <sup>b</sup>Spacers are identified as follows: 0, –(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>; 8, –(CH<sub>2</sub>)<sub>3</sub>NH<sub>2</sub>; 14, threonine; 16, -p-nitrophenyl. <sup>c</sup>Average value for redundant glycans on the v4.0 Glycan Array. <sup>d</sup>Undefined anomeric configuration at reducing terminus.

**Table 7.10** Van der Waals overlaps from a CCG analysis of the CFG array glycans.

ID	CFG ID v4.0	Binders <sup>a</sup>	Sp <sup>b</sup>	Pose 1	Pose 2	Pose 3	Pose 4
1	131	<b>Galβ1-3GalNAca</b>	8	0	0	0	0
2	129	Neu5Acβ2-6( <b>Galβ1-3</b> )GalNAca	8	0	0.7	5.9*	0
3	127	Neu5Acα2-6( <b>Galβ1-3</b> )GalNAca	8	0.1	1.1*	10*	0
4	157,159	Galβ1-4GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAca	8	0	0	13.1*	0
5	125,182	GlcNAcβ1-6( <b>Galβ1-3</b> )GalNAca	8	0	0	5.1*	0
<b>Non-Binders<sup>a</sup></b>							
10	58	Fucα1-2 <b>Galβ1-3</b> GalNAca	8	7.6	1.2	0*	7
11	85	GlcNAcβ1-3 <b>Galβ1-3</b> GalNAca	8	12.4	12.3	0*	8.7
16	166	GlcNAcβ1-2 <b>Galβ1-3</b> GalNAca	8	11.6	3.4	0*	9.8
21	27	(3S) <b>Galβ1-3</b> GalNAca	8	2.8	0*	0*	2.3
23	233	Neu5Acα2-6(Neu5Acα2-3 <b>Galβ1-3</b> )GalNAca	8	17.6	17.7	9.4	12.4
25	232	6S(Neu5Acα2-3 <b>Galβ1-3</b> )GalNAca	8	17.3	16.7	0.7*	14.2
26	214	Neu5Acα2-3 <b>Galβ1-3</b> GalNAca	8	17.4	16.8	0*	12.4
31	91	<b>GalNAcβ1-3</b> GalNAca	8	2.5	0.3*	0*	2
32	133	<b>Galβ1-3</b> GalNAcβ-Sp8 <sup>c</sup>	8	2.2	1.3	3.1	0*
33	134	<b>Galβ1-3</b> GalNAcβ1-4Galβ1-4Glcβ	0	15.6	17	21.1	0*

Only Pose 1 is fully compatible with the experimental specificity data. Poses that lead to incompatibilities with the experimental data denoted with an asterisk (\*). <sup>a</sup>Binding or non-binding classification is based on a mean relative fluorescence signal greater than (binder) or less than (non-binder) 5% of the maximum signal for each concentration in array version 4.0. <sup>b</sup>Sp0: –(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>; Sp8: –(CH<sub>2</sub>)<sub>3</sub>NH<sub>2</sub>. <sup>c</sup>A methyl aglycon was employed in the grafting process to probe the effect of a β-linked spacer.