MODELING CARBOHYDRATES AND THEIR INTERACTIONS WITH PROTEINS

by

XIAOCONG WANG

(Under the Direction of Robert J. Woods)

ABSTRACT

Carbohydrates are ubiquitously expressed at cell surfaces and involved in various cellular interactions. Understanding the shapes and dynamic properties of carbohydrates advances our knowledge of the fundamental properties that determine the specificity and affinity of their interactions with other biomolecules that are ultimately responsible for their biological functions. In this thesis, modeling methods were developed and applied to a number of particularly challenging aspects of carbohydrates and carbohydrate-protein complexes. Firstly, a new set of force field parameters was developed for modeling the highly plastic structures of monosaccharides containing five-membered rings (furanoses) that are constituents of DNA and RNA. Oligo- and polysaccharides of furanoses are also important components of bacterial and fungal pathogen surfaces. This work illustrated that it is no longer necessary to make assumptions about ring conformational preferences in order to interpret experimental NMR data for furanoses; instead they may be determined directly and objectively from molecular dynamics (MD) simulation. Secondly, many oligo- and polysaccharides decorated with chemical modifications, such as sulfation. Sulfated carbohydrates include the well-known example of heparin, as well as may others, including a sulfated fucan isolated from sea urchin *Lytechinus variegatus*. The impact of sulfation on oligosaccharide shape has received relatively study, and

is the focus of a combined NMR and MD analysis here. That research led to the conclusion that the 3D orientation of residues in an oligosaccharide could be influenced significantly by stabilizing hydrogen bonds between sulfate and hydroxyl groups, and by destabilizing electrostatic repulsions between the anionic sulfate groups. Lastly, the binding of a sulfated glycopeptide (PSGL-1) and its analogs to P-Selectin was examined by MD simulation, with the goal of quantifying the contributions made to binding by each of the component amino acids and monosaccharides. The results provide a basis for the rational design of inhibitors for disease-related interactions associated with P-selectin.

INDEX WORDS: Carbohydrate, Force field, Molecular dynamics simulation, 3D structure, Furanoses, Ring conformation, Anomeric effect, Pseudorotation, GLYCAM, AMBER, NMR, Scalar ^3J-coupling constants, Glycan-protein interaction, MM/GBSA, MM/PBSA, Binding free energy, Sulfate fucan, Hydrogen bond, P-Selectin, Drug design.

MODELING CARBOHYDRATES AND THEIR INTERACTIONS WITH PROTEINS

by

XIAOCONG WANG

B.S., University of Science and Technology of China, 2007

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

MODELING CARBOHYDRATES AND THEIR INTERACTIONS WITH PROTEINS

by

XIAOCONG WANG

Major Professor: Robert J. Woods
Committee: James H. Prestegard
    Liming Cai

Electronic Version Approved by:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

TABLE OF CONTENTS

LIST OF FIGURES

ix

the crystallographic data for PSGL-1 (c) and leads to similar binding for GSnP-6 (d). The solvent accessible surface of P-selectin is colored according to the electrostatic potential (acidic region: red; basic region: blue). (e-g) Hydrogen bonds between Fuc (red), Core-2Gal (yellow), and Neu5Ac (purple) and P-selectin residues.

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The work done in this thesis comprise of:

1. Furanose ring conformations in solution and insights to the two-state model with furanose-specific force field parameters in GLYCAM.

2. Influences of sulfation patterns in *Lv* I to its dynamic properties.

3. Comparison of binding affinities of PSGL-1 and its analog when binding with P-selectin with molecular modeling and MM/GBSA calculations.

These topics, including the reviews of the perspective background knowledge and the computational methods applied, are presented as follows:

CHAPTER 2: CLASSICAL FORCE FIELDS, ENERGY CALCULATIONS AND MOLECULAR DYNAMICS (MD) SIMULATION

Chapter 2 provides background knowledge about for classical mechanics force fields, energy minimization, and MD simulation.

CHAPTER 3: MODELING OLIGOSACCHARIDE 3D STRUCTURES AND INTERACTIONS

Chapter 3 covers relevant background information for modeling 3D structures of oligosaccharides and predicting their binding affinities in protein complexes.

CHAPTER 4: INSIGHTS INTO FURANOSE SOLUTION CONFORMATIONS: BEYOND THE TWO-STATE MODEL

Chapter 4 is an original research study undertaken to study the ring conformations of furanoses in solution and provide insights to the two-state model that is commonly employed to interpret

furanose ring conformations from experimental results. The results from this study were submitted as a journal article.

CHAPTER 5: IMPACT OF SULFATION PATTERN ON THE CONFORMATION AND DYNAMICS OF SULFATED FUCAN OLIGOSACCHARIDES AS REVEALED BY NMR AND MD

Chapter 5 is an original research study undertaken to study the influences of sulfation patterns to the dynamic properties of *Lv* I. The theoretical results from this study compose a part of a published article.

Queiroz, I. N. L.; Wang, X. C.; Glushka, J. N.; Santos, G. R. C.; Valente, A. P.; Prestegard, J. H.; Woods, R. J.; Mourao, P. A. S.; Pomin, V. H. *Glycobiology* **2015**, *25*, 535.

Contributions to this article include theoretical calculations, computational data analysis.

CHAPTER 6: COMPUTATIONAL MODELING OF PSGL-1 AND ITS GLYCOPEPTIDE ANALOG

Chapter 6 is an original research study undertaken to predict the binding affinities of the synthetic analog of PSGL-1 bound to P-selectin. The theoretical results from this study compose a part of a published article.

Krishnamurthy, V. R.; Sardar, M. Y. R.; Ying, Y.; Song, X. Z.; Haller, C.; Dai, E. B.; Wang, X. C.; Hanjaya-Putra, D.; Sun, L. J.; Morikis, V.; Simon, S. I.; Woods, R. J.; Cummings, R. D.; Chaikof, E. L. *Nature Communication* **2015**, 6, 12.

Contributions to this article include theoretical calculations, computational data analysis.

CHAPTER 2

CLASSICAL FORCE FIELDS, ENERGY CALCULATIONS AND MD SIMULATION[1]

Carbohydrate Force Field

Molecular modeling has become a powerful tool for studying the biological processes beyond the reach of experimental measure. Quantum mechanics (QM) can describe the physical properties of a model to a high degree of accuracy, at the cost of computing efficiency[1]. However, currently, it is too difficult to acquire solutions from QM for biomolecules with thousands or more atoms, such as biopolymers, within a reasonable amount of time[1]. Moreover, incorporating molecular dynamics into QM adds to the difficulty exponentially[2]. A solution to this dilemma is to employ a model that can approximate the physical properties of molecules with high efficiency, and maintain an adequate level of accuracy to study the structural, dynamical, and energetic properties of large biomolecules. Molecular mechanics (MM) enables such models for biomolecules. MM employs classical (Newtonian) physics to simulate the motion of the model with a force field, which evaluates the potential energies as a function of its 3D structure, and it has been widely adopted in modeling macromolecular systems and shown to generate satisfactory results[3-6].

Classical mechanics force fields present a powerful tool for theoretical approaches to study biomolecules. Currently, MD simulations of biological systems with over 100,000 atoms employing classical force fields can reach a simulation time periods in the microsecond regimen. Simulation scale accessible to MD simulations is constantly increasing with the growing computer power, and most notably the use of graphical processing units (GPUs). Aside from the long and increasing simulation time periods, the accuracy of the intra- and inter-molecular interactions calculated with classical mechanics force fields decides its continuous and future success. Numerous algorithmic improvements, including Particle Mesh Ewald algorithm[7,8] to treat long range electrostatic interactions have been a major contribution towards the current

accuracy. The quality of the force field, however, may still be considered as the most important

determinant of the accuracy for the MM methods.

A typical potential energy evaluation function (force field) is shown in Equation 2.1:

$$E_{potential} = \sum_{i}^{Bonds} \frac{1}{2} k_r (r_i - r_i^0)^2 + \sum_{i}^{Angles} \frac{1}{2} k_\theta (\theta_i - \theta_i^0)^2$$

$$+ \sum_{i}^{Torsions} \sum_{j}^{Coefficients} \frac{1}{2} V_{ij} (1 + cos(j\emptyset - \delta_{ij}))$$

$$+ \sum_{i=1}^{Atoms} \sum_{j=i+1}^{Atoms} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^{6} \right] + \sum_{i=1}^{Atoms} \sum_{j=i+1}^{Atoms} \left(\frac{q_i q_j}{\varepsilon R_{ij}} \right) \qquad (2.1)$$

Potential energies from bond stretching and angle bending are treated within a parabolic

function. Such a harmonic approximation penalizes any distortion away from the equilibrium

geometry within a narrow range; it does not account for anharmonic effects, and precludes the

dissociation of an atom from another. For this reason, classical force fields are generally unable

to model chemical reactions that involve bond making or breaking, but are suited to modeling

molecular structures. The interatomic repulsion and dispersion interactions (van der Waals

interactions) are described by Lennard-Jones (LJ) 6-12 term, and the interatomic electrostatic

interactions were described by a Coulombic term, which depends on the static point charges on

each atom. These charges are not necessarily real as they are often partial atomic charges,

calculated to describe electrostatic features from differences in electron densities. These non-

bond interaction terms control the higher order of the molecular structure and inter-molecular

interactions. Potential energies of torsional rotations are described by cosine functions. Torsion

terms are often mistakenly thought of as internal rotational barriers, but in fact they are added

only to ensure that the rotational energies, which include contributions from all other force field

components, agree with expected values. Thus, torsion terms are corrections for missing non-classical contributions to bond rotational energies.

Many classical force fields share a similar mathematical formulation (such as CHARMM, GROMOS, AMBER), but differ in the constants employed in the equations. These constants are called parameters, and include bond and angle force constants (k_r and k_θ, respectively), and equilibrium values (r_i^0 and θ_i^0, respectively); dihedral energy force constants V_{ij}, together with periodicity and phase angle values (j and δ_{ij}, respectively). Van der Waals parameters between atoms i and j, include the depth of the potential well (ε_{ij}) and the finite distance at which the interatomic potential is zero (σ_{ij}). Lastly, the polar character of the molecule is modeled by assigning partial charges (q_i) to each atom.

Force Field Development in GLYCAM

Carbohydrates are exceedingly important to almost every aspect of life. Not only are they products of photosynthesis and energy sources, they are also involved in numerous aspects of cellular interactions[9-15] through a dense and complex array of oligosaccharides on the surface of all living cell in nature. The diversity of oligosaccharide structures and the similarity of monosaccharide structures add extensive difficulties for experimental approaches to characterize their structural, dynamic, and energetic properties. It is, therefore, immensely important to develop proper carbohydrate force fields, which enable molecular modeling of carbohydrates.

Developing a force field for carbohydrates, comparing to other biomolecules, faces several challenges[16]. First, in most of monosaccharides, all but one carbon center is chiral. This results enormous number of configurational and conformational possibilities. Second, MEPs differ subtly between conformational isomers and even stereoisomers of monosaccharides. It is challenging to represent these similar yet distinct MEPs between stereoisomers with atomic

6

partial charge sets; and it is even more challenging to represent MEPs between conformational isomers of the same monosaccharide with one atomic partial charge set. Third, most common carbohydrates are capable of forming both linear and branch structures because of the abundant hydroxyl groups that can connect to other carbohydrate molecules. This enriches the structural diversity of the oligosaccharides and makes the theoretical characterization of the structural and dynamic properties of oligosaccharides extremely challenging. Last but not least, the anomeric effect and its attendant *exo*-anomeric effect[17-29] determine the preferred conformation other than what is expected solely based on steric considerations[30].

GLYCAM06 is a widely recognized carbohydrate force field[31] that has parameters for all of the common monosaccharides as well as numerous analogs, and therefore can be used to model a great variety of oligosaccharides. The core philosophy of the development process of GLYCAM06 is that parameters should be: (1) be transferable to all carbohydrate ring formations, (2) be self-contained and therefore readily transferable to many quadratic force fields, (3) not require specific atom types for *α*- and *β*-anomers, (4) be readily extendible to carbohydrate derivatives and other biomolecules, (5) be applicable to monosaccharides and complex oligosaccharides, and (6) be rigorously assessed in terms of the relative accuracy of its component terms[31].

In the development of valance parameters in the GLYCAM force field, fragments or analogs of the target molecules are usually chosen to develop the valence terms. Energy profiles of these fragments or analogs were first carried with QM calculation, because QM can deliver results with high accuracy. Then, the valence parameters are derived to fit the MM energies of these fragments or analogs to their QM energy profiles. In the developments of bond length and angle parameters, average bond length or angle values are first obtained from crystallographic

data of the target molecules. The QM energy profiles for the fragments or analogs of the target molecules are created by distorting the bond length or bond angle under development within a narrow range, while optimizing the rest of the structure (Figure 2.1a and 2.1c). The force constants and equilibrium values of bond length or angle parameters are derived with a parabolic function to reproduce the quadratic shape of the QM energy profile and average crystallographic bond length or bond angle values (Figure 2.1b and 2.1d).
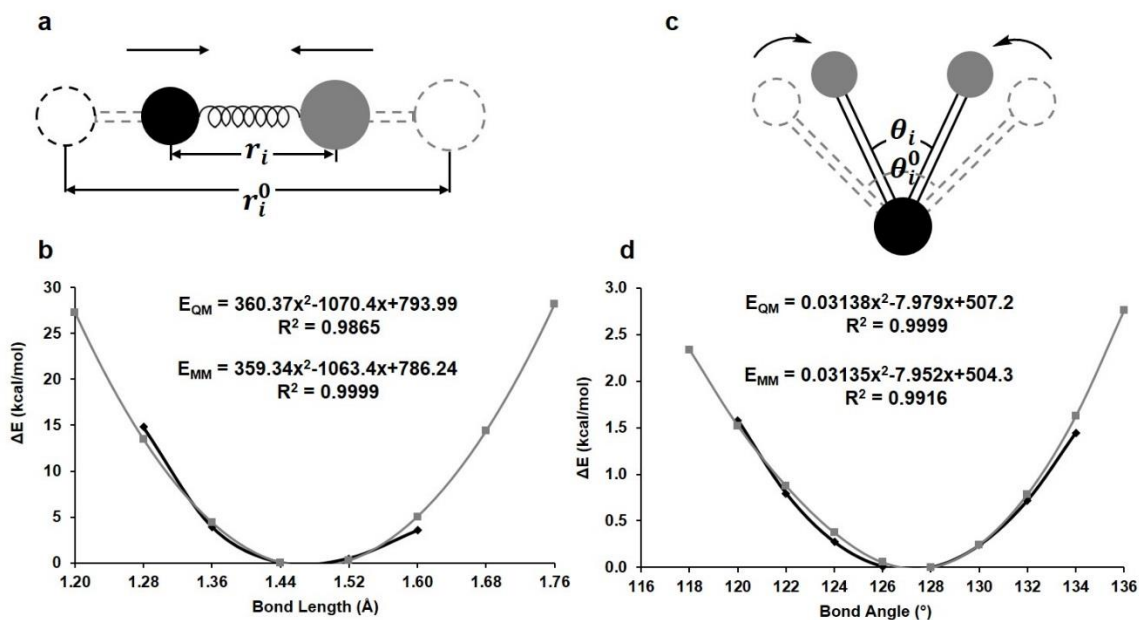
Figure 2.1 Bond length (a) and angle (c) parameters derived from fitting the MM energies (grey) to the QM energy (black) profiles (b and d).

The electrostatic properties of a molecule are strongly affected by its structural and dynamic properties and surrounding environment, therefore, the atomic charge set developed for the target molecules should reproduce their molecular electrostatic potentials (MEP) of different conformations and environment. An ensemble-averaged approach is employed in developing atomic partial charge sets in GLYCAM force field (Figure 2.2). In this approach, a MD

simulation is first perform to sample different conformations of the molecule, especially the exocyclic groups. Then, MEP calculations are performed quantum mechanically on the structures evenly extracted from the MD simulation. Later, the restrained electrostatic potential (RESP)[32] charge fitting methodology is employed to the MEP of each structure to calculate the partial atomic point charges. Finally, the ensemble-averaged atomic charges that represent the MEP of the molecule in solution are achieved by averaging the atomic charges from all frames. The quality of the ensemble-averaged approach depends on not only the accuracy of MEP calculation and RESP fitting methodology, but also the efficiency of the sampling in the MD simulation, especially those low-frequency motions, such as ring flip in pyranoses. To improve the quality of the ensemble-averaged atomic charges, the MD simulations can be performed by restraining the conformation of the molecule to the pre-identified states. Then, the final charge set is computed by weighting charges set of each state with its corresponding population observed from experimental measurements or extensive MD simulations.

Initial structures
(obtained from crystallographic data or QM optimized structures)

↓ explicit solvent MD simulation

ensemble structures
(extracted evenly from MD simulation)

↓ QM MEP calculation + RESP charge fitting

atomic charges of each structure

↓ average

ensemble-averaged atomic charges
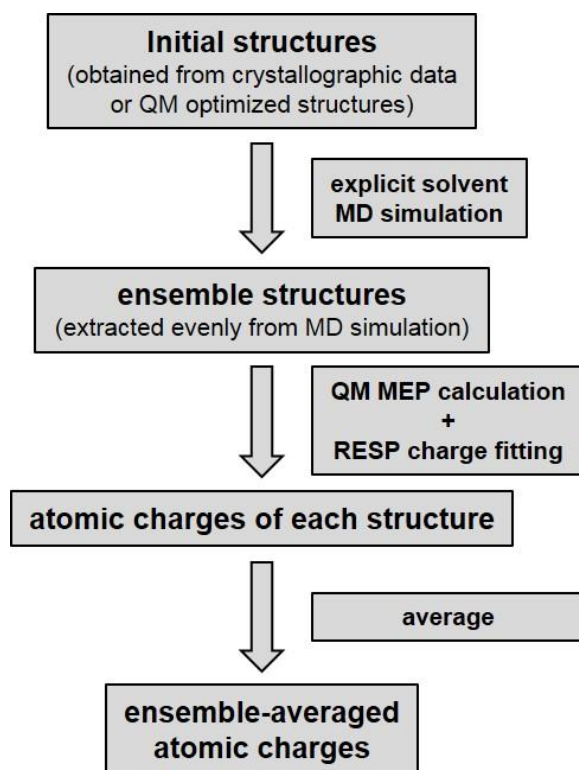
Figure 2.2 Flow chart of calculating ensemble-averaged atomic partial charges in GLYCAM force field parameters.

Torsion angle parameters are derived lastly since torsion terms are corrections that enable classical mechanics force fields to reproduce non-classical effects on bond rotations, such as hyperconjugation. The torsion terms are developed by fitting the difference between the QM and MM energies, computed while rotating the target torsion angle 360 degrees, while allowing the rest of the structure to relax (Figure 2.3). The fitting is achieved by varying the values of the coefficients (V) for the sum of cosine terms with varying periodicities (j) and phases (δ) (Equation 2.1).
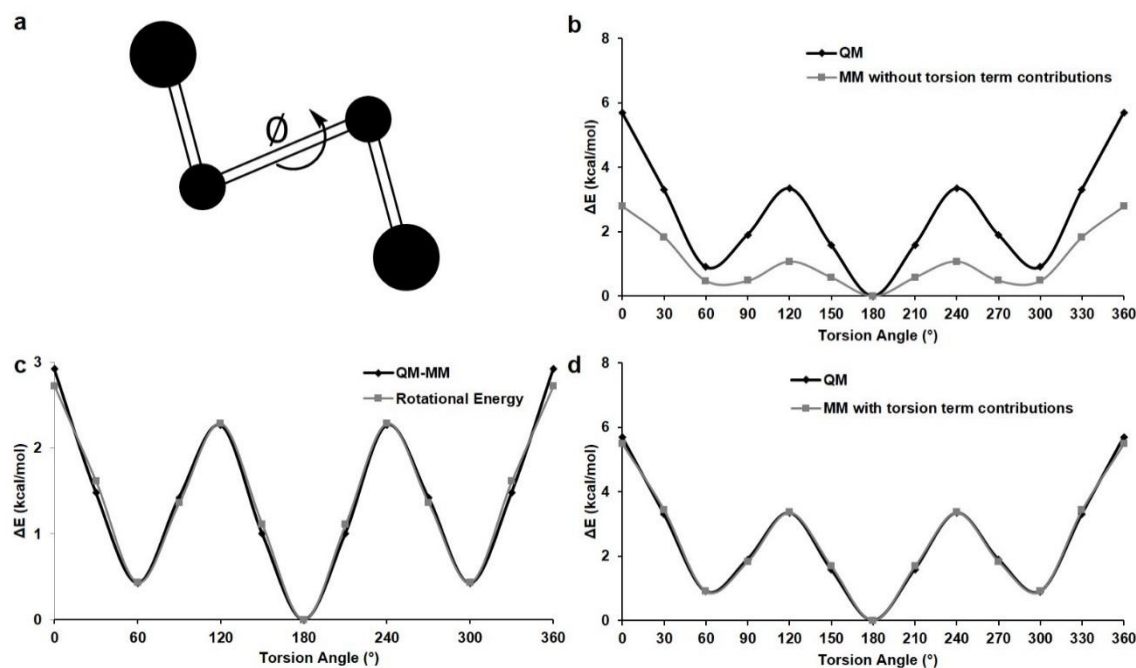
Figure 2.3 Torsion angle parameters (a) derived from fitting rotational energy to the energy difference between QM and MM without torsion term contributions. QM and MM energy profile without torsion term contributions (b); energy difference of QM and MM without torsion

term contributions and rotational energies from torsion terms (c); QM and MM energy profile with torsion term contributions (d).

Energy Minimization

Energy minimization is commonly performed before heating and data collection in MD simulations to eliminate high-energy contacts and improve unrealistic geometries in the starting structures obtained from crystallographic data for biomolecules. Such problems can result in unrealistic inter-atomic forces that destabilize the MD simulation. Two major protocols are commonly used for minimizations of energy functions: Steepest Descent and Conjugate Gradient. In steepest descent, the minimizations steps are taken in the direction of the steepest slope in the potential energy function, and the magnitude of the step is proportional to the value of the slope.

The Steepest Descent method is efficient when the structure is far from the minimum configuration, however, as the structure approaches a minimum, the minimization step size decreases, leading to slow convergence to the minimum. In the Conjugate Gradient method, the energy minimization algorithm produces a set of mutually conjugate directions so that the successive step continually refines the direction towards the minimum. Comparing to steepest descent, more calculations need to be performed to ensure all directions are mutually conjugated, and the time per minimization step is longer. Therefore, energy minimizations with conjugate gradient is usually employed after those with steepest descent.

An energy minimization is usually carried out until either the forces (the slope of the potential energy function) on all atoms in the molecule are below a tolerance that is set in advance, or a preset limit on the number of steps is reached. In large molecules the forces rarely

reach the tolerance (typically 1×10^{-4} kcal/mol•Å) before the number of steps (typically 500 steps

of steepest descent, 24500 steps of conjugate gradient) is reached.
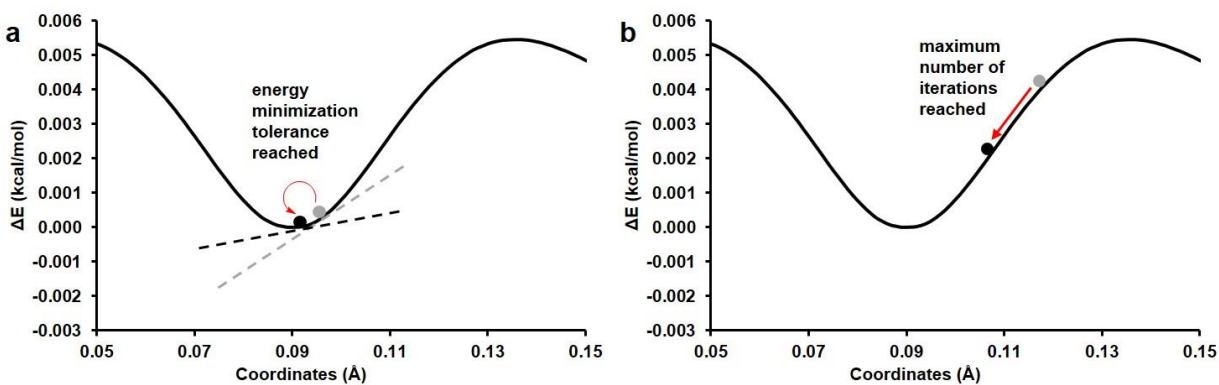
Figure 2.4 One dimensional demonstration of energy minimization ending scenarios: (a) the

gradient is smaller than the preset tolerance; (b) the preset maximum number of iterations has

been reached.

Molecular Dynamics

Molecular dynamics is a computational simulation of physical motions of a molecule while

interacting with its surrounding environment. All molecules are dynamic (exhibiting translation,

rotation or vibration) except at absolute zero degrees. When a molecule changes from one

conformation to another or moves from one location to another, motions of the molecule take

place in a smooth time-dependent manner. In MD simulations, the motions are divided into very

small time steps to guarantee that the discrete mathematical model can represent the highest

frequency motions in the molecule. A Taylor expansion can be employed to predict where a

particle with a position (x) at time t will be after time a time lapse Δt:

$$x(t + \Delta t) = x(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 + \cdots \qquad (2.2)$$

where $\vec{v}$ is the velocity of the particle and $\vec{a}$ is its acceleration.

Different methods have been developed to truncate this infinite Taylor expansion, and the Verlet algorithm[33] is one of the most widely employed, which assumes the acceleration is a constant over the time step. The positions of a particle before and after a time step can be expressed as Equation 2.3 and 2.4, respectively:

$$x(t - \Delta t) = x(t) - \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 \qquad (2.3)$$

$$x(t + \Delta t) = x(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 \qquad (2.4)$$

Equation 2.3 and 2.4 can be combined to remove $\vec{v}(t)$:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \vec{a}(t)\Delta t^2 \qquad (2.5)$$

Thus the position of a particle after time Δt can be calculated if knowing its current and previous positions and acceleration. The atomic accelerations (a_i) can be computed from the forces $\vec{F}_i$ on each atom, as determined from the gradient of the energy function ΔE_i (force field), by applying Newton's second law of motion:

$$\vec{F}_i = -\frac{\partial E_i}{\partial x_i} = \frac{\Delta E_i}{\Delta x_i}|_{t-\Delta t}^{t} = m_i\vec{a}_i \qquad (2.6)$$

where m_i is the mass of particle i.

Iterative application of Equations 2.5 and 2.6 permits the motion of the atoms in a system to be predicted on the basis of the forces generated from the force field. Smaller time steps improve simulation accuracy, however larger time steps result in more efficient sampling of the motion. The time step must be smaller than the frequency of the most rapid motion in the molecule. The highest frequency motions are the bond stretching vibrations, especially bonds with hydrogen atoms; yet these vibrations are generally of minimal interest in the study of biomolecular structure and function. Therefore, algorithms have been introduced that constrain these bonds to their equilibrium lengths[34] thereby increasing the acceptable time step and improving the efficiency of MD simulations. When hydrogen-containing bonds are not

constrained, a time step of 1 fs or less is typical, where as with constraints this value may be doubled. Doubling the time step halves the real time required for a simulation, the importance of this factor is obvious given that a typical biomolecular simulation may run for over a month.

Temperature and Pressure Control

The most common simulation condition for biological systems is the isobaric-isothermal ensemble[35] (NPT), in which the thermodynamic state is characterized by a fixed number of atoms, N, fixed pressure, P, and fixed temperature, T. The energy computed from this ensemble corresponds to the Gibbs free energy. Several thermostat algorithms have been proposed to control the temperature of the system. According to kinetic theory of ideal gas, the temperature of a system is proportional to the average kinetic energy of its constituent microscopic particles:

$$T = \frac{1}{3k_B} \langle mv^2 \rangle \tag{2.7}$$

where k_B is Boltzmann's constant, m is the mass of each particle, and v is the velocity of each particle. Therefore, modulating the temperature of a system can be achieved by changing the velocities of the particles at each time step in an MD simulation with an introduced factor λ (Equation 2.8).

$$\Delta T = T_{new} - T_{current} = \frac{\lambda^2}{3k_B} \langle mv^2 \rangle - \frac{1}{3k_B} \langle mv^2 \rangle \tag{2.8}$$

$$\lambda = \sqrt{\frac{T_{new}}{T_{current}}} = \sqrt{\frac{\Delta T}{T_{current}} + 1} \tag{2.9}$$

And, λ can be calculated with Equation 2.9. However, velocity scaling is a non-equilibrium process and does not represent the dynamics correctly[36]. In the Berendsen thermostat[37]

(Equations 2.8) a weak coupling (τ) to an external heating bath at temperature T_0 is introduced that allows the temperature (velocity) of individual particles to fluctuate while keeping the overall temperature (T) controlled.

$$\Delta T = \frac{\Delta t}{\tau}(T_0 - T) \tag{2.10}$$

The weak coupling factor (τ) influences the scale factor (λ), which changes the velocities of particles at each time step (Equations 2.11 and 2.12):

$$\lambda = \sqrt{\frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right) + 1} \tag{2.11}$$

$$\langle v_{current}^2 \rangle = \frac{\langle v_{bath}^2 \rangle}{\frac{\tau_T}{\Delta t}(\lambda^2 - 1) + 1} \tag{2.12}$$

where v_{bath} is the velocity of particles associated with external bath temperature, and $v_{current}$ is the velocity of the current time step.

An alternative is the Langevin thermostat[38], in which the solute molecule is much bigger comparing to the solvent molecule. In Langevin thermostat, the dynamics of the solvent molecules are controlled by imposing frictional drag forces on their motions, and the dynamics of solute is influenced by random collisions from solvent molecules. Langevin thermostat is more commonly employed in the MD simulations of biomolecular systems, because large biomolecules, such as proteins, are embedded in a bath of much smaller water molecules in explicit solvent MD simulations.

Similar to temperature, the pressure of a system is influenced by the average kinetic energy of its constituent microscopic particles. It is also influenced by the volume (V) of the system (Equation 2.13).

$$P = \frac{N}{3V}\langle mv^2 \rangle \tag{2.13}$$

15

where N is the number of particles.

In Berendsen barostat (Equations 2.14 and 2.15), similar to Berendsen thermostat, the system is weakly coupled (τ) to an external bath (P_0) to allow local pressure (P) fluctuation while controlling the overall pressure.

$$\frac{dP}{dt} = \frac{1}{\tau}(P_0 - P) \tag{2.14}$$

$$\Delta P = \frac{\Delta t}{\tau}(P_0 - P) \tag{2.15}$$

Similar to Berendsen thermostat, the weak coupling factor (τ) also influences the scaling factor (λ):

$$\lambda = 1 - \kappa\frac{\Delta t}{\tau}(P - P_0) \tag{2.16}$$

$$\frac{1}{V_{current}} = (\lambda^3 - 1)\frac{1}{\frac{\Delta t}{\tau_P}\langle mv^2\rangle\frac{N}{3}} + \frac{1}{V_{bath}} \tag{2.17}$$

where κ is the isothermal compressibility; V_{bath} is the volumn associated with external bath pressure, and $V_{current}$ is the volumn of the current time step.

Convergence in MD Simulations

The ergodic hypothesis states that with an infinite trajectory, all accessible states should be sampled in a statistically converged manner. Whether or not MD simulations will reach convergence within a finite time period always uncertain. In general, to ensure accuracy, the simulation period should exceed the relaxation time of the property being studied, and a statistically relevant number of fluctuations should be observed. For example, the C5-C6 bond (ω) in methyl α-D-glucoside interconverts between different rotamers, but each rotamer may be stable for more than 1 ns (Figure 2.5). To achieve convergence in the population distribution among these rotamers, MD simulations therefore need to extend beyond 1 ns; as seen in Figure

2.5, convergence requires at least 10 ns. The relaxation time of a property depends on the type of property, thermodynamic state, and the type of molecule. Some slow motion properties have a relaxation time that is exceedingly long for MD simulations, such as ring flip in pyranoses[39], in contrast, such motions in furanoses occur rapidly[40].

It is worth noting that, aside from long simulation time, the accuracy of the molecular model heavily relies on the quality of the selected force field. Small energy errors caused by the inaccuracy in force field can lead to significant errors in the predicted populations between different states. The quality of the molecular model also depends on several other factors, such as starting geometry, the type of solvent molecules, heating conditions, ion concentration, and protonation states of acidic/basic groups[41-45].

In this thesis, the development of a furanose-specific force field parameter set in GLYCAM that is derived for the same energy evaluation function mentioned in Equation 2.1, is introduced in Chapter 4. In addition to the new parameter set in GLYCAM, MD simulations performed in this thesis also employ GLYCAM06 and AMBER force field parameters. Energy minimization, thermostat and barostat techniques are employed prior to and during the MD simulations, respectively. This chapter provides the basic knowledge for understanding MM and MD simulations in this thesis.
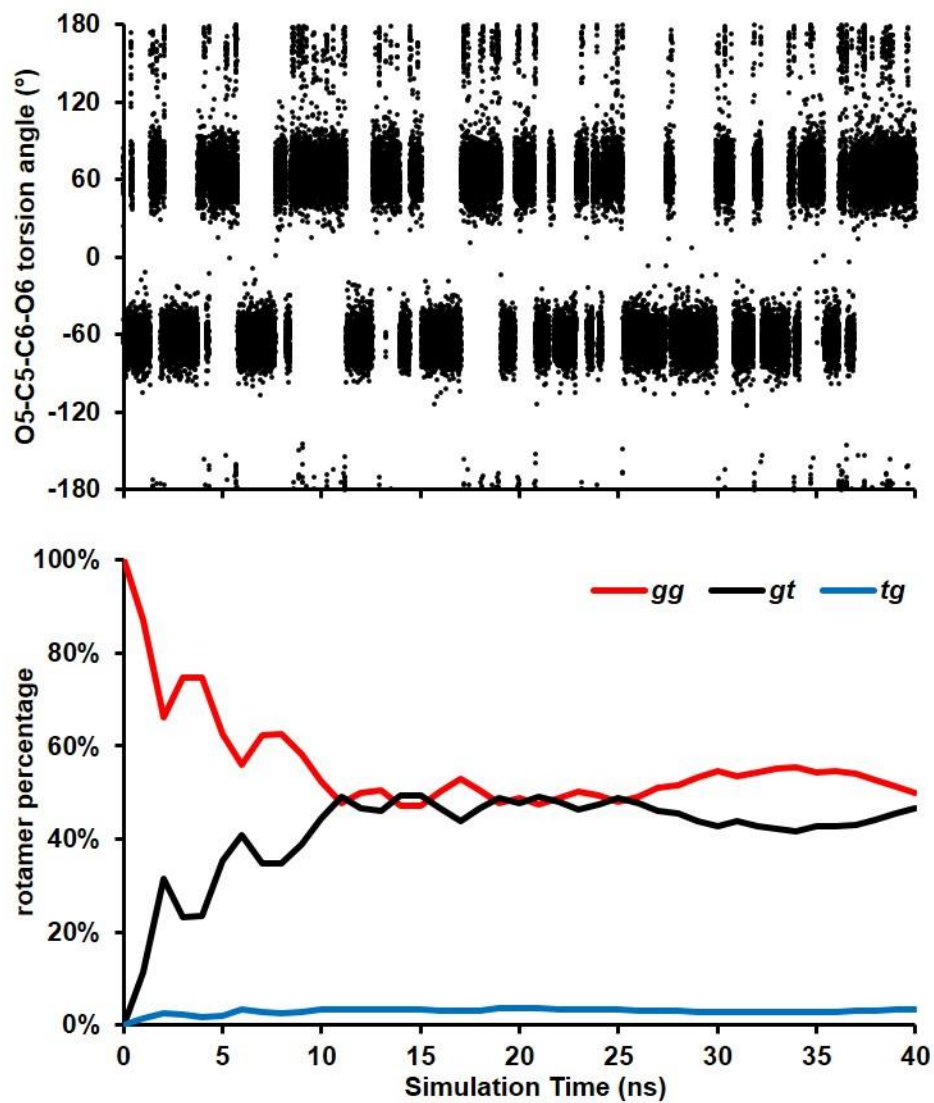
Figure 2.5 Values of ω bond for methyl α-D-glucoside in MD simulation (upper) and rotamer percentage along the course of simulation (bottom).

CHAPTER 3

LITERATURE REVIEW: MODELING OLIGOSACCHARIDE 3D STRUCTURES AND

INTERACTIONS

Carbohydrate Interactions in Biology

Carbohydrates are essential to every aspect of lives. Besides the fact that they are products of

photosynthesis and energy sources, oligosaccharides attached to a variety of biomolecules, as

known as glycan, are involved in numerous aspects of cellular interactions in part due to the

surface of all living cells in nature is covered with a dense and complex glycan array[46,47]. Unlike

the template-driven synthesis for nucleic acids or proteins, oligosaccharide synthesis depends on

the concerted actions of multiple enzymatic reactions[48]. The variability inherent to

oligosaccharide synthesis contributes the basis of the diversity and complexity of their

structures[48-50]. Oligosaccharides, which can form both linear and branch structures, are the

components of the most diverse species of molecules among the four fundamental building

blocks of life: nucleic acids, amino acids, lipids, and glycans[51]. The diversity of unique defined

3D structures endows oligosaccharides with variety and specificity of binding. There is

abundant evidence for the formation of well-defined 3D structures of oligosaccharides on heavily

glycosylated proteins and these structures form ligands for specific glycan-binding proteins[46].

Glycans are the ideal molecules for mediating molecular recognition and binding at the

cell surface, not only due to their structural diversity and uniqueness, but also because of their

quantity and seemly affinity. The high diversity of glycan in conjunction with each of their

unique spatial orientations provide high specificity for glycan-protein interaction[46]. In spite of their high abundance on cell surface, the interactions involving glycan only take place when both correct composition and conformation of glycans are present.

The high abundance of glycans in tissue and on cell surface increases the probability of the occurrence of glycan-involved interactions. Glycan-mediated cell interactions are often found in fluid media and under shear force, therefore, glyco-binding proteins must be in close proximity to their ligands to achieve binding. The presence of multiple replicas of glycan structures would enhance the probability of a glyco-binding protein to encounter and bind to its glycan counterpart.

The binding affinity of a single glycan-protein interaction is usually low ($mM - \mu M$ K_d values)[52,53]. However, this low affinity grants reversibility of the binding and allows quick release of a mismatch target with a small energy penalty. Moreover, this low affinity requires multivalent binding to generate biologically relevant interactions, which contributes to both reversibility and specificity of glycan-protein interactions.

Sulfate Oligosaccharide in Biology

The biological functions of oligosaccharides are usually augmented by additional structural features, such as the presence (or absence) of sulfate moieties[54]. Sulfated oligosaccharides, of which synthesis is carried out in the lumen of Golgi apparatus and catalyzed by Golgi-associated sulfotransferases, are found throughout all animals and in some plants[55,56], and ubiquitously expressed at cell surfaces and in the extracellular matrix. Due to their high negative charges, sulfated oligosaccharides interact with a multitude of proteins. Their bindings to proteins serve a variety of functional purposes, from simple immobilization or protection against proteolytic degradation to distinct modulation of biological activity[57].

The presence of sulfate groups significantly influence the structural[58-61] and dynamic[62,63] properties of oligosaccharides. Sulfate groups import net negative charges in the chain of oligosaccharides, and form strong electrostatic interactions with surrounding charged groups and polar moieties. In heparin sulfate, the strong electrostatic repulsions and possible cross-residue interactions contribute the rigidity of the glycosidic linkages[59-61]. In addition, the absence of a sulfate group would cause significant change in the dynamic motion of the molecule. For example, the non-sulfated NA-domain of heparan sulfate has higher flexibility compared to its S-domains[62]. Experimental approaches, including NMR spin relaxation studies, have been developed to probe molecular motions of oligosaccharides; molecular modeling and MD simulations with atomic-level resolution can help to rationalize experimental data and provide key information of critical interactions that determine structural and dynamical properties[63].

The presence of the anionic sulfate groups in oligosaccharides is also essential for their binding abilities. The unsulfated precursors of these oligosaccharides needs to be sulfated to become fully active sequences[64]. The binding affinity of sulfated oligosaccharides vary with regard to specificity, and often seem to depend primarily on charge density rather than on strict carbohydrate sequence[64]. Yet, there are also an increasing body of evidence that sulfation pattern and spatial distributions in oligosaccharides are also involved in glycan-protein interactions[63,65,66]. Some proteins only require the minimum specific motifs of sulfate substitutions present, and different sulfation patterns can be accommodated in the same binding site. Some proteins require a specific length of unsulfated residues to separate the sulfated binding regions. Some proteins have a preference for a specific pattern of sulfated substituents. For example, removing a key sulfate group from a synthetic pentasaccharide, employed to study

the importance of 3-O- and 6-O-sulfated glucosamine residues within the heparin octasaccharide to bind human antithrombin, decreases the binding affinity by more than 1×10^4 folds[67].

Due to their critical biological roles and ubiquity of existence, drugs targeting their interactions with proteins may be useful in treatment of disease conditions as diverse as cancer, inflammatory bowel disease, and Alzheimer's disease. Rational drug design requires a firm understanding of the pathophysiological role of a given sulfated oligosaccharide-protein interaction, as well as the aspect of specificity.

Modeling Oligosaccharide 3D structures

MD simulations have been widely employed in predicting oligosaccharide structures in solution. Linkages in oligosaccharides usually stay in a stable conformation in solution in part due to the stabilization provided by the *endo-* and *exo-*anomeric effects in carbohydrates[68]. Therefore, a reasonable model for the structure of an oligosaccharide can be generated based on its composition and linkage conformational preferences[69]. While the majority of two-bond linkages prefer a single conformation, in which the φ- and ψ-angles of the linkage can vary within ~30°, some linkages can adopt two or three stable conformations[31]. Thus, a single 3D shape may not be comprehensively representative of spatial properties of an oligosaccharide in solution; and predicting oligosaccharide structures in solution requires MD simulations reproducing all possible shapes with the correct population of each possible shape.

To evaluate the quality of a molecular model for predicting 3D structures of oligosaccharides in solution, structural properties derived from MD simulations are often compared to those properties measured experimentally. For example, the ring conformations of monosaccharide can be determined by the scalar ^3J-coupling constants from the coupled ring

hydrogen atoms. The calculation of ^3J-values from MD simulation trajectory can be performed by either employing a Karplus-type relationship to convert torsion angle values of coupled spins to ^3J-values[70-78], or weighting QM computed ^3J-values of all identified states with the population of each state from MD simulations[79]. Then, these MD derived ^3J-values are compared with the experimentally measured one. In addition to direct comparisons of ^3J-values, populations of carbohydrate ring distribution can be also derived from experimental values with the assumption made with regard to the number and conformation of states present in solution[80-82]. Then, the populations derived from experimental values can compare to those observed in MD simulation. Inter-proton distances can be generated from the intensity of Nuclear Overhauser Effect (NOE) measured from NMR, and these can provide further insight into the conformation of the oligosaccharide.

Molecular Modeling of Glycan-Protein Interactions (MM-GB/PBSA)

Studying glycan-protein interaction helps to understand the mechanisms of various cellular interactions, such as cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction, endocytosis and so on[83-86]. Motivated by the need for quantifying the strength of glycan-protein binding, many experimental approaches have been developed: isothermal titration calorimetry (ITC)[87], affinity capillary electrophoresis[88], surface plasmon resonance (SPR)[89], frontal affinity chromatography[90], and more recently biolayer interferometry (BLI)[91]. However, these methods are complex time- and material-consuming. Recent emerging glycan array methods[92,93] have higher throughput binding specificity and require less material, but lack the ability of quantifying binding affinity. Hence, molecular modeling methods have been devised as an important and efficient tool for generating insight into the structural and energetic properties of glycan-protein complexes at atomic level. Besides providing atomic-level

resolution to identify key residues responsible for binding, molecular models of glycan-protein complexes can also quantify the binding affinity.

The binding free energy for complex formation can be estimated from the difference of the solution-phase free energies of complex ($\Delta G_{complex,solution}$), and those of ligand ($\Delta G_{ligand,solution}$) and receptor ($\Delta G_{receptor,solution}$) (Equation 3.1). The binding free energy can also be decomposed into molecular mechanical (MM) binding free energies without solvent present (in vacuo) ($\Delta G_{bind,vacuo}$) augmented with an estimate of the desolvation energy by the difference of solvation energy of complex ($\Delta G_{solvation}^{complex}$) and that of ligand ($\Delta G_{solvation}^{ligand}$) and receptor ($\Delta G_{solvation}^{receptor}$) (Equation 3.2).
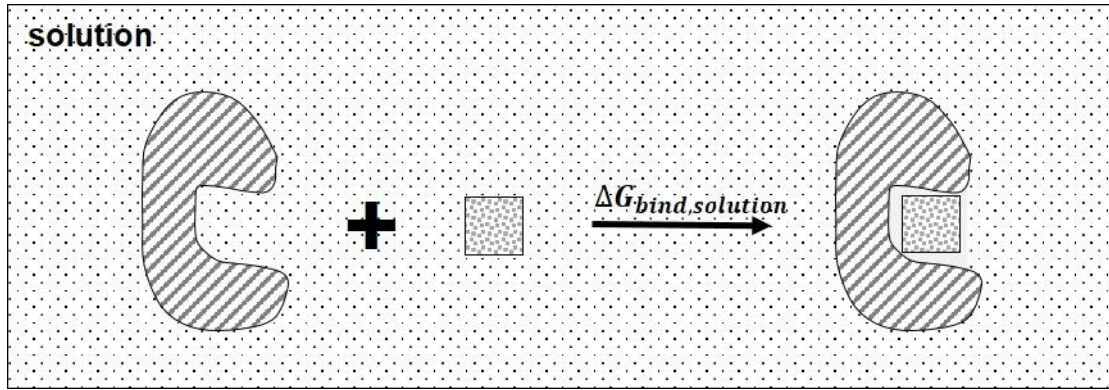
Figure 3.1 Demonstration of binding free energies of a receptor/ligand complex in solution.

$$\Delta G_{bind,solution} = \Delta G_{complex,solution} - (\Delta G_{receptor,solution} + \Delta G_{ligand,solution}) \quad (3.1)$$

$$\Delta G_{bind,solution} = \Delta G_{bind,vacuo} + [\Delta G_{solvation}^{complex} - (\Delta G_{solvation}^{receptor} + \Delta G_{solvation}^{ligand})] \quad (3.2)$$

$$\Delta G_{bind,vacuo} = \Delta H - T\Delta S \approx \Delta E_{MM} - T\Delta S \quad (3.3)$$

The $\Delta G_{bind,vacuo}$ can be calculated with Equation 3.3. The enthalpic contribution (ΔH) to binding is usually approximated by the difference of MM energies between the complex and those for protein and ligand (ΔE_{MM}), and it can be calculated with the difference of the internal

energies ($\Delta E_{internal}$) (bond, angle, and dihedral energies), and that of the non-bond interaction energies (ΔE_{ele} and ΔE_{vdw}) (Equation 3.4):

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw} \qquad (3.4)$$

In calculating the binding free energies, a single trajectory approach is often employed, in which snapshots taken from a single MD simulation trajectory are employed for calculation. $\Delta E_{internal}$ is cancelled since the geometries of the receptor and ligand are the same as those in complex. Thus, in a single trajectory decomposition, ΔE_{MM} depends only on the van der Waals and electrostatic interaction energies.

Entropic contributions ($-T\Delta S$) to binding energy are usually estimated by the normal-mode analysis of vibrational frequencies. However, it only considers entropy arising from changes in the vibrational frequencies of the component molecules that occur upon binding. Changes in populations of conformational states are not included in such an approximation, but may be treated separately[94]. A further challenge with the normal-mode approach is that it is extremely compute-intensive. For this reason, it is common practice to limit the number of MD snapshots included in the analysis to a small fraction of the total trajectory, leading to non-converged entropy estimates.

The energy required to remove solvent (desolvation energy) from the interface of the ligand and receptor is always unfavorable for carbohydrate binding. This is not surprising since carbohydrates, and their binding sites, are typically highly polar. However the direct electrostatic interactions that are formed upon complexation are highly favorable. In many cases, the desolvation energy is similar in magnitude, and opposite in sign, to the direct electrostatic energy, and therefore, the approximations employed when computing desolvation energies can have significant impacts on the predicted binding free energy.

The Poisson-Boltzmann Surface Area (PBSA)[3,95-103] and Generalized Born Surface Area (GBSA)[104-108] are popular implicit solvent models that have been widely used to compute desolvation free energies (ΔG_{sol}) (Equation 3.5). Explicit water models, although often employed in MD simulations, are not employed in binding energy calculations because of the strong dependence of the water-solute energy on the configurations (positions and orientations) of the water molecules.

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA} \tag{3.5}$$

where $\Delta G_{PB/GB}$ is the electrostatic contributions to solvation energies, which are computed using a PB or GB model; ΔG_{SA} is the surface-area-dependent non-electrostatic contribution term.

Combinations of the energies from MM and GBSA or PBSA methods (MM-GB/PBSA) have been applied to numerous protein-ligand[109-117] and protein-protein[118-120] systems, yet, their performance is system-dependent[99,121]. None of the current GB/PB models was specifically developed to treat the amphiphilic properties of carbohydrate binding[122]. In addition, MM/PBSA and MM/GBSA are sensitive to simulation protocols, such as the efficiency of sampling, entropy calculation methods, charge models, selected force fields, the solute dielectric constant, and radius parameters in continuum solvent models[123]. Therefore, predicting absolute binding free energies is extremely challenging and often suffers from considerable inaccuracies.

By virtue of the cancellation of systematic errors, comparing relative binding energies among related ligands with a receptor is generally more accurate than computing direct binding energies. The differences of entropic contributions from similar ligands can be estimated by the structural differences between ligands, such as the number of rotatable bonds[124,125] or derived from the conformational states observed in independent simulations of the ligand[94]. Previous

studies have shown that including conformational entropy was crucial for calculating the absolute binding energies, but not for ranking the binding affinities of similar ligands[126]. Similar to entropic contributions, differences of solvation energies between similar ligands and their binding complexes can also be predicted based on their structural differences[126]. Therefore, according to Equations 3.2 and 3.3, the accuracy of the ranking largely depends on ΔE_{MM}. Since $\Delta E_{internal}$ is cancelled in the single trajectory approach, ΔE_{ele} and ΔE_{vdW} may determine the ranking. Previous studies[126] also demonstrate that the ranking results were sensitive to the solute dielectric constant (ε_{in}), which has a profound impact on ΔE_{ele}.

Modeling sulfated oligosaccharides

Employing molecular models to accurately represent the dynamic properties and binding interactions of sulfated oligosaccharides with molecular models is a challenging task, due in part to high number of charges from sulfate groups.

In molecular modeling, partial atomic charges are commonly employed to describe the electrostatic properties of a molecule. The design of accurate partial charges is thus an important requirement for a good force field, especially when charged species are present. In the case of sulfated oligosaccharides, the partial charges should be capable of representing the average electrostatic properties of the molecule in solution. The partial charges should also properly represent the interactions of sulfate groups with surrounding moieties, such as electrostatic repulsions with other anionic groups and attractions through hydrogen-bonds or salt bridges. In principle, the charges in the sulfate groups could polarize neighboring moieties in either bound and unbound situations, however, polarization effects are not typically included in classical force fields. Polarizable force fields are still under development[127]. When calculating the binding interaction energies of polarizing molecules with MM/PBSA or MM/GBSA, it has been argued

tha the internal dielectric constant may be treated as an adjustable parameter, to account to some extent for the change in electrostatics that would arise from polarization. Clearly, the choice of the dielectric value is system dependent and empirical[128].

This chapter provides basic and general knowledge of 3D structures of oligosaccharide, which lays the ground for the studies in Chapter 5, in which the impact of sulfation patterns to the structural and dynamic properties of *Lv* I is studied. This chapter also provides detailed background knowledge of predicting binding free energies with MM-GB/PBSA methods, which have been employed to calculate the binding affinities of P-selectin ligands in Chapter 6.

CHAPTER 4

INSIGHTS INTO FURANOSE SOLUTION CONFORMATIONS: BEYOND THE TWO-

STATE MODEL[1]

Abstract

A two-state model is commonly used for interpreting ring conformations of furanoses based on NMR scalar ^3J-coupling constants, with the ring populating relatively narrow distributions in the North and the South of the pseudorotation itinerary. The validity of this simple approach has been questioned, and is examined here in detail employing MD simulations with a new GLYCAM force field parameter set for furanoses. Theoretical ^3J-coupling constants derived from unrestrained MD simulations with the new furanose-specific parameters agreed with the experimental coupling constants to within 1 Hz on average. The results confirm that a two state model is a reasonable description for the ring conformation in the majority of methyl furanosides. However, in the case of methyl α-D-arabinofuranoside the ring populates a continuum of states from North to South via the eastern side of the pseudorotational itinerary. Two key properties are responsible for these differences. Firstly, East and West regions in β- and α-anomers, respectively, are destabilized by the absence of the anomeric effect. And, secondly, East or West conformations can be further destabilized by repulsive interactions among vicinal hydroxyl groups and ring oxygen atoms when the vicinal hydroxyl groups are in *syn*-configurations (such as in ribose and lyxose) more so than when in *anti* (arabinose, xylose).

Introduction

It is difficult to overestimate the biological importance of furanoses. β-D-ribose and 2'-deoxy-β-D-ribose are essential components in the backbone of RNA and DNA chains, which provide the genetic codes of every organism. Other pentofuranoses are widely spread in domains of life ranging from bacteria to protozoa, fungi and plants[40]. As all biological function depends ultimately on 3D properties, a thorough understanding of the relationship between furanose structure and function necessitates the ability to determine the conformational properties of these

molecules. However, in contrast to pyranoses, furanoses exhibit a high level of internal ring flexing[129-131]. As a result, furanoses can interconvert between multiple ring conformations[132], unlike pyranoses that are usually found in single, energy-favorable chair conformations separated by large energy barriers[133]. In solution, the conformational analysis of furanose rings relies heavily on deconvoluting NMR scalar ^3J-coupling constants[80-82], and often assumes that the rings populate only two states[134,135], referred to as North and South on the pseudorotational itinerary (Figure 4.1). Here we employ MD simulations to probe this assumption and conclude that it is not valid in all cases.

The two-state model was introduced by Altona et al.[134] to interpret the ring conformations of the ribofuranosyl ring in nucleic acids. However, the validity of two-state model has been questioned in other furanoses, particularly arabinofuranose[70,71,131,136]. An alternative approach that has shown promise, is to perform MD simulations of the furanose[137], and then back calculate the NMR ^3J-values directly from the MD data[71,136,138]. This approach eliminates the need for assumptions regarding states, but requires an accurate force field and adequate conformational sampling. The calculation of NMR ^3J-values from MD data can be performed in one of two ways, either by employing a Karplus-type relationship to convert torsion angles between coupled spins into ^3J-values[70-78], or by computing the ^3J-values quantum mechanically for each state[79], thus eliminating the approximations implicit in Karplus-relationships. The latter method has not yet been applied to furanoses, although it has shown benefits when comparing pyranose MD data to experimental ^3J-couplings[79].
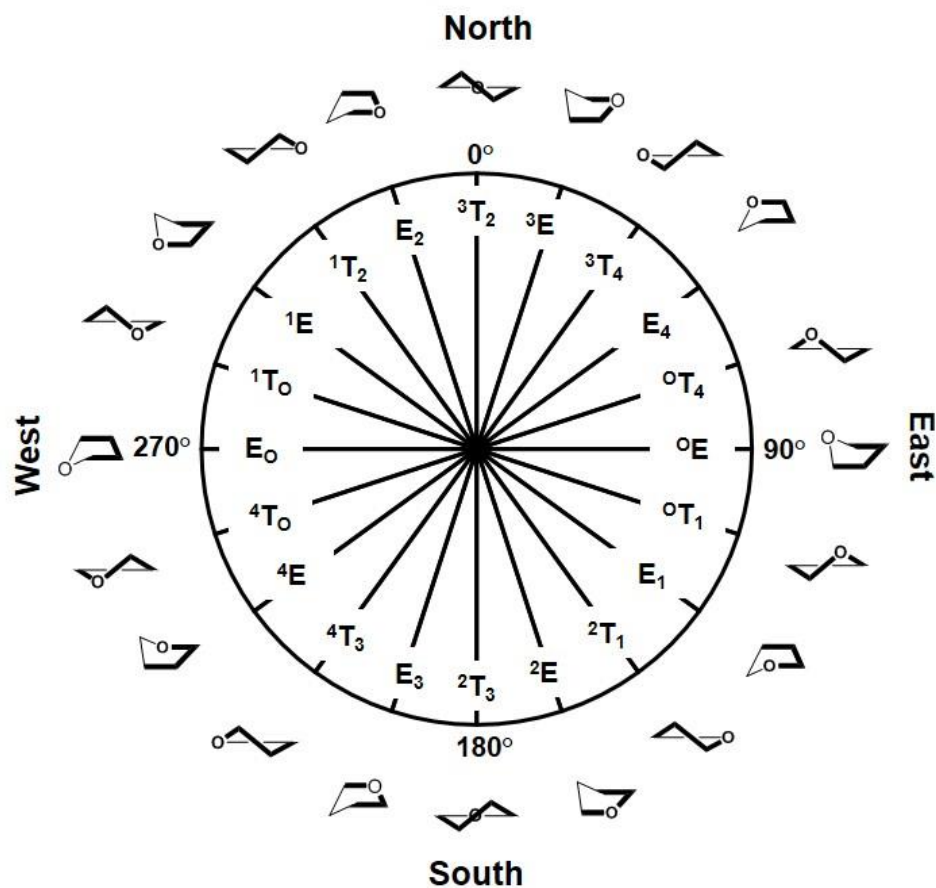
Figure 4.1 Pseudorotational itinerary of furanoses depicting different Envelope (E) and Twist (T) ring conformations with associated conformational phase angle (P) values in degrees.

Previously, the GLYCAM06 force field[31] was created to be applicable broadly to a range of carbohydrates, including pyranoses and furanoses, however, MD simulations of furanoses have not resulted in consistently acceptable reproduction of solution NMR properties[131]. Therefore, developing a reliable force field for furanoses, which is capable of describing their conformational properties in solution, is essential for both theoretical and experimental studies of these molecules. Notably, the fluxional properties of furanosyl rings leads to unique challenges in developing a force field for furanoses, particularly as regards the generation of partial charges

the need for long MD simulation times to ensure converged conformational sampling, and the need to carefully fit torsion terms specific to the five-membered ring.

To address the question of the apparent discrepancy between MD-derived and NMR-derived ring populations in furanoses, we begin by first developing a set of furanose-specific parameters that are compatible with the GLYCAM06 force field[31], which are able to reproduce the quantum-computed relative energies of the ring as it traverses the pseudorotational itinerary. Having derived a suitable set of parameters, MD simulations were performed in explicit water for both α and β anomers of the four methyl D-aldofuranosides: methyl D-arabinofuranoside (**1**), methyl D-lyxofuranoside (**2**), methyl D-ribofuranoside (**3**), methyl D-xylofuranoside (**4**), as well as methyl 2-deoxy-β-D-ribofuranoside (**5β**) (Figure 4.2). ^3J-coupling constants derived from the MD simulations of the examined furanosides with this new set of parameters were found to be in agreement with NMR ^3J-values to within an average error of less than 1 Hz. More importantly, ring conformations from the MD simulations have shown the two-state model is a reasonable description of the ring properties in the majority of the furanosides, but not in **1α**. Further analysis indicated that, for the two-state model to be applicable, barriers in the East and West separate low-energy regions in the northern and southern quadrants of the pseudorotational surface. The relative stability of northern and southern conformations has been rationalized in terms of the numbers of stabilizing gauche effects between vicinal oxygen atoms[18,19,139,140] as well as the presence of the anomeric effect[18,19,23-25,133,139-143]. Here, we find that whether or not there is a barrier in the eastern or western regions depends on the absence of an anomeric effect, as well as the presence of unfavorable *syn*-orientations of vicinal hydroxyl groups. In the unique case of **1α**, there is no barrier in the eastern quadrant due to the absence of any conformation in

which hydroxyl groups can adopt *syn*-orientations, enabling the furanose to populate a continuum of states from North to South.

The present analysis provides an independent method for interpreting furanose conformational properties that does not rely on experimental NMR data as constraints, nor on assumptions of preferred states. In addition, QM calculations have been employed here to quantify for the first time the strength of the anomeric and *exo*-anomeric effects in furanosides.

α: $R_1 = H$ $R_2 = OCH_3$ α: $R_1 = H$ $R_2 = OCH_3$
β: $R_1 = OCH_3$ $R_2 = H$ β: $R_1 = OCH_3$ $R_2 = H$

α: $R_1 = H$ $R_2 = OCH_3$ α: $R_1 = H$ $R_2 = OCH_3$
β: $R_1 = OCH_3$ $R_2 = H$ β: $R_1 = OCH_3$ $R_2 = H$
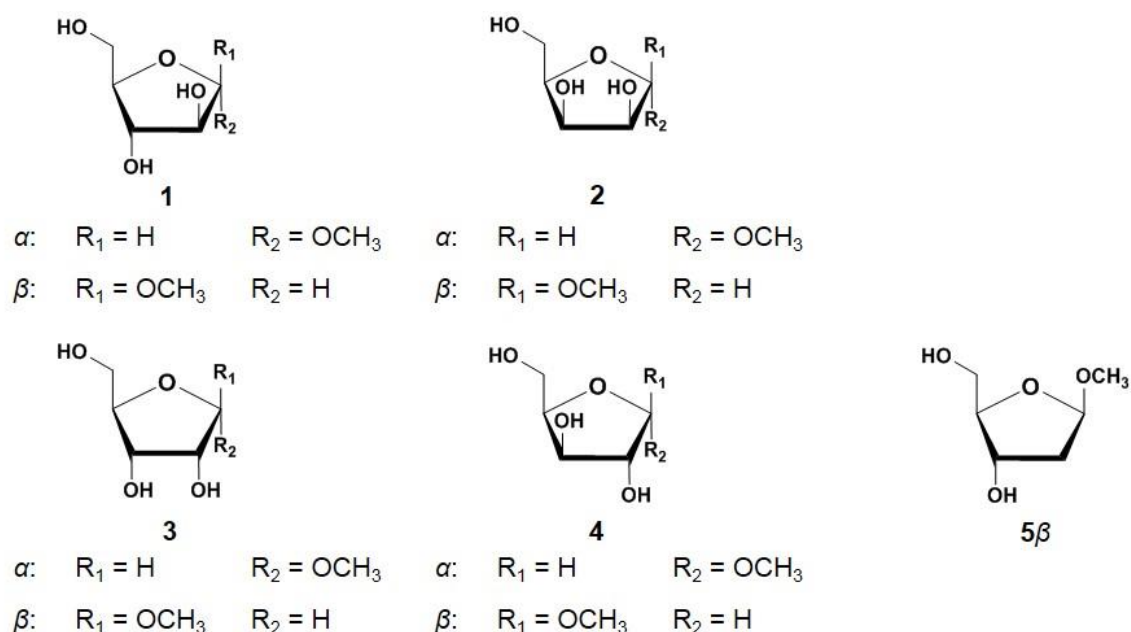
Figure 4.2 Methyl D-furanosides examined in present study.

Results & Discussion

Parameters

New atom type

In order to develop the parameters specific for furanoses, which are orthogonal with the current parameters for pyranoses, two new atom types were introduced: "Cf", representing ring carbon

34

atoms, and "Of", representing the ring oxygen atom (Figure S4.1). Van der Waals terms for Cf and Of were assumed to be the same as for the corresponding atoms in pyranoses[31]. Ensemble-averaged partial atomic charges for the new atom types, as well as all other atoms (Table S4.1), were computed according to the standard GLCYAM protocol[31], discussed in METHODS.

Bond length and bond angle parameters

The average bond lengths for C-C and C-O bonds in furanose ring are nearly identical with those in pyranose ring[144-149]. Therefore, the bond length parameters for furanose ring were imported from GLYCAM06, and simulations with these parameters demonstrated good agreement with the corresponding crystallographic values (Table S4.2).

In earlier bond angle parameter development for pyranoses in GLYCAM06[31], acyclic molecules were distorted from their equilibrium geometries in order to determine the force constants and equilibrium angle values. Similarly, bond lengths were stretched or compressed to determine appropriate bond parameters. However, to determine these values for furanoses, an alternative procedure was adopted, in which the ring was distorted by twisting the relevant torsion angles over a range of +/- 50 degrees. QM energy profiles for valence bond and angle parameter fitting were generated by optimizing tetrahydrofuran (THF) structures with ring torsion angles changing from -50° to 50° in 5° increments. Due to the symmetry of THF, only three energy profiles were necessary, namely for the C1-C2-C3-C4, C2-C3-C4-O4 and C3-C4-O4-C1 torsion angles. Parameters for the Cf-Cf-Cf, Cf-Cf-Of and Cf-Of-Cf angles were derived simultaneously by minimizing the differences between QM and MM ring distortion energies. To assess the performance of the new parameters in reproducing the QM energy profiles, the errors between QM and MM energies were computed over the entire range of the distortion curves ($<|Error|>_{all}$), as well as for the low energy conformations ($<|Error|>_{low}$). The new parameters

reproduced the gas-phase relative energies for THF markedly better than GLYCAM06 (Figure S4.2). The $<|Error|>_{all}$ for the new parameters for the three energy profiles in Figure S4.2 is 0.1 kcal/mol, and in contrast to GLYCAM06, the new parameters better reproduced the QM energies in high energy regions.

Average values for the C-C-C, C-C-O, and C-O-C angles were computed from 300 ns MD simulations and compared to relevant crystallographic values (Table S4.2). Although both C-C-O and C-O-C angles showed larger differences from the crystallographic values, they are within the standard deviation.

Torsion angle parameters

Torsion terms associated with the five-membered ring were optimized based on their ability to reproduce the QM pseudorotational energies for discrete conformations of the ring generated at 1 ° increments in the ring phase angle (P). In this procedure several contributing torsion terms could be simultaneously optimized. Simultaneous parameter fitting has been applied recently in the generation of a set of force field parameters for protein modeling[150]. Model structures were selected that nevertheless enabled the fitting of as few simultaneous torsion terms as possible, in order to minimize the number of new torsion terms, maximize parameter transferability, and provide insight into the underlying structural preferences (Table S4.3). All other torsion terms were generated directly by fitting to internal rotational energies. In order to reduce energy variations originating from interactions involving exocyclic moieties, the conformation of each of the exocyclic moieties was restrained throughout the MM and QM pseudorotation energy minimizations (Table S4.3).

To assess the performance of the torsion angle parameters, the average errors between QM and MM energies were computed for the entire energy curves, $<|Error|>_{pseudo}$, as well as for

36

the minima, <|Error|>minima. The parameters resulting from the development protocol discussed below are presented in Table S4.4.

THF

The torsion terms (Cf-Cf-Cf-Cf, Cf-Cf-Cf-Of and Cf-Cf-Of-Cf) derived simultaneously from fitting to the pseudorotational energies for THF (**6**) are fundamental to all furanoses. The agreement between the MM and QM energies (Figure 4.3) for the pseudorotational energy of THF computed with the optimized torsion parameters was within 0.1 kcal/mol for both <|Error|>pseudo and <|Error|>minima. It is notable that the energy minima for the THF pseudorotation was observed to lie in the East/West quadrants, rather than North/South, as typically observed for furanoses. This observation suggests that the exocyclic groups in furanoses alter the conformational preferences of the five-membered ring, although the parries to interconversion are less than approximately 0.5 kcal/mol. As can be seen in Figure 4.3, the energy minima for the five-membered ring in THF occur at $P = 90$ and $270°$, corresponding to East and West pseudorotational quadrants, respectively. In these conformations the C1-C2-C3-C4 torsion angle is $0°$. However, in this conformation hydroxyl groups at C2 and C3 can be eclipsed (as in lyxose and ribose), disfavoring East/West conformations.
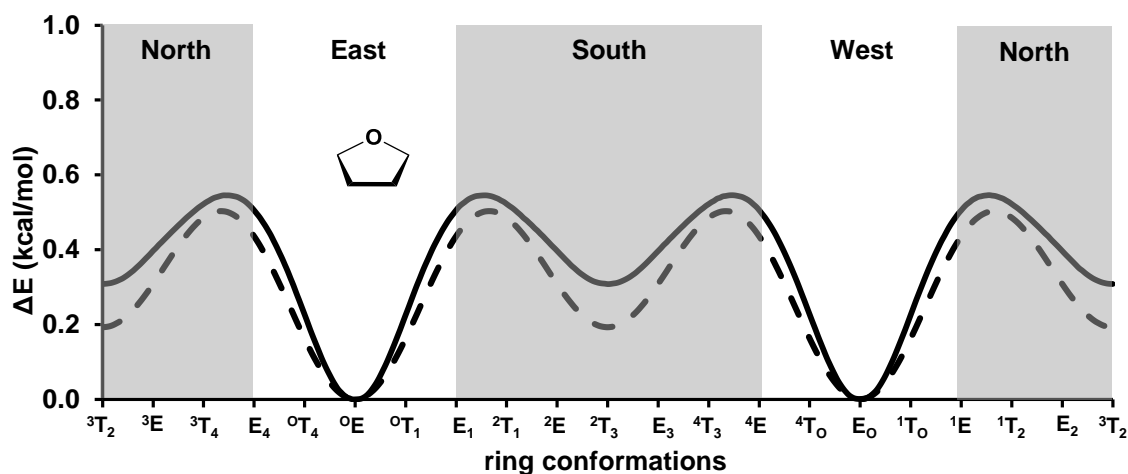
Figure 4.3 Pseudorotational energy curves for **6**. Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.

Models for mono- and dihydroxy substituents in furanoses

The ring conformation of furanoses is greatly affected by the configurations of the hydroxyl groups[70,71,73,151-153], and presumably arises from a combination of steric and electrostatic interactions. These interactions may be between hydroxyl groups, as well as between the ring atoms and the hydroxyl groups. To reproduce the effect of these interactions on the ring psudorotational energies, torsion terms were developed for the Oh/Of-Cf-Cf-Oh, and Cf/Hc-Cf-Cf-Oh sequences, employing the model compounds *cis*- (**7**) and *trans*-tetrahydrofuran-3,4-diol (**8**) and tetrahydrofuran-3-ol (**9**). As shown in Figure S4.3, the new parameters reproduced the shapes of the QM pseudorotational energy curves for both molecules. In the case of **9**, the errors in the fits were similarly small. With the introduction of the vicinal hydroxyl groups, the minima now occur in the North/South regime, as expected in general for furanoses.

Models for pentofuranoses

To extend the model to pentofuranoses, a hydroxymethyl group was introduced into THF, which corresponds to the C4-substituent in pentofuranoses, employing 2-(hydroxymethyl)tetrahydrofuran (**10**) as the model compound. This molecule enabled us to develop the Cf-Cf-Cf-Cg and Cf-Of-Cf-Cg torsion terms (pertinent to the C2-C3-C4-C5 and C1-O4-C4-C5 atomic sequences). The fitting to the pseudorotational energies for **10** was performed for each of the three rotamers of the hydroxymethyl substituent (Figure S4.4). While the average errors were all less than 0.5 kcal/mol, not all low-energy states were well reproduced, however in the worst case the error was less than 1 kcal/mol. Similar agreements were obtained for *cis*- and

trans-2-(hydroxymethyl)tetrahydrofuran-3-ol (**11** and **12**, respectively), which tested the ability of the parameters to reproduce the relative energies of the C4-C5 rotamers in the presence of a vicinal hydroxyl group at C3 (Figure S4.5 and Figure S4.6).

Models for methyl furanosides

The simplest model for the effect of anomeric configuration on five-membered ring conformational energy is (S)-2-methoxytetrahydrofuran (**13**), and, by analogy to the preceding section, the effect of interactions between the anomeric substituent and the adjacent hydroxyl group at C2 can be inferred from an analysis of *cis*- and *trans*-2-methoxytetrahydrofuran-3-ol (**14** and **15**, respectively). Torsion terms derived from fitting to **13** (Cf-Of-Cf-Os and Cf-Cf-Cf-Os), implicitly include any contributions from the anomeric effect[18,19,23-25,139,140]. In GLYCAM06, both the *endo*- and *exo*-anomeric torsion terms were parameterized by fitting to small acyclic molecular fragments, leading to the same set of terms for both applications. Here the torsion terms were derived for intact five-membered rings, with furanose-specific atom types for the ring atoms (Cf and Of), leading to unique terms for the *endo*- (Cf-Of-Cf-Os, and Cf-Cf-Cf-Os) and *exo*-anomeric (Of-Cf-Os-Cf/Cg and Cf-Cf-Os-Cf/Cg) sequences.

As shown in Figure S4.7, the torsion angle terms for *endo*-anomeric sequence reproduced the lowest energy conformation, as well as the overall shape of pseudorotational energy curve. It should be noted that the new parameters lead to a local energy minimum at $P = 99°$, while a similar minimum appears at $P = 71°$ in the QM energy curve. The torsion angle parameters derived from **14** and **15** gave rise to slightly larger errors, reflecting in part the challenge of fitting highly asymmetric energy curves using a sum of symmetric cosine functions. The torsion angle terms for *exo*-anomeric sequence reproduced the rotational energy profiles of **13** in both

northern ($P = 0\,°$) and southern ($P = 180\,°$) ring conformations (Figure S4.8). Features of the *endo-* and *exo-*anomeric effects in furanoses are discussed in the following sections.

Endo-anomeric effect

In carbohydrates, the anomeric effect relates to the preference of the endocyclic C-O-C-O torsion angle to adopt the *gauche* orientation over the *anti* conformation[24,26,154], and has been estimated from QM calculations for pyranoses[22,155] and related acyclic fragments[26,156] to be 1-2 kcal/mol. Its presence in furanoses has been invoked as contributing to ring conformational preferences[18,19,23-25,133,139-143], and is consistent with the observation that electronegative substituents at C1 in furanoses prefers a pseudoaxial over a pseudoequatorial configuration[29-34], but it has not been quantified computationally. We address this here by noting that, over the course of the pseudorotational itinerary for **13**, the endocyclic C4-O4-C1-O torsion angle spans a range of 80 to 160 °, or pseudo-*gauche* to pseudo-*anti*.

Thus the strength of the *endo*-anomeric effect in furanoses can be estimated to be 3.2 kcal/mol at the B3LYP/6-31G* level. In contrast, methoxycyclopentane (**16**) shows no such effect (Figure 4.4). The *endo*-anomeric effect stabilizes eastern ring conformations (C4-O4-C1-O angle is less than 100 °) but not western conformations (C4-O4-C1-O angle is greater than 140 °). Therefore, western conformations in *α*-anomers (or eastern conformation in *β*-anomers) display high potential energies.

Exo-anomeric effect

The preference of the exocyclic C1-O bond in pyranoses to adopt *gauche* orientations is a direct corollary to the *endo*-anomeric effect, and this *exo*-anomeric effect has been extensively studied in pyranoses and acyclic molecules[26,27,31,157].
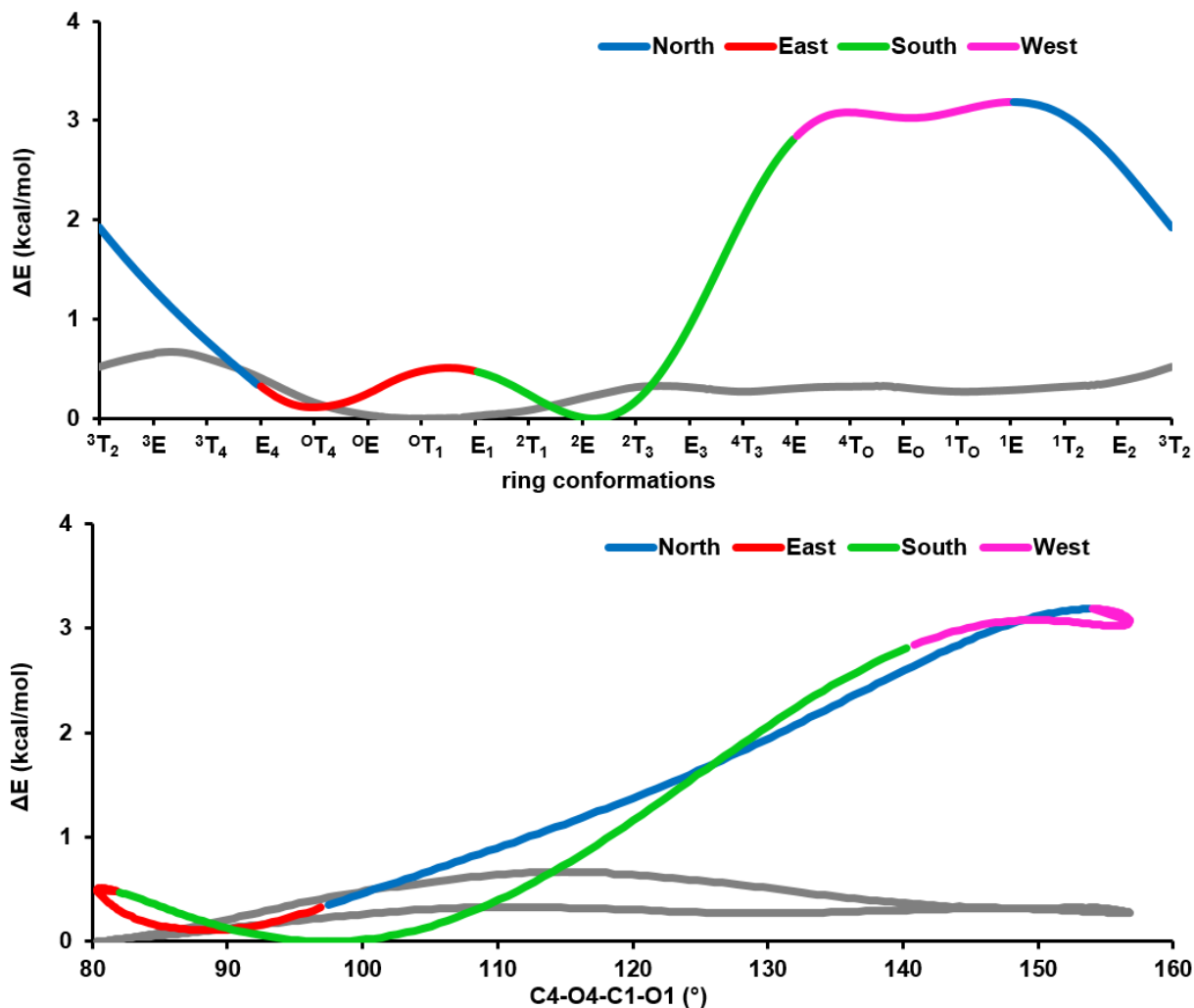
Figure 4.4 Upper: pseudorotational energy curves for **13** with each quadrant of pseudorotation color coded and **16** (grey) computed at the B3LYP/6-31G* level, with the orientation of the exocyclic O4-C1-O-CH3 torsion angle restrained at 60°. Lower: the relative energy of **13** and **16** as functions of the C4-O4-C1-O and C4-C5-C1-O torsions, respectively.

To determine the strength of the *exo*-anomeric effect in furanoses, the rotational energies of the O4-C1-O-CH3 torsion angle in **13** were computed for both North ($P = 0°$) and South ($P = 180°$) conformations. The *exo*-anomeric energy stabilizes the *gauche* orientations of the C1-O bond in **13**, relative to the cyclopentane analog **16**, by approximately 4 kcal/mol, which is

41

comparable to that reported for pyranoses[31] (Figure 4.5). This contribution is particularly important for defining the 3D structure of polymers of furanoses, as it reduces the conformational freedom of the C1-O bond to effectively a single state.
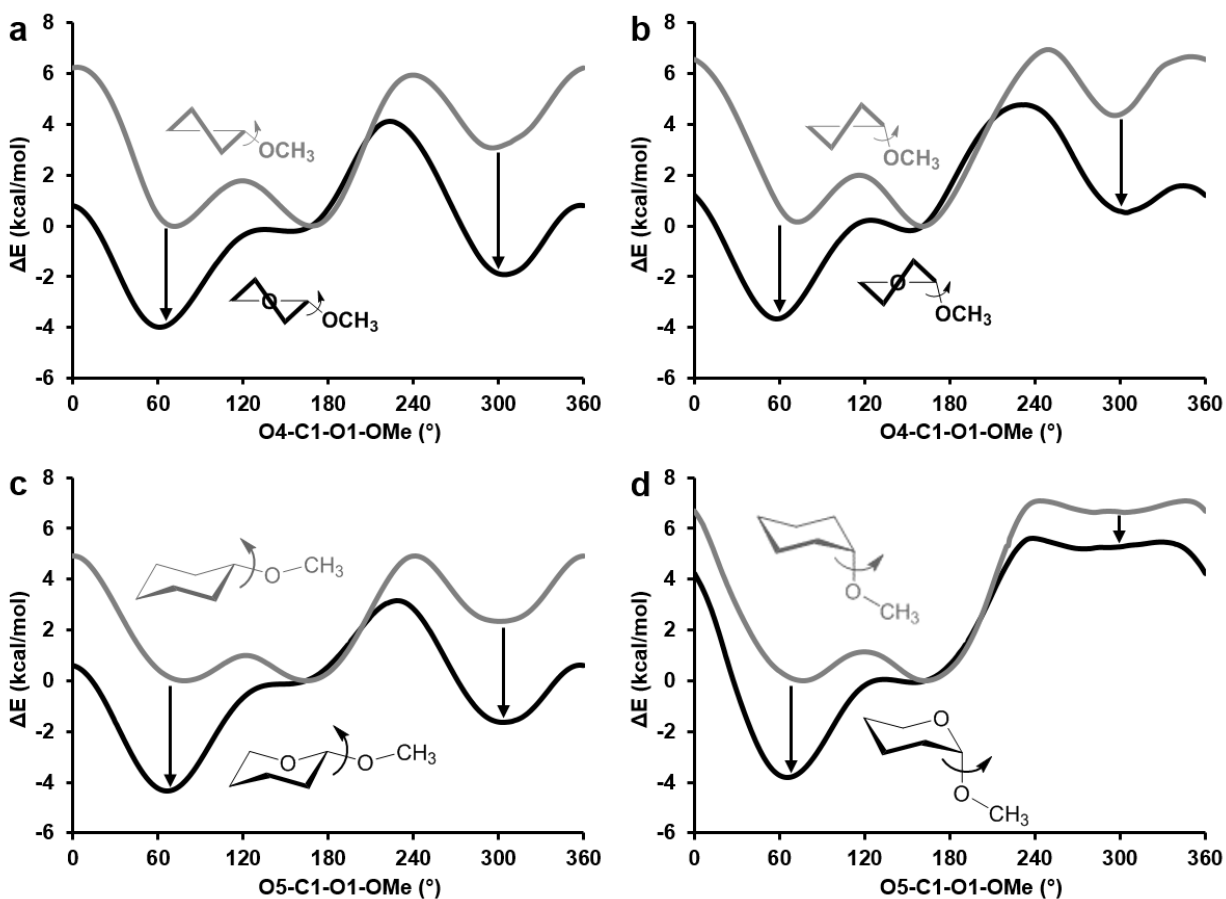
Figure 4.5 Rotational energy curves of O4-C1-O-CH3 angle in **13** (black) and C5-C1-O-CH3 angle in **16** (grey) computed at B3LYP/6-31G* level, while maintaining the ring conformation at $P = 0$ (a) and $180\,^\circ$ (b); rotational energy curves of O5-C1-O-CH3 angle in (S)-2-methoxytetrahydropyran (**17**) (black) and C6-C1-O-CH3 angle in methoxycyclohexane (**18**) (grey) computed at B3LYP/6-31G* level, while maintaining the ring conformation at $^1C_4$ (c) and $^4C_1$ (d). Regions stabilized by the *exo*-anomeric effect are indicated by vertical arrows.

Assessments of solution populations

The performance of the new parameters was assessed by comparing NMR ^3J-coupling constants for conformations observed in MD simulations of furanosides with their experimental values. By employing QM ^3J-coupling constants corresponding to the states observed in the MD simulation, it was possible to avoid employing empirical Karplus equations[70-78] or invoking approximations associated with the two-state model. The average error between the computed and experimental ^3J-coupling constants for **1-4** (*α* and *β*) and **5*β*** was within 1.0 Hz (Table 4.1). It is notable that, while the overall agreement with experiment is good, in the case of **2*α***, the theoretical value for the $^3J_{C1,H4}$ coupling constant (0.7 Hz) is significantly below that reported experimentally (3.7 Hz). However, this coupling constant is less than 0.5 Hz in all other *α*-anomers, suggesting a potential error in the experimental data. In the case of **5*β*** there is also a significant difference between the theoretical (9.3 Hz) and experimental (~5.7 Hz) values for the $^3J_{H2R,H3}$ coupling constant, however, the experimental report noted uncertainty in that particular measurement[153].

The rotamer distributions for the C4-C5 were also generated, and the average error between theoretical and experimental ^3J-values associated with this bond for **1** (*α* and *β*), **3** (*α* and *β*) and **5*β*** was 0.5 Hz; while a that for **2** (*α* and *β*) and **4** (*α* and *β*) was 1.3 Hz (Table 4.2). The origin of the larger average error in these latter cases is uncertain, but may relate to the fact that the parameters for the C4-C5 bond were imported from GLYCAM06 and not re-derived herein. The overall correlation (R^2) between the theoretical and experimental ^3J-values for coupling constants related to ring conformation was computed to be 0.72 (see Figure S4.9).

Table 4.1 MD derived, Karplus equation[73,77] derived, and experimental[70,71,73,151-153] ^{1}H-^{1}H and ^{13}C-^{1}H coupling constants for **1-4** (α and β) and **5β**.

	$^3J_{H1,H2}$			$^3J_{H2,H3}$			$^3J_{H3,H4}$			$^3J_{C1,H4}$		
	MD/QM[a]	MD/Karplus[b]	Expt	MD/QM	MD/Karplus	Expt	MD/QM	MD/Karplus	Expt	MD/QM	MD/Karplus	Expt
1α	1.6	2.1[c]	1.7[d]	3.4	2.8[c]	3.4[d]	5.3	4.5[c]	5.8[d]	0.8	1.0[e]	<0.5[f,g]
1β	5.3	5.6[c]	4.5[h]	5.7	4.6[c]	7.9[h]	5.1	4.6[c]	6.7[h]	3.9	1.8[e]	3.2[f]
2α	3.3	4.5[c]	3.6[i]	5.4	4.8[c]	4.8[i]	5.5	4.9[c]	4.3[i]	0.7	0.7[e]	3.7[f]
2β	6.1	5.6[c]	4.8[i]	5.6	4.9[c]	5.0[i]	4.8	4.6[c]	4.6[i]	1.8	0.4[e]	2.2[f]
3α	4.8	5.5[c]	4.3[j]	6.2	5.3[c]	6.2[j]	5.9	5.4[c]	3.4[j]	0.7	1.0[e]	<0.5[f,g]
3β	0.5	1.2[c]	1.2[j]	5.7	4.9[c]	4.6[j]	8.3	7.1[c]	6.9[j]	2.8	0.9[e]	2.7[f]
4α	4.2	5.0[c]	4.5[i]	2.6	2.1[c]	5.5[i]	5.1	4.4[c]	6.1[i]	0.1	0.3[e]	<0.5[f,g]
4β	0.3	1.2[c]	-[i]	1.4	1.2[c]	1.7[i]	4.6	4.7[c]	5.1[i]	1.4	0.4[e]	2.4[f]
5β	1.7/6.2[k]	2.2/6.5[c,k]	2.6/5.4[k,l]	7.9/9.3[k]	7.7/7.0[c,k]	6.7/~5.7[k,l]	5.4	5.3	4.2			

[a]The standard derivations of all MD derived ^3J-values are within approximately 0.3 Hz. [b]The standard derivations of all Karplus equation derived ^3J-values are within approximately 0.1 Hz. [c]Karplus equation from reference 77. [d]Reference 71. [e]Karplus equation from reference 73. [f]Reference 73. [g]A value of 0 Hz was employed in calculations of average error between the computed and experimental ^3J-values. [h]Reference 70. [i]Reference 152. [j]Reference 151. [k]Reference are reported for couplings involving H2S and H2R (S/R). [l]Reference 153.

Table 4.2 MD derived, Karplus equation[77] derived, and experimental[70,71,151-153] $^3J_{H4,H5S}$ and $^3J_{H4,H5R}$ values for **1-4** (α and β) and **5β**.

	$^3J_{H4,H5S}$			$^3J_{H4,H5R}$		
	MD/ QM[a]	MD/ Karplus[b,c]	expt	MD/ QM	MD/ Karplus	expt
1α	4.2	3.5	3.3[d]	5.1	4.5	5.8[d]
1β	4.3	3.6	3.4[e]	6.1	5.4	6.7[e]
2α	3.6	3.2	4.4[f]	8.6	7.6	6.7[f]
2β	3.3	3.1	4.5[f]	9.1	7.9	7.6[f]
3α	3.5	2.9	3.3[g]	4.3	4.2	4.6[g]
3β	3.6	3.1	3.1[g]	6.6	5.8	6.6[g]
4α	3.9	3.5	3.8[f]	8.8	7.5	6.0[f]
4β	3.5	3.3	4.4[f]	9.0	7.7	7.6[f]
5β	3.7	3.4	4.6[h]	6.4	5.7	7.0[h]

Table title spanning: ^{1}H-^{1}H coupling constants (Hz)

[a]The standard derivations of all MD derived ^3J-values are within approximately 0.5 Hz. [b]The standard derivations of all Karplus equation derived ^3J-values are within approximately 0.4 Hz. [c]Karplus equation from reference 77. [d]Reference 71. [e]Reference 70. [f]Reference 152. [g]Reference 151. [h]Reference 153.

Two-state equilibrium model

The ring conformation distributions of **1-4** (α and β) and **5β** in explicit solvent MD simulations are shown in Figure 4.6. The majority of the furanoses populated conformations predominantly in the northern and southern regions, and could therefore be described as approximately satisfying a two-state distribution. A notable exception to this was seen for **1α**, which populated a continuum of states from North to South via the eastern side of the pseudorotational itinerary. This is in contrast both to the other furanoses, and to earlier studies of **1α**, performed with GLYCAM06 that indicated a preponderant population of states in the North[131]. The overall agreement between the theoretical and experimental NMR data from the present simulations provides compelling support for the present simulation results.
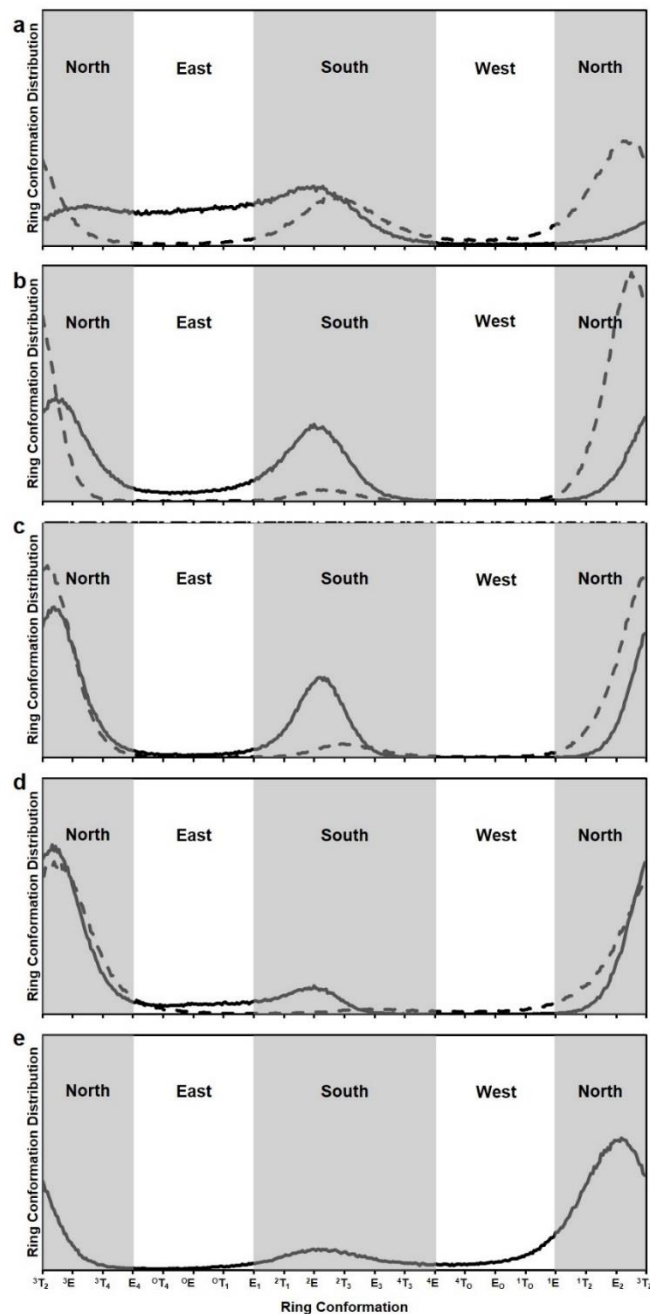
Figure 4.6 Ring conformation distribution for **1-4** (*α* and *β*) and **5*β*** from MD simulations (300 ns each). a: solid/dashed lines correspond to *α* and *β* anomers of **1**, respectively; b: solid/dashed correspond to *α* and *β* anomers of **2**; c: solid/dashed correspond to *α* and *β* anomers of **3**; d: solid/dashed correspond to *α* and *β* anomers of **4**, e: solid line for **5*β***.
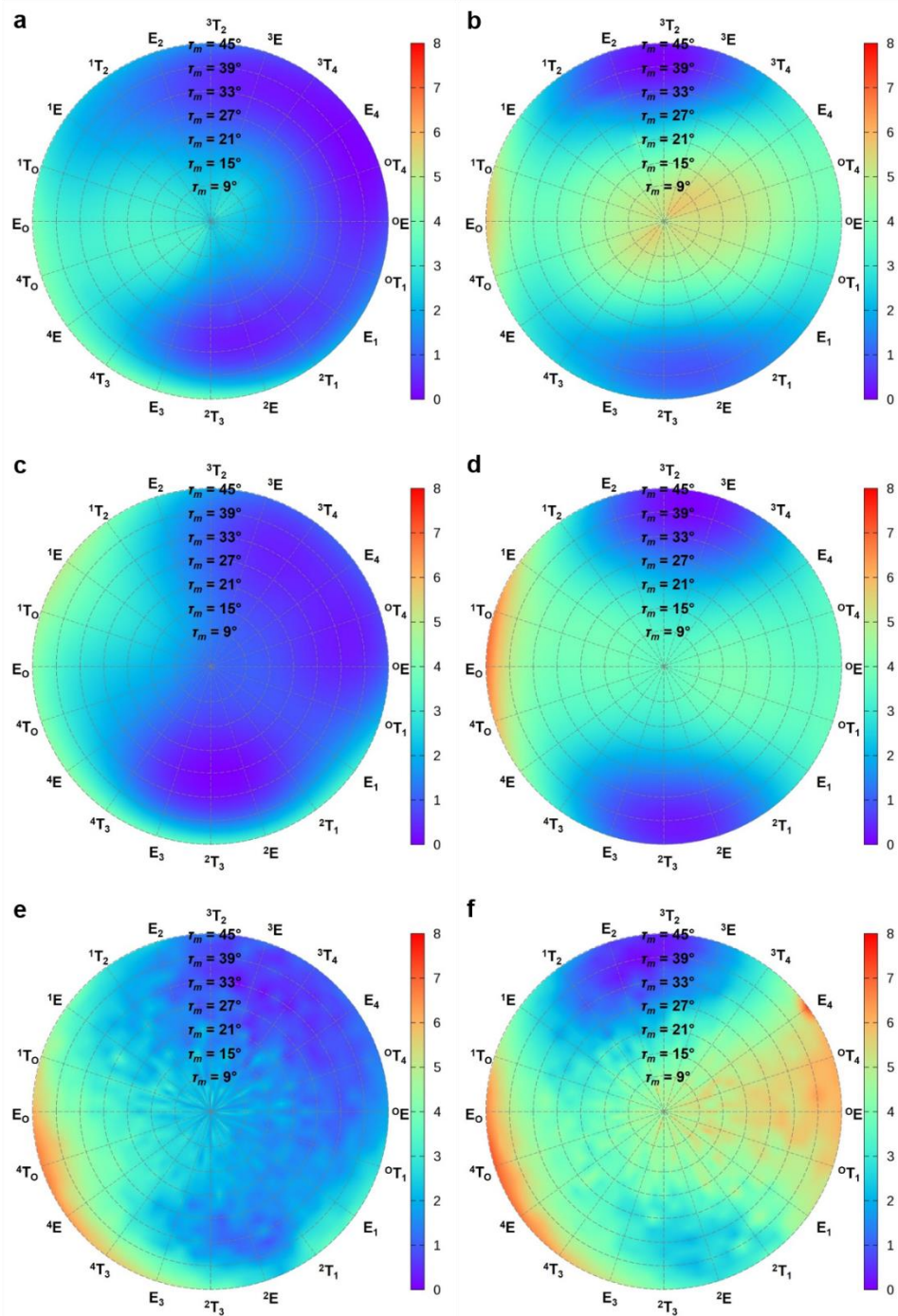
Figure 4.7 Pseudorotational potential energy surfaces for **1α** (left) and **3β** (right). a, b: energies computed at the B3LYP/6-31G* level; c, d: energies computed with new parameters after energy minimization in the gas phase; e, f: energies computed with new parameters for conformations

observed in explicitly solvated MD simulations, without energy minimization. Details provided in METHODS.

By plotting conformational energies onto the pseudorotational surface (Figure 4.7), it is possible to conveniently visualize the presence or absence of energy barriers between northern and southern states. In the case of a furanoside such as **3β**, that prefers conformations in the North and South, both quantum- and classically-computed conformational energies confirm the presence of barriers in the East and West, resulting in a clear division of the conformational space into northern and southern quadrants (Figure 4.7). In contrast, in the case of **1α**, both quantum and classical mechanical methods indicate a barrier only in the West, enabling the ring to populate a continuous distribution of conformations from North to South via the East. Previous NMR[71,152,158], as well as QM[19,20] studies, have also been shown to be consistent with conformations in the North, East, and South and it has been proposed that the anomeric effect[17,19,23-25,139,140,154] and the *gauche* effect[18,19,133,139-143] among the vicinal hydroxyl groups and the ring oxygen atom were the key influences on ring conformation in furanoses. In the case of **1α**, it has been argued that the anomeric and *gauche* effects both stabilize the South conformation[18-20,133]. From the analysis of the anomeric effect presented in Figure 4.4, we can conclude that the barrier in the western quadrant in **1α** is due in part to the absence of stabilizing contributions from the anomeric effect. In the case of other furanoses, for example **3β**, in which there are *syn*-interactions between hydroxyl groups, an examination of the interaction energies between the hydroxyl groups and the ring oxygen atom (Figure 4.8) leads to the conclusion that repulsions between these groups introduce a barrier in the eastern quadrant. Thus in furanoses other than **1α** (that has no *cis*-hydroxyl groups), either the lack of an anomeric effect, or the

presence of repulsive oxygen-oxygen interactions results in barriers in both the East and West, effectively establishing a preference for the populations in the North and South.
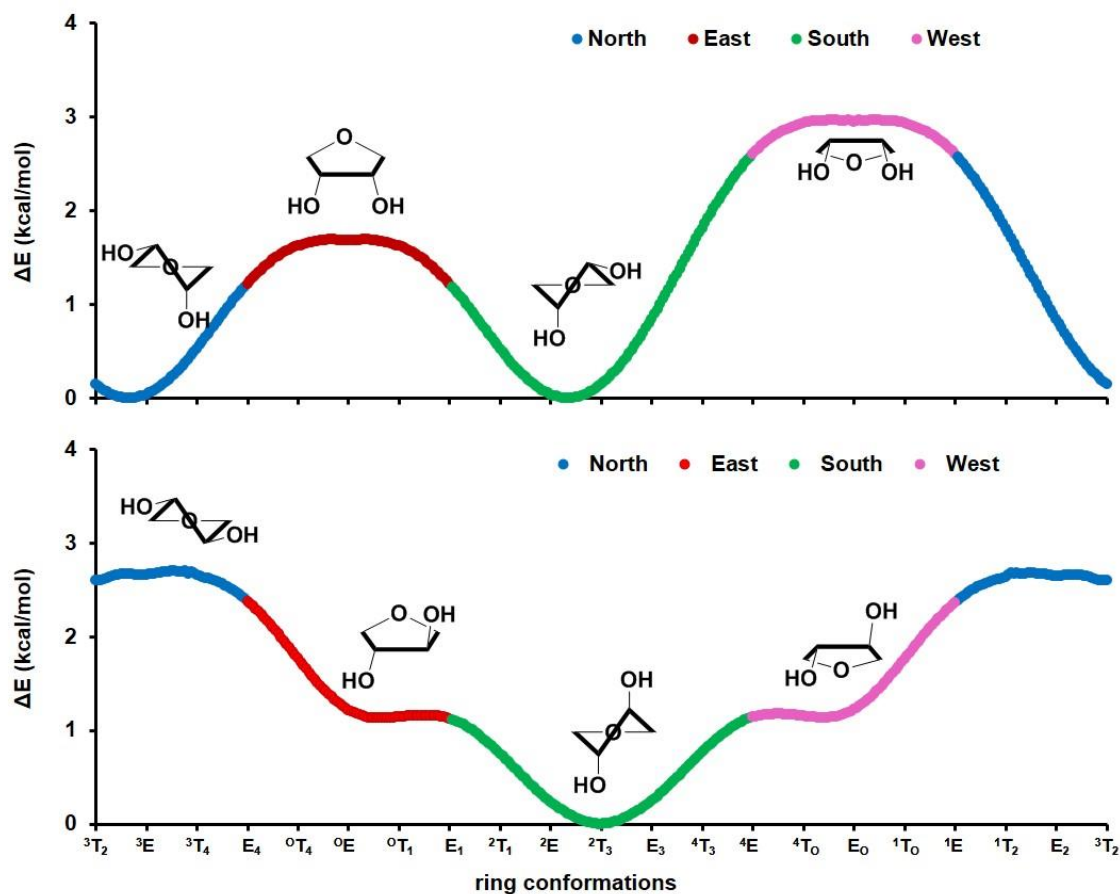
Figure 4.8 Pseudorotational energy curves for **7** (upper) and **8** (bottom) computed at the B3LYP/6-31G* level, with the orientation of the vicinal hydroxyl groups (C2-C3-O3-H3O and C3-C2-O2-H2O) restrained at 180 °. Curves in each quadrant of pseudorotation are color coded.

Conclusion

The present study confirms that the two-state model, commonly employed in interpreting ring conformations of furanoses in solution, is not applicable to all furanoses. Results from MD simulation indicated that **1**α exhibited a continuum distribution from North to South via East of

pseudorotation, which agreed with the low energy pathway found in its pseudorotational potential energy surface, and gave rise to experimentally-consistent scalar coupling values for the ring protons. Nevertheless, the two-state model is a reasonable description for other furanoses, to the extent that their ring conformations broadly populate the northern and southern quadrants. The division (or not) of the populations into two states can be explained by the presence (or absence) energy penalties in East and West regions of the pseudorotation itinerary. These barriers are found to arise from the absence of a stabilizing anomeric effect and unfavorable interactions among vicinal hydroxyl groups and the ring oxygen atom.

Notable, the new force field parameters permit accurate simulations of furanoses to be performed (average difference between theoretical and experimental NMR ^3J-values was within 1 Hz), making the need to adopt an assumption about state preferences obsolete. This is likely to be particularly valuable when examining complex oligofuranosides that may contain additional chemical modifications, as are typical in bacterial[159-162] or fungal pathogen[163-168] surfaces. The ability to characterize the conformational properties of such structures is an important component in understanding the mechanisms of disease infection[169,170], as well as immune response[171,172].

CHAPTER 5

IMPACT OF SULFATION PATTERN ON THE CONFORMATION AND DYNAMICS OF

SULFATED FUCAN OLIGOSACCHARIDES AS REVEALED BY NMR AND MD[58]

Introduction

Sulfated fucans in the jelly of the sea urchin eggs bind to sperm receptors are composed of α-L-fucopyranose (Fuc$p$) uinits and induce the acrosome reactions[173,174]. These sulfated fucans in different sea urchin species possess unique and well-defined chemical structures[175] and their structural differences of sulfated fucans in sea urchin egg jelly may drive their species-speciation, as a result of fertility incompatibility or even evolutionary causes[176,177]. It is, therefore, essential to understand the structural and dynamic properties of these sulfated fucans in both free and bound states[63,178,179].

The structurally defined sulfated oligofucan, namely *Lytechinus variegatus* I (*Lv* I) was produced by mild acid hydrolysis of a longer polysaccharide for the first time, which was synthesized by the sea urchin *Lytechinus variegatus*[180,181]. Experimental approaches have confirmed that *Lv* I was composed of 3-linked α-L-Fuc$p$ units with the sequence of D2-B2-C2-A2-D1-B1-C1-A' (Figure 5.1), in which A' is non-sulfated Fuc$p$, A1, A2, C1 and C2 are 2-sulfated Fuc$p$, B1 and B2 are 2,4-disulfated Fuc$p$, D1 and D2 are 4-sulfated Fuc$p$. Dynamic properties of *Lv* I have also been unveiled by NMR experiments[58]. However, it is challenging to correlate the dynamic properties with its structural characters, such as sulfation patterns, and provide insights to the origins of its dynamic properties, solely with the NMR results.
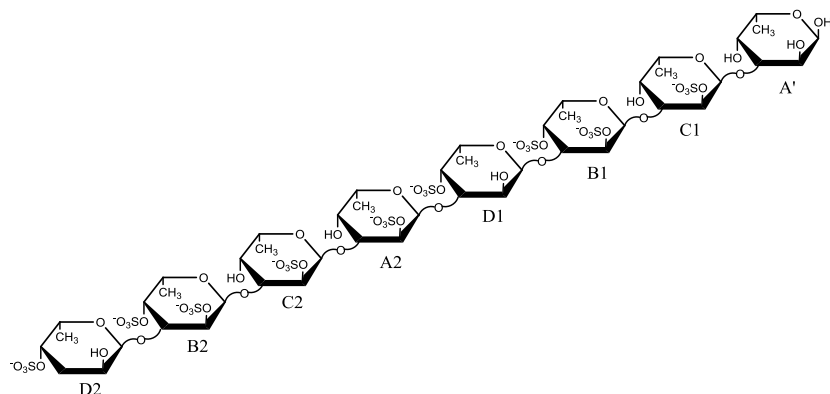
Figure 5.1 Structural representation of *Lv* I with label of each residue.

In this work, we have built the molecular model of *Lv* I with GLYCAM06 force field[31] (version h). The model of *Lv* I was validated by reproducing the NOE distance and scalar ^3J-coupling constants from NMR experiments in MD simulations with explicit solvent. The correlation between its dynamic properties and structural characters, especially sulfation patterns, was examined through MD simulations. To understand the dynamic impacts of sulfation pattern, MD simulations were performed on molecular models of difucoses with corresponding sulfation patterns in *Lv* I. It is revealed that intermolecular hydrogen bonds between sulfate and hydroxyl groups increased the stability of *Lv* I, while intermolecular repulsion between sulfate groups promotes the internal dynamic mobility of *Lv* I. Both increased stability and mobility were seen from the distribution of φ/ψ angles.

Method

MD simulation setup

The initial structures of *Lv* I and sulfated difucoses were built from carbohydrate builder on GLYCAM website (http://glycam.org). The structure and atomic charges of *Lv* I and sulfated difucoses were described by employing the GLYCAM06 force field parameters[31] (version h) and

then solvated with an 8Å TIP3P water[44] buffer in octahedral box using LEaP program in AMBER14 package[182]. Sodium ions were added to neutralize the solvated system.

The energy minimizations for the solvated *Lv* I and sulfated difucoses were performed separately under constant volume (500 steps steepest descent, followed by 24500 steps of conjugate-gradient minimization). Each system was then heated to 300K over a period of 50 ps and followed by an equilibration at 300K for 0.5 ns using nPT ensemble, with the Berendsen thermostat[36] for temperature control in AMBER. All covalent bonds involving hydrogen atoms were constrained using SHAKE algorithm[34], allowing a simulation time step of 2 fs throughout the simulations. After the equilibration, each simulation was carried out with GPU implementation[183] from AMBER14 and trajectory frames were collected at every 1 ps for the total of 450 ns. Non-bonded interaction cutoff was set to 8Å and none of 1-4 non-bonded interactions were scaled.[184]

MD-based scalar proton-proton coupling constants

Representative structures of *Lv* I extracted from the simulation were 10 trajectory frames with the lowest RMSD values comparing with the average structure of *Lv* I from the simulation. Each of the fucoses in these representative structures of *Lv* I was isolated, then the missing atoms for each fucose were added by GaussView03[185] software package. The QM spin-spin coupling constants calculations of each Fuc*p* were computed at B3LYP/HIII-SU3//HF/6-31++G(2d,2p) level of theory[186] in Gaussian09 software package[187] with the fucose ring constrained at its conformation in the MD simulation.

The Karplus equation derived spin-spin coupling constants were calculated with a generic Karplus equation[78]. The electronegativity parameters for the generic Karplus equation were determined by the atom attached on carbon atom in the ring. The torsion angle values of H1-C1-

C2-H2 and H2-C2-C3-H3 for each fucose in *Lv* I were applied to the generic Karplus equation with appropriate electronegativity parameters and the final results were averaged over the 450,000 trajectory frames.

Result and Discussion

Validation of molecular model of *Lv* I

In order to validate the molecular model of α-L-fucose of representing the structural and dynamic features of *Lv* I in solution, the scalar ^3J-coupling constants of ring hydrogen atoms in fucoses and the proton-proton distances across the glycosidic linkages obtained from the MD simulation were compared with their experimental results.

In α-L-Fcu*p*, ^3J-value for the vicinal proton pair of H2 and H3 is more sensitive than that for the pair of H1 and H2. When α-L-Fcu*p* is in $^1C_4$ chair conformation, the torsion angle of H2-C2-C3-H3 ($\theta_{H2,H3}$) is nearly 180°, and H2-H3 produces a coupling constant ($^3J_{H2H3}$) around 10 Hz; when α-L-Fcu*p* is in $^4C_1$ chair conformation, $\theta_{H2,H3}$ is around 60°, and $^3J_{H2H3}$ is around 4 Hz. However, $\theta_{H1,H2}$ is around 60° when α-L-Fcu*p* is in both $^1C_4$ and $^4C_1$ chair conformation, and $^3J_{H2H3}$ changes from 0 to 4 Hz. Therefore, $^3J_{H2H3}$ in α-L-Fcu*p* is advantageous to interpret its ring conformation. Experimental NMR results show that $^3J_{H2H3}$ for all α-L-Fcu*p* in *Lv* I were all above 10 Hz[58], which indicate that in solution all α-L-Fcu*p* in *Lv* I maintained a dominant $^1C_4$ chair conformation. This agreed with the observations from the MD simulation of *Lv* I. In the simulation, all α-L-Fcu*p* maintained $^1C_4$ chair conformation through the entire course of the simulation and no ring flips have been observed. To further valid the molecular model of *Lv* I, QM calculated ^3J-values from representative structures of α-L-Fcu*p* in *Lv* I from the simulation were compared to their corresponding experimental results (Table 5.1). The average difference between QM and experimental $^3J_{H2H3}$ values is 1.1 Hz. All Karplus equation derived $^3J_{H2H3}$ from

the simulation were 3Hz smaller than the experimental values (Table 5.1), but these values indicate *anti* configuration between H2 and H3, which only exist in $^1C_4$ chair conformation of α-L-Fcu*p*.

Table 5.1 NMR[a], averaged QM-calculated and Karplus-derived $^3J_{H,H}$[b] coupling constants for residues in *Lv* I.

	$^3J_{H1\text{-}H2}$ (Hz)			$^3J_{H2\text{-}H3}$ (Hz)		
	NMR	QM	Karplus derived	NMR	QM	Karplus derived
A'	4.5 ±0.9	4.9 ±0.8	2.9 ±0.5	11.7 ±2.3	10.0 ±0.5	7.1 ±0.3
C1	4.5 ±0.9	4.1 ±1.2	3.1 ±0.5	10.2 ±2.0	10.0 ±0.8	7.0 ±0.4
B1	5.0 ±1.0	3.9 ±0.6	3.0 ±0.5	12.1 ±2.4	9.1 ±0.3	7.1 ±0.4
D1	4.0 ±0.8	4.6 ±0.6	3.1 ±0.5	11.7 ±2.3	9.7 ±0.6	7.1 ±0.4
A2	5.0 ±1.0	3.6 ±0.6	2.8 ±0.5	10.4 ±2.1	10.1 ±0.5	7.1 ±0.3
C2	4.5 ±0.9	4.1 ±0.9	3.0 ±0.5	10.2 ±2.0	10.0 ±0.6	7.0 ±0.4
B2	5.0 ±1.0	3.8 ±0.3	3.0 ±0.5	12.1 ±2.4	9.6 ±0.5	7.1 ±0.4
D2	4.0 ±0.8	5.0 ±0.4	3.1 ±0.5	11.7 ±2.3	9.7 ±0.5	7.0 ±0.4

[a]Ref 58. [b]in Hz. Karplus equation from reference 77.

Both intra- and inter-residue proton-proton distances in *Lv* I were also calculated from the MD simulation of *Lv* I to validate its molecular model with regard to representing the properties of its glycosidic linkages, by comparing to experimental NOE results of hydrogen atoms across the linkages. The dynamic properties of *Lv* I were determined by the properties of its glycosidic linkages, since all the residues in *Lv* I stayed dominantly in $^1C_4$ chair conformation. The average distance for each pair of intra-residue protons from the simulation agreed with its NOE derived value (Table 5.2), which demonstrated that the molecular model of *Lv* I reproduced its properties of glycosidic linkages in solution.

Table 5.2 Theoretical and calculated NOEs (intra- and inter-residues), intensity ranges, and interproton distances measured for *Lv* I.

Type	Unit and ^{1}H-^{1}H NOE	Intensity[a]	Internuclear distance (Å)	
			Experimental[b]	MD
Intra	D2 H1-H2	Strong	2.40	2.4 ±0.1
	A2 H1-H2	Strong	2.43	2.4 ±0.1
	B1 H1-H2	Strong	2.44	2.4 ±0.1
	B2 H1-H2			2.4 ±0.1
	C1 H1-H2	Medium	2.66	2.4 ±0.1
	C2 H1-H2			2.4 ±0.1
	A2 H1-H3	Medium	3.36	3.8 ±0.1
	B1 H1-H3	Medium	3.50	3.8 ±0.1
	B2 H1-H3			3.8 ±0.1
	A1 H1-H3	Medium	3.32	3.8 ±0.1
	A1 H1-H2	Strong	2.31	2.4 ±0.1
	D1 H1-H2	Strong	2.50	2.4 ±0.1
Inter	A2 H1-D1 H3	Medium	2.68	2.6 ±0.4
	C2 H1-A2 H3	Medium	2.63	2.6 ±0.2
	C2 H1-A2 H4	Medium	2.89	2.3 ±0.3
	B1 H1-C1 H3	Strong	2.48	2.5 ±0.3
	B2 H1-C2 H3			2.6 ±0.3
	D1 H1-B1 H3	Strong	2.11	2.5 ±0.3

[a]NOE intensities were set as strong (≤2.5 Å) and medium (>2.5 and ≤3.7 Å). [b]The internuclear distances were obtained using the formula $NOE = NOE_{ref} \, (r_{ref}/r)^{1/6}$, in which r denotes the distances between the proton pairs, NOE is the intensity of the peaks and ref is the referential values used for normalization. Since the studied units are Fuc$p$ at the exclusive $^1C_4$ ring configuration, the H1–H2 intra-residual NOE-based interproton distance must be 2.4 Å. The D2 H1–H2 distance of 2.4 Å was used for normalization. Ten NOESY spectra were collected under the same conditions for generating average and error values.

Influence of sulfation pattern on the conformation of *Lv* I

The dynamic properties of *Lv* I in solution were determined by the conformational characters of the glycosidic linkages between residues. As seen from Figure 5.2, the φ/ψ angles of glycosidic linkages in *Lv* I presented different distribution patterns, although all residues in *Lv* I are 1-3 linked. Thus, the sulfation patterns at the glycosidic linkages greatly influence the dynamic properties of *Lv* I.
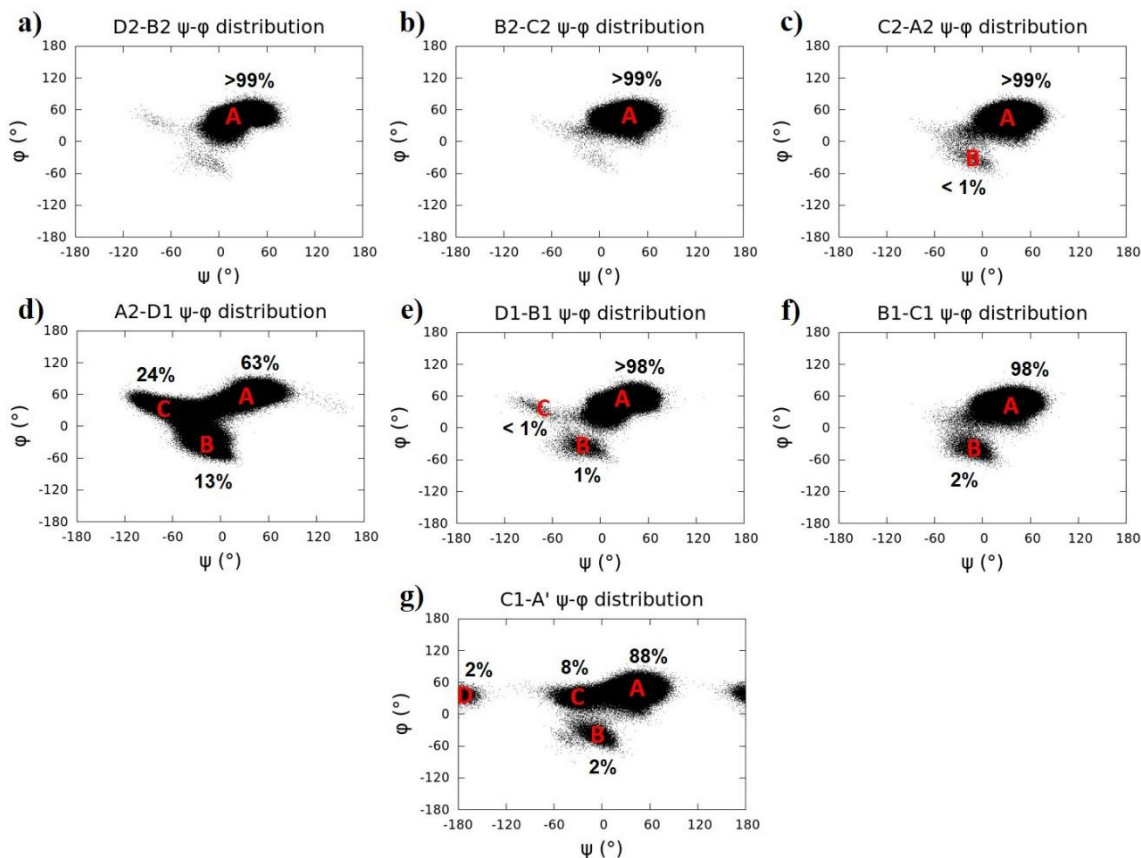
Figure 5.2 Dihedral angle (ψ/ϕ) distribution of the glycosidic linkage in *Lv* I. Each dot in every panel represents a trajectory frame in MD simulation of a disaccharide block of *Lv* I.

In order to study the influence of sulfate patterns on the dynamical properties of *Lv* I, MD simulations of difucose (Figure 5.3) with different sulfation patterns were carried out: difucose A, α-L-Fuc*p*-2(OSO3⁻)-1,3-α-L-Fuc*p*-2(OSO3⁻) (Figure 5.3a); difucose B, α-L-Fuc*p*-2(OSO3⁻)-1,3-α-L-Fuc*p*-4(OSO3⁻) (Figure 5.3c); difucose A, α-L-Fuc*p*-1,3-α-L-Fuc*p* (Figure 5.3f). As seen from Figure 5.3b, difucose A showed a dominant glycosidic linkage conformation (State A). Although the simulation of difucose C explored five glycosidic linkage conformations (State A, B, C, D and E in Figure 5.3g), the total occupancy of three of the states (State B, D and E) was less than 5%. The anomeric and *exo*-anomeric effects provide significant resistance to the

57

glycosidic linkage[16,68], which prefers a tight φ/ψ angle distribution. In addition, hydrogen bond interactions between the sulfate group and hydroxyl group across the glycosidic linkage in difucose A were observed in the MD simulation, which contributed to a tighter distribution of the φ/ψ angle comparing to that in difucose C. As a result, difucose A adopted a single glycosidic linkage conformation for most of the simulation, which matches a low energy conformation of the Carbohydrate Intrinsic (CHI) energy functions[68]. In difucose B, one sulfate group is at C4 position of reducing-end residue and the other one is at C2 position of non-reducing-end residue. The average distance between these two sulfate groups were larger than hydrogen bond lengths in difucose A and difucose C. Comparing to difucose A, the glycosidic linkage in difucose B explored two additional conformations during the course of simulation. As shown in Figure 5.3e, the distances between sulfate groups from structures in state B and C were greater than that from the structure in state A. It is very likely that the different conformational states of glycosidic linkage in difucose B are the results of the competitions between *exo*-anomeric effect and repulsions[188] between the sulfate groups. In state A, the exo-anomeric effect overcomes the repulsions between sulfate groups, while in states B and C, the repulsions between sulfate groups force the ψ angle away from 60°.
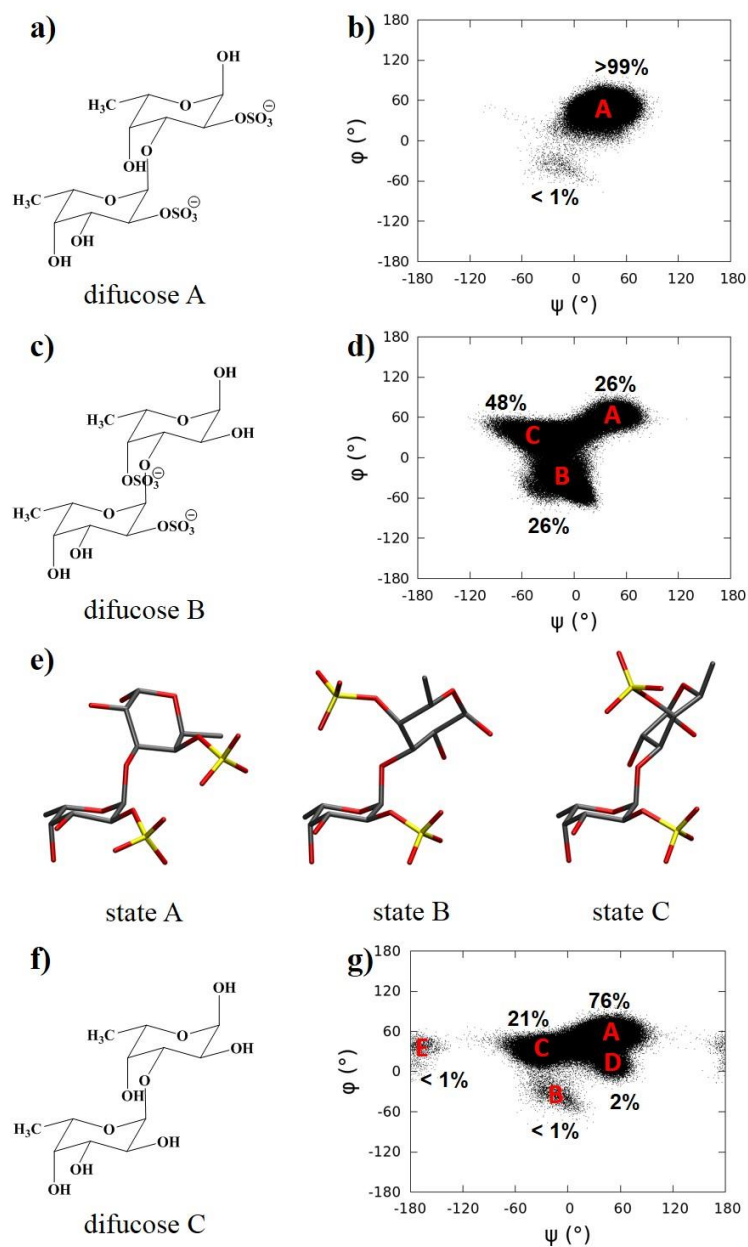
Figure 5.3 Chemical structures of difucose A (a) and its glycosidic linkage conformation (φ/ψ) distribution in the MD simulation (b). Chemical structures of difucose B (c) and its glycosidic linkage conformation (φ/ψ) distribution in the MD simulation (d) with the most representative structure in each state (e). Chemical structures of difucose C (f) and its glycosidic linkage conformation (φ/ψ) distribution in the MD simulation (g).

Based on the observations from the simulations of three difucoses, it is obvious that when sulfation pattern allowed inter-residual hydrogen bonds, glycosidic linkage between fucoses adopted a tight conformation distribution and the structure gained rigidity; when sulfation pattern introduced repulsions, glycosidic linkage adopted more conformations and the dynamics of the structure was enhanced. The sulfation patterns in the first and last four residues (A'-C1, C1-B1, B1-D1 and A2-C2, C2-B2, B2-D', respectively) of *Lv* I (Figure 5.4) allow the inter-residual hydrogen bonds (Table 5.3), and as expected, tight conformation distributions of glycosidic linkages were found; the sulfation pattern in the center of *Lv* I, between D1 and A2 residues (Figure 5.4), likely produces the repulsions between the sulfate groups, thus, generates more motions inside. These motions are later amplified at the end-residues, due to the rigidities in *Lv* I generated by the inter-residual hydrogen bonds. It is also worth noting that the glycosidic linkage between A' and C1 exhibits more conformational states than other linkages. It is possible that the reducing-end residue, A', gained more interactions with the solvent molecules in the simulation, which could contribute to the extra dynamics of the glycosidic linkage.

Table 5.3 Atomic distances in sulfate groups, key inter-residual hydrogen bonds and their occupancies in *Lv* I.

Sulfate group		Hydroxyl group		Distance[a] (Occupancy[b]) Å (%)
Residue	Atom	Residue	Atom	
C1-2SO$_3^-$	O[c]	A'	O4	2.8 ± 0.3 (63)
B1-2SO$_3^-$	O	C1	O4	2.8 ± 0.3 (50)
B1-4SO$_3^-$	O	D1	O1	2.7 ± 0.3 (88)
C2-2SO$_3^-$	O	A2	O4	2.8 ± 0.3 (61)
B2-2SO$_3^-$	O	C2	O4	2.8 ± 0.3 (51)
B2-4SO$_3^-$	O	D2	O2	2.7 ± 0.3 (100)
Sulfate group		Sulfate ester		Distance (Å)
Residue	Atom	Residue	Atom	
A2-2SO$_3^-$	O	D1	O4	5.3 ± 0.7

[a]In Å. [b]Occupancy is calculated in percentage (%), based on a distance between oxygen atoms in sulfate groups and hydroxyl groups of less than 3.5Å. Each of the occupancies of the interaction listed is the sum

of all the individual hydrogen bonds between the oxygen atoms in sulfate group with hydroxyl group and the distance is the average of all the individual hydrogen bonds, unless it is otherwise noted. [c]Oxygen in SO_3^- group.

Conclusion

In this work, theoretical conformational and dynamic analyses of *Lv* I have been reported. Both experimental NMR results and theoretical analyses have shown that the $^1C_4$ chair configuration dominates in all the composing Fuc*p* units and the flexibility of *Lv* I does not originate from the changes in the ring conformation. The MD simulation analyses have pointed to a very selective dynamical process right at the middle of the chain in the *Lv* I, likely due to sulfation-dependent electrostatic repulsive forces and steric hindrance effects. These effects are suggested to be promoted by spatially proximal axial 4-sulfate and equatorial 2-sulfate groups flanking the glycosidic bond of the central disaccharide. This shows curiously that the mobility of *Lv* I is at the center of the molecule. Moreover, it helps to explain, aside from more accessibility to solvent, the amplified motions at the terminal non- and 4-sulfated Fuc*p* units as observed by spin relaxation data (Table 5.2) and plots of dihedral angle distribution (Figure 5.2). The internal units (mostly 2-sulfated) exhibit dynamics strongly regulated by hydrogen bonds. A picture outlining the key hydrogen bonds and the center dynamical spot in the *Lv* I octasaccharide structure is shown at Figure 5.4. A scheme showing the distances of atoms nearby the interglycosidic vectors of disaccharides with sulfation-caused hydrogen bonds and electrostatic repulsion forces is depicted in Figure 5.4.
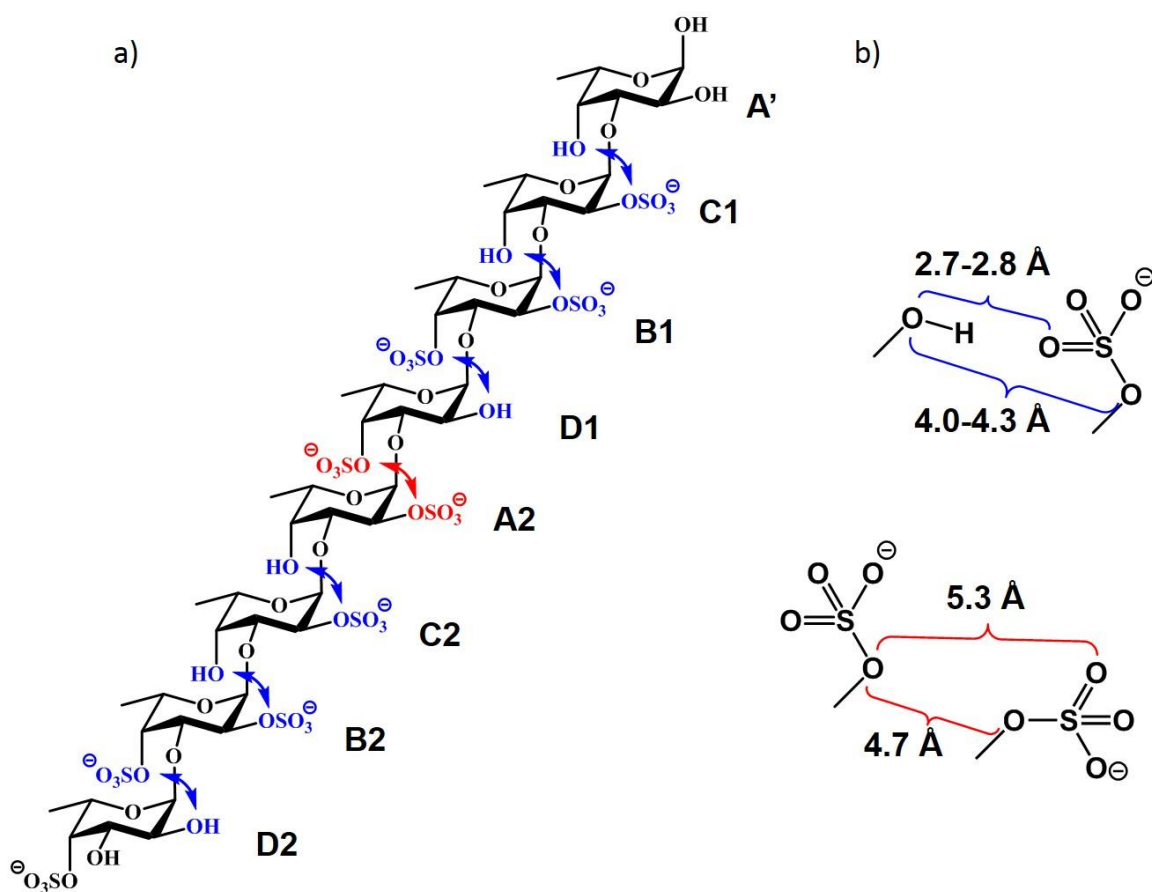
Figure 5.4 Schematic representation for explaining the sulfation pattern-related hydrogen bonds (in blue) and repulsive effects (in red) at (a) the *Lv* I octasaccharide structure, and (b) atomic distances of units. (a) Note that inter-residual hydrogen bonds play a role to decrease dynamics on the two tetrasaccharide repeating units as opposed to the amplified motions occasioned by the repulsive forces between the sulfate groups in A2-D1 disaccharide located right between the two tetrasaccharide sequences. (b) The atomic distances of sulfate-related groups and key inter-residual hydrogen bonds of adjacent units in *Lv* I octasaccharide are reported in Table 5.3.

CHAPTER 6

COMPUTATIONAL MODELING OF PSGL-1 AND ITS GLYCOPEPTIDE ANALOG[213]

Introduction

P-selectins[189,190] and their glycoconjugates, usually found on activated platelets and vascular endothelium, function as mediators of the recruitment of leukocytes to sites of inflammation; and they are rapidly translocated to the cell surface within minutes of an inflammatory stimulus. Excessive trafficking of leukocytes to extravascular location could cause tissue injury that contributes to the development of inflammatory bowel disease, chronic obstructive pulmonary disease, atherosclerosis and postthrombotic syndrome, among a variety of other disorders. Therefore, as a mediator of early adhesion and intracellular signaling events in the inflammation[191], P-selectin represents a promising target for the design of agents that limit adverse inflammatory responses.

P-selectin-glycoprotein-ligand-1[192] (PSGL-1), possesses key functions in multiple biologic processes, is the most well characterized ligand among various glycoprotein counter-receptors which bind selectins with high affinity. It binds with P-selectin expressed on endothelial cell, and forms the initial 'capture and rolling' step in the leukocyte-endothelial cell adhesion process[193]. The interaction of PSGL-1 with P-selectin on activated platelets promotes formation of leukocyte-platelet aggregates that contributes to adhesion and infiltration of inflammatory cells and both activated platelets and soluble P-selectin promote leukocyte infiltration[194-196]. More importantly, the engagement of PSGL-1 to P-selectin activates

intracellular signaling pathways that induces the b2-integrin LFA-1 to adopt an extended conformation associated with the intermediate affinity state, which supports leukocyte deceleration and cell arrest onto the endothelium[197]. PSGL-1 also activates the expression of intracellular protein kinases, such as Rho/Rock kinase, which mediates cell migration, and MAPK kinase that controls expression of pro-inflammatory cytokines[198,199]. Therefore, blocking P-selectin/PSGL-1 complex binding possesses significant potential for the treatment of disorders due to maladaptive acute or chronic inflammatory responses[200-202].

Due to the important biological roles of PSGL-1 in a number of diseases, various biologics, small molecules and glycopeptide mimics have been developed to target these interactions. However, these synthetic blocking antibodies are expensive to manufacture and suffer limited shelf-life, low potency and off-target toxicity[203-205]. Most existing P-selectin inhibitors have been designed to mimic the sLex moiety containing core-2 O-glycan, yet failed to account for the crucial contributions of multiple tyrosine sulfates[204,206-210], due to the limitation of acid sensitivity of tyrosine sulfates[211,212] in synthesis process.

An efficient synthetic approach of generating a diverse set of glycopeptide mimics of PSGL-1 has been reported by collaborators[213]. Key features of this synthesis include an efficient stereo-selective route that has led to multi-gram scale synthesis of the C2 O-glycan and replacement of hydrolytically sensitive tyrosine sulfates with stable, isosteric sulfonate analogues affording compounds with high affinity to P-sel (K_d 14-22 nM). In the process, GSnP-6, a high affinity and chemically stable compound that blocks PSGL-1/P-sel interactions in vitro and in vivo, was identified. In this work, molecular models of both PSGL-1 and GSnP-6 bound to P-selectin were built. MD simulations and MM/GBSA calculations of both protein complexes have been performed and analyzed. Parameters, including suitable GB model and internal

dielectric constant, were developed by reproducing experimental binding affinities of PSGL-1/P-selectin complex in theoretical studies. Molecular models of P-selectin complexes with PSGL-1 and GSnP-6 not only successfully predicted the relative binding affinities between GSnP-6 and PSGL-1, but also provided insights of the origin of the affinity in terms of contributions from the structural components of PSGL-1 and GSnP-6.

Methods

Force field parameters for the oligosaccharide and SO_3^- moiety in the YS and YCS residues were taken from the GLYCAM06 (version h) parameter set[31], while those for the protein came from AMBER12[214] (ff99sb). Parameters for the linkage between the SO_3^- group and the amino acid side chains were approximated from existing terms in two parameter sets (Table S6.1). Ensemble-averaged partial atomic charges for the YS and YCS residues were developed according to the standard GLYCAM protocol[31], from a collection of 300 snapshots extracted at 0.1 ns intervals from MD simulations (30 ns) performed in explicit solvent (TIP3P[44]) for the zwiterionic forms of each amino acid. The initial coordinates for the charge calculations were based on the crystal structure of a monomer of the P-selectin/PSGL-1 complex obtained from the Protein Data Bank (PDB entry code: 1G1S[215]). The coordinates for the tyrosine sulfonate (YCS) residues were generated by replacing the phenolic oxygen atom in the YS residue with a methylene group. Molecular electrostatic potentials were computed at the HF/6-31G*//HF/6-31G* level with the Gaussian03 software package[216], and restrained electrostatic potential charge fitting was performed using the RESP procedure with a restraint weight of 0.01. During charge fitting, the amino acid backbone charges were constrained to the standard values employed in ff99sb[214]. The MD simulations were initiated with RESP charge sets computed for single conformations of each residue. The MD simulations were performed under nPT conditions (12

Å TIP3P[44] water buffer in a cubic box, covalent bonds involving hydrogen atoms constrained using the SHAKE algorithm[34], a time step of 2 fs with heating from 5 to 300K over a period of 50 ps controlled by the Berendsen thermostat[36]). Prior to MD simulation, the systems were subjected to energy minimization under nVT conditions (500 steps steepest descent (SD), followed by 24500 steps of conjugate-gradient (CG) minimization).

Prior to simulation, water molecules and sodium ions in the crystal structure were removed, and the strontium ion substituted by magnesium, because parameters for strontium ion are not available in AMBER. Missing residue E604 was added to Y605 at the N-terminus of the ligand using the LEaP module in AMBER12[214] with the backbone conformation copied from the E154-Y155 sequence. Sodium counter ions were added to each protein-glycopeptide complex to achieve neutrality using LEaP, prior to solvation with TIP3P[44] water (8 Å buffer in an octahedral box). Energy minimization of the solvated complexes was performed in two-steps under nVT conditions. Initially, the positions of water molecules and counter ions were minimized (500 steps SD followed by 24500 steps CG), during which all other solute atoms were restrained (100 kcal/mol·Å^2). In the second step, all restraints were removed with the exception of those on the protein backbone, and the minimization cycle was repeated. Subsequently, heating to 300K was performed over 50 ps (nVT) with a weak restraint (10 kcal/mol·Å^2) on the backbone atoms of P-Selectin only. Systems were then equilibrated at 300 °K for 0.5 ns (nPT ensemble, with the Berendsen thermostat) prior to production MD, under the same conditions (covalent bonds involving hydrogen atoms constrained using SHAKE[34], 2 fs time step). Production MD simulations were for 250 ns performed with the GPU implementation[183] of PMEMD from AMBER12. In all MD simulations, a non-bonded cut-off of 8 Å was applied to van der Waals interactions, with long-range electrostatics treated with the particle mesh Ewald

approximation, and mixed 1-4 non-bonded scale factors applied, as recommended for systems containing both carbohydrates[184] and proteins[217] (SCEE=SCNB=1.0 for the oligosaccharide and SCEE=2.0 and SCNB=1.2 for the protein).

MM/GBSA[3,218] calculations were carried out on 10,000 snapshots extracted evenly from the entire simulation trajectory using the single-trajectory method with the MMPBSA.py.MPI[214] module. The ability to correctly predict the relative per-residue contributions to affinity is essential if the MM-GBSA calculations are to be employed in the rational design of inhibitors based on the PSGL-1 structure. There are several parameterizations of the GBSA model, none of which has been established as optimal for use in analyzing glycopeptide-protein binding. Further, estimating the affinity of PSGL-1/P-selectin interactions faces the additional challenge of quantifying the potentially highly polarizing influence of multiple charge-charge interactions. In a non-polarizable classical force field, one approach to addressing this limitation is to employ an internal dielectric constant (ε) greater than unity in the MM-GBSA analysis. Again, as in the case of the GB approximation, no optimal value for ε has been reported for this type of interaction, although typical values are below 4.0 D. Lastly, entropic effects, arising from changes in conformational flexibility may be estimated separately, but may require very long MD simulations in order to achieve convergence, and are frequently omitted when computing estimates of relative affinity[219]. Because of the novel nature of interactions involving YS and YCS, the suitability of five different GB implementations was examined, specifically: 1) the Hawkins, Cramer, Truhlar pairwise generalized Born model with parameters described by Tsui and Case (GB^{HCT} model, igb=1)[106,220,221] ; the modified GB model developed by Onufriev, Bashford and Case, with the following values for α, β and γ: GB$_1^{OBC}$ (α=0.8, β=0.0, γ=2.909125, igb=2) and GB$_2^{OBC}$ (α=1.0, β=0.8, γ=4.8, igb=5)[222] ; the GBn models described by Mongan,

Simmerling, McCammon, Case and Onufriev: GBn_1, igb=7; GBn_2, igb=8[223,224]. In addition, six

different internal dielectric (ε) values[118] (ε = 1.0, 1.5, 2.0, 3.0, 4.0 and 5.0) were considered in the

MM/GBSA calculations.

Result and Discussion

MD simulations

Computational simulations were used to predict key structural residues responsible for P-

selectin/GSnP-6 interactions. MD simulations were first validated by reproducing all structural

attributes of the PSGL-1/P-selectin interaction, as defined by crystallographic data and

experimental investigations. The positions of the glycan and the peptide components of PSGL-1

were monitored over the course of simulating its interaction with P-selectin (Figure S6.1).

Simulations confirmed that the ligand remained stable in the binding site, with the glycan

displaying less positional variation than the peptide. In addition, all experimentally observed

hydrogen bonds and salt bridges between the ligand and the protein were observed during the

MD simulations. Of note, the dynamic motions of the system weakened some of these

interactions relative to others (Table S6.3). In particular, while the hydrogen bonds associated

with the glycan appeared markedly stable, the interactions between the tyrosine sulfate residues

and the P-selectin surface residues were relatively unstable and depended heavily on the

protonation state of H114. This is consistent with a recent report of Cao et al., who demonstrated

that the protonation state of H114 impacts the binding affinity of PSGL-1 to P-selectin[225]. As

expected, full protonation of this histidine enhanced the stability of these interactions,

particularly those associated with tyrosine sulfate (Y607) and H114[225]. Similarly, the structure

(Figure 6.1) and stability (Figure S6.1) of GSnP-6 in the fully-protonated H114 complex was

comparable to that observed for the native PSGL-1 sequence, and retained all of the key non-covalent interactions (Table S6.2). In the case of neutral H114, the interactions between the tyrosine sulfonates and the protein were unstable, leading to orientation disruption of the peptide component in GSnP-6. The frequency of the rebinding interaction was too low to observe statistical convergence in the MD simulation (Figure S6.1). Thus, only simulations with fully-protonated H114 were employed in the subsequent analysis of interaction energies.

The crystal structure of the PSGL-1 ligand is shown in red with a splined representation of the peptide backbone, sulfated amino acid positions in green and a stick representation for monosaccharide rings. Assuming that H114 is fully protonated leads to optimal reproduction of the crystallographic data for PSGL-1 (**c**) and leads to similar binding for GSnP-6 (**d**). The solvent accessible surface of P-selectin is colored according to the electrostatic potential (acidic region: red; basic region: blue). (**e-g**) Hydrogen bonds between Fuc (red), Core-2Gal (yellow), and Neu5Ac (purple) and P-selectin residues.

Interaction energies

Combinations of five GB parameterizations and protein dielectric constant values (ε) ranging from 1 to 5 were initially evaluated in order to determine which, if any, set of conditions reproduced the relative contributions of PSGL-1 constituent features to P-selectin binding affinity[209]. Optimal conditions were then employed to assess the interaction energies of GSnP-6 with P-selectin, as well as the contributions of distinct structural features of the analogue. The MM/GBSA calculations using either the GB_2^{OBC} and GB_1^{OBC} model produced comparable results with $\varepsilon = 4.0$, in agreement with relative experimental values[209] (Table S6.3). Since the GB_1^{OBC} parameterization has been successfully employed to compute interaction energies in
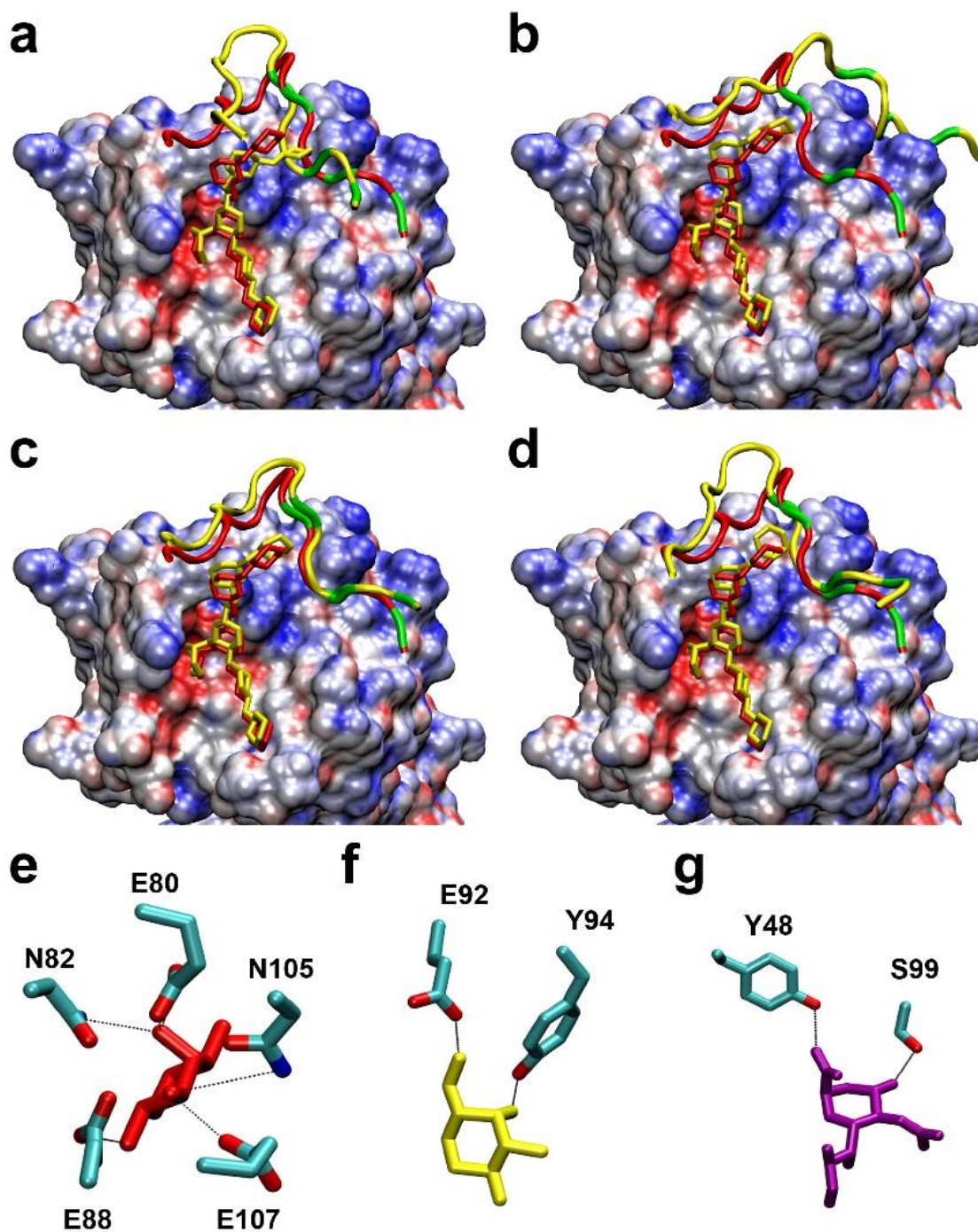
Figure 6.1 Interactions of the N terminus of PSGL-1 and GSnP-6 bound to P-selectin, as a function of the protonation state of H114. Conformation of PSGL-1 (**a**) and GSnP-6 (**b**) ligands most similar to the average shape acquired from MD simulations performed with neutral H114.

other carbohydrate-protein systems involving charged moieties[128], it was selected over the GB$_2$^{OBC} parameterization (Figure S6.2). This combination was able to correctly rank the relative contributions to binding, with Fuc ≈ total sulfate > Neu5Ac, and identified that the second sulfate made a larger contribution (by -1.3 to -1.6 kcal/mol) than the other two.

The experimental interaction energies of P-selectin with PSGL-1 ligand and the analogue GSnP-6 were similar at -9.8 kcal/mol (K_d = 73 nM) and -10.5 kcal/mol (22 nM), respectively. Using the combination of GB$_1$^{OBC} with ε = 4.0, the absolute interaction energies of P-selectin with PSGL-1 and GSnP-6 were predicted to be -45.1 ± 3.4 kcal/mol and -44.6 ± 3.1 kcal/mol, respectively (Table 6.1). As anticipated, due to the omission of entropic penalties, predicted interaction energies were almost 4-fold larger than experimental values. Nonetheless, interaction energies were statistically equivalent, which is in agreement with experimental data. To facilitate detection on binding microarrays, GSnP-6 was chemically derivatized with a 7-amino-4-methylcoumarin (MCA) tag, which alters net charge, as well as local properties of the terminal lysine. The MCA tag was not included in the modeling, which employed only a lysine residue. The remaining residues in the peptide sequence of PSGL-1 and GSnP-6 render essentially identical contributions to the interaction energies. Due to the similarities of the two ligands, the relative contributions of the glycans to binding affinity were also indistinguishable within standard deviations. Likewise, the individual contributions from each sulfate group were comparable for each ligand, despite chemical differences (Error! Reference source not found.). An unexpected and potentially significant observation concerns the second tyrosine sulfate residue (607), whose strong affinity (-8.4 to -9.5 kcal/mol) appears to arise predominantly from van der Waals interactions (-6.6 to -7.0 kcal/mol). In summary, computational analysis reproduced salient structural and energetic features of PSGL-1/P-selectin interactions and

predicted that GSnP-6 would behave comparably, consistent with observed affinity data (Figure 6.1).

Table 6.1 Per-residue MM/GBSA interaction energies[a] for interactions of P-selectin with residues in PSGL-1 and GSnP-6.

	PSGL-1 Sulfates	GSnP-6		PSGL-1 Carbohydrates	GSnP-6
Y/YC 605[b]	-1.4 ±0.6	-1.3 ±0.6	Neu5Ac	-2.7 ±0.8	-2.6 ±0.8
Y/YC 607[b]	-5.5 ±0.7	-6.0 ±0.6	Core-2 Gal	-4.2 ±0.7	-4.2 ±0.7
Y/YC 610[b]	-0.8 ±0.4	-0.6 ±0.3	GlcNAc	-3.3 ±0.6	-3.2 ±0.5
SO_3^- 605[c]	-1.9 ±0.8	-2.3 ±0.9	Fuc	-4.8 ±1.3	-4.7 ±1.3
SO_3^- 607[c]	-2.9 ±0.9	-3.5 ±0.8	GalNAc	-0.6 ±0.3	-0.4 ±0.2
SO_3^- 610[c]	-0.8 ±0.6	-0.8 ±0.4	Gal	0.1 ±0.1	0.1 ±0.1
Subtotal	-13.3 ±1.7	-14.1 ±1.5	Subtotal	-15.5 ±1.8	-15.0 ±1.8
	Amino Acids				
K603[e]	N/A	1.3 ±0.1	L613	-4.1 ±1.1	-3.4 ±1.2
E604	0.1 ±0.2	-0.6 ±0.1	P614	-1.9 ±1.2	-1.3 ±1.0
E606	-1.2 ±0.4	-1.2 ±0.3	E615	-0.9 ±0.2	-0.9 ±0.2
L608	-2.2 ±1.2	-3.3 ±0.8	T616[d]	-1.4 ±0.3	-1.3 ±0.4
D609	-1.8 ±0.7	-1.9 ±0.7	E617	-0.6 ±0.3	-0.5 ±0.3
D611	-0.7 ±0.1	-0.7 ±0.1	P618	-0.4 ±0.1	-0.4 ±0.2
F612	-1.2 ±0.6	-0.9 ±0.5			
			Subtotal	-16.3 ±2.3	-15.1 ±2.1
			Total Interaction Energy	-45.1 ±3.4	-44.6 ±3.1

[a]All results are in kcal/mol. The entropy contributions are not included in these results. [b]Contribution from tyrosine sulfate or tyrosine sulfonate not including the SO3[-] group. [c]SO3[-] is counted as a residue in the energy decomposition, instead of –O-SO$_3^-$ or –CH$_2$-SO$_3^-$. [d]Glycosylation site. [e]Numbering based on the crystal structure.

Conclusion

In this work, molecular models of P-selectin complexes with both PSGL-1 and GSnP-6 have been built and insights of binding affinities have been provided by these models. Both models reproduced the key intermolecular hydrogen bonds observed in crystallographic data. Binding free energies of P-selectin/PSGL-1 complex were calculated with MM/GBSA methodology. A combination of GB parameterization (GB$_1$[OBC]) and protein dielectric constant value ($\varepsilon = 4.0$)

correctly reproduced the relative binding affinity contributions from different residues in PSGL. The same set of parameters was employed in predicting the binding free energies of P-selectin/GSnP-6 complex, which occurred to be statistically equivalent to those of P-selectin/PSGL-1 complex. In addition, the relative contributions from glycans and SO_3^- groups in GSnP-6 were indistinguishable with the corresponding ones in PSGL-1. Both models defined the origin of P-selctin binding affinities in terms of key binding resiudes: Fuc, sulfate/sulfonate groups and Neu5Ac.

CHAPTER 7

CONCLUSIONS

Molecular modeling along with MD simulations have become an important tool for understanding the structural basis and functionality of carbohydrates and related biological molecules. Molecular modeling can provide atomic-level of structural detail as a function of time, which enables its ability to present interactions between specific atoms or moieties and suggest their influences to the entire molecule. Thus, molecular models could help to interpret experimental results in terms of structural behaviors. In Chapter 5, the ^{13}C-based T_1 relaxation analysis from NMR experiments suggested the higher mobility for terminal residues in *Lv* I. The molecular model suggested the difference in dynamics among residues originated from the sulfation pattern in *Lv* I, which led to inter-residue hydrogen bonds or repulsions among different residues. Molecular models could be employed to quantify binding affinities in protein/ligand complexes. In Chapter 6, the molecular models of P-selectin complexes with PSGL-1 and GSnP-6 suggested that both ligands were statistically equivalent in terms of binding affinity, which corresponded to the experimental measurements. The molecular models of both complexes also identified the essential interactions from key residues for binding, and laid ground work for future rational design of PSGL-1 analogs. The accuracy of molecular modeling necessitates proper force field parameters for target systems. In Chapter 4, molecular models of furanoses with a furanose-specific force field parameters reproduced their ring conformational properties in solution and identified the two-state model is not valid for all furanoses.

Molecular modeling methods with classical mechanics and molecular dynamics simulations have achieved tremendous improvements in the last decades, but they are still limited by inherent constraints and approximations in different employed methodologies.

Whether or not an MD simulation can achieve convergence has been debated for years, even though more conformational spaces of biomolecules can be explored with the increase of the computing power and advances in the sampling methodologies. In addition, after an MD simulation explored conformational spaces of a biomolecule, statistically categorizing conformations of biomolecules and binning trajectory frames into different categories, which require multidisciplinary expertise, are also extremely challenging.

Molecular mechanics in general is unable to model chemical reactions, in part due to the harmonic approximations to the bond stretching and angle bending energy potentials, which could not model bond dissociation and formation. This limits the ability of MM methods for studying mechanisms of enzymatic reactions. Recent developments of QM/MM methods, which treat the reaction region with QM methods, could potentially overcome this disadvantage.

In predicting binding affinities, implicit solvent model is commonly employed to calculate the desolvation energies for the binding site. This approximation may have gained efficiency, but its accuracy depends on the quality of the implicit solvent models and the biomolecular system being studied. Including explicit solvent molecules in calculating desolvation energies, however, is slow and suffers large variations in the outcomes, because the conformations of the solvent molecules have significant impacts on the values of calculated desolvation energies. Calculation of entropy is also demanding area for improvements. Although normal-mode analysis has been used for entropy calculation, it suffers from the low efficiency. Without accurate entropy values, predicting binding affinities has been limited to

ranking similar ligands. Polarization effect is not commonly considered in predicting binding affinities, either. Static atomic partial charges employed in MM methods are usually assigned to atoms in a molecule despite of the influence from its surrounding environment. Polarization effects, which redistribute electrons, are commonly observed in binding site of a protein complex, especially when charged moieties are involved. Employing polarizable force fields is a feasible approach to include polarization effect in MM methods, however, polarizable force field is still under development.

In summary, molecular modeling of carbohydrates has become an indispensable tool for studying 3D structures of oligosaccharides and their interactions with other biomolecules. With continuing advances in the methodologies and speed of computing power, molecular modeling is being extended to study larger systems, greater conformation changes and longer time scales of all biomolecular systems. Efforts are also dedicated to make molecular modeling as a tool for experimentalists to use routinely and help them to interpret and understand experimental results on daily basis.

REFERENCES

1. Friesner, R. A. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6648.

2. Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sánchez, M. L.; Villà, J. *Accounts of Chemical Research* **2002**, *35*, 341.

3. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research* **2000**, *33*, 889.

4. Mackerell, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584.

5. Bao, G.; Suresh, S. *Nature Materials* **2003**, *2*, 715.

6. Kapp, L. D.; Lorsch, J. R. *Annual Review of Biochemistry* **2004**, *73*, 657.

7. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Journal of Chemical Physics* **1995**, *103*, 8577.

8. Sagui, C.; Darden, T. A. *Annual Review of Biophysics and Biomolecular Structure* **1999**, *28*, 155.

9. Goldstein, I. J.; Hayes, C. E. *Advances in carbohydrate chemistry and biochemistry* **1978**, *35*, 127.

10. Ashwell, G.; Harford, J. *Annual Review of Biochemistry* **1982**, *51*, 531.

11. Hakomori, S. I. *Advances in Cancer Research* **1989**, *52*, 257.

12. Lee, Y. C.; Lee, R. T. *Accounts of Chemical Research* **1995**, *28*, 321.

13. Koch, K. E. *Annual Review of Plant Physiology and Plant Molecular Biology* **1996**, *47*, 509.

14. Lis, H.; Sharon, N. *Chemical Reviews* **1998**, *98*, 637.

15. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364.

16. Foley, B. L.; Tessier, M. B.; Woods, R. J. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012**, *2*, 652.

17. Richards, M. R.; Bai, Y.; Lowary, T. L. *Carbohydrate Research* **2013**, *374*, 103.

18. Callam, C. S.; Lowary, T. L. *Journal of Organic Chemistry* **2001**, *66*, 8961.

19. D'Souza, F. W.; Ayers, J. D.; McCarren, P. R.; Lowary, T. L. *Journal of the American Chemical Society* **2000**, *122*, 1251.

20. Gordon, M. T.; Lowary, T. L.; Hadad, C. M. *Journal of the American Chemical Society* **1999**, *121*, 9682.

21. Polak, M.; Mohar, B.; Kobe, J.; Plavec, J. *Journal of the American Chemical Society* **1998**, *120*, 2508.

22. Senderowitz, H.; Parish, C.; Still, W. C. *Journal of the American Chemical Society* **1996**, *118*, 2078.

23. Ellervik, U.; Magnusson, G. *Journal of the American Chemical Society* **1994**, *116*, 2340.

24. Cossé-Barbi, A.; Watson, D. G.; Dubois, J. E. *Tetrahedron Letters* **1989**, *30*, 163.

25. Cosse-Barbi, A.; Dubois, J. E. *Journal of the American Chemical Society* **1987**, *109*, 1503.

26. Jeffrey, G. A.; Pople, J. A.; Binkley, J. S.; Vishveshwara, S. *Journal of the American Chemical Society* **1978**, *100*, 373.

27. Lemieux, R. U.; Koto, S. *Tetrahedron* **1974**, *30*, 1933.

28.   Jeffrey, G. A.; Pople, J. A.; Radom, L. *Carbohydrate Research* **1972**, *25*, 117.

29.   Rees, D. A.; Scott, W. E. *Journal of the Chemical Society B: Physical Organic* **1971**, 469.

30.   Goekjian, P. G.; Wu, T. C.; Kishi, Y. *Journal of Organic Chemistry* **1991**, *56*, 6412.

31.   Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *Journal of Computational Chemistry* **2008**, *29*, 622.

32.   Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *Journal of Physical Chemistry* **1993**, *97*, 10269.

33.   Verlet, L. *Physical Review* **1967**, *159*, 98.

34.   Vangunsteren, W. F.; Berendsen, H. J. C. *Molecular Physics* **1977**, *34*, 1311.

35.   Martyna, G. J.; Tobias, D. J.; Klein, M. L. *Journal of Chemical Physics* **1994**, *101*, 4177.

36.   Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684.

37.   Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684.

38.   Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523.

39.   Sattelle, B. M.; Hansen, S. U.; Gardiner, J.; Almond, A. *Journal of the American Chemical Society* **2010**, *132*, 13132.

40.   Taha, H. A.; Richards, M. R.; Lowary, T. L. *Chemical Reviews* **2013**, *113*, 1851.

41.   Gu, W.; Wang, T. T.; Zhu, J.; Shi, Y. Y.; Liu, H. Y. *Biophysical Chemistry* **2003**, *104*, 79.

42.   Wang, J.; Tan, C. H.; Tan, Y. H.; Lu, Q.; Luo, R. *Communications in Computational Physics* **2008**, *3*, 1010.

43.   Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins-Structure Function and Genetics* **1998**, *32*, 399.

44.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926.

45.   Cao, Z. X.; Zhang, X. M.; Liu, L.; Zhao, L. L.; Li, H. Y.; Wang, J. H. *International Journal of Molecular Sciences* **2015**, *16*, 14291.

46.   Cohen, M.; Varki, A. In *International Review of Cell and Molecular Biology, Vol 308*; Jeon, K. W., Ed.; Elsevier Academic Press Inc: San Diego, 2014; Vol. 308, p 75.

47.   Varki, A.; Angata, T. *Glycobiology* **2006**, *16*, 1R.

48.   Van Breedam, W.; Pohlmann, S.; Favoreel, H. W.; de Groot, R. J.; Nauwynck, H. J. *Fems Microbiology Reviews* **2014**, *38*, 598.

49.   Varki, A. *Essentials of glycobiology*; Cold Spring Harbor Laboratory Press, Plainview, New York 11803, USA, 1999.

50.   Taylor, M. E.; Drickamer, K.; Taylor, M. E.; Drickamer, K. *Introduction to glycobiology*; Oxford University Press, Great Clarendon Street, Oxford, Oxon, OX2 6DP, UK, 2003.

51.   Marth, J. D. *Nature Cell Biology* **2008**, *10*, 1015.

52.   Cohen, M. *Biomolecules* **2015**, *5*, 2056.

53.   Collins, B. E.; Paulson, J. C. *Current Opinion in Chemical Biology* **2004**, *8*, 617.

54.   Honke, K.; Taniguchi, N. *Medicinal Research Reviews* **2002**, *22*, 637.

55.   Dietrich, C. P.; Nader, H. B.; Depaiva, J. F.; Santos, E. A.; Holme, K. R.; Perlin, A. S. *International Journal of Biological Macromolecules* **1989**, *11*, 361.

56.   Medeiros, G. F.; Mendes, A.; Castro, R. A. B.; Bau, E. C.; Nader, H. B.; Dietrich, C. P. *Biochimica Et Biophysica Acta-General Subjects* **2000**, *1475*, 287.

57.   Lindahl, U.; Kusche-Gullberg, M.; Kjellén, L. *Journal of Biological Chemistry* **1998**, *273*, 24979.

58.   Queiroz, I. N. L.; Wang, X. C.; Glushka, J. N.; Santos, G. R. C.; Valente, A. P.; Prestegard, J. H.; Woods, R. J.; Mourao, P. A. S.; Pomin, V. H. *Glycobiology* **2015**, *25*, 535.

59.   Coombe, D. R.; Kett, W. C. *Cellular and Molecular Life Sciences* **2005**, *62*, 410.

60.   Rabenstein, D. L. *Natural Product Reports* **2002**, *19*, 312.

61.   Shen, J.; Wang, H. J.; Xia, Y. M. *Structural Chemistry* **2014**, *25*, 1423.

62.   Mobli, M.; Nilsson, M.; Almond, A. *Glycoconjugate Journal* **2008**, *25*, 401.

63.   Mulloy, B.; Forster, M. J. *Glycobiology* **2000**, *10*, 1147.

64.   Mulloy, B. *Anais Da Academia Brasileira De Ciencias* **2005**, *77*, 651.

65.   Allen, B. L.; Filla, M. S.; Rapraeger, A. C. *Journal of Cell Biology* **2001**, *155*, 845.

66.   Allen, B. L.; Rapraeger, A. C. *Journal of Cell Biology* **2003**, *163*, 637.

67.   Atha, D. H.; Lormeau, J. C.; Petitou, M.; Rosenberg, R. D.; Choay, J. *Biochemistry* **1985**, *24*, 6723.

68.   Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J. *Journal of Computational Chemistry* **2014**, *35*, 526.

69.   www.glycam.org, Complex Carbohydrate Research Center, the University of Georgia, Athens, Georgia, U.S.

70.   Taha, H. A.; Roy, P. N.; Lowary, T. L. *Journal of Chemical Theory and Computation* **2011**, *7*, 420.

71.   Taha, H. A.; Castillo, N.; Sears, D. N.; Wasylishen, R. E.; Lowary, T. L.; Roy, P. N. *Journal of Chemical Theory and Computation* **2010**, *6*, 212.

72.   Zhao, H.; Pan, Q.; Zhang, W.; Carmichael, I.; Serianni, A. S. *The Journal of Organic Chemistry* **2007**, *72*, 7071.

73.   Houseknecht, J. B.; Lowary, T. L.; Hadad, C. M. *Journal of Physical Chemistry A* **2003**, *107*, 372.

74.   Stenutz, R.; Carmichael, I.; Widmalm, G.; Serianni, A. S. *The Journal of Organic Chemistry* **2002**, *67*, 949.

75.   Cloran, F.; Carmichael, I.; Serianni, A. S. *Journal of the American Chemical Society* **1999**, *121*, 9843.

76.   Bose, B.; Zhao, S.; Stenutz, R.; Cloran, F.; Bondo, P. B.; Bondo, G.; Hertz, B.; Carmichael, I.; Serianni, A. S. *Journal of the American Chemical Society* **1998**, *120*, 11158.

77.   Altona, C.; Francke, R.; Dehaan, R.; Ippel, J. H.; Daalmans, G. J.; Hoekzema, A.; Vanwijk, J. *Magnetic Resonance in Chemistry* **1994**, *32*, 670.

78.   Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.

79.   Gonzalez-Outeirino, J.; Kirschner, K. N.; Thobhani, S.; Woods, R. J. *Canadian Journal of Chemistry-Revue Canadienne Chimie* **2006**, *84*, 569.

80.   Sun, G.; Voigt, J. H.; Filippov, I. V.; Marquez, V. E.; Nicklaus, M. C. *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 1752.

81.   Padrta, P.; Sklenar, V. *Journal of Biomolecular NMR* **2002**, *24*, 339.

82.   De Leeuw, F. A. A. M.; Altona, C. *Journal of Computational Chemistry* **1983**, *4*, 428.

83.   Brewer, C. F.; Miceli, M. C.; Baum, L. G. *Current Opinion in Structural Biology* **2002**, *12*, 616.

84.   Crocker, P. R. *Current Opinion in Structural Biology* **2002**, *12*, 609.

85.   Houzelstein, D.; Goncalves, I. R.; Fadden, A. J.; Sidhu, S. S.; Cooper, D. N. W.; Drickamer, K.; Leffler, H.; Poirier, F. *Molecular Biology and Evolution* **2004**, *21*, 1177.

86.   Rabinovich, G. A.; Baum, L. G.; Tinari, N.; Paganelli, R.; Natoli, C.; Liu, F. T.; Iacobelli, S. *Trends in Immunology* **2002**, *23*, 313.

87.   Doyle, M. L. *Current Opinion in Biotechnology* **1997**, *8*, 31.

88.   Heegaard, N. H. H.; Nilsson, S.; Guzman, N. A. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **1998**, *715*, 29.

89.   Mann, D. A.; Kanai, M.; Maly, D. J.; Kiessling, L. L. *Journal of the American Chemical Society* **1998**, *120*, 10575.

90.   Hirabayashi, J.; Arata, Y.; Kasai, K. *Journal of Chromatography A* **2000**, *890*, 261.

91.   Abdiche, Y.; Malashock, D.; Pinkerton, A.; Pons, J. *Analytical Biochemistry* **2008**, *377*, 209.

92.   Liang, C. H.; Wu, C. Y. *Expert Review of Proteomics* **2009**, *6*, 631.

93.   Wu, C. Y.; Liang, P. H.; Wong, C. H. *Organic & Biomolecular Chemistry* **2009**, *7*, 2247.

94.   Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.

95.   Cieplak, P. *Molecular Simulation* **2002**, *28*, 173.

96.   Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H. *Journal of Computational Chemistry* **2001**, *22*, 1830.

97.   Lee, M. R.; Baker, D.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 1040.

98.   Lee, M. R.; Kollman, P. A. *Structure* **2001**, *9*, 905.

99.   Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *Journal of Medicinal Chemistry* **2005**, *48*, 4040.

100.   Simonson, T. *Current Opinion in Structural Biology* **2001**, *11*, 243.

101. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *Journal of the American Chemical Society* **1998**, *120*, 9401.

102. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annual Review of Biophysics and Biomolecular Structure* **2001**, *30*, 211.

103. Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 5221.

104. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *Journal of the American Chemical Society* **1990**, *112*, 6127.

105. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *Journal of Physical Chemistry A* **1997**, *101*, 3005.

106. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chemical Physics Letters* **1995**, *246*, 122.

107. Jayaram, B.; Sprous, D.; Beveridge, D. L. *Journal of Physical Chemistry B* **1998**, *102*, 9571.

108. Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51*, 129.

109. Hou, T. J.; Guo, S. L.; Xu, X. J. *Journal of Physical Chemistry B* **2002**, *106*, 5527.

110. Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. *Journal of Medicinal Chemistry* **2002**, *45*, 1412.

111. Brown, S. P.; Muchmore, S. W. *Journal of Chemical Information and Modeling* **2006**, *46*, 999.

112. Lepsik, M.; Kriz, Z.; Havlas, Z. *Proteins* **2004**, *57*, 279.

113. Kuhn, B.; Kollman, P. A. *Journal of Medicinal Chemistry* **2000**, *43*, 3786.

114. Hou, T. J.; Zhu, L. L.; Chen, L. R.; Xu, X. J. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 273.

115. Hou, T. J.; Yu, R. *Journal of Medicinal Chemistry* **2007**, *50*, 1177.

116. Stoica, I.; Sadiq, S. K.; Coveney, P. V. *Journal of the American Chemical Society* **2008**, *130*, 2639.

117. Wang, W.; Kollman, P. A. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 14937.

118. Wang, W.; Kollman, P. A. *Journal of Molecular Biology* **2000**, *303*, 567.

119. Hou, T. J.; Chen, K.; McLaughlin, W. A.; Lu, B. Z.; Wang, W. *Plos Computational Biology* **2006**, *2*, 46.

120. Gohlke, H.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 238.

121. Pearlman, D. A. *Journal of Medicinal Chemistry* **2005**, *48*, 7796.

122. Woods, R. J.; Tessier, M. B. *Current Opinion in Structural Biology* **2010**, *20*, 575.

123. Wang, J. M.; Hou, T. J.; Xu, X. J. *Current Computer-Aided Drug Design* **2006**, *2*, 287.

124. Chang, C.-e. A.; Chen, W.; Gilson, M. K. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 1534.

125. Doig, A. J.; Sternberg, M. J. *Protein Science : A Publication of the Protein Society* **1995**, *4*, 2247.

126. Hou, T.; Wang, J.; Li, Y.; Wang, W. *Journal of Computational Chemistry* **2011**, *32*, 866.

127. Baker, C. M. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 241.

128. Hou, T.; Wang, J.; Li, Y.; Wang, W. *Journal of Chemical Information and Modeling* **2011**, *51*, 69.

129. Levitt, M.; Warshel, A. *Journal of the American Chemical Society* **1978**, *100*, 2607.

130.   Bartenev, V. N.; Kameneva, N. G.; Lipanov, A. A. *Acta Crystallographica Section B-Structural Science* **1987**, *43*, 275.

131.   Seo, M.; Castillo, N.; Ganzynkowicz, R.; Daniels, C. R.; Woods, R. J.; Lowary, T. L.; Roy, P. N. *Journal of Chemical Theory and Computation* **2008**, *4*, 184.

132.   Angyal, S. J. *Advances in Carbohydrate Chemistry and Biochemistry* **1984**, *42*, 15.

133.   Houseknecht, J. B.; Lowary, T. L.; Hadad, C. M. *Journal of Physical Chemistry A* **2003**, *107*, 5763.

134.   Altona, C.; Sundaral.M *Journal of the American Chemical Society* **1972**, *94*, 8205.

135.   Altona, C.; Sundaralingam, M. *Journal of the American Chemical Society* **1973**, *95*, 2333.

136.   Hendrickx, P. M. S.; Corzana, F.; Depraetere, S.; Tourwe, D. A.; Augustyns, K.; Martins, J. C. *Journal of Computational Chemistry* **2010**, *31*, 561.

137.   Hatcher, E.; Guvench, O.; MacKerell, A. D. *Journal of Physical Chemistry B* **2009**, *113*, 12466.

138.   Raman, E. P.; Guvench, O.; MacKerell, A. D. *The Journal of Physical Chemistry B* **2010**, *114*, 12981.

139.   Plavec, J.; Tong, W.; Chattopadhyaya, J. *Journal of the American Chemical Society* **1993**, *115*, 9734.

140.   Thibaudeau, C.; Chattopadhyaya, J. *Nucleosides & Nucleotides* **1997**, *16*, 523.

141.   Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *Journal of the American Chemical Society* **1994**, *116*, 6558.

142.   Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J. *Journal of the American Chemical Society* **1994**, *116*, 8033.

143. Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *Pure and Applied Chemistry* **1996**, *68*, 2137.

144. Ishii, T.; Senoo, T.; Kozakai, T.; Fukada, K.; Sakane, G. *Acta crystallographica. Section E, Crystallographic communications* **2015**, *71*, o139.

145. Ishii, T.; Ohga, S.; Fukada, K.; Morimoto, K.; Sakane, G. *Acta crystallographica. Section E, Structure reports online* **2014**, *70*, o569.

146. Ohanessian, J.; Avenel, D.; Kanters, J. A.; Smits, D. *Acta Crystallographica Section B-Structural Science* **1977**, *33*, 1063.

147. Ohanessian, J.; Gillierpandraud, H. *Acta Crystallographica Section B-Structural Science* **1976**, *32*, 2810.

148. Longchambon, F.; Avenel, D.; Neuman, A. *Acta Crystallographica Section B-Structural Science* **1976**, *32*, 1822.

149. McDonald, T. R. R.; Beevers, C. A. *Acta Crystallographica* **1952**, *5*, 654.

150. Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. *Journal of Chemical Theory and Computation* **2014**, *10*, 4515.

151. Kline, P. C.; Serianni, A. S. *Journal of the American Chemical Society* **1990**, *112*, 7373.

152. Serianni, A. S.; Barker, R. *Journal of Organic Chemistry* **1984**, *49*, 3292.

153. Church, T. J.; Carmichael, I.; Serianni, A. S. *Journal of the American Chemical Society* **1997**, *119*, 8946.

154. Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48*, 5019.

155. Lii, J.-H.; Chen, K.-H.; Allinger, N. L. *Journal of Computational Chemistry* **2003**, *24*, 1504.

156. Lii, J. H.; Chen, K. H.; Johnson, G. P.; French, A. D.; Allinger, N. L. *Carbohydrate Research* **2005**, *340*, 853.

157. Tvaroska, I.; Bleha, T. In *Tipson, R. S. And D. Horton* 1989, p 45.

158. Angyal, S. J. *Carbohydrate Research* **1979**, *77*, 37.

159. Besra, G. S.; Khoo, K. H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J. *Biochemistry* **1995**, *34*, 4257.

160. Brennan, P. J.; Nikaido, H. *Annual Review of Biochemistry* **1995**, *64*, 29.

161. Crick, D. C.; Mahapatra, S.; Brennan, P. J. *Glycobiology* **2001**, *11*, 107R.

162. Bhamidi, S.; Scherman, M. S.; Rithner, C. D.; Prenni, J. E.; Chatterjee, D.; Khoo, K. H.; McNeil, M. R. *Journal of Biological Chemistry* **2008**, *283*, 12992.

163. Latge, J. P.; Kobayashi, H.; Debeaupuis, J. P.; Diaquin, M.; Sarfati, J.; Wieruszeski, J. M.; Parra, E.; Bouchara, J. P.; Fournet, B. *Infection and Immunity* **1994**, *62*, 5424.

164. Fontaine, T.; Simenel, C.; Dubreucq, G.; Adam, O.; Delepierre, M.; Lemoine, J.; Vorgias, C. E.; Diaquin, M.; Latge, J. P. *Journal of Biological Chemistry* **2000**, *275*, 27594.

165. Leitao, E. A.; Bittencourt, V. C. B.; Haido, R. M. T.; Valente, A. P.; Peter-Katalinic, J.; Letzel, M.; de Souza, L. M.; Barreto-Bergter, E. *Glycobiology* **2003**, *13*, 681.

166. Costachel, C.; Coddeville, B.; Latge, J. P.; Fontaine, T. *Journal of Biological Chemistry* **2005**, *280*, 39835.

167. Beauvais, A.; Schmidt, C.; Guadagnini, S.; Roux, P.; Perret, E.; Henry, C.; Paris, S.; Mallet, A.; Prevost, M. C.; Latge, J. P. *Cellular Microbiology* **2007**, *9*, 1588.

168. Lamarre, C.; Beau, R.; Balloy, V.; Fontaine, T.; Hoi, J. W. S.; Guadagnini, S.; Berkova, N.; Chignard, M.; Beauvais, A.; Latge, J. P. *Cellular Microbiology* **2009**, *11*, 1612.

169.	Golgher, D. B.; Colli, W.; Soutopadron, T.; Zingales, B. *Molecular and Biochemical Parasitology* **1993**, *60*, 249.

170.	Carapito, R.; Imberty, A.; Jeltsch, J. M.; Byrns, S. C.; Tam, P. H.; Lowary, T. L.; Varrot, A.; Phalip, V. *Journal of Biological Chemistry* **2009**, *284*, 12285.

171.	Predy, G. N.; Goel, V.; Lovlin, R.; Donner, A.; Stitt, L.; Basu, T. K. *Canadian Medical Association Journal* **2005**, *173*, 1043.

172.	Cooper, P. D.; Petrovsky, N. *Glycobiology* **2011**, *21*, 595.

173.	Pomin, V. H.; Mourao, P. A. S. *Glycobiology* **2008**, *18*, 1016.

174.	Vilela-Silva, A.-C. E. S.; Hirohashi, N.; Mourao, P. A. S. *International Journal of Developmental Biology* **2008**, *52*, 551.

175.	Pomin, V. H. *Biopolymers* **2009**, *91*, 601.

176.	Biermann, C. H.; Marks, J. A.; Vilela-Silva, A.; Castro, M. O.; Mourao, P. A. S. *Evolution & Development* **2004**, *6*, 353.

177.	Mourao, P. A. S. *Brazilian Journal of Medical and Biological Research* **2007**, *40*, 5.

178.	Becker, C. F.; Guimaraes, J. A.; Mourao, P. A. S.; Verli, H. *Journal of Molecular Graphics & Modelling* **2007**, *26*, 391.

179.	Pomin, V. H. *Progress in Biophysics & Molecular Biology* **2014**, *114*, 61.

180.	Mulloy, B.; Ribeiro, A. C.; Alves, A. P.; Vieira, R. P.; Mourao, P. A. S. *Journal of Biological Chemistry* **1994**, *269*, 22113.

181.	Cinelli, L. P.; Castro, M. O.; Santos, L. L.; Garcia, C. R.; Vilela-Silva, A.-C. E.; Mourao, P. A. *Glycobiology* **2007**, *17*, 877.

182.	Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.;

Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu X.; Kollman, P. A. *Amber14*; University of California: San Francisco, CA, 2014.

183. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542.

184. Kirschner, K. N.; Woods, R. J. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 10541.

185. Dennington, R. I.; Keith, T.; Millam, J. M.; Eppinnett, K.; Hovell, W. L. *Semichem, Inc.* **2003**, *Shawnee Mission KS*.

186. Lutnæs, O. B.; Ruden, T. A.; Helgaker, T. *Magnetic Resonance in Chemistry* **2004**, *42*, S117.

187. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V.

G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.

188. Zsiska, M.; Meyer, B. *Carbohydrate Research* **1993**, *243*, 225.

189. Palabrica, T.; Lobb, R.; Furie, B. C.; Aronovitz, M.; Benjamin, C.; Hsu, Y. M.; Sajer, S. A.; Furie, B. *Nature* **1992**, *359*, 848.

190. Mayadas, T. N.; Johnson, R. C.; Rayburn, H.; Hynes, R. O.; Wagner, D. D. *Cell* **1993**, *74*, 541.

191. Zarbock, A.; Ley, K.; McEver, R. P.; Hidalgo, A. *Blood* **2011**, *118*, 6743.

192. Carlow, D. A.; Gossens, K.; Naus, S.; Veerman, K. M.; Seo, W.; Ziltener, H. J. *Immunological Reviews* **2009**, *230*, 75.

193. McEver, R. P.; Cummings, R. D. *Journal of Clinical Investigation* **1997**, *100*, 485.

194. Wang, H.-B.; Wang, J.-T.; Zhang, L.; Geng, Z. H.; Xu, W.-L.; Xu, T.; Huo, Y.; Zhu, X.; Plow, E. F.; Chen, M.; Geng, J.-G. *Nature Immunology* **2007**, *8*, 882.

195. Sreeramkumar, V.; Adrover, J. M.; Ballesteros, I.; Isabel Cuartero, M.; Rossaint, J.; Bilbao, I.; Nacher, M.; Pitaval, C.; Radovanovic, I.; Fukui, Y.; McEver, R. P.; Filippi, M.-D.; Lizasoain, I.; Ruiz-Cabello, J.; Zarbock, A.; Moro, M. A.; Hidalgo, A. *Science* **2014**, *346*, 1234.

196. Martins, P. D.; van den Berk, N.; Ulfman, L. H.; Koenderman, L.; Hordijk, P. L.; Zwaginga, J. J. *Arteriosclerosis Thrombosis and Vascular Biology* **2004**, *24*, 193.

197. Ma, Y. Q.; Plow, E. F.; Geng, J. G. *Blood* **2004**, *104*, 2549.

198. Fox, R.; Nhan, T. Q.; Law, G. L.; Morris, D. R.; Liles, W. C.; Schwartz, S. M. *Embo Journal* **2007**, *26*, 505.

199. Hidari, K.; Weyrich, A. S.; Zimmerman, G. A.; McEver, R. P. *Journal of Biological Chemistry* **1997**, *272*, 28750.

200. Dong, Z. M.; Brown, A. A.; Wagner, D. D. *Circulation* **2000**, *101*, 2290.

201. Sato, C.; Shikata, K.; Hirota, D.; Sasaki, M.; Nishishita, S.; Miyamoto, S.; Kodera, R.; Ogawa, D.; Tone, A.; Kataoka, H. U.; Wada, J.; Kajitani, N.; Makino, H. *Diabetes* **2011**, *60*, 189.

202. von Bruehl, M.-L.; Stark, K.; Steinhart, A.; Chandraratne, S.; Konrad, I.; Lorenz, M.; Khandoga, A.; Tirniceriu, A.; Coletti, R.; Koellnberger, M.; Byrne, R. A.; Laitinen, I.; Walch, A.; Brill, A.; Pfeiler, S.; Manukyan, D.; Braun, S.; Lange, P.; Riegger, J.; Ware, J.; Eckart, A.; Haidari, S.; Rudelius, M.; Schulz, C.; Echtler, K.; Brinkmann, V.; Schwaiger, M.; Preissner, K. T.; Wagner, D. D.; Mackman, N.; Engelmann, B.; Massberg, S. *Journal of Experimental Medicine* **2012**, *209*, 819.

203. Baert, F.; Noman, M.; Vermeire, S.; Van Assche, G.; D'Haens, G.; Carbonez, A.; Rutgeerts, P. *New England Journal of Medicine* **2003**, *348*, 601.

204. Chang, J.; Patton, J. T.; Sarkar, A.; Ernst, B.; Magnani, J. L.; Frenette, P. S. *Blood* **2010**, *116*, 1779.

205. Beeh, K. M.; Beier, J.; Meyer, M.; Buhl, R.; Zahlten, R.; Wolff, G. *Pulmonary Pharmacology & Therapeutics* **2006**, *19*, 233.

206. Kranich, R.; Busemann, A. S.; Bock, D.; Schroeter-Maas, S.; Beyer, D.; Heinemann, B.; Meyer, M.; Schierhorn, K.; Zahlten, R.; Wolff, G.; Aydt, E. M. *Journal of Medicinal Chemistry* **2007**, *50*, 1101.

207.   Huang, A.; Moretto, A.; Janz, K.; Lowe, M.; Bedard, P. W.; Tam, S.; Di, L.; Clerin, V.; Sushkova, N.; Tchernychev, B.; Tsao, D. H. H.; Keith, J. C., Jr.; Shaw, G. D.; Schaub, R. G.; Wang, Q.; Kaila, N. *Journal of Medicinal Chemistry* **2010**, *53*, 6003.

208.   Leppanen, A.; Mehta, P.; Ouyang, Y. B.; Ju, T. Z.; Helin, J.; Moore, K. L.; van Die, I.; Canfield, W. M.; McEver, R. P.; Cummings, R. D. *Journal of Biological Chemistry* **1999**, *274*, 24838.

209.   Leppanen, A.; White, S. P.; Helin, J.; McEver, R. P.; Cummings, R. D. *Journal of Biological Chemistry* **2000**, *275*, 39569.

210.   Leppanen, A.; Yago, T.; Otto, V. I.; McEver, R. P.; Cummings, R. D. *Journal of Biological Chemistry* **2003**, *278*, 26391.

211.   Koeller, K. M.; Smith, M. E. B.; Huang, R. F.; Wong, C. H. *Journal of the American Chemical Society* **2000**, *122*, 4241.

212.   Huang, K.-T.; Wu, B.-C.; Lin, C.-C.; Luo, S.-C.; Chen, C.; Wong, C.-H.; Lin, C.-C. *Carbohydrate Research* **2006**, *341*, 2151.

213.   Krishnamurthy, V. R.; Sardar, M. Y. R.; Ying, Y.; Song, X. Z.; Haller, C.; Dai, E. B.; Wang, X. C.; Hanjaya-Putra, D.; Sun, L. J.; Morikis, V.; Simon, S. I.; Woods, R. J.; Cummings, R. D.; Chaikof, E. L. *Nature Communication* **2015**, *6*, 12.

214.   Case, D. A.; Darden, T. A.; Cheatham, I., T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-

Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber12*; University of California: San Francisco, CA, 2012.

215. Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T. *Cell* **2000**, *103*, 467.

216. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2009.

217. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668.

218. Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. *Journal of Biomolecular Structure and Dynanics* **1998**, *16*, 671.

219. Genheden, S.; Ryde, U. *Physical Chemistry Chemical Physics* **2012**, *14*, 8662.

220.   Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Journal of Physical Chemistry* **1996**, *100*, 19824.

221.   Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275.

222.   Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.

223.   Nguyen, H.; Roe, D. R.; Simmerling, C. *Journal of Chemical Theory and Computation* **2013**, *9*, 2020.

224.   Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *Journal of Chemical Theory and Computation* **2007**, *3*, 156.

225.   Cao, T. M.; Takatani, T.; King, M. R. *Biophysical Journal* **2013**, *104*, 292.

226.   Evdokimov, A. G.; Kalb, A. J.; Koetzle, T. F.; Klooster, W. T.; Martin, J. M. L. *Journal of Physical Chemistry A* **1999**, *103*, 744.

227.   Evdokimov, A.; Gilboa, A. J.; Koetzle, T. F.; Klooster, W. T.; Schultz, A. J.; Mason, S. A.; Albinati, A.; Frolow, F. *Acta Crystallographica Section B: Structural Science* **2001**, *57*, 213.

228.   Podlasek, C. A.; Stripe, W. A.; Carmichael, I.; Shang, M. Y.; Basu, B.; Serianni, A. S. *Journal of the American Chemical Society* **1996**, *118*, 1413.

229.   Houseknecht, J. B.; Altona, C.; Hadad, C. M.; Lowary, T. L. *Journal of Organic Chemistry* **2002**, *67*, 4647.

230.   Case, D. A.; Darden, T. A.; Cheatham; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D.

H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *Amber9*; University of California: San Francisco, CA, 2009.

231.   Kirschner, K. N.; Woods, R. J. *Proceedings of the National Academy of Sciences* **2001**, *98*, 10541.

232.   Han, D.; Wang, L. N.; Zhong, W. H.; Meng, X. B.; Li, Z. J. *Acta Crystallographica Section E-Structure Reports Online* **2007**, *63*, O3168.

233.   Constable, E. C.; Housecroft, C. E.; Mahmood, A.; Neuburger, M.; Schaffner, S. *Acta Crystallographica Section E-Structure Reports Online* **2006**, *62*, O3151.

234.   Temeriusz, A.; Anulewicz‑Ostrowska, R.; Paradowska, K.; Wawer, I. *Journal of Carbohydrate Chemistry* **2003**, *22*, 593.

235.   Bombicz, P.; Czugler, M.; Tellgren, R.; Kálmán, A. *Angewandte Chemie International Edition* **2003**, *42*, 1957.

236.   Ferrieres, V.; Gelin, M.; Boulch, R.; Toupet, L.; Plusquellec, D. *Carbohydrate Research* **1998**, *314*, 79.

237.   Podlasek, C. A.; Stripe, W. A.; Carmichael, I.; Shang, M.; Basu, B.; Serianni, A. S. *Journal of the American Chemical Society* **1996**, *118*, 1413.

238.   Koziol, A. E.; Lis, T. *Acta Crystallographica Section C:Crystal Structure Communications* **1991**, *47*, 2076.

239.   Czugler, M.; Kálmán, A.; Kovács, J.; Pintér, I. *Acta Crystallographica Section B: Structural Science* **1981**, *37*, 172.

240.   Poppleton, B. J. *Acta Crystallographica Section B: Structural Science* **1976**, *32*, 2702.

241.   Groth, P.; Hammer, H. *Acta Chemica Scandinavica* **1968**, *22*, 2059.

242.   Furberg, S.; Mostad, A. *Acta Chemica Scandinavica* **1962**, *16*, 1627.

APPENDIX

SUPPLEMENTARY INFORMATION CHAPTER 4

Methods

Generation of pseudorotational geometries for five-membered rings

The geometries for five-membered rings were defined using the Altona-Sundaralingam (AS) method[134] that employs two structure-related parameters in the model to characterize the pseudorotation: the pseudorotational phase angle, P, which distinguishes the particular ring conformation among the continuum of conformations; and the amplitude of the ring puckering, τ_m, which describes the magnitude of the ring puckering off the plane formed by three or four ring atoms. In the AS model, all of the ring torsion angles (θ) can be defined in terms of a combination of P and τ_m:

$$\theta_i = \tau_m \cos(P + i \times 144°) \tag{S4.1}$$

where i identifies the ring torsion angle as:
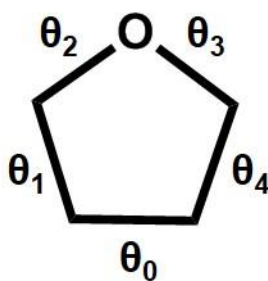
Based on Equation S4.1, P can be defined in terms of the five ring torsion angles (Equation S4.2):

$$tanP = \frac{(\theta_2 + \theta_4) - (\theta_1 + \theta_3)}{2\theta_0 \sin(36° + 72°)} \tag{S4.2}$$

The pseudorotational geometries were generated for ring conformations corresponding to P values from $0°$ to $360°$ in $1°$ steps with τ_m fixed at the ideal value of $39°$.[226-229] These geometries were then energy minimized, while restraining the ring torsion angles to the corresponding values calculated from Equation S4.1.

QM calculation

All of the QM calculations were performed with the Gaussian 09 software package.[187] Structural optimizations were performed at the HF/6-31G* level of theory; to be consistent with parameter development in GLYCAM06[31] single point energy calculations were computed at the B3LYP/6-31G* level.

MM calculation

The AMBER9 software package[230] was used to compute all MM energies associated with the parameter development; all other MM and MD calculations were performed with AMBER14.[182] As is standard with GLYCAM,[31] none of the 1-4 non-bonded interactions were scaled.[231] Torsion angles were restrained (5000 kcal/mol•rad^2) at their desired values, as required. A 12Å cut-off for non-bonded interactions was applied.

Molecular Dynamics Simulation

The initial structures of furanoses **1-4** (α and β) and **5β** (Figure 2) were obtained from QM energy minimization. All methyl furanosides were solvated with TIP3P water[44] using a 12 Å buffer in a cubic box, using the LEaP module in the AMBER14 software package.[182] The energy minimizations for these solvated furanoses were performed separately under nVT condition (500 steps steepest descent, followed by 24500 steps of conjugate-gradient minimization). Each system was, subsequently, heated to 300K over a period of 50 ps, followed by equilibration at

300K for a further 0.5 ns using nPT condition, with the Berendsen thermostat[36] for temperature control. All covalent bonds involving hydrogen atoms were constrained using SHAKE algorithm,[34] allowing a simulation time step of 2 fs throughout the simulations. After the equilibration, production simulations were carried out with the GPU implementation[183] of the PMEMD.MPI module and trajectory frames collected at every 1 ps from the total of 300 ns. None of the 1-4 non-bonded interactions were scaled[231] and a non-bonded cut-off of 8 Å was applied to van der Waals interactions, with long-range electrostatics treated with the particle mesh Ewald approximation.

Atomic charge calculation

The atomic partial charges of these molecules were derived by employing the restrained electrostatic potential (RESP) charge fitting methodology with a weak hyperbolic charge restraint weight of 0.0005.[32] The molecular electrostatic potentials of all the molecules employed for parameter development were computed at the HF/cc-pVTZ//HF/6-31G* level of theory.[31] To avoid any conformational bias arising from employing a single conformation for charge calculations, average charges were derived from all of the 360 ring conformations ($P = 0$ to 360 ° at 1 ° increments).

Ensemble-averaged partial atomic charges were computed for the intact furanosides from 100 snapshots for each of the 20 ring conformations (P values vary from 0 ° to 360 ° with 18 ° increments and τ_m fixed at 39 °) selected from 10 ns explicitly-solvated MD simulations at each P value, employing a torsion angle restraint of 300 kcal/mol•rad^2. During the MD simulations the torsion angles were restrained at conformations corresponding to the desired P and τ_m values calculated from Equation S4.1. For consistency with GLYCAM06[31], RESP fitting was performed with a restraint weight of 0.01, employing MEPs computed at the HF/6-31G* level.

Prior to charge calculations, the geometries were optimized at the HF/6-31G* level, while freezing all torsion angles at the values obtained from the MD simulation. In order to be compatible with GLYCAM06[31], the partial charges for the carbon and oxygen atoms in the OMe group were set to 0.196 and -0.458 a.u., respectively.

MD derived scalar ^3J-coupling constants

Population-weighted NMR scalar couplings were computed for each pseudorotational state P as:

$$<J> = \sum_{P=0}^{360} J(P)\rho(P) \qquad (S4.3)$$

where $J(P)$ is the coupling constants of the measured hydrogen atoms for a certain P value of the ring; and $\rho(P)$ is the population of that ring conformation in the simulation.

Structures for P values ranging from 0 to 360 in 6° increments were optimized at the HF/6-31G* level with the furanose ring was constrained at the desired P value and with the pucker amplitude (τ_m) fixed at 39°. Coupling constants were computed at the B3LYP/HIII-SU3 level[186] for each of the 60 conformational states. The populations ($\rho(P)$) of each P state were determined from unconstrained MD simulations of each furanose, by binning the observed ring conformations according to their P values (± 3°).

The coupling constants for H4-H5S and H4-H5R were similarly derived as the MD population-weighted average of each of the staggered rotamers (r) of the C4-C5 bond:

$$<J> = \sum_{i=1}^{3} J(r_i)\rho(r_i), \qquad (S4.4)$$

The QM-computed coupling constants for H4-H5S and H4-H5R for each furanose in **1-4** (α and β) and **5β** were generated for 7 representative structures from northern ($P = 0$°, 6°, 12°, 18°, 342°, 348° and 354°) and southern ($P = 162$°, 168°, 174°, 180°, 186°, 192° and 198°) conformations. All three C4-C5 rotamer conformations were considered for each P value.

Pseudorotational potential energy surface calculation

While calculating the pseudorotational potential energy surfaces, the ring puckering amplitude (τ_m) was changed from planar ($0°$) to a maximum of $45°$ in $3°$ increments, while the pseudorotational phase angle (P) was varied from 0 to $360°$ in $6°$ increments, giving rise to a total of 900 conformational states. Ring torsion angles for every combination of P and τ_m were restrained to the corresponding values calculated from Equation S4.1, in additions, when computing the gas phase pseudorotational potential energy surfaces, the torsion angles associated with the exocyclic groups in each furanose were restrained (CMe-O1-C1-C2 = C3-C2-O2-H2O = C2-C3-O3-H3O = C3-C4-C5-O5 = C4-C5-O5-H5O = $180°$). When computing the solution phase pseudorotational potential energy surfaces, a 10 ns simulation for each of the 900 combinations of P and τ_m was performed with restraints only on the ring torsion angles. The MM energies were subsequently computed without energy minimization and averaged for 10000 frames extracted from the MD simulations. All energies are presented relative to the lowest energy for each furanoside.
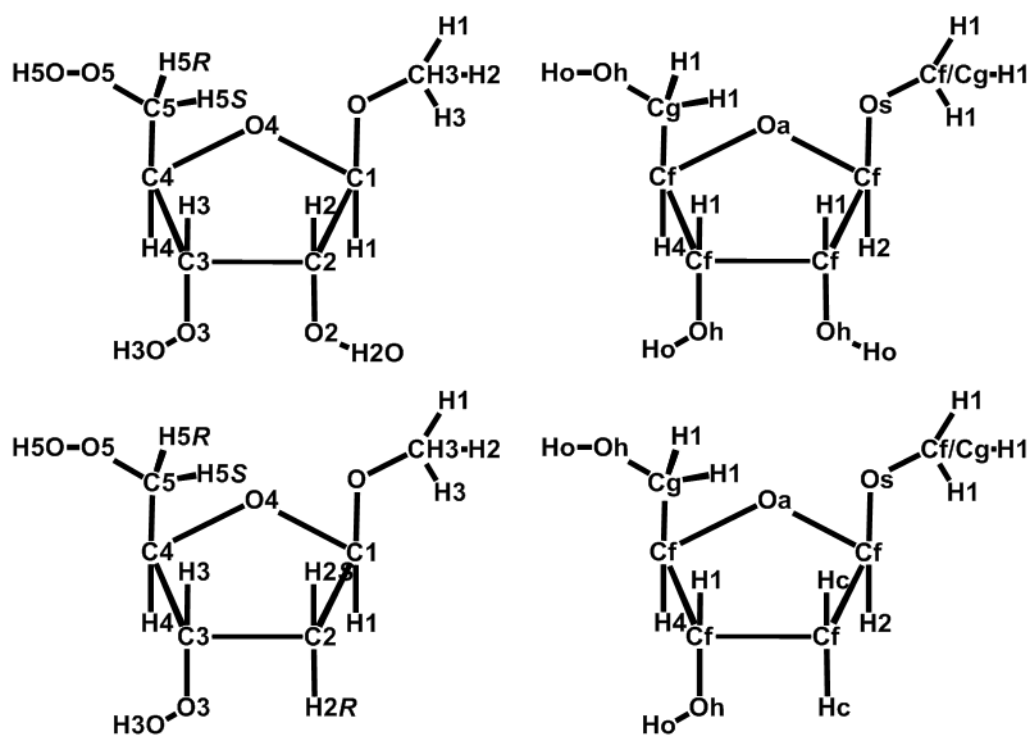
Figure S4.1 Atom names (a) and atom types (b) of **3β**, and atom names (c) and atom types (d) of

5β.
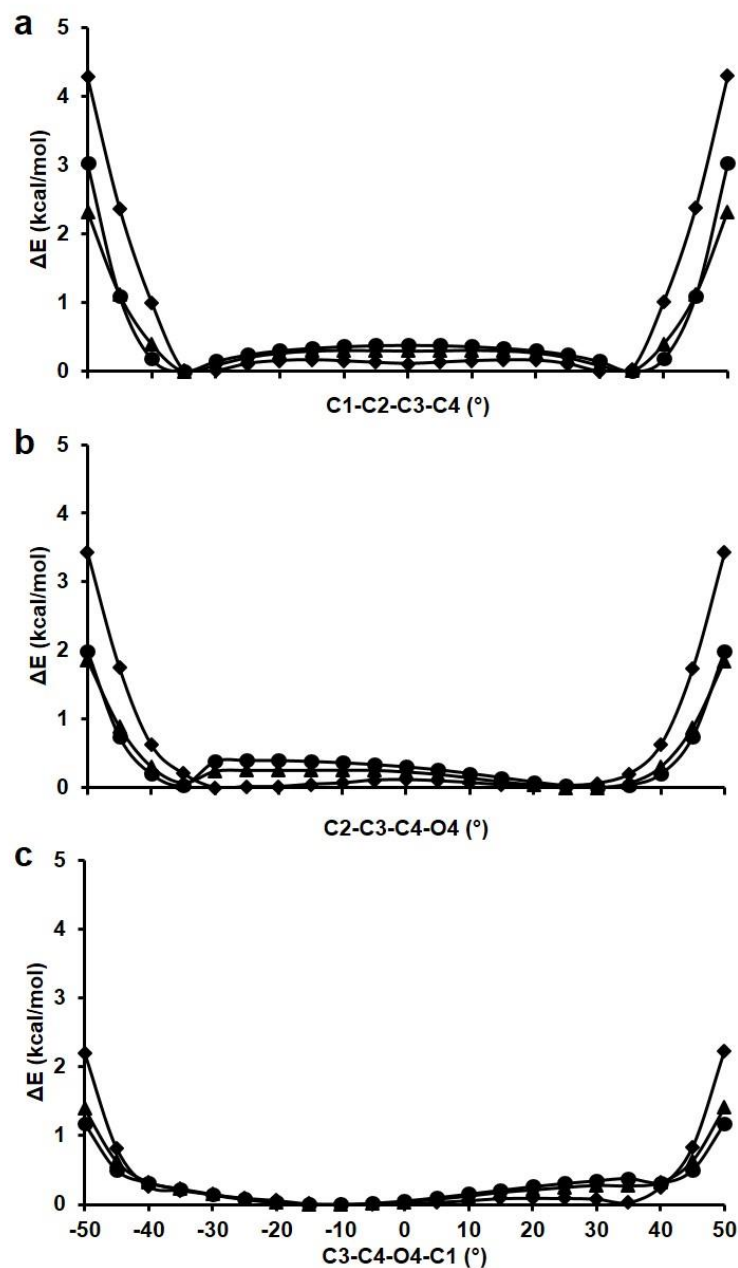
Figure S4.2 Energy curves for various torsion angles within THF: C1-C2-C3-C4 (a), C2-C3-C4-O4 (b) and C3-C4-O4-C1 (c). B3LYP/6-31G* (●), the new parameters (▲), and GLYCAM06 (♦). The energy errors between QM and MM with the new parameters are: $<|Error|>_{all} = 0.1$ and $<|Error|>_{low} = 0.1$ kcal/mol; the energy errors between QM and MM with GLYCAM06 are: $<|Error|>_{all} = 0.4$ and $<|Error|>_{low} = 0.2$ kcal/mol.

Figure S4.3 Pseudorotational energy curves for **7** (a), **8** (b), and **9** (c). Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters. Atomic numbering corresponds to that for furanoses, rather than for THF analogs.
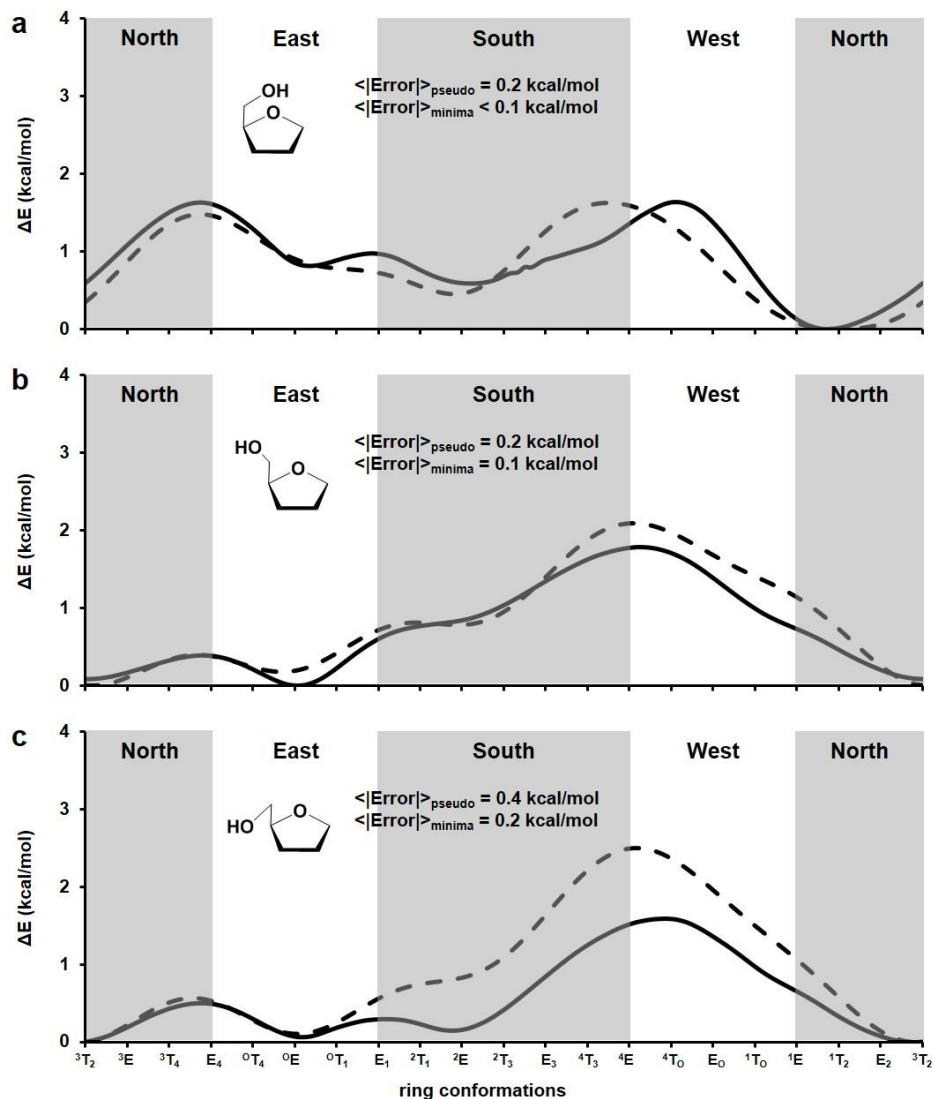
Figure S4.4 Pseudorotational energy curves for **10**, with C3-C4-C5-O5 torsion angle of the hydroxymethyl group held in the staggered orientations (60 ° (a), 180 ° (b), and 300 ° (c), also referred to as the *gauche-gauche* (*gg*), *gauche-trans* (*gt*), and *trans-gauche* (*tg*) states, where the first letter refers to the relative orientation of O5 to O4, and the second refers to O5 to C3). Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.
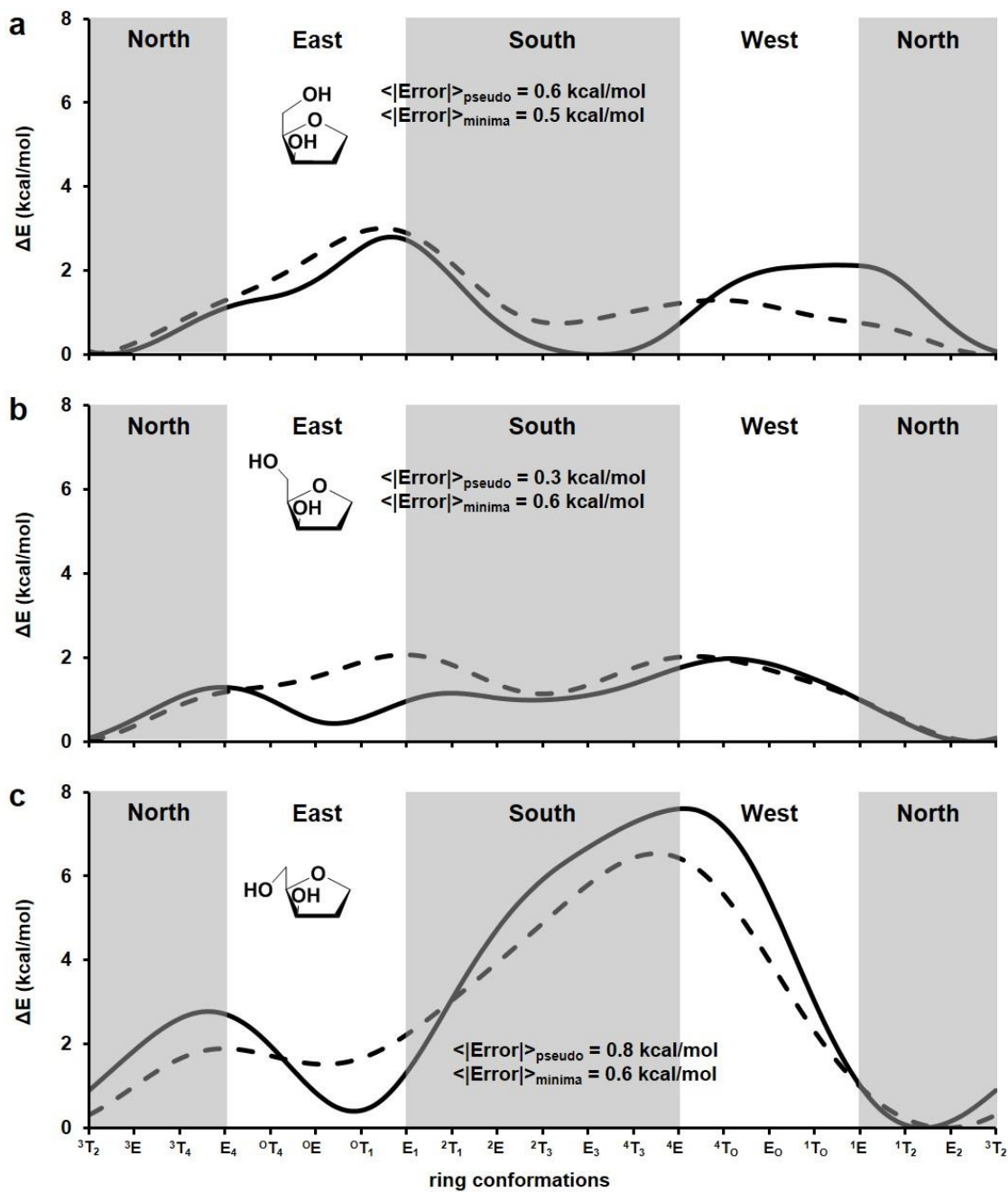
Figure S4.5 Pseudorotational energy curves for **11**, with C3-C4-C5-O5 torsion angle of the hydroxymethyl group restrained at *gg* (a), *gt* (b), and *tg* (c) conformations. Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.

106
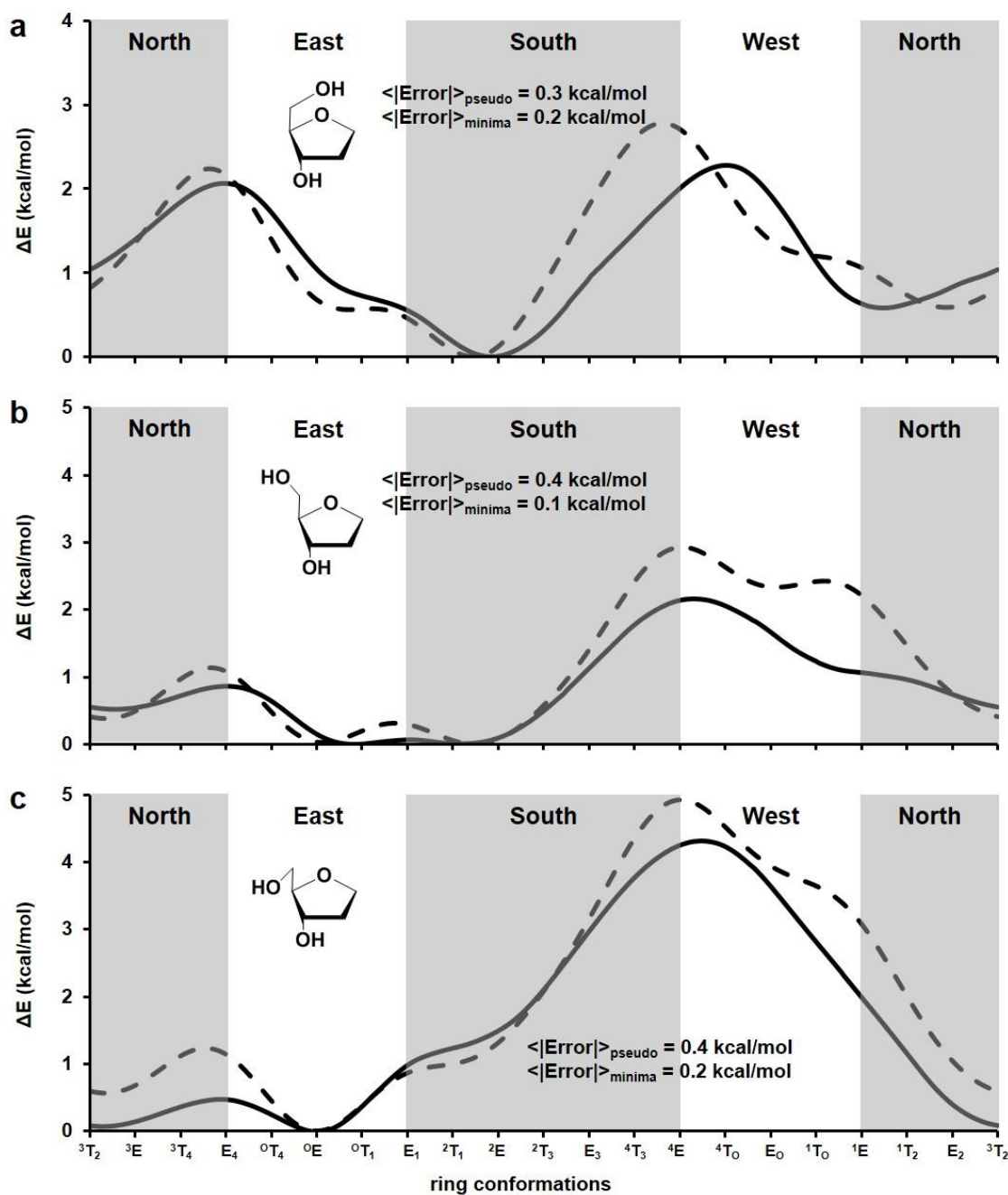
Figure S4.6 Pseudorotational energy curves for **12**, with C3-C4-C5-O5 torsion angle of the hydroxymethyl group restrained at *gg* (a), *gt* (b), and *tg* (c) conformations. Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.
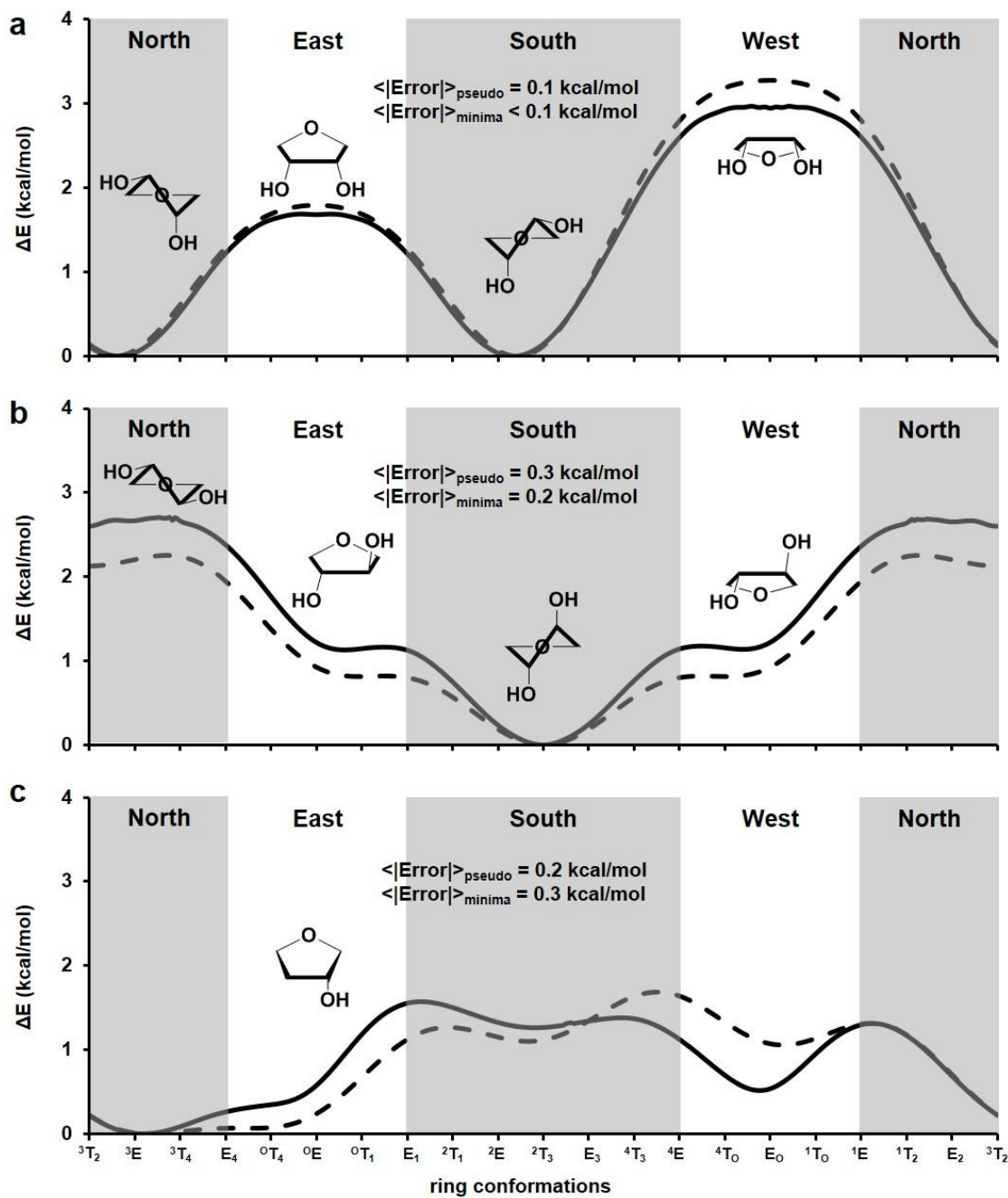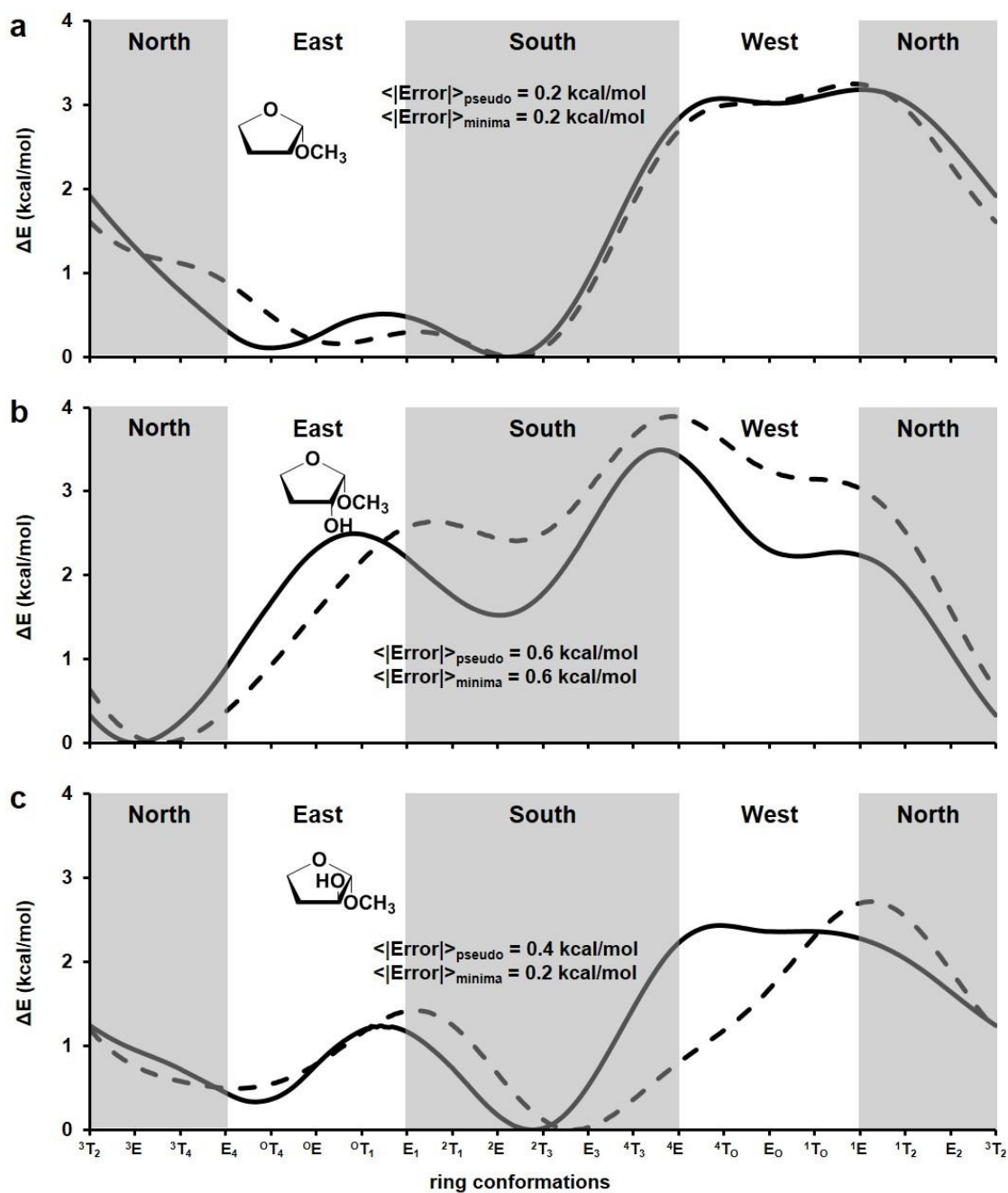
Figure S4.7 Pseudorotational energy curves for **13** (a), **14** (b) and **15** (c). Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.
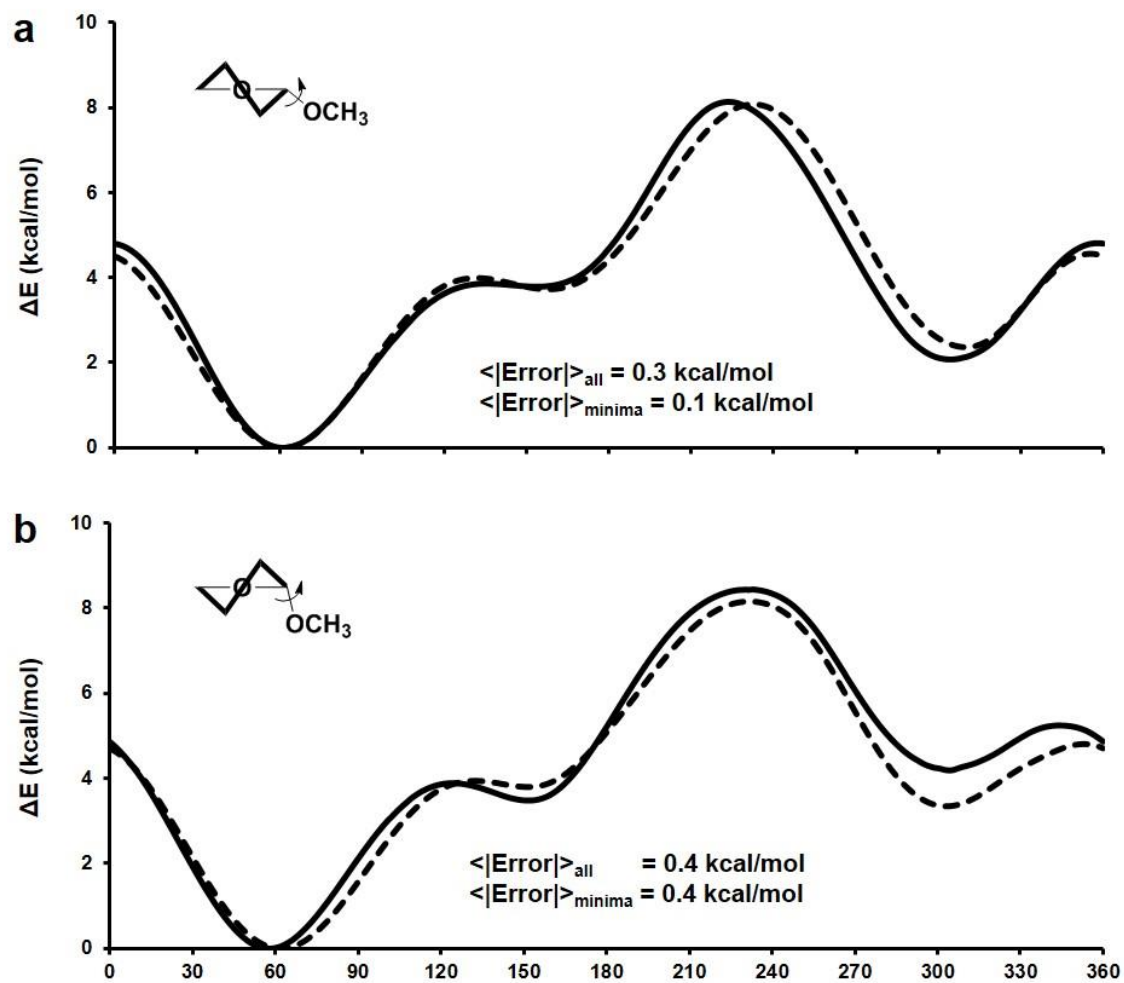
Figure S4.8 Rotational energy curves of *exo*-anomeric linkage for **13**, with the ring restrained at northern (a) and southern (b) conformations (*P* = 0 and 180 °, respectively). Solid lines: energies computed at the B3LYP/6-31G* level; dashed lines: energies computed with new parameters.
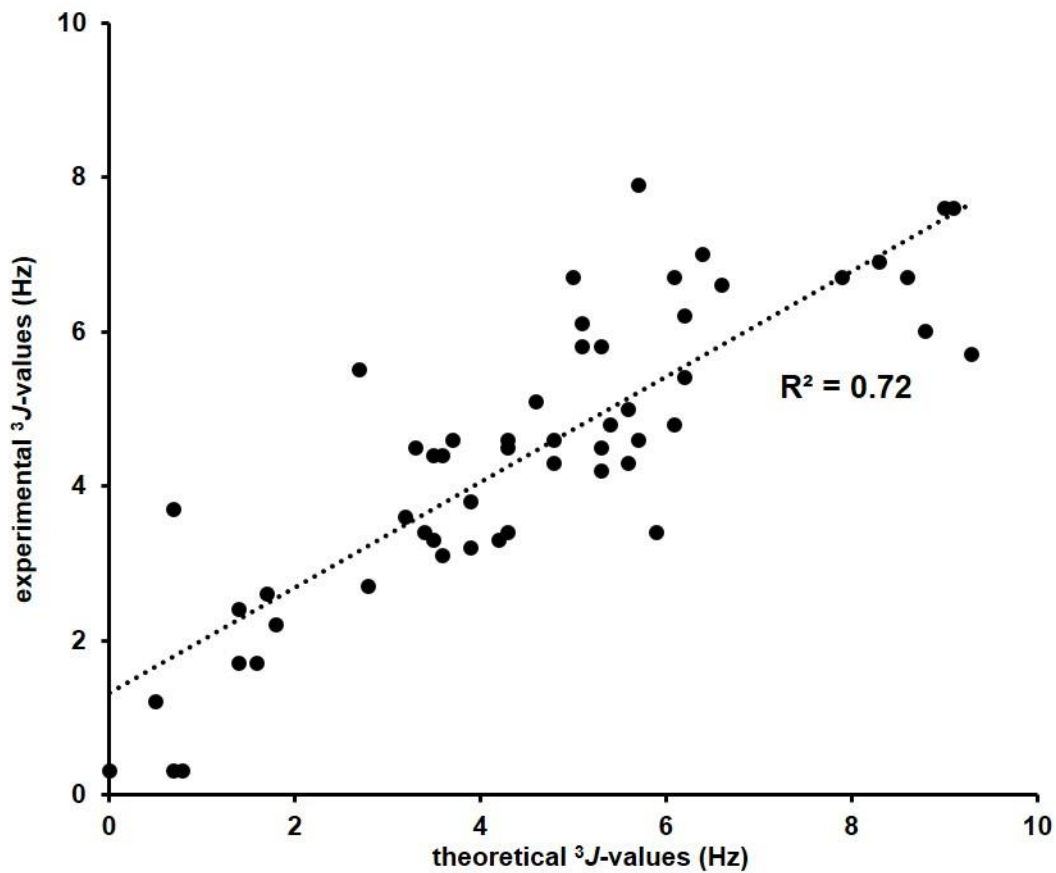
Figure S4.9 The overall correlation between the theoretical and experimental ^3J-values for **1-4** (α and β) and **5β**.

Table S4.1 Ensemble-averaged atomic charge sets for **1-4** (α and β) and **5β**.
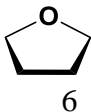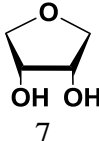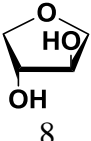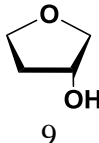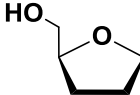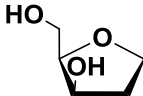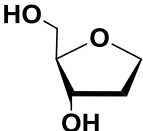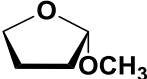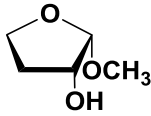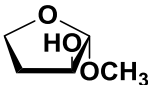
Atom	1α	1β	2α	2β	3α	3β	4α	4β	Atom	5β
					methyl D-furanosides					
CH3	0.264	0.264	0.264	0.264	0.264	0.264	0.264	0.264	CMe	0.264
O	-0.458	-0.458	-0.458	-0.458	-0.458	-0.458	-0.458	-0.458	OMe	-0.458
C1	0.381	0.392	0.406	0.402	0.393	0.417	0.381	0.391	C1	0.381
C2	0.308	0.299	0.245	0.245	0.269	0.256	0.297	0.327	C2	0.035
O2	-0.707	-0.693	-0.656	-0.647	-0.659	-0.663	-0.692	-0.711	C3	0.248
H2O	0.427	0.422	0.417	0.417	0.417	0.419	0.420	0.427	O3	-0.693
C3	0.293	0.280	0.238	0.240	0.223	0.241	0.286	0.260	H3O	0.420
O3	-0.718	-0.726	-0.682	-0.675	-0.679	-0.681	-0.725	-0.708	C4	0.313
H3O	0.427	0.434	0.428	0.429	0.427	0.423	0.434	0.429	O4	-0.485
C4	0.281	0.276	0.243	0.195	0.327	0.289	0.256	0.216	C5	0.231
O4	-0.476	-0.480	-0.467	-0.469	-0.479	-0.489	-0.463	-0.462	O5	-0.669
C5	0.235	0.247	0.289	0.329	0.218	0.239	0.262	0.291	H5O	0.413
O5	-0.673	-0.674	-0.675	-0.683	-0.692	-0.669	-0.674	-0.674		
H5O	0.416	0.417	0.408	0.411	0.429	0.412	0.412	0.408		

Table S4.2 Average bond lengths[a] and bond angles[b] of ring in **1-4** (α and β) and **5β** from MD simulations (300 ns each) and crystallographic data[c].

		1α	**1β**	**2α**	**2β**	**3α**	**3β**	**4α**	**4β**	**5β**	average MD	average X-ray[226,227,232-242] (n=16)
Bond	C1-C2	1.54	1.55	1.54	1.55	1.55	1.54	1.55	1.55	1.53	1.54	1.53
	C2-C3	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.52	1.54	1.53
	C3-C4	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.53
	C4-O4	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.44
	O4-C1	1.47	1.47	1.47	1.47	1.47	1.47	1.47	1.47	1.46	1.47	1.42
Angle	C1-C2-C3	104	102	102	103	102	100	102	103	103	102	102
	C2-C3-C4	104	103	102	101	103	103	102	103	104	103	102
	C3-C4-O4	107	107	106	107	107	107	106	107	107	107	105
	C4-O4-C1	107	107	107	108	107	107	106	107	107	107	109
	O4-C1-C2	108	107	107	106	107	107	107	108	107	107	105

[a]in Å. All the average bond lengths have standard derivations of approximately 0.03 Å. [b]in degree. All the average bond angles have standard derivations of approximately 3 °. [c]All average bond lengths and bond angles from crystallographic data have standard derivations of approximately 0.01 Å and 1 °, respectively.

Table S4.3 Torsion angle parameters and corresponding molecules employed for their developments and validations, with specific restraints for exocyclic groups.

Molecule	Torsion terms	restraints for exocyclic groups[a]
6	Cf-Cf-Cf-Cf Cf-Cf-Cf-Of Cf-Cf-Of-Cf	
7 **8**	Oh-Cf-Cf-Oh Cf-Cf-Cf-Oh Of-Cf-Cf-Oh	C3-C2-O2-H2O = 180 ° C2-C3-O3-H3O = 180 °
9	Hc-Cf-Cf-Oh	C3-C2-O2-H2O = 180 °
10	Cf-Of-Cf-Cg Cf-Cf-Cf-Cg	C3-C4-C5-O5 = 60 °, 180 °, 300 ° C4-C5-O5-H5O = 180 °
11 **12**	Validation	C3-C4-C5-O5 = 60 °, 180 °, 300 ° C4-C5-O5-H5O = 180 ° C2-C3-O3-H3O = 180 °
13	*Endo*-anomeric terms Cf-Cf-Cf-Os Cf-Of-Cf-Os *Exo*-anomeric terms Of-Cf-Os-Cf/Cg Cf-Cf-Os-Cf/Cg	O4-C1-O-CH3 = 60 ° Ring of THF is restrained to $^3T_2$ and $^2T_3$ conformation
14 **15**	Os-Cf-Cf-Oh H2-Cf-Cf-Oh H2-Cf-Cf-H1	O4-C1-O-CH3 = 60 ° C3-C2-O2-H2O = 180 °
16	Validation	Ring is restrained to $^3T_2$ or $^2T_3$ conformation
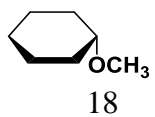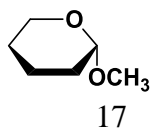
17 18		Validation	Ring maintaines at $^1C_4$ or $^4C_1$ conformation

[a]Numbering according to methyl furanoside rather than THF.

114

Table S4.4 Furanose-specific force field parameters.

Bond	$K_r{}^b$	$r_{eq}{}^c$	Bond	K_r	r_{eq}
Cf-Cf[a]	310.0	1.520	Cf-Hc[a]	340.0	1.090
Cf-Cg[a]	310.0	1.520	Cf-Of[a]	285.0	1.460
Cf-H1[a]	340.0	1.090	Cf-Os[a]	285.0	1.460
Cf-H2[a]	340.0	1.090	Cf-Oh[a]	320.0	1.430

Angle	$K_\theta{}^d$	$\theta_{eq}{}^e$	Angle	K_θ	θ_{eq}
Cf-Cf-Cf	47.0	106.7	Cf-Os-Cf/g[a]	50.0	111.6
Cf-Cf-Cg[a]	45.0	113.5	Cg-Cf-H1[a]	45.0	111.0
Cf-Cf-H1[a]	45.0	111.0	Cg-Cf-Of[a]	70.0	108.5
Cf-Cf-H2[a]	45.0	111.0	H1-Cf-H1[a]	45.0	109.5
Cf-Cf-Hc[a]	45.0	112.6	H1-Cf-Of[a]	60.0	110.0
Cf-Cf-Of	61.0	106.7	H1-Cf-Oh[a]	60.0	110.0
Cf-Cf-Oh[a]	70.0	107.5	H1-Cf-Os[a]	60.0	110.0
Cf-Cf-Os[a]	70.0	108.5	H2-Cf-Of[a]	60.0	110.0
Cf-Cg-H1[a]	45.0	111.0	H2-Cf-Os[a]	60.0	110.0
Cf-Cg-Oh[a]	70.0	107.5	Hc-Cf-Hc[a]	40.0	109.5
Cf-Of-Cf	45.0	106.7	Of-Cf-Os[a]	100.0	112.0
Cf-Oh-Ho[a]	55.0	109.5	Of-Cg-Oh[a]	100.0	112.0

Torsion	$V_n{}^f/2$	γ^g, n^h	Torsion	$V_n/2$	γ, n
Cf-Cf-Cf-Cf	-0.90	0.0, 3	Os-Cf-Cf-Oh	0.40	0.0, 3
Cf-Cf-Cf-Of	0.90	0.0, 3		-1.60	0.0, 1
	0.30	0.0, 2	H1-Cf-Cf-Os	0.00	0.0, 3
Cf-Cf-Of-Cf	-0.20	0.0, 3	H2-Cf-Cf-Oh	-0.40	0.0, 3
	0.30	0.0, 1		-0.20	0.0, 2
Cf-Cf-Cf-H1	0.00	0.0, 3		0.20	0.0, 1
Cf-Of-Cf-H1	0.00	0.0, 3	H1-Cf-Cf-H2	-0.60	0.0, 3
Hc-Cf-Cf-H1	0.00	0.0, 3		-0.20	0.0, 2
Hc-Cf-Cf-Of	0.00	0.0, 3		0.20	0.0, 1
Hc-Cf-Cf-Hc	0.00	0.0, 3	Cf-Of-Cf-Cg	-0.40	0.0, 3
Cf-Cf-Cf-Hc	0.00	0.0, 3		-0.20	0.0, 2
Cf-Cf-Cf-Oh	-0.50	0.0, 2		0.20	0.0, 1
	-0.50	0.0, 1	Cf-Cf-Cf-Cg	-0.60	0.0, 3
Of-Cf-Cf-Oh	0.10	0.0, 3		-0.80	0.0, 2

Torsion	Value	Phase, n	Torsion	Value	Phase, n
	0.20	0.0, 2		-0.80	0.0, 1
	-1.05	0.0, 1	Of-Cf-Cg-Oh[a]	0.25	0.0, 2
H1-Cf-Cf-Oh	0.00	0.0, 3		-1.10	0.0, 1
H1-Cf-Cf-H1	0.00	0.0, 3	Cf-Cf-Cg-Oh[a]	0.10	0.0, 3
H1-Cf-Cf-Of	0.00	0.0, 3	H1-Cf-Cg-Oh[a]	0.05	0.0, 3
Cf-Cf-Oh-Ho[a]	0.18	0.0, 3	Of-Cf-Cg-H1[a]	0.05	0.0, 3
H1-Cf-Oh-Ho[a]	0.18	0.0, 3	Cf-Cf-Cg-H1[a]	0.15	0.0, 3
Hc-Cf-Cf-Oh	-0.80	0.0, 2	H1-Cf-Cg-H1[a]	0.17	0.0, 3
Oh-Cf-Cf-Oh	-0.10	0.0, 3	Oh-Cf-Cf-Cg	0.00	0.0, 3
	-0.10	0.0, 2	H1-Cf-Cf-Cg	0.00	0.0, 3
	0.10	0.0, 1	Cf-Cg-Oh-Ho[a]	0.18	0.0, 3
Cf-Of-Cf-Os	-0.40	0.0, 3	H1-Cg-Oh-Ho[a]	0.18	0.0, 3
	1.20	0.0, 2	Cg-Cf-Cf-Hc[a]	0.00	0.0, 3
	0.40	0.0, 1	Of-Cf-Os-Cf/g	0.80	0.0, 2
Cf-Cf-Cf-Os	0.20	0.0, 3		-0.40	0.0, 1
	-0.80	0.0, 1	Cf-Cf-Os-Cf/g	-0.20	0.0, 2
Cf-Of-Cf-H2	0.00	0.0, 3		-0.60	0.0, 1
Cf-Cf-Cf-H2	0.00	0.0, 3	H2-Cf-Os-Cf/g	0.00	0.0, 3
Cf-Os-Cf-Cf	-0.20	0.0, 2	Cf-Os-Cf/g-H1[a]	0.27	0.0, 3
	-0.60	0.0, 1	Cf-Cf-Cg-Os[a]	0.10	0.0, 3
Cf-Os-Cf/g-H1[a]	0.27	0.0, 3	Of-Cf-Cg-Os[a]	0.25	0.0, 2
Cf-Os-Cf-Of	0.80	0.0, 2		-1.10	0.0, 1
	-0.40	0.0, 1	H1-Cf-Cg-Os[a]	0.05	0.0, 3
Cf-Os-Cf-H2	0.00	0.0, 3	Cf-Cg-Os-Cf/g[a]	0.16	0.0, 3
Hc-Cf-Cf-Os	0.00	0.0, 3	H1-Cg-Os-Cf/g[a]	0.27	0.0, 3
Hc-Cf-Cf-H2	0.00	0.0, 3	Cg-Os-Cf/g-H1[a]	0.27	0.0, 3

[a]Valence terms incorporated from the GLYCAM06 parameter sets. [b]Bond stretching force constant (kcal/mol•Å^2). [c]Bond length (Å). [d]Angle-bending force constant (kcal/mol•rad^2). [e]Angle (degrees). [f]Indicates relative barrier to rotation (kcal/mol). [g]Phase factor (degrees). [h]Periodicity or n-fold term in Fourier series expansion.
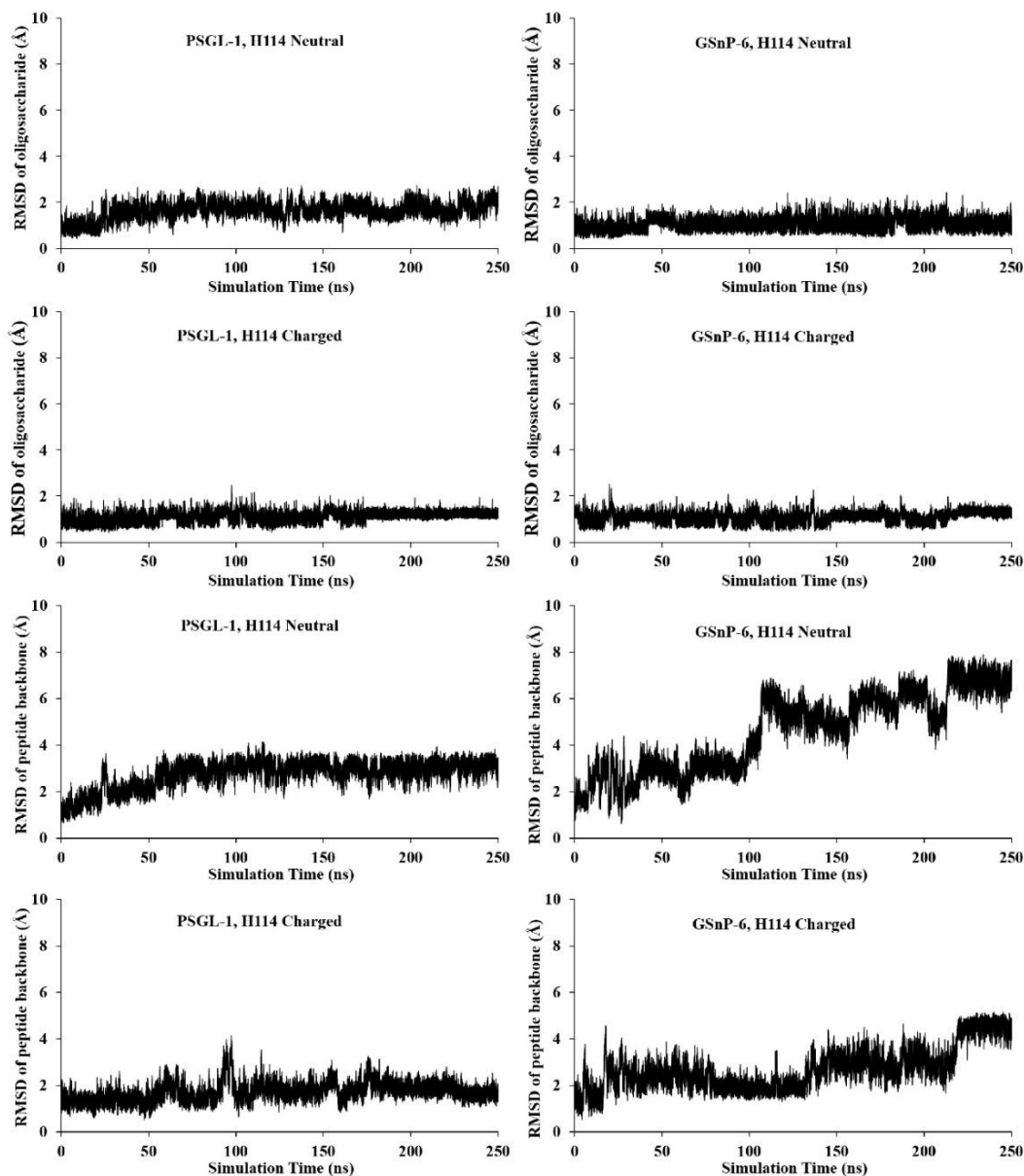
Figure S6.1 Root mean square deviations (RMSD) of the oligosaccharide and peptide components of the PSGL-1 (**left**) and GSnP-6 (**right**) ligands relative to their initial position in the P-selectin complex, over the course of the MD simulations, as a function of the protonation state of histidine H114.
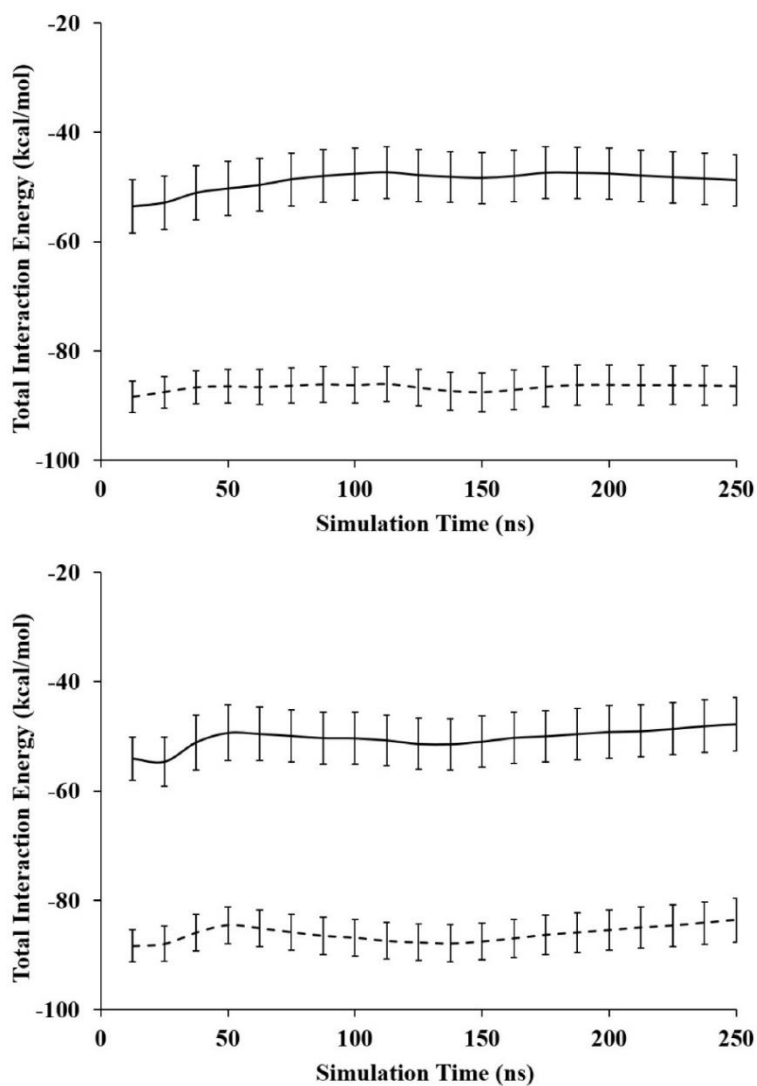
Figure S6.2 Computed interaction energies (running average) for the PSGL-1 (**upper**) and GSnP-6 (**lower**) ligands with P-selectin (H114 charged), with the GB_1^{OBC} solvation model with internal dielectric constants of 1.0 (**solid line**) and 4.0 (**dashed line**).

Table S6.1 Parameters employed for the linkage between the SO_3^- group and the amino acid side chains in YS and YCS.

Bond Angles	k_θ	θ_{eq}		Source
S -CT-C	60.0	109.50		OS-CT-C[a]
CT-C -CA	63.0	117.00		CT-C –CT[a]
O2-S -CT	104.0	106.87		O2-S –Os[b]
Os-C -CA	59.5	119.00		Os-Ck-Ck[b]
S -Os-C	50.0	118.88		S -Os-Cg[b]
Torsion Angles	$V_1/2$	$V_2/2$	$V_3/2$	
C -Os-S -O2	0.00	0.00	0.18	Cg-Os-S -O2[b]
CA-C -Os-S	-1.20	0.00	0.00	Cg-Cg-Os-S[b]

[a]From ff99sb. [b]From GLYCAM06 (version h).

Table S6.2 Key intermolecular hydrogen bond distances and occupancies between PSGL-1 and P-selectin.

Ligand Residue	Atom	P-selectin Atom	x-ray	PSGL-1 Ligand MD H114 neutral	PSGL-1 Ligand MD H114 charged	GSnP-6 MD H114 neutral	GSnP-6 MD H114 charged
Fuc	O2	E88-Oε^a	2.6^b	$2.6 \pm 0.1 (100)^c$	$2.7 \pm 0.1 (100)$	$2.7 \pm 0.1 (100)$	$2.7 \pm 0.1 (100)$
	O3	N105-Nδ2	2.9	$3.2 \pm 0.1 (29)$	--	$3.2 \pm 0.1 (91)$	$3.3 \pm 0.1 (70)$
	O3	E107-Oε^a	2.7	$2.7 \pm 0.1 (100)$	$2.8 \pm 0.2 (98)$	$2.7 \pm 0.1 (100)$	$2.6 \pm 0.1 (82)$
	O4	E80-Oε	2.6	$2.8 \pm 0.1 (100)^d$	$2.9 \pm 0.2 (100)^d$	$2.8 \pm 0.1 (100)^d$	$2.8 \pm 0.2 (100)^d$
	O4	N82-Nδ2	3.0	$2.8 \pm 0.1 (98)$	$2.8 \pm 0.1 (94)$	$3.0 \pm 0.1 (15)$	$2.9 \pm 0.1 (35)$
Core-2 Gal	O4	Y94-Oη	2.7	$2.8 \pm 0.1 (90)$	$2.8 \pm 0.1 (82)$	$2.8 \pm 0.1 (82)$	$2.9 \pm 0.2 (72)$
	O6	E92-Oε	2.5	$2.9 \pm 0.2 (100)^d$	$2.9 \pm 0.2 (100)^d$	$3.0 \pm 0.2 (100)^d$	$2.9 \pm 0.2 (100)^d$
Neu5Ac	O1	Y48-Oη	2.5	$2.8 \pm 0.3 (100)^d$	$2.8 \pm 0.3 (100)^d$	$2.9 \pm 0.3 (100)^d$	$2.8 \pm 0.3 (100)^d$
	O4	S99-Oγ	3.1	$2.9 \pm 0.2 (87)$	$2.8 \pm 0.2 (63)$	$2.8 \pm 0.2 (97)$	$2.9 \pm 0.2 (90)$
SO$_3^-$ 605	O^e	K8-Nζ	--	$2.9 \pm 0.2 (20)$	$2.9 \pm 0.2 (19)$	$2.9 \pm 0.2 (23)$	$3.0 \pm 0.2 (26)$
	O^e	K112-Nζ	--	$2.9 \pm 0.2 (25)$	$2.9 \pm 0.2 (15)$	--	$2.9 \pm 0.2 (9)$
SO$_3^-$ 607	O^e	S46-Oγ	3.3	$2.7 \pm 0.2 (41)$	$2.7 \pm 0.2 (47)$	$2.9 \pm 0.2 (6)$	$2.8 \pm 0.2 (50)$
	O^e	S47-Oγ	3.0	$2.8 \pm 0.2 (20)$	$2.9 \pm 0.3 (100)^d$	$2.8 \pm 0.2 (36)$	$2.9 \pm 0.3 (100)^d$
	O^e	H114-Nε2	2.7	$3.0 \pm 0.2 (14)$	$2.8 \pm 0.1 (99)$	$3.0 \pm 0.2 (9)$	$2.8 \pm 0.1 (100)$
SO$_3^-$ 610	O^e	R85-Nη1	2.7	$3.0 \pm 0.2 (100)^d$	$3.0 \pm 0.3 (24)$	$3.0 \pm 0.2 (22)$	$3.0 \pm 0.2 (15)$
	O^e	R85-Nη2	3.7	$3.0 \pm 0.2 (100)^d$	$2.9 \pm 0.2 (48)$	$3.0 \pm 0.2 (22)$	$2.9 \pm 0.2 (20)$

[a]Only one of the oxygen atom in carboxylate groups interacts with Fuc. [b]In Å. [c]Percentage (%) based on a distance between non-hydrogen atoms of less than 3.5Å. When multiple hydrogen bonds are formed between two heavy atoms through different hydrogens, the occupancy of the interaction listed is the sum of all the individual hydrogen bonds and the distance is the average of all the individual hydrogen bonds. [d]The occupancy of the interactions between two heavy atoms, calculated as the sum of all the individual hydrogen bonds through different hydrogens, is greater than 100%. [e]Oxygen atom in SO3- group of tyrosine sulfate or tyrosine sulfonate.

Table S6.3 MM/GBSA energy[a] decompositions with different GB models and internal dielectric values for the interactions of P-selectin with the PSGL-1 ligand.

GB^{HCT}(igb = 1)						
ε_{int}	1.0	1.5	2.0	3.0	4.0	5.0
Sulfate Groups						
SO$_3^-$ 605[b]	0.3 $\pm$ 1.2	-0.9 $\pm$ 0.9	-1.5 $\pm$ 0.8	-2.1 $\pm$ 0.9	-2.4 $\pm$ 1.0	-2.6 $\pm$ 1.0
SO$_3^-$ 607[b]	-4.2 $\pm$ 1.4	-3.7 $\pm$ 1.0	-3.5 $\pm$ 0.9	-3.3 $\pm$ 0.8	-3.2 $\pm$ 0.9	-3.2 $\pm$ 0.9
SO$_3^-$ 610[b]	-1.8 $\pm$ 1.9	-1.4 $\pm$ 1.3	-1.2 $\pm$ 1.0	-1.0 $\pm$ 0.7	-0.9 $\pm$ 0.6	-0.8 $\pm$ 0.5
Subtotal	-5.7 $\pm$ 2.6	-6.0 $\pm$ 1.9	-6.2 $\pm$ 1.6	-6.4 $\pm$ 1.4	-6.5 $\pm$ 1.5	-6.6 $\pm$ 1.4
Monosaccharides						
Neu5Ac	0.6 $\pm$ 1.0	-0.9 $\pm$ 0.7	-1.6 $\pm$ 0.7	-2.3 $\pm$ 0.8	-2.7 $\pm$ 0.8	-2.9 $\pm$ 0.8
Core-2 Gal	-1.7 $\pm$ 0.8	-2.5 $\pm$ 0.7	-2.9 $\pm$ 0.7	-3.3 $\pm$ 0.7	-3.5 $\pm$ 0.7	-3.6 $\pm$ 0.8
GlcNAc	-1.6 $\pm$ 1.2	-2.4 $\pm$ 0.8	-2.8 $\pm$ 0.7	-3.2 $\pm$ 0.6	-3.4 $\pm$ 0.6	-3.5 $\pm$ 0.6
Fuc	7.2 $\pm$ 1.8	3.2 $\pm$ 1.4	1.3 $\pm$ 1.3	-0.7 $\pm$ 1.3	-1.7 $\pm$ 1.3	-2.3 $\pm$ 1.3
GalNAc	-0.5 $\pm$ 0.4	-0.5 $\pm$ 0.4	-0.6 $\pm$ 0.3	-0.6 $\pm$ 0.3	-0.6 $\pm$ 0.3	-0.6 $\pm$ 0.3
Gal	0.3 $\pm$ 0.1	0.2 $\pm$ 0.1	0.1 $\pm$ 0.1	0.1 $\pm$ 0.1	0.1 $\pm$ 0.1	0.1 $\pm$ 0.1
Subtotal	4.3 $\pm$ 2.5	-2.9 $\pm$ 1.9	-6.5 $\pm$ 1.8	-10.0 $\pm$ 1.8	-11.8 $\pm$ 1.8	-12.8 $\pm$ 1.9
Amino Acids						
E604	0.7 $\pm$ 0.4	0.4 $\pm$ 0.2	0.3 $\pm$ 0.2	0.2 $\pm$ 0.2	0.1 $\pm$ 0.2	0.1 $\pm$ 0.2
Y605[c]	-1.3 $\pm$ 0.9	-1.5 $\pm$ 0.8	-1.5 $\pm$ 0.7	-1.6 $\pm$ 0.7	-1.6 $\pm$ 0.7	-1.6 $\pm$ 0.7
E606	-0.8 $\pm$ 1.0	-1.0 $\pm$ 0.7	-1.1 $\pm$ 0.6	-1.2 $\pm$ 0.5	-1.2 $\pm$ 0.4	-1.3 $\pm$ 0.4
Y607[c]	-4.8 $\pm$ 1.0	-5.3 $\pm$ 0.8	-5.5 $\pm$ 0.8	-5.8 $\pm$ 0.8	-5.9 $\pm$ 0.7	-6.0 $\pm$ 0.7
L608	-2.1 $\pm$ 1.2	-2.2 $\pm$ 1.2	-2.2 $\pm$ 1.2	-2.2 $\pm$ 1.2	-2.3 $\pm$ 1.2	-2.3 $\pm$ 1.2
D609	-0.9 $\pm$ 0.6	-1.3 $\pm$ 0.6	-1.5 $\pm$ 0.6	-1.7 $\pm$ 0.7	-1.8 $\pm$ 0.7	-1.9 $\pm$ 0.7
Y610[c]	-0.8 $\pm$ 0.5	-0.8 $\pm$ 0.5	-0.8 $\pm$ 0.4	-0.8 $\pm$ 0.4	-0.8 $\pm$ 0.4	-0.8 $\pm$ 0.4
D611	-0.4 $\pm$ 0.1	-0.5 $\pm$ 0.0	-0.6 $\pm$ 0.1	-0.7 $\pm$ 0.1	-0.7 $\pm$ 0.1	-0.7 $\pm$ 0.1
F612	-0.7 $\pm$ 0.5	-0.9 $\pm$ 0.6	-1.0 $\pm$ 0.6	-1.2 $\pm$ 0.6	-1.2 $\pm$ 0.6	-1.3 $\pm$ 0.7
L613	-3.7 $\pm$ 1.1	-3.9 $\pm$ 1.1	-4.0 $\pm$ 1.1	-4.1 $\pm$ 1.1	-4.1 $\pm$ 1.2	-4.1 $\pm$ 1.2
P614	-3.7 $\pm$ 1.9	-2.9 $\pm$ 1.5	-2.5 $\pm$ 1.3	-2.1 $\pm$ 1.2	-1.9 $\pm$ 1.2	-1.8 $\pm$ 1.2
E615	-0.2 $\pm$ 0.3	-0.6 $\pm$ 0.2	-0.7 $\pm$ 0.2	-0.9 $\pm$ 0.2	-1.0 $\pm$ 0.2	-1.0 $\pm$ 0.2
T616[d]	-1.3 $\pm$ 0.3	-1.3 $\pm$ 0.3	-1.3 $\pm$ 0.3	-1.4 $\pm$ 0.3	-1.4 $\pm$ 0.3	-1.4 $\pm$ 0.3
E617	-0.1 $\pm$ 0.3	-0.3 $\pm$ 0.2	-0.4 $\pm$ 0.2	-0.5 $\pm$ 0.3	-0.6 $\pm$ 0.3	-0.6 $\pm$ 0.3
P618	-0.2 $\pm$ 0.2	-0.3 $\pm$ 0.1	-0.3 $\pm$ 0.1	-0.4 $\pm$ 0.1	-0.4 $\pm$ 0.1	-0.4 $\pm$ 0.1
Subtotal	-20.3 $\pm$ 3.2	-22.4 $\pm$ 2.8	-23.1 $\pm$ 2.6	-24.4 $\pm$ 2.6	-24.8 $\pm$ 2.6	-25.1 $\pm$ 2.6
Total Interaction Energy	-21.7 $\pm$ 5.9	-31.3 $\pm$ 4.8	-35.8 $\pm$ 4.4	-40.8 $\pm$ 3.5	-43.1 $\pm$ 4.4	-44.5 $\pm$ 3.5

ε_{int}	1.0	1.5	2.0	3.0	4.0	5.0
GB$_1^{OBC}$ (igb = 2)						
Sulfate Groups						
SO$_3^-$ 605[b]	2.6 ±1.6	0.6 ±0.9	-0.4 ±0.6	-1.4 ±0.7	-1.9 ±0.8	-2.2 ±0.9
SO$_3^-$ 607[b]	-2.6 ±1.9	-2.7 ±1.3	-2.8 ±1.0	-2.8 ±0.9	-2.9 ±0.9	-2.9 ±0.9
SO$_3^-$ 610[b]	-1.6 ±1.9	-1.2 ±1.3	-1.1 ±1.0	-0.9 ±0.7	-0.8 ±0.6	-0.8 ±0.5
Subtotal	-1.6 ±3.1	-3.3 ±2.0	-4.3 ±1.5	-5.1 ±1.3	-5.6 ±1.3	-5.9 ±1.4
Monosaccharides						
Neu5Ac	0.7 ±0.9	-0.8 ±0.7	-1.6 ±0.7	-2.3 ±0.8	-2.7 ±0.8	-2.9 ±0.9
Core-2 Gal	-4.5 ±0.8	-4.4 ±0.7	-4.3 ±0.6	-4.2 ±0.7	-4.2 ±0.7	-4.2 ±0.7
GlcNAc	-1.2 ±1.2	-2.1 ±0.8	-2.6 ±0.6	-3.0 ±0.6	-3.3 ±0.6	-3.4 ±0.6
Fuc	-5.7 ±2.1	-5.3 ±1.5	-5.1 ±1.3	-4.9 ±1.3	-4.8 ±1.3	-4.7 ±1.3
GalNAc	-0.4 ±0.5	-0.5 ±0.4	-0.5 ±0.3	-0.6 ±0.3	-0.6 ±0.3	-0.6 ±0.3
Gal	0.3 ±0.1	0.2 ±0.1	0.2 ±0.1	0.1 ±0.1	0.1 ±0.1	0.1 ±0.1
Subtotal	-10.8 ±2.7	-12.9 ±2.0	-13.9 ±1.7	-14.9 ±1.8	-15.5 ±1.8	-15.7 ±1.9
Amino Acids						
E604	0.8 ±0.4	0.5 ±0.2	0.4 ±0.2	0.2 ±0.2	0.1 ±0.2	0.1 ±0.2
Y605[c]	-0.5 ±1.0	-0.9 ±0.8	-1.1 ±0.7	-1.3 ±0.7	-1.4 ±0.6	-1.5 ±0.6
E606	-0.5 ±0.8	-0.8 ±0.6	-0.9 ±0.5	-1.1 ±0.4	-1.2 ±0.4	-1.2 ±0.4
Y607[c]	-3.3 ±0.9	-4.3 ±0.8	-4.8 ±0.8	-5.3 ±0.7	-5.5 ±0.7	-5.7 ±0.7
L608	-2.0 ±1.1	-2.1 ±1.2	-2.1 ±1.2	-2.2 ±1.2	-2.2 ±1.2	-2.2 ±1.2
D609	-0.7 ±0.6	-1.2 ±0.6	-1.4 ±0.6	-1.7 ±0.7	-1.8 ±0.7	-1.9 ±0.7
Y610[c]	-0.7 ±0.6	-0.8 ±0.5	-0.8 ±0.4	-0.8 ±0.4	-0.8 ±0.4	-0.8 ±0.4
D611	-0.4 ±0.1	-0.5 ±0.0	-0.6 ±0.1	-0.7 ±0.1	-0.7 ±0.1	-0.7 ±0.1
F612	-0.5 ±0.5	-0.8 ±0.5	-1.0 ±0.6	-1.1 ±0.6	-1.2 ±0.6	-1.2 ±0.6
L613	-3.5 ±1.0	-3.7 ±1.1	-3.9 ±1.1	-4.0 ±1.1	-4.1 ±1.1	-4.1 ±1.2
P614	-3.4 ±1.8	-2.7 ±1.4	-2.4 ±1.3	-2.0 ±1.2	-1.9 ±1.2	-1.8 ±1.1
E615	-0.1 ±0.4	-0.5 ±0.2	-0.7 ±0.2	-0.8 ±0.2	-0.9 ±0.2	-1.0 ±0.2
T616[d]	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3
E617	-0.1 ±0.3	-0.3 ±0.2	-0.4 ±0.2	-0.5 ±0.3	-0.6 ±0.3	-0.6 ±0.3
P618	-0.2 ±0.1	-0.3 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1
Subtotal	-16.5 ±3.1	-19.8 ±2.7	-21.5 ±2.6	-23.1 ±2.6	-24.0 ±2.5	-24.4 ±2.5
Total Interaction Energy	-28.9 ±6.0	-36.0 ±4.8	-39.7 ±4.4	-43.1 ±4.3	-45.1 ±3.4	-46.0 ±4.3

GB$_2$^{OBC} (igb = 5)						
ε_{int}	1.0	1.5	2.0	3.0	4.0	5.0
Sulfate Groups						
SO$_3^-$ 605[b]	2.8 ±1.8	0.7 ±1.0	-0.3 ±0.7	-1.3 ±0.7	-1.8 ±0.8	-2.1 ±0.9
SO$_3^-$ 607[b]	-3.7 ±2.2	-3.5 ±1.5	-3.3 ±1.2	-3.2 ±0.9	-3.1 ±0.9	-3.1 ±0.9
SO$_3^-$ 610[b]	-2.2 ±2.6	-1.6 ±1.7	-1.4 ±1.3	-1.1 ±0.9	-1.0 ±0.7	-0.9 ±0.6
Subtotal	-3.1 ±3.9	-4.4 ±2.5	-5.0 ±1.9	-5.6 ±1.5	-5.9 ±1.4	-6.1 ±1.4
Monosaccharides						
Neu5Ac	0.3 ±1.0	-1.1 ±0.8	-1.8 ±0.7	-2.5 ±0.8	-2.8 ±0.8	-3.0 ±0.9
Core-2 Gal	-5.4 ±0.9	-5.0 ±0.7	-4.8 ±0.6	-4.5 ±0.7	-4.4 ±0.7	-4.4 ±0.7
GlcNAc	-1.4 ±1.3	-2.2 ±0.9	-2.7 ±0.7	-3.1 ±0.6	-3.3 ±0.6	-3.4 ±0.6
Fuc	-8.9 ±2.3	-7.4 ±1.7	-6.6 ±1.4	-5.9 ±1.3	-5.5 ±1.3	-5.3 ±1.3
GalNAc	-0.5 ±0.6	-0.6 ±0.4	-0.6 ±0.4	-0.6 ±0.3	-0.6 ±0.3	-0.6 ±0.3
Gal	0.3 ±0.1	0.2 ±0.1	0.2 ±0.1	0.1 ±0.1	0.1 ±0.1	0.1 ±0.1
Subtotal	-15.6 ±3.0	-16.1 ±2.2	-16.3 ±1.9	-16.5 ±1.8	-16.5 ±1.8	-16.6 ±1.9
Amino Acids						
E604	0.9 ±0.5	0.6 ±0.3	0.4 ±0.2	0.2 ±0.2	0.1 ±0.2	0.1 ±0.2
Y605[c]	-0.3 ±1.2	-0.8 ±0.9	-1.0 ±0.8	-1.2 ±0.7	-1.4 ±0.7	-1.4 ±0.7
E606	-0.5 ±1.0	-0.8 ±0.7	-0.9 ±0.6	-1.1 ±0.4	-1.2 ±0.4	-1.2 ±0.4
Y607[c]	-3.1 ±1.0	-4.2 ±0.9	-4.7 ±0.8	-5.2 ±0.8	-5.5 ±0.7	-5.6 ±0.7
L608	-1.9 ±1.1	-2.0 ±1.1	-2.1 ±1.2	-2.2 ±1.2	-2.2 ±1.2	-2.2 ±1.2
D609	-0.9 ±0.7	-1.3 ±0.6	-1.5 ±0.7	-1.7 ±0.7	-1.8 ±0.7	-1.9 ±0.8
Y610[c]	-1.0 ±0.8	-0.9 ±0.6	-0.9 ±0.6	-0.9 ±0.5	-0.9 ±0.4	-0.9 ±0.4
D611	-0.4 ±0.1	-0.5 ±0.1	-0.6 ±0.1	-0.7 ±0.1	-0.7 ±0.1	-0.7 ±0.1
F612	-0.4 ±0.5	-0.8 ±0.5	-0.9 ±0.6	-1.1 ±0.6	-1.2 ±0.6	-1.2 ±0.6
L613	-3.2 ±1.0	-3.6 ±1.0	-3.7 ±1.1	-3.9 ±1.1	-4.0 ±1.1	-4.1 ±1.1
P614	-3.8 ±2.0	-3.0 ±1.6	-2.6 ±1.4	-2.2 ±1.2	-2.0 ±1.2	-1.9 ±1.2
E615	-0.1 ±0.4	-0.4 ±0.3	-0.6 ±0.2	-0.8 ±0.2	-0.9 ±0.2	-1.0 ±0.2
T616[d]	-1.5 ±0.4	-1.5 ±0.3	-1.5 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3
E617	-0.1 ±0.3	-0.3 ±0.2	-0.4 ±0.2	-0.5 ±0.3	-0.6 ±0.3	-0.6 ±0.3
P618	-0.2 ±0.2	-0.3 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1
Subtotal	-16.5 ±3.4	-19.8 ±2.9	-21.4 ±2.8	-23.1 ±2.6	-24.1 ±2.6	-24.4 ±2.6
Total Interaction Energy	-35.2 ±6.9	-40.3 ±5.2	-42.7 ±4.7	-45.2 ±4.3	-46.5 ±4.3	-47.1 ±4.3

GBn_1 (igb = 7)						
ε_{int}	1.0	1.5	2.0	3.0	4.0	5.0
Sulfate Groups						
SO_3^- 605[b]	1.5 ±1.4	-0.1 ±0.8	-0.9 ±0.6	-1.7 ±0.7	-2.1 ±0.8	-2.4 ±0.9
SO_3^- 607[b]	-6.0 ±2.4	-5.0 ±1.6	-4.5 ±1.2	-4.0 ±0.9	-3.7 ±0.9	-3.5 ±0.9
SO_3^- 610[b]	-2.3 ±2.3	-1.7 ±1.6	-1.4 ±1.2	-1.1 ±0.8	-1.0 ±0.7	-0.9 ±0.6
Subtotal	-6.8 ±3.6	-6.8 ±2.4	-6.8 ±1.8	-6.8 ±1.4	-6.8 ±1.4	-6.8 ±1.4
Monosaccharides						
Neu5Ac	-1.0 ±1.4	-1.9 ±1.0	-2.4 ±0.8	-2.9 ±0.8	-3.1 ±0.8	-3.2 ±0.8
Core-2 Gal	-9.5 ±1.3	-7.7 ±0.9	-6.8 ±0.7	-5.9 ±0.7	-5.4 ±0.7	-5.1 ±0.7
GlcNAc	-1.2 ±1.4	-2.1 ±0.9	-2.6 ±0.7	-3.0 ±0.6	-3.3 ±0.6	-3.4 ±0.5
Fuc	-20.5 ±3.2	-15.1 ±2.2	-12.4 ±1.7	-9.7 ±1.4	-8.3 ±1.4	-7.5 ±1.3
GalNAc	-0.5 ±0.5	-0.6 ±0.4	-0.6 ±0.4	-0.6 ±0.3	-0.6 ±0.3	-0.6 ±0.3
Gal	0.5 ±0.2	0.3 ±0.1	0.2 ±0.1	0.2 ±0.1	0.1 ±0.1	0.1 ±0.1
Subtotal	-32.2 ±4.0	-27.1 ±2.8	-24.6 ±2.2	-21.9 ±1.9	-20.6 ±1.9	-19.7 ±1.8
Amino Acids						
E604	0.7 ±0.5	0.4 ±0.3	0.3 ±0.2	0.1 ±0.2	0.1 ±0.2	0.0 ±0.2
Y605[c]	-1.5 ±0.8	-1.5 ±0.7	-1.6 ±0.7	-1.6 ±0.7	-1.6 ±0.6	-1.7 ±0.6
E606	-0.7 ±1.2	-0.9 ±0.8	-1.1 ±0.6	-1.2 ±0.5	-1.2 ±0.4	-1.3 ±0.4
Y607[c]	-5.4 ±1.2	-5.7 ±0.9	-5.8 ±0.9	-6.0 ±0.8	-6.0 ±0.8	-6.1 ±0.7
L608	-2.4 ±1.3	-2.4 ±1.2	-2.4 ±1.2	-2.3 ±1.2	-2.3 ±1.2	-2.3 ±1.2
D609	-1.3 ±0.9	-1.6 ±0.8	-1.7 ±0.8	-1.9 ±0.8	-1.9 ±0.8	-2.0 ±0.8
Y610[c]	-0.9 ±0.6	-0.9 ±0.5	-0.9 ±0.4	-0.9 ±0.4	-0.9 ±0.4	-0.8 ±0.4
D611	-0.6 ±0.1	-0.6 ±0.1	-0.7 ±0.1	-0.7 ±0.1	-0.7 ±0.1	-0.8 ±0.1
F612	-0.5 ±0.5	-0.8 ±0.5	-0.9 ±0.5	-1.1 ±0.6	-1.2 ±0.6	-1.2 ±0.6
L613	-3.9 ±1.1	-4.0 ±1.1	-4.1 ±1.1	-4.1 ±1.2	-4.2 ±1.2	-4.2 ±1.2
P614	-3.6 ±2.0	-2.9 ±1.6	-2.5 ±1.4	-2.1 ±1.2	-1.9 ±1.2	-1.8 ±1.2
E615	-0.3 ±0.4	-0.6 ±0.3	-0.8 ±0.2	-0.9 ±0.2	-1.0 ±0.2	-1.0 ±0.2
T616[d]	-1.5 ±0.4	-1.5 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3	-1.4 ±0.3
E617	-0.4 ±0.2	-0.5 ±0.2	-0.6 ±0.2	-0.6 ±0.3	-0.7 ±0.3	-0.7 ±0.3
P618	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.5 ±0.1	-0.5 ±0.1	-0.5 ±0.1
Subtotal	-22.7 ±3.5	-23.9 ±2.9	-24.6 ±2.7	-25.2 ±2.7	-25.4 ±2.6	-25.8 ±2.6
Total Interaction Energy	-61.7 ±7.4	-57.8 ±5.5	-56.0 ±4.8	-53.9 ±3.6	-52.8 ±4.4	-52.3 ±3.5

GBn$_2$ (igb = 8)						
ε_{int}	1.0	1.5	2.0	3.0	4.0	5.0
Sulfate Groups						
SO$_3^-$ 605[b]	10.5 ±3.9	5.9 ±2.3	3.5 ±1.5	1.2 ±0.9	0.0 ±0.7	-0.7 ±0.7
SO$_3^-$ 607[b]	7.3 ±2.4	3.8 ±1.6	2.1 ±1.3	0.4 ±1.0	-0.5 ±1.0	-1.0 ±0.9
SO$_3^-$ 610[b]	-0.7 ±1.8	-0.7 ±1.2	-0.6 ±0.9	-0.6 ±0.6	-0.6 ±0.5	-0.6 ±0.5
Subtotal	17.1 ±4.9	9.0 ±3.0	5.0 ±2.2	1.0 ±1.5	-1.1 ±1.3	-2.3 ±1.2
Monosaccharides						
Neu5Ac	-1.4 ±1.2	-2.2 ±0.8	-2.6 ±0.7	-3.0 ±0.7	-3.2 ±0.8	-3.3 ±0.8
Core-2 Gal	-6.1 ±0.9	-5.4 ±0.6	-5.1 ±0.6	-4.7 ±0.6	-4.6 ±0.7	-4.5 ±0.7
GlcNAc	-0.6 ±1.3	-1.7 ±0.9	-2.3 ±0.7	-2.8 ±0.6	-3.1 ±0.6	-3.3 ±0.6
Fuc	-6.0 ±2.0	-5.5 ±1.5	-5.2 ±1.3	-5.0 ±1.2	-4.8 ±1.2	-4.8 ±1.3
GalNAc	-0.1 ±0.3	-0.3 ±0.3	-0.4 ±0.2	-0.5 ±0.2	-0.5 ±0.2	-0.6 ±0.2
Gal	0.2 ±0.1	0.1 ±0.1	0.1 ±0.1	0.1 ±0.1	0.1 ±0.1	0.0 ±0.1
Subtotal	-14.0 ±2.8	-15.0 ±2.0	-15.5 ±1.8	-15.9 ±1.6	-16.1 ±1.7	-16.5 ±1.8
Amino Acids						
E604	0.4 ±0.4	0.2 ±0.2	0.2 ±0.2	0.1 ±0.2	0.0 ±0.2	-0.0 ±0.2
Y605[c]	-1.2 ±0.8	-1.4 ±0.7	-1.5 ±0.7	-1.5 ±0.7	-1.6 ±0.7	-1.6 ±0.7
E606	-0.9 ±1.0	-1.1 ±0.7	-1.1 ±0.6	-1.2 ±0.4	-1.3 ±0.4	-1.3 ±0.4
Y607[c]	-5.7 ±1.1	-5.9 ±0.9	-6.0 ±0.9	-6.1 ±0.8	-6.1 ±0.8	-6.1 ±0.8
L608	-1.9 ±1.1	-2.1 ±1.1	-2.1 ±1.1	-2.2 ±1.2	-2.2 ±1.2	-2.2 ±1.2
D609	-1.0 ±0.9	-1.4 ±0.7	-1.6 ±0.6	-1.8 ±0.6	-1.9 ±0.7	-1.9 ±0.7
Y610[c]	-1.0 ±0.6	-0.9 ±0.5	-0.9 ±0.5	-0.9 ±0.4	-0.9 ±0.4	-0.9 ±0.4
D611	-0.6 ±0.1	-0.7 ±0.1	-0.7 ±0.1	-0.8 ±0.1	-0.8 ±0.1	-0.8 ±0.1
F612	-0.6 ±0.5	-0.8 ±0.5	-1.0 ±0.6	-1.1 ±0.6	-1.2 ±0.6	-1.3 ±0.6
L613	-3.7 ±1.0	-3.9 ±1.1	-4.0 ±1.1	-4.1 ±1.1	-4.1 ±1.1	-4.1 ±1.1
P614	-3.2 ±1.8	-2.6 ±1.4	-2.3 ±1.3	-2.0 ±1.2	-1.8 ±1.2	-1.7 ±1.1
E615	-0.4 ±0.3	-0.7 ±0.2	-0.8 ±0.2	-0.9 ±0.2	-1.0 ±0.2	-1.0 ±0.2
T616[d]	-1.1 ±0.3	-1.2 ±0.3	-1.2 ±0.3	-1.3 ±0.3	-1.3 ±0.3	-1.3 ±0.3
E617	-0.3 ±0.2	-0.5 ±0.2	-0.5 ±0.3	-0.6 ±0.3	-0.6 ±0.3	-0.7 ±0.3
P618	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.4 ±0.1	-0.5 ±0.1	-0.5 ±0.1
Subtotal	-21.6 ±3.2	-23.4 ±2.7	-23.9 ±2.6	-24.8 ±2.6	-25.3 ±2.6	-25.4 ±2.5
Total Interaction Energy	-18.5 ±7.2	-29.4 ±5.3	-34.4 ±4.7	-39.7 ±3.4	-42.5 ±4.2	-44.2 ±3.3

^{a}In kcal/mol. The entropy contributions are not included in these results. ^{b}SO$_3^-$ is counted as a residue in the energy decomposition, instead of $-$O-SO$_3^-$ or $-$CH$_2$-SO$_3^-$. ^{c}Contribution from tyrosine sulfate or tyrosine sulfonate not including the SO$_3^-$ group. ^{d}Glycosylation site.