COMPUTATIONAL ELUCIDATION OF THE STRUCTURAL BASIS FOR THE

ANTIGENICITY OF BACTERIAL CAPSULAR POLYSACCHARIDES:

Neisseria meningitidis

by

BERINYUY YONGYE AUSTIN

(Under the direction of Robert J. Woods)

ABSTRACT

Neisseria meningitidis (*N. meningitidis*) is the leading cause of bacterial meningitis in neonates. The capsule comprises of polysaccharides that serve as key virulence factors that mask vulnerable epitopes and impart antiphagocytic properties. Thirteen strains are known, based on antibody recognition of the capsular polysaccharides (CPS). Only five strains: A, B, C, W-135 and Y are responsible for meningococcal disease. Currently, CPS-conjugate vaccines are available only against types A, C, W-135 and Y. Type B is weakly immunogenic because its CPS, α-(2,8)-linked polysialic acid, is a self-antigen present in the gangliosides of human neural cell-adhesion molecules. Attempts to boost the immunogenicity of CPSs from type B have involved chemical modifications, with the primary goal being the induction of cross-reactivities between antibodies raised against the modified CPS with the bacterial CPS.

Herein, computational protocols are employed to elucidate the conformational properties of haptens from the CPS of *N. meningitidis* B, and to probe the effects of chemical group modifications on their conformational properties. A generalizable biomolecular force field, GLYCAM06, aimed at modeling neutral and charged carbohydrates, such as the *N. meningitidis*

CPSs, has been developed. The force field is tested on its ability to reproduce the experimental and QM vibrational frequencies of carbohydrates, experimental rotational energy barriers and quantum mechanics (QM) rotational energy curves of small molecules. Solution phase tests determine whether explicit-solvent molecular dynamics (MD) simulations, employing GLYCAM06, can reproduce experimental NMR observables such homonuclear scalar ^3J-coupling constants, nuclear Overhauser enhancement distances and the populations of rotational isomeric states.

Explicit-solvent thermodynamic integration MD simulations are employed to determine whether GLYCAM06 can be utilized to predict the binding modes of carbohydrate-protein interactions. The experimental relative affinities of a monoclonal antibody for a series of trisaccharides from the capsule of *Shigella flexneri* variant Y are employed as benchmarks. For structural properties, a pentasaccharide is docked to crystal structures and to a comparative model of the antibody. The docking simulations underscore the importance of employing high quality protein structures in generating theoretical models of protein-carbohydrate complexes. Refining the theoretical complexes via explicit-solvent MD simulations, significantly improves quality in terms of reproducing experimental hydrogen bonds.

COMPUTATIONAL ELUCIDATION OF THE STRUCTURAL BASIS FOR THE

ANTIGENICITY OF BACTERIAL CAPSULAR POLYSACCHARIDES:

Neisseria meningitidis

by

BERINYUY YONGYE AUSTIN

B.S., The University of Buea, Cameroon, 2000

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

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

COMPUTATIONAL ELUCIDATION OF THE STRUCTURAL BASIS FOR THE

ANTIGENICITY OF BACTERIAL CAPSULAR POLYSACCHARIDES:

Neisseria meningitidis

by

BERINYUY YONGYE AUSTIN

Major Professor:		Robert J. Woods
Committee:		Robert A. Scott
		Kelley W. Moremen
		Lionel A. Carreira

Electronic Version Approved:

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

DEDICATION

In memory of my late father Mr. John-Bosco Barah and my late older brother Mr. Louis Lendzemo Yongye, two distinguished men who exemplified leadership, discipline, perseverance and hardwork.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1

Introduction – Literature review

1.1. Introduction

Neisseria meningitidis (*N. meningitidis*) is a gram negative organism naturally found as one of the microflora in the nasopharyngeal tract of humans.[1] *N. meningitidis* together with *Streptococcus pneumoniae, Haemophilus influenzae* type b and *Escherichia coli* (*E. coli*) K1, are the major causes of septicemia and meningitis in neonates worldwide.[2,3] Infection occurs when the bacteria adhere to pili in the mucosal surface and cross the epithelial cells through endocytosis.[1] Annually, about 500,000 cases of meningococcal meningitis are reported worldwide. Typically, 10% result in deaths while survivors may suffer from deafness, seizures, cerebral infarctions or neurological disorders.[3]

N. meningitidis has been classified into 13 serogroups, with the main strain distinguishing factors being the composition and glycosidic linkages between the basic repeat units of the capsular polysaccharides (CPS), Table 1.1. With the exception of type D, all the other strains express polyanionic CPSs that constitute the *O*-antigens extending from their cell walls.[4] Historically, only five of the encapsulated strains, A, B, C, W-135 and Y have been responsible for invasive meningococcal disease. The CPSs are the key virulence factors, which protect the bacterium from intracellular phagocytosis by obstructing cationic antimicrobial proteins[5] and from extracellular lysis by down-regulating complement adhesion.[6] However, the location of the CPSs at the surface makes them susceptible targets of the hosts' humoral immune system – primarily antibodies.

Although the CPSs from *N. meningitidis* bear notable structural and sequence similarities with each other, there are profound differences in their interaction with antibodies, reflected in the apparent size and shape of their antigenic determinants or epitopes. In general, antibodies raised against each CPS are highly specific, and show no cross-reactivity with the CPSs of other strains.[7] For example, the only difference between the CPSs of strains W-135 and Y is epimerization at C4 of α-D-galactopyranose in W-135 to give α-D-glucopyranose in Y, Figure 1.1. Yet antibodies are specific only to their respective homologous strains.[7] The CPSs from strains B and C are both linear, homopolymers of α-5-*N*-acetylneuraminic acid (Neu5Ac), but differ only in the linkage positions along their propagated chains.[8] The monomers are linked via α-(2,8) and α-(2,9) glycosidic bonds in strains B and C, respectively, Figure 1.1. Inhibition studies have established that a minimum oligosaccharide length of 10 residues is required to inhibit binding of a monoclonal antibody (mAb) to the native strain B CPS, while the minimum requirement for the C strain is only 2-3, with maximum inhibition occurring between 4 to 5 residues.[9] Based on inhibitory studies of dextran and anti-dextran antibodies, Kabat[10] suggested that an antibody's binding site can only accommodate a maximum of six monosaccharide units. Because of the unusually large degree of polymerization (DP), 10, required to inhibit binding of the anti-B homologous antiserum to the native strain B CPS, compared to the dextran-antidextran inhibition system, a conformational epitope has been proposed for this CPS, while the smaller number of DP, 5, observed for the strain C CPS has led to its classification as a conventional epitope.[9]

Another intriguing observation was noted in cross-reactivity studies between the CPSs of types B and C, with antibodies raised against the CPS of *E. coli* K92. The CPS of *E. coli* K92 is also a linear homopolymer of α-Neu5Ac, but consists of alternating α-2,8 and α-2,9 linkages.[11]

Antibodies elicited against the K92 CPS cross-reacted with strain C but failed to precipitate strain B haptens.[12,13] The observed cross-reactivity has been attributed to similar and different solution conformational properties between the C and K92 CPSs, and the B and K92CPS, respectively. Similar arguments have been proposed to explain the cross-reactivity of a human monoclonal antibody (IgM[NOV]) with poly-α-(2,8)-Neu5Ac, poly(adenine) and other polynucleotides.[14] In this case the CPS and poly(adenine) displayed similar helical parameters as well as spatial distribution of charges between the carboxyl groups of the B polymer and the phosphate groups of poly(adenine).[15] When the carboxyl groups in the α-(2,8) polymer were reduced to the hydroxymethyl groups (sialitol), the modified CPS required only six residues to inhibit binding to its homologous antiserum.[16] This modification effectively switched the native B CPS from a conformational to a conventional epitope. It was not determined whether the antibodies that were raised against the reduced CPS cross-reacted with the native CPS. Nonetheless, this suggests that any polysaccharides intended to be employed in eliciting cross-reactive responses with the CPS from strain B should preserve the characteristic conformational epitope of the strain B CPS.

Vaccine development strategies against *N. meningitidis* involve immunizing with purified CPSs from each strain. Currently, there are two CPS-based vaccines available: a divalent preparation consisting of purified CPSs from strains A and C, and a tetravalent vaccine comprising of CPSs from strains A, C, W-135 and Y.[17] A vaccine against strain B is currently not available, mostly due to the immuno-biology of its CPS. It is weakly immunogenic, a property that has been associated to immunologic tolerance of its CPS because similar poly-α-(2,8)-linked Neu5Ac oligosaccharides are expressed in the gangliosides of neural cell adhesion molecules during human fetal development.[18-21] In addition, these CPSs are T-lymphocyte

independent type II antigens whose immunologic responses lack T-lymphocyte memory and require repeated immunizations. There is an urgency to develop vaccines that are highly immunogenic and can induce long term immunologic memory responses. In general, immunologic memory can be induced in an immune response against carbohydrate antigens by conjugating the carbohydrate to carrier peptides.[17] Efforts to boost immune responses against strain B have typically involved chemical modifications of the 5-*N*-acetyl moiety to *N*-glycolyl, *N*-propionyl, *N*-butanoyl and *N*-propyl groups,[16,22-26] with limited success.

The foremost goal of this research is to provide a rationale, via computational methods, for the structural features of antigenic oligosaccharides from the surfaces of *N. meningitidis,* which are responsible for mediating the specificity and antigenicity of their interactions with antibodies. There have not been any such studies that relate the immunological properties (epitope size, immunogenicity, antibody cross-reactivity, antigenic sub-group specificities and antibody affinities) of the antigens to carbohydrate sequence in the *N. meningitidis* bacterium. These structural analyses would provide a basis for the rational development of antibacterial vaccines. The focus has been on the conformational properties of haptens derived from the CPS of the B strain for which no effective vaccine is currently available. Particular attention has also been given to the computational protocols that can be utilized in studies involving antigenic carbohydrate-antibody complexes.

1.2. Carbohydrate conformational properties

The involvement of carbohydrates in numerous biological functions and human pathology has been well established.[27,28] Because of the myriad of roles carbohydrates play in biological systems, and the conformation-function correlation associated with biomolecules,

there is a continually increasing interest for the accurate characterizations of the structures and functions of carbohydrates either free or in complex with other biomolecules.

Well established experimental structure determination methods such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been applied to investigate the structural and conformational properties of carbohydrates. But, the inherent flexibility of carbohydrates makes the unequivocal interpretation of their electron densities from X-ray crystallography particularly challenging. Even in systems whereby the flexibility of the carbohydrate is limited due to complexation to a protein, X-ray diffraction data most often include only residues that are closely associated with the protein, because their low mobility facilitates the resolution of their electron density. In addition, X-ray crystallography provides only stationary descriptions of essentially dynamic systems. Alternatively, NMR techniques can be utilized to determine indirect structural observables such as scalar J-couplings, residual dipolar couplings and nuclear Overhauser effects (nOe), which are employed synergistically to derive structural properties. However, the absence and scarcity of inter-glycosidic $^3J_{HH}$ coupling constants and nOe data, respectively, size-limitations, and the uncertainty in the Karplus-type equations utilized to generate an experimental structure, are the major draw-backs of NMR methods. The long NMR time-scales relative to the internal fluctuations in carbohydrates can result in the derivation of a single experimental structure that may not be representative of any meaningful physical state of the carbohydrate.[29] Furthermore, the microheterogeneity of carbohydrate samples from biological sources severely impedes the acquisition of pure samples for NMR spectroscopy or X-ray crystallography.

Computational methods, such as Monte Carlo and molecular dynamics (MD) simulations, present an alternative approach to carbohydrate structure characterization. These methods have

been routinely applied to proteins and nucleic acids for several decades, but only recently have they been applied to carbohydrate-related systems.[30-35] The level of interest in utilizing computational methods to study carbohydrate conformational properties is indicated in the numerous implementations of carbohydrate parameterizations in classical force fields.[36] However, the reliability of a carbohydrate force field is limited by the level of robustness of the parameterization schemes and the stringency of the test cases employed for validation. Nonetheless, with an accurately parameterized and extensively tested force field, the conformational properties of carbohydrates can be investigated with a reasonable level of confidence.

1.3. Challenges in modeling the conformational properties of carbohydrates

Carbohydrates are particularly challenging to model in classical force fields, because of the anomeric and *exo*-anomeric effects,[37,38] the dependence of their conformational properties on the surrounding medium,[34,39-41] their high flexibility, differences in molecular electrostatic potential arising from variations in conformation and configuration (epimerization), and to a lesser extent, *gauche* effects.[38,41] In solution carbohydrates may exist in dynamic equilibrium interchanging between open-chains and five- or six-membered cyclic structures. The most common cyclic forms are six-membered rings known as hexopyranoses. Upon cyclization, the hydroxyl group at the C1 atom in hexopyranoses can adopt either axial or equatorial orientations giving rise to the α- or β-anomers, respectively, Figure 1.2. It has been demonstrated that the relative stabilities of these anomers are quite medium-dependent.[42] In D-glucopyranose, gas-phase quantum mechanics (QM) calculations at the B3LYP/6-311++G**//B3LYP/6-31G* level predict the populations of the α- and β-anomers to be 64% and 36%, respectively.[42] However, employing the QM SM5.4/AM1 solvation model reversed the abundance of the populations with

the α- or β-anomers now 36% and 64%, respectively.[42] These latter populations are in perfect agreement with solution-phase experimental observations.[43] The gas-phase stabilization of the α-anomer is due to anomeric effects that arise from hyperconjugation between the non-bonded lone-pair orbitals (n_p) on the ring oxygen atom and the antibonding orbital (σ^*) of the C1-O1 bond. In going from the gas- to the aqueous-phase the reversal of stability may be rationalized in terms of the larger internal dipole moments of the β-anomer relative to the α-anomer, analogous to dipole moment differences in the equatorial and axial conformers of 2-methoxytetrahydropyran.[44]

Carbohydrates, unlike proteins, may exist as several conformations in equilibrium making them highly flexible molecules, which do not exhibit well-defined secondary or tertiary solution structures. Unlike amino acids that propagate through peptide bonds in a well-defined linear manner, the presence of multiple hydroxyl groups on a single monosaccharide affords the ability to polymerize via glycosidic bonds into both linear and branched structures. Like proteins, their conformations are usually described with the aid of dihedral angles – φ, ψ and ω, between the monosaccharide building blocks, Figure 1.2. This treatment considers the carbohydrate hexo-pyranosyl rings as non-rigid, but the overall conformation is primarily dependent on the dihedral angles adopted by the inter-residue bonds. In theory these dihedral angles can populate a significant combination of staggered rotamers leading to arbitrary multiple conformational families being accessible. Rather, well defined conformations are populated, which are influenced collectively by the steric and electronic properties of the carbohydrate. For example in α-D- and β-L-pyranoses the *exo*-anomeric effect leads to a preference for the *gauche* rotamer of the φ-angle. This *exo*-anomeric effect is also present in the *trans* rotamer. In the β-anomer, the *trans* rotamer is a local minimum. However, in the α-anomer, this rotamer is highly disfavored

because it results in one monosaccharide being situated under the ring of the other giving rise to unfavorable steric interactions between the monosaccharides. In β-D- and α-L-pyranoses the +*gauche* rotamer is preferred, though there is a small presence of the –*gauche* rotamer.[45] The ψ-angle is not restricted by electronic effects and typically oscillates between ±60° with an average value of zero, though there is evidence for the ψ-angle existing in the 180° rotamer.[46]

A further complexity in carbohydrate modeling involves oligosaccharides that polymerize through "1→6" glycosidic linkages by employing the exocyclic hydroxymethyl group of hexoses or sugars containing more than six carbon atoms such as sialic acids. These polysaccharides differ from other carbohydrates in that they possess extra rotatable bonds, ω, Figure 1.2. In pyranoses the ω-angle is defined using the *gauche-gauche* (*gg*) and *gauche-trans* (*gt*) and *trans-gauche* (*tg*) rotamers of the O6-C6-C5-O5 and O6-C6-C5-C4 atomic sequences, respectively, Figure 1.3.

In solution the conformational preferences of these rotamers are heavily dependent on the polarity of the solvent,[40,47] and the stereochemistry of the C4 atom.[48-50] The effect of solvent polarity is demonstrated in methyl 2,3-di-*O*-methyl-α-D-glucopyranoside wherein the *gg*:*gt*:*tg* ratio was 51:49:0 in D_2O, 61:32:7 in $CDCl_3$ and 71:29:0 in benzene-d_6.[40] Differences in populations due to epimerization at the C4 carbon atom are exemplified in methyl α-D-glucopyranoside (α-D-Glc*p*-OMe) and methyl α-D-galactopyranoside (α-D-Gal*p*-OMe). The *gg*:*gt*:*tg* ratios in α-D-Glc*p*-OMe from different studies are 53:47:0[48] and 57:38:5,[50] while in α-D-Gal*p*-OMe, the *gg*:*gt*:*tg* ratios are 14:47:39,[48] 16:75:9[50] and 21:61:18,[49] 13:70:17[49] and 15:69:16.[49] Although the rotamer ratios in each hexose sugar from the different studies are not identical, the overall picture is that in α-D-Glc*p*-OMe and α-D-Gal*p*-OMe the orders of the rotamer populations are *gg* > *gt* > *tg* and *gt* > *tg* ≈ *gg*, respectively, and are different.

These rotamers may also exhibit different lifetimes in solution. Lifetimes in the 5-10 ns regime have been reported from explicit-solvent MD simulations.[41] Therefore, conformational studies of carbohydrates with extra degrees of freedom (DOF) in the glycosidic region, long explicit-solvent MD simulations (~ 50 ns) are required in order to achieve experimentally consistent rotamer sampling.[41] In terms of conformational DOF, the linkage regions between the monosaccharides that constitute the *O*-antigens of *N. meningitidis* are relatively more complex, consisting of four rotatable bonds in type B, and five rotatable bonds in types A and C, Figure 1.1. Therefore, in principle, longer conventional MD simulation times are required to achieve statistically significant rotamer sampling for the accurate interpretation of the experimental data. Alternatively, enhanced-sampling techniques may be utilized to achieve statistical sampling over a shorter time-frame. One common approach is the replica-exchange molecular dynamics (REMD) simulation.[51] This involves the parallel simulation of several replicas of a system at successively increasing temperatures. At specific user-defined intervals a pair of adjacent replicas may swap their temperatures based on an exchange probability.[52] The outcome is the transfer of heat energy to systems at lower temperatures, which results in the low energy system overcoming rotational barriers in the potential energy surface consequently leading to more sampling of conformational space.

Besides the stereo-electronic influences on the conformational properties of carbohydrates, the conformational properties are also phase-dependent. In the gas-phase the hydroxyl groups preferentially form intramolecular hydrogen bonds (H-bonds), some of which may extend across a glycosidic bond.[36,53] These inter-glycosidic H-bonds restrict the conformational space of the glycosidic dihedral angles and have led to the notion that carbohydrates are generally more rigid in the gas-phase relative to the aqueous-phase. In an

aqueous-environment, solvent molecules compete with the intramolecular H-bonds potentially leading to the sampling of different conformational spaces. Solvent molecules may also impact the conformational dynamics of carbohydrates, specifically mannan, cellulose, xylan and hyaluronan, by disfavoring conformations that result in unfavorable entropic bridging of water molecules between adjacent monosaccharides.[34] This phase-influence on conformational properties has been demonstrated in the adiabatic and potential of mean force (φ,ψ)-energy surfaces of α-D-xylopyranosyl-(1,4)-α-D-xylopyranose.[36,54] The global minima in the gas- and aqueous-phases were completely different with (φ,ψ)-values of (0, -40) and (-37, 52), respectively . In addition, the gas-phase global minimum was totally absent in the aqueous-phase, and vice-versa.

Another challenging aspect of carbohydrate modeling involves characterizing the conformational properties of charged sugars, namely ulosonic acids. Examples of carbohydrates that fall under this classification are sialic acids, whose most common member is Neu5Ac, Figure 1.4. Neu5Ac is widely distributed in humans and higher animals.[55] Therefore, it is imperative that its conformational properties are accurately modeled by any carbohydrate force field of biological relevance. A peculiarity of ulosonic acids is the presence of a charged carboxyl group at the anomeric center, which may influence the rotamer preferences of the φ-angle. Unlike other carbohydrates, in solution the φ-angle of α-Neu5Ac primarily populates two major rotamers, *–gauche* and *trans* in an approximate 1:1 ratio.[56,57] In the aqueous- and gas-phases the *–gauche* rotamer is stabilized by hyperconjugation between the non-bonded n_p orbital of the exocyclic oxygen atom and the antibonding σ*-orbital of the C1-C2 bond. In the *trans* rotamer the orientation of the lone pair orbitals on the exocyclic oxygen atom precludes hyperconjugative interactions with the C1-C2 antibonding orbitals. However, the increased

stabilization of the *trans* rotamer observed in solution may be due to the higher dipole moment of this rotamer, which was shown here via QM calculations performed on 2-carboxy-2-methoxytetrahydropyran at the B3LYP/6-31++G(2d,2p)//HF/6-31G* level employing the PCM solvation model.[58] The dipole moments of the *trans* and –*gauche* rotamers were predicted to be 12.6D and 9.9D, respectively.

The preference for the *gauche* rotamer between the two oxygen atoms in the O-C-C-O motif has been ascribed to the well known *gauche* effect.[41] The exocyclic glyceryl sidechain of Neu5Ac features two O-C-C-O motifs, designated ω_8 and ω_9, Figure 1.4. In Neu5Ac, the ω_9 angle defines the hydroxymethyl group and populates all three staggered rotamers in solution. Contrary to the *gauche* effect, ω_8 predominantly exists as the *trans* rotamer when O8 is not involved in a glycosidic bond.[15,57,59-62] This *trans* rotamer is known to be stabilized by a pair of intramolecular H-bonds between the hydroxyl group at C8 and the carboxyl group at C1, and between the hydroxyl group at C7 and the carbonyl group of the acetyl moiety at C5.[59] Therefore, force field parameters developed to model the O-C-C-O sequence in the glyceryl sidechain of Neu5Ac, must reproduce this observed stabilization of the *trans* rotamer. It should be noted that the O-C-C-O motif is predominantly present as the *gauche* rotamer in the pyranose rings of monosaccharides. Developing a force field that over stabilizes the *trans* rotamer of ω_8 may induce a *trans* preference for the O-C-C-O dihedral angles in pyranose rings, resulting in unrealistic ring flips.

The requirement of selectively stabilizing different rotamers of the O-C-C-O motif with the same set of force field parameters, coupled with all the aforementioned stereoelectronic and structural intricacies in carbohydrates, introduce uncharacteristically unique challenges in the

development of a carbohydrate force field. Part of this research involves the development, testing and implementation of such a carbohydrate force field, GLYCAM06.[45]

1.4. Modeling the binding domains of antibodies: from sequence to 3-D structure

The structural interpretation of many biochemical processes requires knowledge of the 3-D structures of biomolecules. These structures are fundamental to providing atomic-resolution insights, such as key regions or residues that are vital for structural stability, and ligand binding or inhibition. In the specific case of antibodies, these structural insights may be employed to guide the rational design of vaccines or to engineer more potent antibodies that can be utilized in immunizations. Experimental structural techniques such as X-ray crystallography and NMR spectroscopy have been the methods of choice for obtaining high-resolution structures of biomolecules. Though these methods have been widely successful some major difficulties are still encountered when deriving experimental structures, which have stimulated wide interests in applying theoretical methods in the prediction of the 3-D structures of biomolecules. While advances in the experimental determination of 3-D structures have been slow, rapid advances in the genome sequencing project have greatly increased the availability of amino acid sequences, evidenced in the millions of sequences available as opposed to only tens-of-thousands of 3-D structures.[60-62] This indicates that a significant number of proteins do not have accessible 3-D structures, and will not be available in a timely manner for atomic-level resolution structural studies. For these proteins, computational methods may provide an alternative approach to generating 3-D models from their available primary sequences. Given the strong dependence of biological function on structure it is imperative that high quality theoretical models, of comparable accuracy as experimental structures, be generated for any structural investigations. Often times the fastest, though not necessarily the most reliable, way to elucidate the structure-

function relationship of a newly identified protein is from a computational model of an experimental homolog.[60] This section will focus on the key issues that must be addressed when generating a comparative model of an antibody.

Antibodies or immunoglobulins are highly conserved, multi-domain proteins with an overall Y-topology consisting of β-sheets that are the structurally conserved region (SCR), held together by disulfide bonds and loops,[61] Figure 1.5. The stem region is denoted "Fc," the letter "c" indicating the ease of crystallizing this fragment, while the arms comprise the antigen-binding fragments, Fab. A single polypeptide chain forms half of the Fc and Fab domains and is known as the "heavy" chain. A "hinge" region comprising a loop structure is found between the Fab and the Fc domains. In the Fab an additional polypeptide, known as the "light chain" complements the heavy chain. Four distinct domains constitute the Fab. Two of these are the constant-light (C_L) and constant-heavy (C_H) domains, denoted "constant" because the amino acid sequences in these domains are highly conserved across immunoglobulin structures. The light and heavy chains are linked together by a disulfide bond between the C_L and C_H domains. The other two domains of the Fab are the variable-light (V_L) and variable-heavy (V_H) domains, which collectively make up the Fv domain. Although amino acid sequences are highly conserved amongst immunoglobulins, sequence variations occur in the Fv domain, with the greatest variation occurring at its apex. This region is primarily involved in antigen binding, by displaying a surface that is complementary to that of the antigen,[61] and is therefore called the complementarity determining region (CDR).

The CDR is formed by six loops, namely: L1, L2, and L3 from the V_L domain, and H1, H2 and H3 from the V_H domain. The main antigen-binding properties reside in the CDRs of the Fab. Therefore, the majority of theoretical and experimental antigen-antibody binding and

structural studies are carried out with either the Fab or Fv domains. In the case of the Fv domain, an artificial linkage connects the V_L and V_H domains, giving rise to a chimeric single-chain Fv domain, ScFv. Theoretical studies typically employ the Cartesian coordinates of the Fv domain in order to reduce computational cost.

Despite the high sequence variability displayed by the CDRs, it has been reported that five of them, namely: L1, L2, L3, H1 and H2 only adopt only a small repertoire of main-chain conformations known as canonical structures.[61,62] These conformations are determined by the length of the loop and the presence of key residues in both the loop and framework regions. The occurrence of only a few canonical structures makes computational methods easily applicable to antibody modeling. The H3 loop, however, cannot be easily classified into any of the common canonical structures because it is the most variable in sequence, structure and size amongst antibodies.[63,64] This loop occupies the central position of the CDR, which implies it is crucial for antigen specificity, involved in various interactions with the V_L domain, the other loops, framework residues and ligand.[61,63,64] Therefore, the generation of this loop in comparative modeling necessitates considerable attention.

There have attempts to characterize the possible conformations of the CDRH3 loop. An approach that separates the H3 loop into a "torso" and "head" region has been utilized to investigate the relationship between its sequence and structure.[65] Using the Kabat numbering scheme, Cys-92 and Gly-104 were utilized to define the boundaries of CDRH3. The first four and last six residues adjacent to the β-sheet framework constituted the "torso" region, while the remainder of the loop made up the "head" region.[65] From a survey of 42 immunoglobulin structures, two general structural patterns emerged in the "torso" region: non-bulge (TNB) and β-bulge (TB) at residue 101.[65,66] In the TNB structures, residues 27 and 32 in the H1 loop

interacted with an aromatic residue of H3 at position 96(d) or 99(f), while the side-chain carboxyl group of Asp-101 formed a H-bond with the indole group of Trp-103.[65] The TNB conformation could be either type I or type II when the aromatic group was at position 96(d) or 99(f), respectively. The TNB fold was also observed when basic residues were simultaneously present at positions 93 and 94.[64,65] All TNB structures studied had short H3 loops, which simplified the problem of predicting their structures to that of determining the structure of the "torso" region.[65] It was also noted that the "head" regions of the short H3 loops could be predicted using rules that govern the structures of short hair-pins.[65] In the TB structure the H1 loop made hydrophobic contacts with the hydrophobic region of the residue 94. The TB structure was stabilized by a salt-bridge between the carboxyl group of Asp-101 and a positively charged residue at position 94, and a H-bond between the carbonyl group of the residue preceding Asp-101 with the indole group of Trp-103, respectively.[64,65] When the H3 loop had up to 16 or more residues the first two and last three residues flanking the "head" region were used to search the database for loops with similar residues.[65]

It should be pointed out that there are exceptions to the rules noted above.[64] For instance, the TB fold has been observed in an antibody (pdbid: 2FBJ) wherein an aspartate residue was absent at position 101.[64] In another antibody, 26-10 (pdbid: 1IGI and 1IGJ), the CDRH3 loop adopted the TB motif even though there no basic amino acid was present at position 94. Instead a glycine residue occupied position 94.[64] Antibodies Yst9.1 (pdbid: 1MAM) and 17/9 (pdbid: 1HIL) contains only Arg-94 and no Asp-101. Yet it displays neither the TB nor TNB folds. Though these exceptions are few, they highlight some of the potential pitfalls that may be encountered when applying canonical structure rules to model the fold of the CDRH3 loop.

A variety of methods has been proposed for modeling the CDRs of immunoglobulins, which broadly include: *ab initio* or *de novo* loop conformational search methods,[72-75] empirical knowledge-based schemes,[62,65-67] or a combination both.[68] In *ab initio* or *de novo* approaches a large number of loop conformations are generated via conformational search methods, followed by the selection of the lowest energy conformation calculated via empirical force field equations, as the overall structure of the CDR. Often, coarse-grain methods that include only Cα atoms in the conformational search stage are employed in order to reduce computational cost. The backbone atoms are next re-constructed prior to implementing an all-atom force-field for refinement and scoring. The selection of the best structure is still problematic, and the majority of methods depend on selecting a representative structure at the centers of large clusters.[69] However, a very large number of "near-native" conformations must be generated and accurately clustered. Nonetheless, amongst theoretical techniques *ab initio* methods are the only methods that could potentially give rise to new protein folds.[60] Differences in *ab initio* structure modeling stem from the methods employed in searching the conformational space, which include simulated annealing,[70] uniform sampling[71] and random conformation dynamics.[72] While these methods may result in appreciable conformational sampling, they completely neglect all the structural information available from NMR or X-ray structures. In addition, the scoring functions may not be completely accurate in terms of ranking the native conformation of the loop as the lowest energy structure. The performance of *ab initio* methods are also limited by the exponential growth in the number of conformations that must be sampled with increase in the length of the loop.[73]

Knowledge-based approaches involve performing unrestricted database searches for loops whose "stem" regions match that of the target sequence. The "stem" region comprises the

backbone atoms of residues that come before and after the loop. These methods derive from the observation that loops do not exist in random conformations, but adopt defined main-chain conformations.[73] Possible template structures are then ranked based on sequence similarities or geometric constraints between the target and template.[73] The chosen templates are next superimposed on the "stem" region of the growing model followed by annealing. While experimental structural data are included in the modeled loops, the main draw-back is the paucity of sufficient structural information.

Combined approaches implement both *ab initio* and knowledge-based methods. In brief, possible template loops are identified via database searches and screened, followed by conformational searches in the presence of the target protein. These methods have been successfully applied in modeling the CDRs of two antibodies.[68] Given the intricacies involved in determining the CDRH3 loop, neither knowledge-based nor *ab initio* methods when used independently will lead to an accurate model. Hence, this work utilizes some aspects of both *ab initio* and knowledge-based methods. First, high-resolution experimental structures that possess both canonical conformations in the CDRs and high sequence similarities as the unknown structure are selected as potential templates. The loops from the template are then grafted unto the SCR of the model, followed by conformational sampling via explicit-solvent MD simulations.

The most common and widely used knowledge-based method is comparative modeling. It is based on the notion that proteins with high sequence similarities will most likely adopt the same fold. Only in very rare cases, have proteins with high sequence similarities adopted different structures.[74] Consequently, the 3-D model of a target protein can be constructed from a template protein that shares a high sequence similarity with the target.[74] Till date, comparative

modeling remains the best method for predicting high quality structures of proteins.[74,75] However, novel folds cannot be predicted using this method. In addition, functional regions are not modeled with better accuracy unless they share more structural conservation than other regions. The protocol in comparative modeling involves the following sequential steps: one or more templates with high sequence similarities to the target are selected; the best alignment between the templates and target is determined; SCR's and variable regions are identified, and the target structure is built; the model is evaluated for both structural and energetic fitness; lastly, the model is refined through force-field based minimizations or MD simulations.

The most critical steps that determine model quality are the correct selection of templates and the sequence-to-structure alignment of the target and templates.[74,75] The traditional method for identifying potential templates is based on sequence similarities between the target and proteins in the database. The search programs of choice are typically BLAST,[76] and FASTA.[77] More sophisticated search algorithms that take into account structural information have been reported,[76,80] but shown to offer no significant benefits compared to BLAST in situations of high-sequence similarities.[75] Once suitable templates have been identified, consensus models are built from multiple templates or by splicing peptide fragments from several models to afford the final model. The main advantage of using multiple templates is that a large percentage of the coordinates of the resultant model are transferred from experimental structures. However, particular attention must be exercised in the treatment of spliced segments. In automated-protein prediction methods, problems of optimally combining the structures of multiple templates often results in the utilization of only one template structure.[78] Recently, an automated procedure, multiple mapping method with multiple templates (M4T), has been recently reported.[78] Templates structures are clustered based on sequence similarities within the templates, as well as

between the templates and target. The unique contributions from each template and their experimental resolutions are also taken into consideration. Sequence to structure mapping is performed by combining different segments of aligned regions that best fit the structural environment of the template. In general, modeling of long loop regions that are not aligned in the template still remains a challenge, and are sometimes built through *ab initio* predictions.[74] Errors in sidechain rotamers typically arise from aligned-non-identical residues and/or shifts in the backbone, which may be corrected via refinements. Typically, refining models derived from templates leads to deterioration in quality due to limitations in force fields in predicting backbone conformations, or the bias of force fields towards certain secondary structural elements.[79] Nonetheless, some improvements in the accuracy of refined comparative models have been reported.[80]

Regardless of the approach employed to generate a comparative model, the coordinates of the model are obtained from pre-existing high-resolution experimental structures, known as templates. These templates are obtained by searching protein-databanks for similarities between the amino acid sequences of existing structures and that of the unknown or target. For antibodies it is sufficient to include only the residues of the V_L and V_H domains. A heuristic search method such as BLAST[76] implementing the BLOSUM62 sequence scoring matrix, can be employed to search the database and align the sequences obtained from the search. A good template requires at least 90% sequence identity between the V_L domains and 70% between the V_H domains. The sequences between the target and template will not be completely identical. For aligned, identical amino acids, the conformations of the side chains can be directly obtained from the coordinates of the template. For aligned, but non-identical residues, the backbone coordinates are transferred

from the template, while those of the side chains are incorporated from a database of rotamer libraries.

A problem often encountered in comparative modeling, is the occurrence of unequal numbers of amino acid residues between some of the CDRs of the template and target sequences. A possible solution involves either insertions or deletions followed by splicing of residues. However, these can potentially introduce structural errors in the CDR. An alternative entails the utilization of coordinates from the loop region of a different immunoglobulin template that has a similar canonical structure and an identical number of residues i.e. using the correct sequence length and not highest sequence identity. Both the first and last four residues before and after the CDR are then aligned structurally before grafting the new loop unto the model. It is quite common to identify several structures from the databank, which have high sequence similarities and share common canonical structure motifs with the target sequence. Therefore, several initial V_L and V_H domains can be generated.

In the majority of cases, the templates employed to generate the coordinates of the V_L and V_H domains are from different immunoglobulin structures. In order to generate the Fv domain, V_L and V_H domains must be spatially oriented relative to each other. In this instance, the protein databank is searched for an Fv domain whose V_L and V_H chains share a high structural similarity with those of the comparative model. The Fv domain of the comparative model may be generated by superimposing different combinations of the V_L and V_H chains unto those of the selected Fv template. The structural quality of the modeled Fv domains are subsequently determined by performing tests with statistical analytical tools such as WHATIF and PROCHECK. The Fv model with the best score is selected and subjected to refinements via empirical force field-based methods such as energy minimizations or MD simulations. Energy

minimizations may only relieve bad steric interactions and reach a local minimum structure. For good rotamer sampling, MD simulations at room temperature are preferred due to the overcoming of high energy barriers on the potential energy surface. The final structure from the MD simulation is annealed to 5K followed by a final minimization stage. In a test case, the quality of refinement is judged by comparing root-mean-squared deviations (RMSD) between the Cα, backbone and side chain atoms of the model and the experimental structure.

1.5. Ligand docking – generating antigen-antibody complexes

Predicting the structures of ligand-receptor complexes is central to understanding numerous biological processes. In the framework of antigen-antibody recognition, the complexes can provide the molecular basis for antigen recognition and specificity. Despite all the technological advances in the structural determination of biomolecules, the structural characterization of carbohydrate antigen-antibody complexes experimentally is still not a trivial task. For systems that are experimentally intractable, automated docking of ligands to receptors can be a practical alternative to generating these complexes, if the structure of the receptor has been solved either experimentally or theoretically. Docking entails predicting the binding modes between ligands and receptors, by exploring the orientations and conformations of a ligand on the surface of a receptor, followed by an assessment of the interaction energy between the two molecules. The highly conserved 3-D folds of antibodies and the precise knowledge of the locations of their binding sites make docking to antibodies relatively straightforward. Successful applications of automated docking of general functional ligands to antibodies have been reported.[81] Several docking software packages are currently available. The more commonly utilized ones include: Autodock,[82] GOLD,[83] DOCK,[84] and FlexX.[85] The implementation of the docking tools differs in three aspects: search algorithms employed to explore the ligand-receptor

binding modes; treatment of the ligand or protein, which could be rigid or flexible, all-atom or united atom for computational efficiency; and the energy functions employed to score or rank the resultant complexes. Extensive reviews of these methods are available.[86,87]

The main conformational search approaches include: MD simulations, Monte Carlo, complementarity, genetic algorithms, distance geometry, tabu and systematic techniques.[86] Any successful algorithm utilizing any of these search methods must be able to sample all the translational and rotational DOF of the ligand, as well as the conformational DOF of both the ligand and the protein, including the experimental binding mode.[86] However, because of the very large total search space involved for both the protein and large flexible ligands most search algorithms often will not find the experimental binding mode in the docking process.[88] To reduce the dimensionality of the search space, the majority of docking packages treat the protein receptor as a rigid entity. Some exceptions are Autodock 4.0 (the most current version of Autodock) and GOLD, which include total side chain flexibility and flexibility in serine and threonine, respectively. Other attempts to simulate protein flexibility involve the utilization of multiple protein conformations generated via MD simulations.[89] Nonetheless, the receptor is still rigid during the docking process. It should be noted that the rigid-receptor approximation is not completely accurate, especially in scenarios whereby the conformation of the receptor changes upon ligand binding. In antibodies, for example, alterations such as side chain adjustments,[90] over loop rearrangements[91] and entire domain movements[92] have been observed. In order to account for these induced-fit effects, auxiliary post-docking refinement methods such as explicit-solvent MD simulations are typically employed.[93]

The scoring functions employed in docking simulations range from force field-based methods such as GLYCAM,[45] AMBER,[79] CHARMM,[94] OPLS;[95] knowledge-based[96] or

empirical functions.[97] A robust scoring function must be able to distinguish the experimental binding mode from all the other complexes generated during the search process. They may be employed fully to search and score a ligand-receptor pose, or in a two-stage process whereby a reduced function is employed for rapid conformational searches followed by the scoring of each resultant pose with the full function.[86] A major inadequacy of most scoring functions, with the exception of some implementations of Autodock,[82,97] is the omission of solvation effects. Some methods attempt to include solvation by including a solvent model only during the scoring stage. Nonetheless, the scored pose is from a gas-phase conformational search.[86]

In this work, Autodock 3.0.5 has been employed to generate the carbohydrate-antibody complexes. Autodock performs automated docking of flexible ligands to rigid receptors. A united-atom model is employed for the ligand whereby non-polar aliphatic hydrogen atoms are merged to their central carbon atoms, while the receptor is modeled as a three-dimensional affinity grid, consisting of regularly spaced points, which store the electrostatic and van der Waals interaction energies between each atom-type in the ligand and all the atoms in the receptor. The Lamarckian Genetic Algorithm (LGA) with the local search algorithm of Solis and Wets[98] are the search methods. The LGA illustrates a correlation between the ligand and protein by translation, orientation and conformation of the ligand. It is similar to the Lamarckian notion of off-springs gaining adaptive qualities of their parents' genotype. Each new docking cycle incorporates conformational results from the previous trial ligand in order to improve the positioning of the next trial. The process of determining a favorable complex is carried out through a series of cycles, each cycle having a user-defined number of steps. During each step, small, random perturbations are performed on the ligand's degrees of freedom: translation, rotation and orientation, resulting in a different geometry. After each operation, the ligand-

receptor interaction energy is calculated and compared with that of the preceding step. If the new energy is lower, this structure is accepted. If the new energy is higher, the structure can be accepted based on a probability function, which depends on the temperature and energy difference between the two configurations.[82] After a set number of user-defined acceptances or rejections, the simulation progresses to the next cycle at a scaled-temperature. The scoring function employed to rank each pose is empirically derived and takes into account the solvation or surface accessibility of atoms in the receptor. Although Autodock is very effective when the receptor's binding surface has a deep groove or buried pocket, it may not consistently reproduce the binding modes of complexes wherein the surface of the receptor is shallow, and/or both the ligand and the receptor are highly charged species.

1.6. Computing antibody antigenicity

An important aspect of predicting antibody-antigen interactions is the correct determination of their interaction energies. The scoring function employed in Autodock is generic and has been parameterized for carbohydrates, but there remain problems.[97] For instance, it is common to have unusual glycosidic dihedral angles from a docking simulation. There has been one previous attempt to empirically optimize the scoring function of Autodock to reproduce interaction energies of carbohydrate-containing complexes.[97] While the energy function showed improvements in predicting carbohydrate-protein binding free energies, energetic penalties arising from deviations from low-energy carbohydrate dihedral angles were not included in the optimization. A more robust approach involves the utilization of an all-atom force field's energy function to compute the ligand-receptor interaction energies. Two such methods include: thermodynamic integration (TI) and molecular mechanics generalized-Born surface area (MM-GBSA) calculations.

Thermodynamic integration methods are more applicable in determining the relative affinities of a receptor for closely and structurally related ligands.[99] A functional group in the ligand is chemically modified (perturbed) to another functional group in a series of explicit-solvent MD simulation steps and the free energy change associated with the perturbation is determined. This perturbation is performed for the ligand in solution and also in complex with the protein, Figure 1.6. For example the relative affinities of a receptor R for two ligands, L and L*, can be determined experimentally by taking the difference between ΔG_1 and ΔG_2, known as $\Delta\Delta G$. In theory, $\Delta\Delta G$ can be determined computationally by bringing together L and L* with R. However, this would entail significant restructuring of the receptor, the ligands and solvent, such that statistical sampling of phase-space would be extremely intensive computionally.[100] This major difficulty is circumvented by determining the values of ΔG_3 and ΔG_4, each representing the alchemical processes of "mutating" L to L* both unbound and complexed to the protein, respectively. Since ΔG is a state function, $(\Delta G_4 - \Delta G_3) = (\Delta G_2 - \Delta G_1) = \Delta\Delta G$. If $(\Delta G_4 - \Delta G_3)$ is less than zero, then L* is a better ligand, and the reverse is true.

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(p^N, r^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (1.1)$$

ΔG_3 and ΔG_4 are determined from computational calculations by solving the integral above, equation 1.1. A series of MD simulations are performed for values of λ from 1 to 0, representing L and L*, respectively. The increments are selected to ensure maximum overlaps between the successive states that are modeled as linear combinations of the initial and final states. For each value λ the Hamiltonian is calculated and the value stored. The final free energy change is determined from the area under the graph of the values of the Hamiltonian against the λ values.[100] In this work, a two-stage methodology is employed to mutate L to L*. In the first stage

only the coulombic properties of L, primarily partial atomic charges, are gradually "mutated" to those of L*, and in the second stage the van der Waals properties of L are transformed to those of L*, both stages performed through a series of MD simulation steps. Though computationally expensive, this method has the advantage that the binding affinity is decomposed into contributions from solvation, electrostatics (charges and dipoles) and sterics (van der Waals and hydrophobic packing). It should be noted that only relative free energies can be computed through TI calculations.

The MM-GB/SA approach is a more common method employed to determine relative and absolute free energies of binding. In brief, explicit-solvent MD simulations are performed with the selected ligand-receptor complexes. Conformational snapshots are selected from the trajectory at regular intervals after the simulation has reached statistical equilibration. Solvent molecules are next stripped off the snapshots. Employing a suitable force field and solvation model, the ligand-receptor interactions are computed for each snapshot and averaged, with the coordinates of the structures fixed at their MD-determined values. The individual average interaction energies of different ligand-receptor systems are compared to determine relative affinities. The contribution from solvation is estimated from the utilization of the generalized Born implicit solvation model. The atomic interaction energies can also be decomposed into contributions from solvation, electrostatics, sterics and entropy, equation 1.2. The entropic contributions can be calculated from a normal mode analysis of energy-minimized coordinates of the molecules.

$$<\Delta G_{Bind}> = <G_{Complex}> - (<G_{Receptor}> + <G_{Ligand}>)$$

where $<G> = <E_{MM}> + <G_{Solv}> - T<S_{MM}>$

$<E_{MM}> = <E_{EE}> + <E_{vdW}>$; Total molecular mechanical energy

$<G_{Solv}> = <G_{Pol}> + <G_{Non-pol}>$; Solvation free energy including polar and non-polar components

$T<S_{MM}>$ = Temperature * (sum of rotational, vibrational and translational entropies of the solute)

Given that the solvent is treated implicitly, a potential problem of the MM-GB/SA approach is the treatment of deeply-buried water molecules, which may be involved in solvent-mediated interactions between the ligand and the receptor. It is also apparent that entropic contributions that result from the displacement of solvent molecules upon complex formation are also neglected. In addition, the choice of solvation model and internal dielectric constant (ε_{int}) must be carefully considered. In a previous study in which values of 1 and 4, $\varepsilon_{int} = 4$ were utilized, both gave reasonable interaction energies. However, $\varepsilon_{int} = 1$ consistently overestimated the interactions,[101] which may be due to the use of non-polarizable force fields and solvent models. Nonetheless, MM-GB/SA calculations have been successfully applied to provide insights on the stabilities[102] of antigen-antibody complexes and affinity maturation[103] of antibodies.

Table 1.1 Capsular polysaccharide structures from *N. meningitidis*

Serotype	Oligosaccharide repeat units (anionic residues in bold)	Acetylation
A	-6)-α-D-ManNAc-(1-**OPO3**-	3 *O*
B	-8)-**α-D-Neu5Ac**-(2-	------
C	-9)-**α-D-Neu5Ac**-(2-	7/8 *O*
W-135	-6)-α-D-Gal-(1-4)-**α-D-Neu5Ac**-(2-	(7/9)[a] *O*
Y	-6)-α-D-Glc-(1-4)-**α-D-Neu5Ac**-(2-	(9 *O*)
X	-4)-α-D-GlcNAc-(1-**OPO3**-	------
H	-4)-α-D-Gal-(1-2)-Gro-(3-**OPO3**-	------
Z	-3)-α-D-GalNAc-(1-1)-Gro-(3-**OPO3**-	------
I	-3)-**β-D-ManNAcA**-(1-4)-**α-L-GulNAcA**-(1-	4 *O*
K	-3)-**β-D-ManNAcA**-(1-4)-**β-D-ManNAcA**-(1-	4 *O*
L	-3)-**β-D-ManNAcA**-(1-4)-β-D-GlcNAc-(1-3)-α-D-GlcNAc-(1-**OPO3**-	------
29e	-3)-α-D-GalNAc-(1-7)-**β-Kdo**-(2-	------

[a]Parenthesis denote partial *O*-acetylation
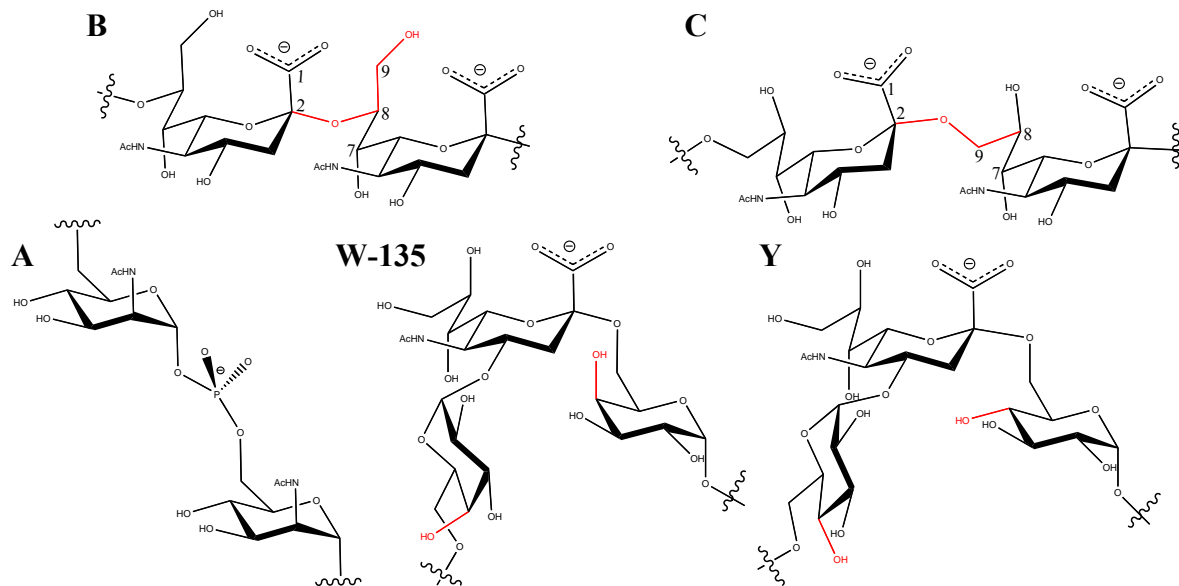
Figure 1.1. Basic repeat units from the CPSs of the five most virulent strains of *N. meningitidis*
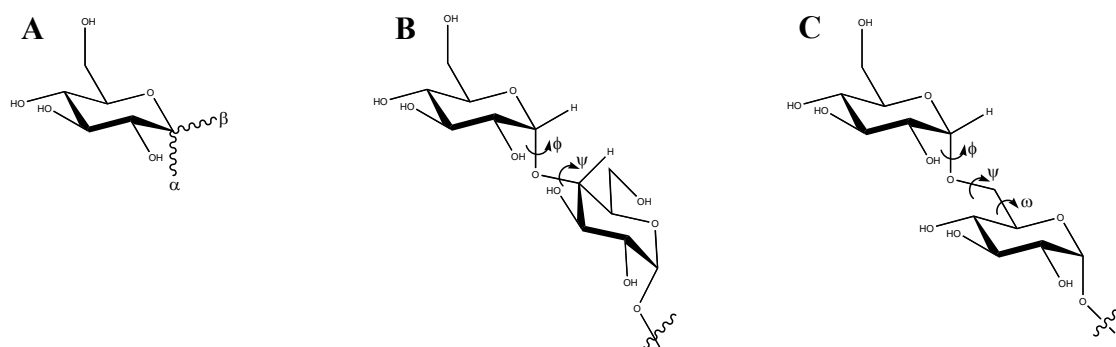
Figure 1.2. Different configurations (α, β) of the anomeric center in D-glucose, A. Glycosidic dihedral angles relevant in determining the conformational properties of a carbohydrate with α-(1,4)- and α-(1,6)-linkages, B and C, respectively. In the 1,4-linkage, the φ- and ψ-angles are defined by H_1-C_1-$O_{4'}$-$C_{4'}$ and C_1-$O_{4'}$-$C_{4'}$-$H_{4'}$, respectively. In the 1,6-linkage the φ-, ψ- and ω-angles comprise of H_1-C_1-$O_{6'}$-$C_{6'}$ and C_1-$O_{6'}$-$C_{6'}$-$C_{5'}$ and $O_{6'}$-$C_{6'}$-$C_{5'}$-$O_{5'}$, respectively
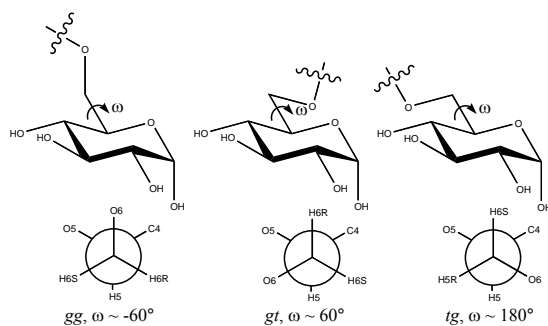
Figure 1.3. A schematic representation of the three possible rotamers of the ω torsion angle, defined using the three staggered rotamers of the O6-C6-C5-O5 and O6-C6-C5-C4 atomic sequences; *gauche-gauche* (*gg*), *gauche-trans* (*gt*), and *trans-gauche* (*tg*)
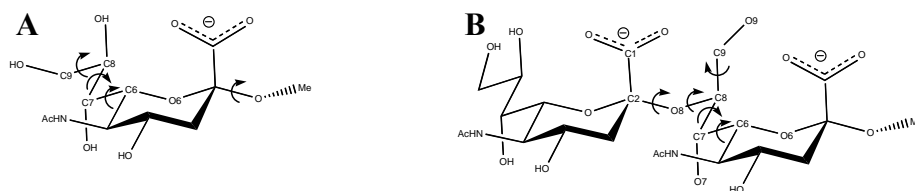
Figure 1.4. Important exocyclic torsion angles in methyl α-Neu5Ac (A), methyl α-Neu5Ac-(2,8)-α-Neu5Ac (B). The nomenclature is: φ (C_1-C_2-O_8'-C_8'), ψ (C_2-O_8'-C_8'-C_7'), ω_9 (O_9-C_9-C_8-O_8), ω_8 (O_8-C_8-C_7-O_7) and ω_7 (O_7-C_7-C_6-O_6)
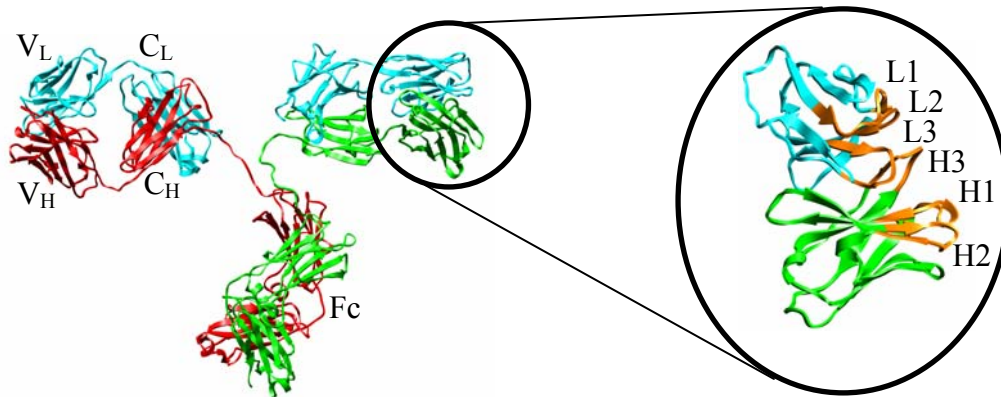
C_L = Constant region of light chain
V_L = Variable region of light chain
C_H = Constant region of heavy chain
V_H = Variable region of heavy chain
F_c = Stem region

Figure 1.5. Overall Y-topology of an intact antibody (pdbid: 1IGT), formed by multiple domains consisting of β-sheets and loop regions is portrayed. The light chains are shown in cyan and the heavy chains in red and green. The hypervariable loops that form the antigen-binding site, L1-L3 from the light chain and H1-H3 from the heavy chains, are portrayed in the insert

$$\begin{array}{ccc}
\text{R} \quad + \quad \text{L} & \xrightarrow{\quad \Delta G_1 \quad} & \text{R-L} \\[2em]
\Big\downarrow = 0 \quad \Big\downarrow \Delta G_3 & & \Big\downarrow \Delta G_4 \\[2em]
\text{R} \quad + \quad \text{L*} & \xrightarrow{\quad \Delta G_2 \quad} & \text{R-L*}
\end{array}$$
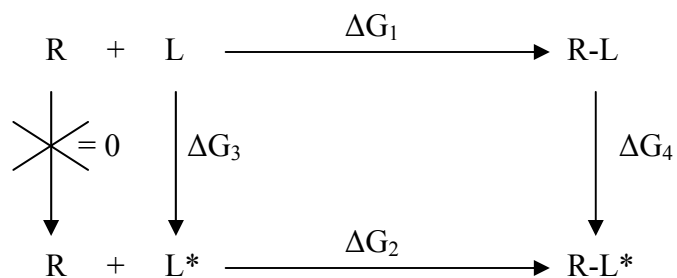
Figure 1.6. Schematic representation of a thermodynamic closed-cycle employed to determine the relative affinities of an antibody (R) for two closely related ligands L and L*. ΔG_1 and ΔG_2 are the changes in free energy of R binding L and L*, respectively. ΔG_3 and ΔG_4 are the changes in solvation and binding free energies, respectively, between L and L*.

CHAPTER 2

GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates[1]

Abstract

A new derivation of the GLYCAM06 force field, which removes its previous specificity for carbohydrates, and its dependency on the AMBER force field and parameters, is presented. All pertinent force field terms have been explicitly specified and so no default or generic parameters are employed. The new GLYCAM is no longer limited to any particular class of biomolecules, but is extendible to all molecular classes in the spirit of a small-molecule force field. The torsion terms in the present work were all derived from quantum mechanical data from a collection of minimal molecular fragments and related small molecules. For carbohydrates, there is now a single parameter set applicable to both α- and β anomers, and to all monosaccharide ring sizes and conformations. We demonstrate that deriving dihedral parameters by fitting to QM data for internal rotational energy curves for representative small molecules generally leads to correct rotamer populations in molecular dynamics simulations, and that this approach removes the need for phase corrections in the dihedral terms. However, we note that there are cases where this approach is inadequate. Reported here are the basic components of the new force field as well as an illustration of its extension to carbohydrates.

In addition to reproducing the gas-phase properties of an array of small test molecules, condensed-phase simulations employing GLYCAM06 are shown to reproduce rotamer populations for key small molecules and representative biopolymer building blocks in explicit water, as well as crystalline lattice properties, such as unit cell dimensions, and vibrational frequencies.

KEYWORDS: carbohydrate, force field, GLYCAM, AMBER, molecular dynamics, parameter development

2.1. Introduction

Carbohydrates are involved in numerous biological functions, such as recognition in axonal growth or path-finding,[104] blood anticoagulation,[105] cell-cell recognition,[106] antibody-antigen interactions,[107] structure factors in extra-cellular matrices[39] and post- or co-translational modifications of polypeptides.[108] Correct glycosylation patterns are essential for normal cell and organism function, and aberrant glycosylation is associated with numerous human diseases.[27,28] In order to provide a structural basis for the multitude of biological roles played by carbohydrates, it is imperative that their spatial and dynamic properties be accurately determined. Experimental structure determination methods such as X-ray crystallography,[30,116,117] NMR spectroscopy[32,35,109,110] and fluorescence energy-transfer spectroscopy[111] have been applied in studies of carbohydrate conformation, either free or complexed with proteins. While NMR spectroscopy has been extensively used to characterize the dynamics of glycans in solution,[36] inter-glycosidic linkage conformations are notoriously difficult to determine by NMR spectroscopy because of the paucity of nuclear Overhauser effects (nOes),[112] the uncertainties in the Karplus-type equations employed to interpret scalar J-coupling constants[113] and the potential for the linkage to populate multiple rotamer states.[33] Moreover, NMR techniques employed to determine the structural properties of polysaccharides or protein-carbohydrate complexes is limited by molecular weight constraints. Alternatively, X-ray crystallography can be a powerful source of structural information. However, the presence of multiple glycoforms often prevents crystallization of glycoproteins, and the inherent flexibility of oligosaccharides is the presumed reason for the notable absence of X-ray structures for any but the smallest systems.

Theoretical methods, such as Monte Carlo and molecular dynamics (MD) simulations are employed increasingly to augment the experimental approaches in determining the

conformational properties of carbohydrates, and biomolecules in general. The level of interest in applying classical simulations to oligosaccharides arises from experimental limitations, and is demonstrated by the numerous force fields and parameter sets that have been derived for carbohydrates (See reference 36 for a review). The performance of several parameter sets, which included the first version of GLYCAM (GLYCAM_93[114]), has been quantitatively evaluated against data from quantum mechanical calculations[115] as well as on a relative basis, using a chemometric analysis from twenty second-generation carbohydrate force fields.[38] The conclusion from those comparisons was that no single parameter set or force field consistently out-performed the others. In a study that focused on the abilities of three prominent biomolecular force fields to reproduce experimental solution data for two related disaccharides, the second version of GLYCAM (GLYCAM2000[41]) with AMBER performed well in terms of conformational predictions.[116] Although the GLYCAM_93[114] and 2000[41] parameter sets, augmented by terms from the AMBER parameters,[79] have been employed frequently in the study of oligosaccharides[34,38-41,124-129] and oligosaccharide-protein complexes,[35,108,109,116,130-134] they have some serious limitations. Earlier GLYCAM versions performed poorly at reproducing diffusion rates in explicit solvent MD simulations, and differed substantially from other carbohydrate parameter sets in the prediction of putative radial pair distribution functions (RPD) between hydroxyl groups and TIP3P[117] water.[116] For example, the RPD for the O3-atom in methyl α-D-isomaltoside computed with GLYCAM2000a resulted in a less structured secondary hydration shell, and a first hydration shell at 3.0 Å, which was higher than the average distance between $O_{carbohydrate}$…O_{water} atoms found in carbohydrate crystals (2.77 Å).[118] The first hydration shell of the O3…O_{water} RPD had a peak density value of 1.0, which compared unfavorably with the experimental value of about 3.0 for the RPD of O_{water}…O_{water}.[116] The computed O3…O_{water} RPD

36

employing the present version of GLYCAM (GLYCAM06) resulted in a first hydration shell at 2.75 Å with a peak density of 1.6, and a well-defined secondary hydration layer at about 5.0-5.5 Å. This improvement in GLYCAM06 was due to the utilization of AMBER-consistent van der Waals parameters for the hydroxyl oxygen. In the case of self-diffusion rates of α-D-isomaltose GLYCAM2000a gave a value that was about twice the experimental value.[116] The self-diffusion rate computed with GLYCAM06 was $3.6 \cdot 10^{-6}$ cm^2/s, which is in agreement with the experimental value of $4.2 \cdot 10^{-6}$ cm^2/s.[116]

The previous versions of GLYCAM were designed with the intention that they would introduce the minimal parameters necessary to add carbohydrate functionality to AMBER, and otherwise to maintain consistency with the AMBER force field development philosophy.[114] Regarding transferability, GLYCAM development followed the general approach employed in biomolecular force fields of defining a single dihedral angle term for each molecular-class-specific linkage. Analogous to the treatment of peptide backbone dihedral angles, the rotational properties associated with glycosidic linkages were characterized by a single dihedral angle term specific to a unique atomic sequence, present only in carbohydrates. All other potentially-contributing terms associated with that linkage were explicitly eliminated.[114] That approach facilitated precise fitting to QM rotational data, but was specific to six-membered ring forms (pyranosides) and to the anomeric configuration. This limited the ability to readily introduce new chemical functionality into GLYCAM, to study ring conformational interconversions, and to apply it to other ring sizes (furanosides, in particular) .[119]

The derivation of a highly consistent and transferable parameter set for modeling carbohydrates and glycoconjugates (GLYCAM06) is the focus of the present work. In light of accumulated applications of GLYCAM_93,[34,38,40,41,115,116,120-129] several areas were targeted for

improvement. In particular the parameter set should 1) be transferable to all carbohydrate ring conformations and sizes, 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, 6) be rigorously assessed in terms of the relative accuracy of its component terms, and 7) avoid the use of 1-4 electrostatic or non-bonded scaling factors.[41] In a study of the ω-angle rotation (O_5-C_5-C_6-O_6) in monosaccharides, we observed that O_6 may interact with either O_4 (in a 1-5 relationship) or O_5 (in a 1-4 relationship) and the use of 1-4 scaling therefore unbalanced these interactions leading to an inability to correctly predict rotamer populations.[41]

Parameterization of GLYCAM06 employed training and test sets of approximately 100 molecules from the chemical families of hydrocarbons, alcohols, ethers, amides, esters, carboxylates, molecules of mixed functional groups, as well as simple ring systems related to cyclic carbohydrates, as outlined in Table 2.1. If a molecule had the potential of forming an internal hydrogen bond during a torsion rotation, two rotational energy curves were examined, one in which a hydrogen bond was allowed, and a second in which the hydrogen bond was disallowed. To maintain consistency with the current AMBER protein parameters, the B3LYP/6-31++G(2d,2p)//HF/6-31G* level of theory was selected as the reference for all valence quantum calculations.[130] Energies at this level may be expected to display minimal basis set superposition error (BSSE),[131] which is not only important for determining accurate interaction energies, but also for locating minima and establishing barrier heights along the valence reaction coordinates. It has been shown that the B3LYP functional performs well in carbohydrates and related molecules,[131-133] and was therefore selected for our force field development. Of particular

importance for dynamics, this level of theory leads frequently to a reduction of barrier heights for internal rotations relative to those obtained with the HF/6-31G* method, resulting in increased conformational sampling for oligosaccharides.

2.2. Methods

2.2.1. Quantum mechanical calculations

Quantum mechanical calculations were performed using the Gaussian98 software package.[134] Geometries were optimized using the Gaussian98 default optimization criteria, unless otherwise noted. The HF/6-31G* level of theory was employed for neutral fragments, whereas for anionic molecules diffuse functions were added. Rotational energy curves were generated by varying the relevant torsion angle from 0-360° in 30° increments while optimizing all other variables. Single point energies were computed at the B3LYP/6-31++G(2d,2p) level of theory.

2.2.2. Molecular mechanics calculations

All molecular mechanics (MM) calculations were performed using either the AMBER7,[135] or AMBER8[136] software packages, with non-bonded and electrostatic scaling factors (SCEE, SCNB) set to unity. The valence parameters were determined by minimizing the error between the energies obtained from the *ab initio* and molecular mechanics calculations in the traditional manner.

2.2.3. Molecular dynamics

Molecular dynamics (MD) simulations were carried out with the explicit inclusion of solvent (TIP3P)[117] under isothermal-isobaric (nPT) conditions. Charged systems were neutralized by adding the appropriate number of counter ions. In general the initial solvent configurations were optimized through 50 cycles of steepest descent and 950 cycles of conjugate

gradient energy minimization with the SANDER module, followed by minimization of the entire system through the same protocol. The entire system was then annealed by heating from 5 to 300K in 50ps, followed by cooling to 5K in another 50ps. Initial velocities were assigned from a Boltzmann distribution at 5K. Prior to the production dynamics stage the entire system was thermally equilibrated by heating again from 5 to 300K in 150ps. A 2-fs time step was used to integrate the equations of motion, using the Verlet Algorithm.[137] Long-range electrostatic interactions were treated using the particle mesh Ewald summation. A dielectric constant of unity was employed in all MD and MM calculations. Bonds containing hydrogen were constrained to their equilibrium lengths using the SHAKE algorithm.[138]

2.2.4. Potential of mean force calculations

The weighted histogram analysis method (WHAM)[139] was used to calculate potentials of mean force (PMF).[140] To ensure maximum sampling along the reaction coordinate, harmonic force constants of 10 and 5 kcal/mol were used for regions of high (maxima) and low (local and global minima) potential energies, respectively. Prior to the PMF calculations, the system was equilibrated following the protocol outlined in the MD section. Umbrella sampling was performed for 140 ps under nPT conditions with a time-step of 2 fs.

2.2.5. Single molecule and ensemble-averaged charge calculations

The atomic partial charges of all molecules employed in the gas phase for the development of valence bond, angle and torsional parameters were derived from the HF/cc-pVTZ//HF/6-31G* molecular electrostatic potential (MEP), employing the restrained electrostatic potential (RESP)[141] charge fitting methodology, with a hyperbolic charge restraint weight of 0.0005, as indicated by Bayly.[141] However, partial charge sets destined for condensed-phase simulations were generated from the HF/6-31G*//HF/6-31G* MEP, appropriate for

TIP3P[117] water simulations. For the condensed-phase charge sets, a RESP charge restraint weight of 0.01 was employed, based on earlier MD simulations of experimental crystal lattices of α-D-glucopyranose,[129] wherein this value led to optimal reproduction of unit-cell dimensions of the crystal.

In GLYCAM06 the issue of charge-conformation coupling was addressed by employing ensemble-averaged (EA) charge sets.[142] For any given monosaccharide, an MD simulation, typically 50-100ns was performed in TIP3P water, with 100-200 structures being selected from the trajectory for individual charge calculations. For each of these snapshots partial charges were calculated by fitting to the HF/6-31G* MEP. Prior to the charge calculations, each structure was optimized at the HF/6-31G* level, with the rotatable exocyclic bonds constrained to their MD conformations.[142] The charges were then averaged for each monosaccharide to afford a final EA charge set, as exemplified in Table 2.2. These sets of charges were thus weighted by the actual occurrence of a particular solution conformation in an MD simulation. For these terminal monosaccharides, the average RESP-derived charge on the aglycon was -0.194 au, with the total charge on the glycoside equal to +0.194 au (Q_{term}), for overall charge neutrality. To generate charges on non-terminal residues, the charge on the linking hydroxyl proton is added to that of the linking oxygen atom to give a new charge, Q_{ol}. Next, Q_{term} is subtracted from Q_{ol} to afford the final charge on the linking oxygen atom. This results in an overall charge of zero for non-terminal residues.

Non-bonded van der Waals parameters were taken directly from the PARM94 parameter set, which are also employed in the current distribution of AMBER parameter sets AMBER9.[143]

2.2.6. Vibrational analysis

Vibrational frequencies were computed for a-D-glucopyranose using the NMODE facility in AMBER9 with the GLYCAM parameter set. The calculation was performed on a monomer and on a 64-unit crystal. In both calculations, the initial structure was adapted from neutron scattering crystallographic data (PDB ID: GLUCSA10).[144] The NMODE calculations were performed with default settings except for the values of the 1-4 non-bonded and 1-4 electrostatic scale factors, which were each set equal to unit as suggested by earlier work.[129] Both monomer and crystal were minimized to ΔRMS=1×10^{-5}. Since periodic boundary conditions are not implemented in NMODE, the 64-mer was minimized as if it were a discrete nanoparticle. Frequency analyses were also performed on the monomer at the HF/6-31G*//6-31G* and B3LYP/6-31++G(2d,2p)// 6-31++G(2d,2p) levels of theory using Spartan '06 (HF) and Gaussian '03 (DFT).

Because of the complexity of the molecular structures, particularly in the 64-mer, a program (DMODES, for ''describe modes'') was written to automatically detect and assign vibrational modes. Although assignments for glucopyranose have often[145-149] been made based on the potential energy distribution (PED) within each vibrational mode, NMODE facilitates use of the eigenvectors (collection of atomic motion vectors). Primarily for this reason, DMODES bases its assignments on the motions and relative kinetic energies of the atoms. However, there are other reasons to make assignments based on motion rather than PED's. While PED assignments are limited to bonded and near neighbor nonbonded interactions, motion analysis can easily consider concerted motions of atoms far removed from each other. For example, motion analysis can assign ring expansions and deformations, assignments that are far more difficult to make using PED analysis. In addition, motion analysis yields assignments that are

intuitive and that use language familiar to spectroscopists. The use of a computer program in general is, of course, also preferred to visual inspection of the eigenvector data because of increases in speed, objectivity, and reliability. At present, DMODES is in the prototype stage and is of limited general usefulness, but a copy of the code can be obtained by contacting the authors.

In brief, DMODES employs the relationships of individual atomic motions with respect to each other and to the molecular geometry to determine an appropriate description of the vibrational motion for a mode. For example, the dot product of a unit motion vector for atom i, $\vec{V}_i$, with a corresponding unit bond vector, $\vec{B}_{ij}$, indicates the extent to which the motion is a stretch along that bond. Since the dot product of two unit vectors yields the cosine of the angle between them, the result of this calculation will be nearly ± 1 if the motion is parallel to the bond (stretching) and ~0 if the motion is perpendicular (bending or torsioning). Similarly, bending and torsional motions of a system i—j—k can be determined using eq. (1):

$$\vec{V}_i \bullet \left(\vec{B}_{ij} \times \vec{B}_{jk} \right) \qquad\qquad 1$$

If this relationship is nearly zero, the motion lies within the plane formed by i, j, and k and is therefore a bend (if also determined not to be a stretch). If the magnitude of the relationship is nearly one, the motion is nearly perpendicular to the plane and is torsional. Intermediate values, of course, correspond to mixtures of torsion and bending. DMODES employs a set of cutoff values to distinguish between motions.

The significance of an individual atom's motion was determined using the relative kinetic energy of the atom. Each atom's motion vector represents its relative velocity at its vibrational potential energy minimum. The vector magnitudes can therefore be converted into relative kinetic energies, which correspond to the significances of the individual atomic motions within the vibrational mode. Although DMODES considered each atomic motion, the primary

vibrational modes, which are presented here, were assigned on the basis of the most significant motion within each motion class. The implications of this choice are discussed in more detail in the results and conclusions section.

Assignments for the 64-mer required additional considerations. Since NMODE does not employ crystal-based symmetry arguments when calculating frequencies, 4608 normal modes (3N, where N = 64 units x 24 atoms per unit) were generated. Also, since an accurate frequency calculation requires a minimization, the positions of the atoms were necessarily displaced somewhat from their perfect-crystal locations. Because of this, the assignments for the 64 units, while generally similar, were each unique. To reduce the 294,912 different motion assignments, the data were treated in the following manner. First, the motions of all 64 units were assigned as usual for each vibrational frequency. During the assignment, the overall importance of each unit's motion to each mode was scored by summing the magnitudes of the motion vectors for each of the atoms in the unit. The motion vectors used for this score had previously been altered to remove any overall translational component (due to lattice-wide motions). The maximum magnitude sum for each frequency was then determined. If, within any frequency, a unit's magnitude was at least 50% of the largest magnitude, the results of its motion analysis were included in the results presented here. This procedure removes from our immediate attention those motions that occur with low intensity, being primarily due to resonance with more energetic motions of nearby neighbors. However, it does not discount resonances that result in motions that are more energetic, if somewhat unexpected. An example of the latter appears in the data and is discussed below.

Modern characterizations of the infrared and Raman spectra of carbohydrates typically employ some form of calculation of the PED when making assignments,[145-152] though they

occasionally use more traditional methods[150,152]. Those that rely significantly[152] or exclusively[153] on traditional methods, as well as the occasional computation,[151] will often employ more descriptive terminology in their assignments. While the assignment method employed here was designed to translate more readily into standard spectroscopic descriptions, it is important, also, to compare with the less-intuitive PED-based systems. To present a balanced comparison with experimental data, we have chosen to compare with four experimental studies.[162,163,165,166] Since the various assignment naming methodologies do not always translate in a one-to-one correspondence (for an illustration, see Fig. 20 in Wells and Atalla [152]), we have chosen categories that are sufficiently broad that they facilitate comparisons, but are also narrow enough that the comparisons remain useful.

Of the four experimental studies considered here, two[151,152] employed traditional assignment nomenclature and two[148,149] assigned based on PEDs. All four used one or more computational method to make assignments, though one[152] also used traditional methods, e.g., isotopic substitution, to guide their assignments. Three[148,149,151] compared their calculations to spectra from earlier studies,[145-147,150,153] but their results are being used here instead of the earlier work because of the detailed assignments they provided. Since our interest is comparison with experiment, only the experimentally observed frequencies from these studies are considered here. Additionally, we have only considered those experimental frequencies for which assignments were offered.

Since our assignments are based on the most energetic motion class within a given mode, we compared with only the most significant (first on the list) potential energy contribution within any PED assignment.[148,149] With no hierarchy of significance apparent in the work of Gregurick and Kafafi,[151] each assignment within a frequency was counted. The ranges from Wells and

Atalla[152] come mostly from their Figure 20, with occasional modifications based on the accompanying text, the latter being necessary for proper correlation to the set of motion classes being considered here.

Since IR and Raman intensities are not calculated natively by NMODE, and since the focus of the present study is the extent to which the force field reproduces vibrational frequencies we will not consider them here. In our comparisons with experimental data below, where both IR and Raman frequencies are available, we have compared with only the former, since the frequencies usually differ by only a few wavenumbers, which is not significant in this work. For modes that are not IR active, we have compared with the Raman frequency.

2.3. Results and Discussion

2.3.1. Atomic partial charges

The accurate reproduction of the electrostatic properties of a molecule is essential in a classical force field that seeks to quantitatively model intermolecular interactions of polar molecules, such as carbohydrates, proteins and nucleic acids in water. The complexity of this requirement is reflected by the number of partial atomic charge derivation schemes reported in the literature. These include empirical methods, in which charges are fit to reproduce liquid densities and heats of vaporization,[95] or a popular alternative, and that used in AMBER, in which the charges are fit to reproduce quantum mechanical MEPs.[79] It should be noted that the concept of a partial charge is artificial but convenient, thus there is no definitive approach to their assignment.

A common drawback of several empirical methods is that they rely on reproducing bulk liquid properties, and cannot be applied readily to solutes. In all MEP partial charge-based models, there are limitations that arise from the omission of a mechanism for incorporating the

dependence of partial charge on molecular conformation.[154,155] In addition, it is not necessarily ideal to locate the partial charges at the nuclear positions,[117,156] nor do all atoms require partial charges.[157] Close examination of ESP-charges computed for biomolecular fragments[154] revealed that aliphatic hydrogen atoms tend to have small magnitude charges, as expected from the low bond polarity of most C-H bonds. Further, charges on all hydrogen atoms showed wide fluctuations coupled to the central carbon atom. However, the net charge of the methyl or methylene groups displayed minimal fluctuations. In GLYCAM_93,[114] the effects of basis set and ESP-fitting algorithm on partial charges in carbohydrates were examined and again indicated the same behavior for aliphatic groups.

A second problematic area is that of charge partitioning. While MEP-fitting has the advantage of reproducing intermolecular interaction energies, caution must be exercised in the selection of atoms that will be included in the fitting procedure. Francl *et al.*[157] performed singular value decomposition analyses on the least-squares matrices used to assign partial charges in a series of compounds, which frequently showed that an optimal fit to the ESP was achievable with fewer charges than the total number of atoms, suggesting that consistent reproduction of the MEP can be obtained by assigning zero net charges to some atoms. Consequently, in GLYCAM06, restraints were employed in the ESP fitting to ensure that the charges on all aliphatic hydrogen atoms were zero, leading to consistent charge sets with minimal degradation in the accuracy of the fit. For example, the omission of partial atomic charges on aliphatic hydrogens in methyl α-D-mannopyranoside had essentially no effect on the molecular dipole moment (4.589D) relative to the all atom charge model (4.581D), and resulted in only an approximate 2.5% increase in the relative error in the fit to the potential.

In contrast to intermolecular interactions, which are well reproduced by MEP-derived partial charges, intramolecular interactions or solution properties are often poorly reproduced, unless conformationally averaged charges are employed. This issue becomes particularly significant when considering the relative energies of conformational isomers. For example, in the case of cyclohexanol, slightly different partial charge sets result from MEP fitting for both the equatorial and axial forms. When separate charge sets are utilized, incorrect relative energies of the ring conformations were obtained. However, accurate prediction of the relative energies between the chair forms, $^4C_1 \leftrightarrow {}^1C_4$ (<1 kcal/mol)[158] necessitated employing a common charge set for both the equatorial and axial species.

Extension of this logic to monosaccharides is challenged by notable variations among partial charge sets as a function of both anomeric and ring-carbon configuration. From the data in Table 2.2, both changes in anomericity (α, β) and hydroxyl group configuration (methyl α,β-D-glucopyranoside - α,β-D-Glc*p*OMe, methyl α,β-D-mannopyranoside - α,β-D-Man*p*OMe and methyl α,β-D-galacotopyranoside - α,β-D-Gal*p*OMe) resulted in partial charge sets that were statistically indistinguishable. Therefore, it is reasonable to generate a single charge set that can be used for either anomer of a given monosaccharide. However, a limitation in employing anomer-averaged charge sets is the loss of precision in the calculated intermolecular properties. Nevertheless, with a single charge set for a given monosaccharide, the relative intramolecular energies of ring conformations in which the anomeric configuration becomes inverted ($^4C_1 \leftrightarrow {}^1C_4$) can now be effectively computed. For an example of the effect of charge protocol on internal energies, consider the relative energies of the α- and β-anomers of Glc*p*OMe. At the HF/6-31G*//HF/6-31G* level of theory the α-anomer is predicted to be 1.46 kcal/mol more stable than the β-anomer, while at the B3LYP/6-31++G(2d,2p)//HF/6-31G* level, this value is 0.77

kcal/mol, which is in good agreement with other recently reported theoretical results.[159] Using GLYCAM06 and the α,β EA (RESP< α,β>) charge set derived from Table 2.2, the α-anomer was predicted to be 1.15 kcal/mol more stable, a value between the HF/6-31G* and B3LYP/6-31++G(2d,2p) results. When the same calculations were performed using the individual EA charge sets, (RESP< α>) for α-D- and (RESP<β>) β-D-GlcpOMe, the relative energy was reversed, with the β-anomer predicted to be the more stable by 4.19 kcal/mol. Illustrating further the concept of a common charge set, the relative energies of twelve low-energy conformers of D-GlcpOMe and D-GalpOMe, computed using GLYCAM06 were compared with their respective HF/6-31G*//HF/6-31G* values, Table 2.3. In the case of D-GlcpOMe, when anomer-specific charges were used, the predicted relative energies were once more reversed relative to the QM-predicted values, with the most stable α-conformer being 4.45 kcal/mol less stable than the most stable β-conformer. When <α,β> EA charge sets were employed, the resulting relative energies were comparable once more with the QM values; the α-anomer being preferred by 1.29 kcal/mol. It can be inferred that this energy difference has an electrostatic origin, since a common parameter set was employed for both anomers. Amongst the β-D-GlcpOMe conformers, the trend in the GLYCAM06 predicted relative energies (β1<β3<β2) did not reproduce that determined by QM (β2<β1<β3). This variance, however, is insignificant given that the higher energy conformers in both the QM and MM calculations are all within 0.6 kcal/mol of the global minima. In the case of D-GalpOMe, the GLYCAM06-computed relative energies employing anomer-specific charge sets resulted in both large energy differences (11.71 kcal/mol) and incorrect anomer preferences (β<α). Once more, the correct trends were predicted when the RESP<α,β> charge set was used, with the most stable α-anomer being 1.58 kcal/mol more stable than the lowest energy β-anomer, which compared favorably with the QM value of 1.88

kcal/mol. This strategy of averaging anomer-specific charges could be extrapolated to obtain a common set of charges for all the glycosides (RESP<overall>). Employing this charge set in D-GlcpOMe and D-GalpOMe, the α-anomer was correctly predicted to be more stable by 1.12 and 1.79 kcal/mol, respectively.

2.3.2. Bond lengths and angles parameters

The stereoelectronic effects that give rise to valence bond and angle variations at the anomeric carbon atom in pyranoses are well known.[160,161] These variations are not only dependent on anomericity, but also influenced by the size of the aglycon.[162] The stereoelectronic effects may be incorporated into Type II classical force-fields through stretch-bend and stretch-torsion cross terms.[163] However, these cross terms do not exist in well-established macromolecular force fields, such as AMBER,[79] CHARMM [94] and GROMOS.[164] In GLYCAM_93, these geometrical variations were incorporated by defining anomer-specific atom types, each with unique values for the bond (K_r) and angle force constants (K_θ), and the equilibrium bond lengths (r_{eq}) and angles (θ_{eq}). While the use of unique atom types for the different anomeric carbon atoms facilitated the inclusion of these stereoelectronic effects, it inhibited the ability of the force field to model processes involving ring inversions, which are equivalent to anomer interconversions. In addition, maintaining non-generalizable anomer-dependent parameters prevented facile extension of the force field to non-standard glycosyl residues. In the present parameterization, a common set of parameters has been derived for both α- and β-anomers employing a single sp^3 atom type (CG). For example, in the previous scheme, a 2-deoxymonosaccharide would require new parameters to be derived, whereas in the current approach all such derivatives are implicitly allowed.

In principle harmonic stretching and bending force constants may be derived from spectroscopic data. However, such data are not available for all molecules of interest to this work. In addition, to be consistent with the use of QM data in the derivation of rotational barriers, we wished to employ only gas-phase force constants. Thus all force constants were derived by fitting a classical harmonic function to a distortion energy curve, computed at the B3LYP/6-31++G(2d,2p) level.

The accurate determination of the bending force constants for atomic sequences involved in glycosidic angles ensures that the flexing and rotational dynamics of the molecule will be correctly reproduced. Values of K_θ for the C-O-C and O-C-O angles were estimated from distortion analyses of the applicable angles in methoxymethane and 1,1-dimethoxyethane, respectively. Fitting to the QM energy curves, computed over the range $\theta_{eq} \pm 10°$, gave K_θ values of 50 kcal/rad^2 and 100 kcal/rad^2 for the C-O-C and O-C-O angles, respectively. The average relative errors in these curve fits were 0.73 and 0.16 kcal/mol, respectively. These angle-bending constants are slightly softer than those employed in GLYCAM_93 C-O-C (60.0 kcal/rad^2) and O-C-O (110.0 kcal/rad^2) .[114]

It should be noted that the experimentally observed bond lengths and valence angles in molecules such as carbohydrates and amino acids should not be employed as equilibrium values. These complex molecules contain internal strain forces that perturb the bond lengths and angles away from their theoretical equilibrium values. Equilibrium values, however, may be estimated from small carefully selected molecular fragments. These equilibrium values may differ significantly from the observed valence angles in larger molecules. In addition, effects associated with intermolecular interactions can lead to considerable differences between the QM values for bonds and angles, and those observed experimentally. For example, in the case of peptide bonds,

the average experimental C-N bond length (1.33 Å) from a survey of crystal structures from the Cambridge Crystallographic Data Bank differs markedly from both the gas-phase electron diffraction, and MP2/6-31G* values of 1.386 Å and 1.365 Å, respectively, computed for *N*-methylacetamide.[165,166] Similarly, the θ_{eq} value for the C-O-C angle in GLYCAM06 (111.6°) is smaller than the average values observed in α-(2→8) trisialoside (117°),[15] and generally in saccharides (116.5°).[167,168] Nonetheless, using tetrahydro-2-methoxy-5-(tetrahydro-2*H*-pyran-2-yloxy)-2*H*-pyran, Figure 2.1, the GLYCAM06 minimized structure gave a glycosidic valence angle of 116.8˚, comparable to that found in experimental structures of glycosides.[168] When methyl α-D-Glc*p*(1→4)α-D-Glc*p* and methyl β-D-Glc*p*-(1→4)β-D-Glc*p*, Figure 2.1, were employed the glycosidic angles were 117.3˚ and 116.9˚, respectively, which compared favorably with experimental observations.[168]

The bond and angle parameters developed in this work are presented in Table 2.4. Only six out of thirty-five bond, and twenty out of ninety-one angle parameters were eventually transferred from the existing GLYCAM_93 parameter set.

2.3.3. Torsion parameters

The GLYCAM_93 force field, as well as other current biomolecular force fields, such as the PARM99[169] parameterization in AMBER, frequently employ a single atomic sequence to define the torsion properties for a given bond. For example, in GLYCAM_93, the rotation associated with the φ-angle was defined by the terms for the O_5-C_1-O_X-C_X sequence only; with a specific torsion term required for each α- and β-anomer. The related terms C_2-C_1-O_X-C_X and H_1-C_1-O_X-C_X were explicitly set to zero. While this approach may lead to accurate reproduction of the QM rotational energy curves, it suffers from a lack of transferability. In view of making the parameters in GLYCAM06 fully transferable we adopted a more general approach wherein all

atomic sequences had an explicitly defined set of torsion terms. Each term was derived by fitting to rotational energy curves for a molecular training set composed of relevant small molecules. Due to the symmetry of these small molecules, no phase angles were required. The omission of phase angles further facilitates parameter transferability. However, as each rotational energy is now the result of the sum of the contributions from the constituent atomic sequences, the overall accuracy is limited by the sum of errors in each term. The benefits in terms of applying to novel molecular classes is particularly important for carbohydrates as they frequently exist in highly derivatized forms *in vivo*.

Together with the non-bonded terms, torsion terms are crucially important for force fields to correctly model the conformational and particularly, the dynamic properties of macromolecules. As opposed to bond and angle parameters, in which only small excursions from equilibrium values are observed, torsional rotations may cover the entire range between 0-360˚, and frequently exhibit more than one energy minimum. In order to achieve correct thermodynamic and kinetic behavior during MD simulations, it is important that both the relative energies of the minima and the barrier heights between them be reproduced. Thus, we not only sought to locate the minima and maxima, but also to reproduce all barrier heights. To assess the extent to which this had been achieved for each molecular class, the errors between the QM and MM rotational energies were computed over the entire range of the curves, $<Error>_{curve}$, and the minima, $<Error>_{minima}$, are presented in Table 2.5.

2.3.3.1. Hydrocarbons

Hydrocarbons provide the foundation for any biomolecular force field. As can be seen from the data in Table 2.5, the force field reproduced the gas-phase relative energies for these

simple molecules very well, with an average error in the energies over the entire rotational energy curve of 0.15 kcal/mol and an error in the $<Error>_{minima}$ of 0.11 kcal/mol.

Rotating about the central C_{sp3}-C_{sp3} (CG-CG) bond in butane allowed us to examine how several torsion parameters performed collectively, namely those associated with the CG-CG-CG-CG, CG-CG-CG-HC, and HC-CG-CG-HC sequences. GLYCAM06 predicted the stabilization of the *trans* conformation, relative to the *gauche* conformation, of 1.16 kcal/mol, in reasonable agreement with both the B3LYP/6-31++G(2d,2p) value computed here of 0.95 kcal/mol and the experimental range of 0.74 to 0.90 kcal/mol.[184-186] For this molecule the GLYCAM06 rotational barriers were determined to be 5.51 kcal/mol at 0° and 3.15 kcal/mol at 120°, which may be compared with the B3LYP/6-31++G(2d,2p) values of 5.80 and 3.33 kcal/mol, respectively, and with the experimental values[170] of 4.56 and 3.62 kcal/mol, respectively.

2.3.3.2. Alcohols

The hydroxyl groups of alcohols make this class of compounds particularly relevant in the development of force fields for nucleic acids and carbohydrates. The overall errors in the fits for alcohols were $<Error>_{curve}$ = 0.45 kcal/mol and $<Error>_{minima}$ = 0.28 kcal/mol. The alcohols with a single hydroxyl group were easier to fit and resulted in $<Error>_{curve}$ = 0.20 kcal/mol, while the fit was slightly degraded for the diols with $<Error>_{curve}$ = 0.73 kcal/mol.

2.3.3.2.1. Mono-alcohols

Mono-alcohols were used for the derivation of parameters that involved hydroxyl groups without the inclusion of other electronegative groups. As relates to the rotational behavior of the hydroxyl protons, the H1-CG-OH-HO rotational barrier in methanol was determined at the B3LYP/6-31++G(2d,2p) level and with GLYCAM06 to be 1.03 and 1.08 kcal/mol, respectively, both in excellent agreement with the experimental value[171] of 1.07 kcal/mol. For ethanol,

GLYCAM06 determined the *trans* conformation about the CG-CG-OH-HO torsion to be 0.34 kcal/mol more stable than the *gauche* conformation, which may be compared with the experimental value[171-173] of 0.40 kcal/mol, while the B3LYP/6-31++G(2d,2p) level overestimated the stability of the gauche conformer by 0.30 kcal/mol. The barrier to rotation between the *gauche* conformations was determined to be 1.11 kcal/mol, which compared well with the B3LYP/6-31++G(2d,2p) value and experimental range[174] of 1.01 kcal/mol and 0.8 – 1.2 kcal/mol, respectively. Similarly, for propan-2-ol, GLYCAM06 determined the *gauche* (defined by H1-CG-OH-HO) conformer to be more stable than the *trans* by 0.11 kcal/mol, which may be compared to the experimental[172,173] and QM values of 0.45 ± 0.21 kcal/mol and 0.24 kcal/mol, respectively. The GLYCAM06 value for the HC-CG-CG-OH rotational barrier in ethanol was found to be 3.09 kcal/mol, compared to the B3LYP/6-31++G(2d,2p) value of 3.07 kcal/mol, which both slightly underestimated the experimental range[171,175] of 3.32 to 3.55 kcal/mol.

Developing parameters for the CG-CG-CG-OH sequence was challenging due primarily to the different conformational preferences of this torsion in open-chain and cyclic systems. In open-chain systems, using n-propanol as an example, the *gauche* conformation computed here at the MP2/6-311++G(2d,2p)//MP2/6-31G** level, was favored over the *trans* by 0.21 kcal/mol. In contrast, the equatorial conformer of cyclohexanol[176] in which the sequence CG-CG-CG-OH is *trans*, is more stable than the axial conformation (CG-CG-CG-OH is *gauche*) by 0.56 kcal/mol. The differences between the open-chain and the cyclic preferences appear to reside in induction effects, which are difficult to account for in a classical mechanical treatment lacking polarizability.[177] Therefore, GLYCAM06 was derived to slightly overestimate the *trans* conformation in n-propanol (0.77 kcal/mol) over the *gauche* in order to optimize the performance of this parameter in glycan rings. As a result, optimal agreement was achieved for

the rotational energy profile of the equatorial conformation of hydroxymethylcyclohexane Figure 2.2, as well as in highly substituted open-chain alcohols, such as 2,2-dimethylpropanol and 1,1,2,2-tetramethylpropanol.

2.3.3.2.2. Diols

Diols mimic the atomic sequence, O-C-C-O prevalent in the vicinal hydroxyl fragments on the rings of common glycosyl residues. The torsional preferences for this linkage affect ring conformations as well as the rotamer preferences of exocyclic side chains, such as in sialic acid. This torsion term is most strongly influenced by the OH-CG-CG-OH parameters, as well as by internal hydrogen bonding. In order to establish the relative contributions from each of these effects, rotational energy curves for the O-C-C-O linkage were computed both allowing, and prohibiting the formation of internal H-bonds.[41] The ability of GLYCAM06 to reproduce these relative energies and trends was illustrated by the average error values $<Error>_{curve}$ and $<Error>_{minima}$, in the rotational energy profiles. For example, in 2,3-butanediol $<Error>_{curve}$ = 0.40 kcal/mol and $<Error>_{minima}$ = 0.35 kcal/mol when internal H-bonding was allowed, while $<Error>_{curve}$ = 0.34 kcal/mol and $<Error>_{minima}$ = 0.21 kcal/mol when internal H-bonding was disallowed.

The presence of the OH-CG-CG-OH dihedral angle in the exocyclic glyceryl side chains of sialic acids such as 5-*N*-acetylneuraminic acid (Neu5Ac) is significant in determining the conformational properties of the side chain, Figure 2.3. The side chain features two dihedral angles, which bear the OH-CG-CG-OH sequence: ω_8 ($O_7C_7C_8O_8$) and ω_9 ($O_8C_8C_9O_9$). The tendency for the OH-CG-CG-OH angle to adopt the *gauche* conformation in simple diols has been referred to as the *gauche* effect.[41] However, in Neu5Ac, while ω_9 populates all three staggered rotamers, ω_8 predominantly exists in the *trans* conformation, as ascertained from both

solution NMR[15,57,59,62] and X-ray crystallography[59,178] studies. A pair of intramolecular hydrogen bonds, between H8O and the oxygen atoms of the carboxylate group, and H7O and the carbonyl group of the amido moiety, appears to stabilize this rotamer. However, the magnitude of this stabilization is unclear, as the same *trans* preference for ω_8 is also observed in the β-anomer,[59] in which the intramolecular H8O-carboxylate H-bond is not possible. Application of the initial parameters for the OH-CG-CG-OH torsion term, generated by fitting to the QM rotational energy curves for compounds in the initial training set of diols, in solvated MD simulations, did not quantitatively reproduce the rotamer populations for the glyceryl side chain in Neu5Ac. In contrast to experimental data, the initial parameters led to a strong predilection for the *gauche* rotamer. Consequently a new approach was adopted in developing torsional parameters for the OH-CG-CG-OH atomic sequence.

The initial values for the OH-CG-CG-OH torsion term were modified by systematically adjusting the V_1 coefficient until explicitly solvated MD simulations of Neu5Ac were able to reproduce the solution populations for ω_8. Because our goal was to increase the sampling of the *trans* rotamer of ω_8, and only the energy function of the V_1 coefficient has a single minimum at 180°, only this term was modified. Next, the effect of the different V_1 values on the stability of the ring conformations in selected monosaccharides was evaluated, because the OH-CG-CG-OH atomic sequence is also present in ring systems. Because the entire range of V_1 values were consistent with predicting the ring conformational preferences, and a series of MD simulations of Neu5Ac indicated that different V_1 values resulted in a predominant occurrence of the ω_8 *trans* rotamer, the most consistent value was finally determined by performing explicitly solvated MD simulations of 1,2-ethanediol, and selecting the value that best reproduced its experimentally determined rotamer populations.

The rotamer sampling of ω_8 in the glyceryl side chain, employing different V_1 values are presented in Table 2.6. Traversing these values from -1.0 to 1.0 kcal/mol resulted in an overall increase in the occurrence of the *trans* rotamer, indicating that an adjustment of this coefficient was necessary in order to reproduce the solution conformational properties of ω_8.

As mentioned earlier, the OH-CG-CG-OH term is also present in monosaccharide rings, most commonly existing in *gauche* relationships, Figure 2.4. Increasing the V_1 coefficient could induce a preference for the *trans* relationships of OH-CG-CG-OH dihedral angles present in ring systems, subsequently introducing artifactual ring flips during MD simulations. To determine whether the different V_1 coefficients had any influence on the stabilities of pyranose chair conformations when the parameters were employed in MD simulations, the MM generalized Born surface area (MMGBSA) implicit solvation model was used to compute the average MM energies of α-D-Man*p* and α-D-idopyranose (α-D-Ido*p*) in both $^4C_1$ and $^1C_4$ conformations. In the experimental solution conformations, α-D-Man*p* displays a single *trans* and two *gauche* O-C-C-O conformations in both the $^4C_1$ and $^1C_4$ chair forms. In contrast, α-D-Ido*p* displays three *trans* and three *gauche* O-C-C-O conformations in the $^4C_1$ and $^1C_4$ chair forms, respectively.

For each monosaccharide, 1-ns explicit solvent MD simulations were performed in which the ring was either in the $^4C_1$ or $^1C_4$ conformation. Snapshots were selected from the last 500 ps at 1 ps intervals from each simulation, and water molecules were removed. Using different V_1 coefficients the average MM energies of the individual chair conformers of each monosaccharide were determined by averaging the MMGBSA computed energies of the 500 selected snapshots. Over the entire range of V_1 values, the calculations predicted a consistent preference for the $^4C_1$ over the $^1C_4$ conformations in α-D-Man*p*, Figure 2.4. This insensitivity of the α-D-Man*p* ring conformation relative to the V_1 coefficient can be attributed to the similar distribution of the

trans and *gauche* O-C-C-O conformations in both chair forms. The preference for the $^4C_1$ over the $^1C_4$ conformer is also expected due to the unfavorable axial configuration of the hydroxymethyl group at C5, in addition to the energetically disallowed 1-3 interaction between the axial hydroxyl group at C3 and the C5 hydroxymethyl group. On the other hand, traversing the range of the V_1 coefficients in the α-D-Ido*p* system, there was a preference for the $^1C_4$ conformation at the negative extreme (-2.0 kcal/mol) with a progressive increase in the stabilization of the $^4C_1$ conformation towards the positive extreme (2.0 kcal/mol). This trend is expected because for negative V_1 coefficients, the V_1 energy profile as a function of dihedral angle has a maximum barrier at 180°, the reverse being true for positive coefficients. Thus the exclusive presence of the O-C-C-O *trans* conformations in the $^4C_1$ structure leads to a destabilization for negative V_1 values, but becomes a stabilizing factor for positive values of V_1.

From MD simulations of methyl α-Neu5Ac, observation of the ω_8 *trans* and *gauche* rotamer populations as a function of V_1, Table 2.6, indicated that the following values: -0.2, 0.0, 0.4, 0.6, 0.8 and 1.0 kcal/mol, would lead to rotamer preferences that were consistent with experimental data. In addition, the MMGBSA calculations showed that these coefficients would not cause spurious ring flips during MD simulations. Therefore, the most consistent coefficient was determined by performing explicit solvent MD simulations of 1,2-ethanediol, with V_1 ranging from -0.5 to 0.5 kcal/mol, in increments of 0.1 kcal/mol. The conformational properties of 1,2-ethanediol have been extensively investigated by both experimental and theoretical studies,[179-190] which predict a predominance of the *gauche* conformation, albeit to different extents, ranging from *gauche:trans* 99:1[188] to 67:33.[187] Experimental NMR data[189] indicate a mixture of 80:20% *gauche:trans* populations in solvents with low dielectric constants. Chidichimo *et al.*[182] reported the existence of only the *gauche* conformation. However, their

studies were done using nematic-lyotropic crystalline solutions. A V_1 value of -0.1 kcal/mol best reproduced the experimental populations obtained in conditions most similar to those of the simulation, with a *gauche:trans* ratio of 81:19%. Using this value of -0.1 kcal/mol a PMF-WHAM analysis of methyl α-Neu5Ac with ω_8 as the reaction coordinate was carried out. This analysis produced an energy profile with the *trans* rotamer being more stable by approximately 1.0 kcal/mol relative to the +*gauche* conformer, and 1.2 kcal/mol with respect to the –*gauche* conformer, Figure 2.5. The rotational barriers to transition between +*gauche* and –*gauche* conformers were approximately 2.4 kcal/mol and 3.4 kcal/mol, respectively, indicating some transition to the +*gauche* rotamer. The MD simulations correctly predicted the solution behavior of the glyceryl ω_8 torsion, Figure 2.5. To assess further the extent to which these new parameters were able reproduce the B3LYP/6-31++G(2d,2p) rotational energies, the rotational energy curves for the O-C-C-O torsions for 2,3-butanediol were determined for comparison, Figure 2.6. As expected, when intramolecular hydrogen bonds are allowed, there is a stabilization of the *gauche* conformers over the *trans* conformer. When intramolecular hydrogen bonds are excluded, by constraining the hydroxyl proton to a *trans* orientation (HO-OH-CG-CG =180˚), or by substituting the hydroxyl protons with methyl groups, electrostatic repulsions between the oxygen atoms reverse the relative stabilities of the rotamers, resulting in the *trans* rotamer being the most stable. The $<Error>_{curve}$ = 0.87 kcal/mol and $<Error>_{minima}$ = 0.35 kcal/mol when internal H-bonding was allowed, and $<Error>_{curve}$ = 1.04 and $<Error>_{minima}$ = 0.61 kcal/mol when internal H-bonding was disallowed. Comparing these error values to those determined using the initial OH-CG-CG-OH parameters, indicated that the present parameters resulted in fits that were slightly less accurate. However, this slight degradation was necessary for the correct prediction of the solution conformational properties of the glyceryl side chain of sialic acid.

2.3.3.3. Ethers

The parameters derived from ethers are important to the accurate description of carbohydrate ring conformational properties, as well as the rotational properties of glycosidic linkages. The overall accuracy of the derived torsion terms is indicated in the <Error>$_{curve}$ and <Error>$_{minima}$ of 0.56 and 0.35 kcal/mol, respectively. In methoxymethane, GLYCAM06 yielded a rotational barrier for the H1-CG-OS-CG torsion of 2.31 kcal/mol, in reasonable agreement with the B3LYP/6-31++G(2d,2p) and experimental values[191] of 2.43 and 2.62 kcal/mol, respectively.

The OS-CG-CG-OS torsion term is particularly important for glycans having a 1→6 linkage. The simplest molecule exhibiting this torsion term is 1,2-dimethoxyethane (DME). A single V$_2$ coefficient of 0.82 kcal/mol for this term led to good overall reproduction of the gas-phase data for the entire set of diethers, with a value of <Error>$_{curve}$ = 0.30 kcal/mol (Table 2.5) for the OS-CG-CG-OS and associated parameters. This term is also significant in branched glycans, which may exhibit glycosidic linkages between vicinal hydroxyl groups within the glycan ring. Incorrect parameters for this term may alter the conformational stability of the glycan ring. To quantify the behavior of the parameters for this linkage, we examined the rotational properties of the central C$_{sp3}$-C$_{sp3}$ (CG-CG) bond in 2,3-dimethoxybutane. As can be seen from the data in Figure 2.6, the parameters well reproduced the QM data for 2,3-dimethoxybutane.

To assess further the applicability of the parameters for the OS-CG-CG-OS sequence to solution phase studies, the conformer populations of DME from a 20-ns explicit solvent MD simulation were compared to those determined from solution phase experimental data. The conformers were characterized by the rotational preferences of the dihedral angles composed of the C$_{sp3}$-O$_{sp3}$ (CG-OS) bonds, and the central CG-CG bond; with T and G denoting the *trans* and

gauche conformations, respectively. Theoretical[192-194] and experimental[197-199] techniques have been employed to elucidate these conformational properties in aqueous solutions at different DME concentrations, and in the gas phase. There was a general preference for the G rotamer around the CG-CG bond, and for the T rotamer around the CG-OS bonds, with the major conformers identified as the TTT, TGT, TGG', TGG and TTG.[195] Notable, were the observations that the populations of the various conformers differed in aqueous solutions relative to those present in neat liquid or in the gas phase. The populations also displayed a dependence on temperature. In the gas phase, the TTT, TGT and TGG' were the major rotamers, while in neat liquid and aqueous solutions, the TGT and TGG were predominant,[195,196] indicating that DME-DME and DME-H_2O interactions influenced rotamer populations. Thus our 20-ns MD-derived conformer populations were compared to those observed experimentally at low DME concentrations. The populations determined from the MD simulations could be ranked as: TGT (56%), TGG (13%), TTT (11%), TGG' (10%), TTG (6%) and others (4%), which concurred qualitatively with experimental observations that the TGT and TGG predominate.[195,196]

2.3.3.4. Anomeric Linkages

Anomeric linkages, represented by the C-O-C-O torsion angle common to all carbohydrates, are often modeled in pyranosides by tetrahydro-2-methoxy-2*H*-pyran (axial) and tetrahydro-2-methoxy-2*H*-pyran (equatorial), corresponding to α- and β-pyranosides, respectively.[114,197] In GLYCAM_93, reproduction of the C-O-C-O rotational energies for the α- and β-linkages was accomplished by using unique torsion parameters for each anomer, which were distinguished by different atom types for the anomeric carbon atom. Here we use a single parameter for both anomeric configurations. Both GLYCAM06 and GLYCAM_93 correctly reproduced the B3LYP/6-31++G(2d,2p) rotational energy profiles, with <Error>$_{curve}$ values of

0.45 kcal/mol (GLYCAM06) and 0.3 kcal/mol (GLYCAM_93), for the α-glycosides, while for the β-glycosides the <Error>$_{curve}$ values were 0.36 kcal/mol (GLYCAM06) and 0.57 kcal/mol (GLYCAM_93). Therefore, employing a torsion term common to both α- and β-glycosides did not lead to degradation in the fit to the relevant rotational energy curves. As depicted in Figure 2.7, GLYCAM06 reproduced the rotational topologies for both α- and β-anomers, showing good quantitative agreement with the B3LYP/6-31++G(2d,2p) energies. The HF/6-31G* rotational energy curve for tetrahydro-2-methoxy-2*H*-pyran (axial) displayed a global minimum at 60°, and a less stable local minimum at 300°, while for tetrahydro-2-methoxy-2*H*-pyran (equatorial), the global minimum was at 300° with a local minimum 5 kcal/mol higher at approximately 60°. Three of these minima are consistent with the minima based on expectations due to the stabilizing *exo*-anomeric effect,[37] and the repulsive steric interactions. However, the minimum of tetrahydro-2-methoxy-2*H*-pyran (axial) located at 300° has not been observed experimentally, even though it displays a stabilizing *exo*-anomeric effect, presumably due to large steric repulsions between the hydrogen atoms of the aglycon and that of C2. This structure was subjected to a subsequent optimization at the MP2/6-311++G(2d,2p) level of theory, during which it collapsed to the global minimum energy structure with the C-O-C-O torsion of approximately 60°.

2.3.3.5. Amides and Esters

Amide and ester parameters are relevant to a carbohydrate force field since many monosaccharides contain *N*- and *O*-acetyl moieties. In addition, *N*-linked glycans are attached via an amide bond to asparagine side chains in glycoproteins.

The exact extent of deviation from planarity in the ground-state amido groups, and the origin of the large barrier to rotations about the N_{sp2}-C_{sp2} (N-C) bond in amides have been the

subject of an extensive study.[198] From their highest level calculation (CCSD(T)/PVTZ) Fogarasi *et al.* found that the optimized structure of methanamide was exactly planar. It is accepted that the large barrier to rotation is associated with the breaking of π-delocalization at 90˚. Langley *et al.*[199] observed that the rotational barrier is sensitive to which proton is rotated and the out-of-plane distortion of the proton not constrained by the torsional rotation. When either H-N-C-O torsion angle for one proton is driven above 90˚, the unconstrained proton lags behind and follows an out-of-plane, rather than a purely torsional, energetic pathway.[199] Consequently, in acetamide the two different H-N-C-O dihedral angles were both varied counterclockwise from 0˚ to 90˚, and from 180˚ to 90˚, resulting in two transition states at the 90˚ angle having rotational barriers of 13.96 kcal/mol and 10.87 kcal/mol, corresponding to the 0→90˚ and 180→90˚ paths, respectively. In both transition states the nitrogen atom is pyramidal. In the lower energy state the unconstrained proton is closer to the carbonyl oxygen ($H_{unconst}$-N-C-O = 22°), while in the higher energy state this proton is farther from the oxygen atom ($H_{unconst}$-N-C-O = -154°). The rotational energies in GLYCAM06 were in line with the B3LYP/6-31++G(2d,2p) values, although GLYCAM06 underestimated the barrier heights by approximately 2.0 kcal/mol in each case, Figure 2.8. In general, the amide parameters performed well, giving $<Error>_{curve}$ = 0.64 kcal/mol and $<Error>_{minima}$ = 0.38 kcal/mol.

To determine the torsion terms relevant to substituted amides, the rotational energy curves of *N*-methylacetamide and *N,N*-dimethylacetamide were utilized. The latter molecule enabled the derivation of a rotational profile consisting solely of the two coupled torsion terms (CG-N-C-O/CG-N-C-CG), while *N*-methylacetamide allowed us to examine the transferability of the CG-N-C-O/CG-N-C-CG and the H-N-C-O/H-N-C-CG coupled torsion pairs. As it is impossible to separate these terms, coupled through a 180° phase-shift, the energy contributions

were derived by simultaneously fitting to the QM-derived rotational energy curve, employing torsional angles of θ° (CG-N-C-O) and θ+180° (CG-N-C-CG), in a non-linear least-squares approach. For *N,N*-dimethylacetamide GLYCAM06 gave the rotational barrier about the N_{sp2}-C_{sp2} bond to be 14.6 kcal/mol, which is lower than the B3LYP/6-31++G(2d,2p) value of 16.9 kcal/mol. However, GLYCAM06 is in better agreement with the NMR determined[200] effective barrier of 15.3 kcal/mol, Figure 2.8. The transferability of the GLYCAM06 parameters was illustrated by the reproduction of the B3LYP/6-31++G(2d,2p) rotational energy curve for the N_{sp2}-C_{sp2} bond in *N*-methylacetamide, in which the *trans* rotamer was correctly predicted to be less stable than the *cis* rotamer, Figure 2.8. *N,N*-dimethylacetamide was also used to parameterize the CG-N-CG-H1/C-N-CG-H1 coupled torsion pair. The C-N-CG-H1 torsion is present in *N*-acetylated carbohydrates, and its parameters are relevant to the C-N-CG-H2 linkage present in *N*-linked sugars. GLYCAM06 gave a rotational barrier of 0.64 kcal/mol for the N-CG bond, which may be compared with the B3LYP/6-31++G(2d,2p) value of 0.59 kcal/mol.

Esters were employed for generating parameters necessary for extending the force field to model the properties of *O*-acetylated carbohydrates such as those common in certain bacterial capsular polysaccharides, and in the sialic acid residues of higher animals.[201] The overall <Error>$_{curve}$ computed for the parameters of the following atomic sequences H1-CG-OS-C, HC-CG-C-OS, CG-OS-C-O, CG-OS-C-CG and CG-CG-OS-C were 0.03, 0.02, 0.18, 0.62 and 0.45 kcal/mol, respectively.

The GLYCAM06 rotational energy barrier for the *O*-methyl group (H1-CG-OS-C) in methyl acetate was 0.91 kcal/mol which can be compared to the B3LYP/6-31++G(2d,2p) and microwave[202] values of 0.87 and 1.22 kcal/mol, respectively. The negligible rotational energy barrier for the acetyl methyl group (HC-CG-C-OS) of 0.25 kcal/mol was also in good agreement

with both B3LYP/6-31++G(2d,2p) (0.23 kcal/mol) and experimental (0.29 kcal/mol) values.[202] In esters, the CG-OS-C-O and CG-OS-C-CG dihedral angles are coupled through a 180° phase-shift. The parameters for these atomic sequences were derived by following the same simultaneous fitting approach employed in developing the coupled parameters (CG-N-C-O/CG-N-C-CG) in amides, with torsional angles of θ˚ (CG-OS-C-O) and θ+180˚ (CG-OS-C-CG). This simultaneous fit resulted in V_2 values of -3.20 and 3.0 kcal/mol for CG-OS-C-O and CG-OS-C-CG atomic sequence, respectively. The GLYCAM06 rotational energy curve for the CG-OS-C-O sequence resulted in a local minimum at 180˚, which was 7.52 kcal/mol less stable than the global minimum at 0˚. This relative energy underestimated the IR spectroscopy[203] value of 8.5 ± 1.0 kcal/mol, but was comparable to the B3LYP/6-31++G(2d,2p) value of 7.48 kcal/mol.

Lastly, *N-tert*-butylacetamide and *N*-(tetrahydro-2*H*-pyran-3-yl) acetamide were utilized to determine the coupled torsion parameters (CG-CG-N-H/CG-CG-N-C) and (H1-CG-N-H/CG-CG-N-C), while ethyl acetate and tetrahydro-2*H*-pyran-3-yl acetate were employed in deriving the C-OS-CG-CG parameters, Figure 2.9. The B3LYP/6-31++G(2d,2p) rotational energy curve for the CG-CG-N-C sequence in *N-tert*-butylacetamide resulted in a three-fold 120° periodic profile, which GLYCAM06 reproduced with <Error>curve = 0.05 kcal/mol. For this rotation in tetrahydro-2*H*-pyran-3-yl acetate, however, both GLYCAM06 and the B3LYP/6-31++G(2d,2p) energies resulted in asymmetric profiles with a broad low energy global minimum at 90°, which was approximately 1.75 kcal/mol more stable than a local minimum at 300°. The minimum energy structure at 90° is consistent with the conformational properties of the acetamido moiety in the solid-phase structures of Neu5Ac methyl ester and its monohydrate,[59,204] as well as the solution-phase NMR structure of β-D-2-deoxy-2-*N*-acetylglucopyranoside,[205] in which the carbonyl group almost eclipses the ring C-H bond. This spatial orientation results in an

approximate *trans* dihedral angle between the amide proton and the ring aliphatic hydrogen atom in the H1-CG-N-H sequence. It should be noted that at the global minimum of the CG-CG-N-C sequence being driven, the H1-CG-N-H dihedral angle is approximately at the *trans* conformation. The ability of these parameters to model their corresponding rotational energy curves was reflected in the $<Error>_{curve}$ and $<Error>_{minima}$ values of 0.52 and 0.08 kcal/mol, respectively, Figure 2.9. However, during a 10-ns MD simulation of *N*-acetyl 2-deoxy-*N*-acetyl-β-D-glucopyranoside employing the QM-derived parameters, rotation about the C_{sp3}-N_{sp2} (CG-N) bond to the local minimum ($\tau_{C-N-C2-H2} \approx 180°$) occurred after approximately 1ns. This rotamer remained stable throughout the remainder of the simulation. A very recent experimental and QM-NMR study of the conformational properties of this group in GlcNAc indicated that the H-C_{sp3}-N_{sp2}-H atomic sequence exists only in the *trans* low energy conformation($\tau_{C-N-C2-H2} \approx 0°$), but displays broad librations around that minimum.[206] Therefore, to increase the stability of the H1-CG-N-H *trans* rotamer, and ultimately prevent this conformational transitions from occurring, V_1 terms were introduced for the H1-CG-N-H and H1-CG-N-C parameters to determine a reasonable value. The V_1 coefficients were varied from 0.0 to 2.0 kcal/mol in increments of 0.5 kcal/mol, and PMF calculations were performed for each value, employing *N*-(tetrahydro-2*H*-pyran-3-yl) acetamide with the C1-C2-N-C dihedral angle as the reaction coordinate. Based on the approximately 2 kcal/mol increase in the relative energy of the local minimum, Figure 2.9, the pair of values 1.0 and -1.0 kcal/mol was selected for the H1-CG-N-H and H1-CG-N-C torsion parameters, respectively. Employing these values, a 50-ns MD simulation performed with *N*-acetyl 2-deoxy-*N*-acetyl-β-D-glucopyranoside did not display any conformational transitions of the acetamido moiety. The origin of the observed solution-phase transitions, when employing the purely QM-derived torsion terms, remains undetermined.

In the case of ethyl acetate, rotating the about the C_{sp3}-O_{sp3} bond in the atomic sequence (CG-CG-OS-C) resulted in a symmetric B3LYP/6-31++G(2d,2p) energy curve, which produced a large barrier (7.45 kcal/mol) at 0˚, with minima at the *trans* configuration, as well as at 90° and 270°. A negligible QM-determined rotational energy barrier (0.83 kcal/mol) separated the global *trans* minimum from the local minima at 90° and 270°. These results were consistent with previous theoretical studies of ethyl acetate.[207] GLYCAM06 reproduced the B3LYP/6-31++G(2d,2p) relative energies over the entire curve with $<Error>_{curve}$ = 0.45 kcal/mol. As for the CG-CG-OS-C rotation in tetrahydro-2*H*-pyran-3-yl acetate, both B3LYP/6-31++G(2d,2p) and GLYCAM06 produced similar asymmetric energy profiles, Figure 2.9. The $<Error>_{curve}$ for this atomic sequence was 0.54 kcal/mol. To assess the extent to which GLYCAM06 could model other *O*-acetylated pyranosyl moieties, rotations were performed around the C_{sp3}-O_{sp3} bond in both tetrahydro-2-methoxy-2*H*-pyran-3-yl acetate and (tetrahydro-2*H*-pyran-2-yl)methyl acetate, Figure 2.10. Unlike the rotational energy profile of tetrahydro-2*H*-pyran-3-yl acetate, the energy profile of tetrahydro-2-methoxy-2*H*-pyran-3-yl acetate resulted in a barrier to rotation at approximately 30°, which could arise from Coulombic repulsions between the carbonyl oxygen and the oxygen atom of the methoxy group. Compared to the QM-value, GLYCAM06 underestimated this repulsion by about 2.31 kcal/mol. Overall, there was a good reproduction of the B3LYP/6-31++G(2d,2p) data ($<Error>_{curve}$ values of 0.89, and 0.50 kcal/mol, respectively).

2.3.3.6. Carboxylates

Under physiological conditions, the extent to which carboxylate groups are ionized depends on their pKa. In the case of the ionic state it should be noted that it will never exist in the absence of a counter ion, such as sodium. It is generally presumed that uronic and ulosonic acids exist in the ionized form, and parameters were derived accordingly, with exceptions noted

below. The B3LYP/6-31++G(2d,2p) rotational energy curve about the C_{sp2}-C_{sp3} (C-CG) bond in 2-methylpropanoate displayed minima at 120˚ and 300˚. For this rotation, GLYCAM06 was fit to the MP2/cc-pVDZ rotational energies, Figure 2.11, with <Error>$_{curve}$ and <Error>$_{minima}$ values of 0.72 and 0.60 kcal/mol, respectively. The QM-determined rotational energy barrier of the carboxyl group was 0.8 kcal/mol. This is a very low rotational barrier, which is less than the room temperature equipartition energy of a molecule (kT). Thus during an MD simulations, free rotations of the carboxylic group about the C-CG bond are expected. To this extent, it was more important to ensure reasonable barriers to rotation, than to identify the minima.

In addition to the O2-C-CG-CG torsion term, modeling uronic acids also required the development of parameters for the OH-CG-CG-C and OS-CG-CG-C sequences. The OH-CG-CG-C parameters were derived using 2-hydroxyproponoate as the model compound, by simultaneously fitting to two sets of curves: with and without the presence of an internal hydrogen bond between the hydroxyl and carboxyl groups. For the OH-CG-CG-C rotations, the <Error>$_{curve}$ was 0.79 kcal/mol and <Error>$_{minima}$ was 0.31 kcal/mol (with the inclusion of intramolecular H-bond), while <Error>$_{curve}$ and <Error>$_{minima}$ were 0.96 mol^{-1} and 0.16 kcal/mol, respectively (when the H-bond was disallowed). The <Error>$_{curve}$ and <Error>$_{minima}$ for the OS-CG-CG-C rotations were 1.15 and 0.09 kcal/mol, respectively.

2.3.3.7. Mixed Functional Groups

Mixed functional groups that are present in carbohydrates include the alcohol-ethers, alcohol-amides, and ether-amides. Alcohol-ethers are required in the derivation of torsion terms that model the torsion angles involving the ring oxygen and the C2 hydroxyl group, as well as the glycosidic ether oxygen and other ring hydroxyl groups. These molecules with mixed functionalities can form intramolecular H-bonds, hence rotational energy curves were derived in

69

the presence and absence of internal H-bonds, as appropriate. The B3LYP/6-31++G(2d,2p) and GLYCAM06 OH-CG-CG-OS rotational energy curves for 2-methoxybutan-3-ol are shown in Figure 2.12. Overall for ether-alcohols, the $<Error>_{curve}$ was 0.61 kcal/mol with a value of 0.59 kcal/mol for the $<Error>_{minima}$.

The highly polar mixed functionality of ether-amides was challenging to parameterize (Figure 2.S1, supplementary materials), as indicated in the overall values of $<Error>_{curve}$ = 1.12 kcal/mol and $<Error>_{minima}$ = 1.10 kcal/mol. In terms of carbohydrates, the OS-CG-N-C parameter is relevant to the modeling carbohydrate-protein linkages. In this set of parameters, the OS-CG-N-C parameter was the least accurate with $<Error>_{curve}$ = 1.93 kcal/mol. The other ether amide parameters performed better as can be seen in Table 2.5.

In order to determine the torsion parameters associated with the OH-CG-CG-N rotation, molecules having both alcohol and amide functionality were employed. The overall values for $<Error>_{curve}$ and $<Error>_{minima}$ were 0.49 and 0.41 kcal/mol, respectively. The performance of these parameters is exemplified by *N*-ethanolacetamide as shown in Figure 2.13.

2.3.3.8. Special Cases

2.3.3.8.1. Ulosonic acids

For GLYCAM06 to correctly model the glycosidic conformational properties of ulosonic acids such as 5-*N*-acetylneuraminic acid (Neu5Ac), a class of sugars frequently present at the termini of eukaryotic cell surface glycans, accurate parameterization of the anomeric angle φ ($C_xC_xO_{x-1}C_{x-1}$) is required, Figure 2.3. Unlike most glycosyl residues, ulosonic acids contain a carboxyl functional group attached to the anomeric carbon atom, which alters the rotational preferences for the φ-angle in this class of compounds. In solution, the φ-angle predominantly exists in two conformations, -*gauche* and *trans*, in an approximate 1:1 ratio.[56,57] The torsion term

(C-CY-OS-CG) associated with the carboxylate group for Neu5Ac contributes significantly to the φ-angle conformation in these carbohydrates. Generating parameters for this torsion term proved challenging; exhibiting ether-carboxylate mixed functionality. Initial MD simulations of the glycoside of methyl α-Neu5Ac employing parameters generated by fitting to gas-phase QM rotational energy profiles of either ionized or protonated 2-methoxypropanoate resulted in the sampling of only the -*gauche* rotamer about the φ-angle of methyl α-Neu5Ac. The inability of the initial parameters to reproduce experimental φ populations may originate from the omission of charge polarizabilities in the force field. Thus, we took the approach of empirically varying the V_1 coefficient of this term until we obtained the proper rotamer distribution for φ from explicitly solvated simulations of methyl α-Neu5Ac. By changing the V_1 value from 1.0 to 8.0 kcal/mol it was possible to shift the preference of the sampling from predominantly –*gauche* to exclusively *trans*. An intermediate value of 3.5 kcal/mol reproduced the experimental rotamer distribution, Figure 2.14. Employing this V_1 term, a PMF analysis was carried out to determine the energy profile of the φ-angle in methyl α-Neu5Ac in solution, Figure 2.14. The PMF results indicated that the -*gauche* and *trans* conformers now displayed similar relative energies in solution, with a stabilization energy of approximately 2.3 kcal/mol over the +*gauche*. The rotational energy barrier between the -*gauche* and *trans* conformers was approximately 1.4 kcal/mol. The small barrier to rotation results in the frequent transitions between the two rotamers, as observed in the MD simulation trajectory, Figure 2.14. To assess whether this new torsional coefficient could model a mixed system of ionized and neutral carboxylic acid groups, the GLYCAM06 rotational energy curve for the rotation around the exocyclic C_{sp3}-O_{sp3} (C_1-O_1) bond was compared to the B3LYP/6-31++G(2d,2p) curves for both ionized and neutral (*R*)-tetrahydro-2-methoxy-2*H*-pyran-2-carboxylic acid, Figure 2.15. Notably, the GLYCAM06

rotational behavior appeared to be a compromise between each of the QM curves in ionized and neutral (R)-tetrahydro-2-methoxy-2H-pyran-2-carboxylic acid.

2.3.3.9. Force field validation

In the course of developing torsion parameters for molecules in the training set, most of the parameters were validated concurrently by determining how well GLYCAM06 was able to reproduce the B3LYP/6-31++G(2d,2p) rotational energy curves for test molecules, or by performing MD simulations and comparing MD-computed rotamer populations to those observed experimentally. In this section we extend the validation by examining how the parameters collectively perform in selected glycans.

Unlike the earlier parameterization of GLYCAM,[114] which was based on carbohydrate-specific analogues, only small representative molecules of a broad diversity were employed in the derivation of GLYCAM06. Therefore, validation of the force field by comparison to carbohydrate molecules provides an independent assessment of the ability of this small-molecule-based approach to model carbohydrate properties. Explicit solvent MD simulations (50 ns) were performed for α-D-GlcpOMe, α-D-GalpOMe and α-D-ManpOMe, and compared to NMR experimental data, such as rotamer populations and scalar ^3J-couplings. While scalar ^3J-couplings can rarely provide a complete picture of the structural properties of carbohydrates, when combined with nOe distances both datasets can provide valuable insights about oligosaccharide conformational properties. Average ring puckering parameters were also computed from the MD simulation data and compared to solid phase neutron diffraction data. The comparison of ring puckering parameters between gas-phase and crystallographic data is not ideal. However, Momany et al.[208] have shown that the geometrical properties of pyranosyl rings from gas-phase B3LYP/6-31G* calculations were remarkably close to those determined in the

solid-phase, suggesting that the packing environment may not have a significant effect on the geometry of the ring.

In solution, the rotamer preferences of the ω-angle (O_5-C_5-C_6-O_6) involving the exocyclic hydroxymethyl group are greatly influenced by the stereochemistry at the C4 position and the polarity of the solvent[40]. The three accessible staggered rotamers about the C_5-C_6 bond are generally designated *gauche-gauche* (*gg*), *gauche-trans* (*gt*) and *trans-gauche* (*tg*) with respect to the dihedral angles O_5-C_5-C_6-O_6 and C_4-C_5-C_6-O_6, respectively. A significant test of a carbohydrate force field is its ability to reproduce the rotamer populations about the C_5-C_6 exocyclic bonds because the populations are sensitive to the energetic balance between the formation of intramolecular and solute-solvent hydrogen bonds.[209] The ability of GLYCAM06 to reproduce the NMR *J*-couplings for the C_5-C_6 bond is reflective of the performance of its dihedral, non-bonded and electrostatic components.

Homonuclear scalar ^3J-couplings between the H_5-C_5-C_6-H_{6R}/H_{6S} hydrogen atoms from a 50-ns explicit solvent MD simulation of α-D-Glc*p*OMe were computed, employing an experimentally parameterized Karplus equation.[210] Values computed over the entire trajectory resulted in coupling constants of 5.4 ± 1.7 Hz and 2.9 ± 2.0 Hz for $^3J_{H5H6R}$ and $^3J_{H5H6S}$, respectively, which compared favorably with experimental values,[48-50] Table 2.7. In the course of the MD simulation, rotations about the exocyclic C_5-C_6 bond resulted in each of the rotational states of the ω-angle being populated. The MD *gg*:*gt*:*tg* populations, 62:36:2, were in good agreement with the experimentally observed populations of 57:38:5[50] and 53:47:0.[48]

Similar *J*-calculations were performed for α-D-Gal*p*OMe, the C-4 epimer of Glc*p*. From the MD data, the $^3J_{H5H6R}$ and $^3J_{H5H6S}$ coupling constants were 7.9 ± 1.6 Hz and 3.7 ± 1.8 Hz respectively. Several experimental values have been reported for the $^3J_{H5H6R/S}$ coupling constants,

and the *gg:gt:tg* populations of α-D-Gal*p*OMe, Table 2.7. The MD-computed values for $^3J_{H5H6R}$ and $^3J_{H5H6S}$ fell within the experimental ranges of 7.8-8.6 Hz and 3.7-6.0 Hz, respectively.[48-50] With respect to the *gg:gt:tg* populations, the MD values of 6:76:18 showed good agreement with the more recent data.

In addition to bond rotational properties, pyranose ring puckering parameters, ring inter-proton torsion angles and distances provide another means of assessing the accuracy of the ring geometry. Analyses of the Cremer and Pople[211] ring puckering parameters over the 50-ns MD simulations of α-D-Glc*p*OMe, α-D-Gal*p*OMe and α-D-Man*p*OMe indicated stable $^4C_1$ ring chair forms, with (Q,θ) values of (0.55 ± 0.04, 6 ± 10), (0.57 ± 0.04, -4 ± 8) and (0.55 ± 0.04, 1 ± 10), respectively, which were in good agreement with the values determined from solid state neutron diffraction experiments of (0.57, 2.3),[212] (0.57, 4.9)[213] and (0.56, 0), respectively.[212]

Ring inter-proton distances and dihedral angles are a source of NMR-restraints employed in determining the ring geometries of glycans. In addition, the distances provide internal calibration references employed in the elucidation of glycan conformational properties from nOe data. Thus, the feasibility of employing GLYCAM06 to predict glycan ring conformational properties was determined by its ability to reproduce experimental inter-proton distances and dihedral angles. A comparison of inter-proton distances and dihedral angles computed using GLYCAM06, and those determined from neutron diffraction data[228,229] for α-D-Glc*p*OMe, α-D-Gal*p*OMe and α-D-Man*p*OMe is presented in Table 2.8. In order to make a direct comparison between experimental and theoretical data, the coupling constants between vicinal ring-hydrogen atoms were computed from solvated MD simulations of α/β-D-Glc*p*OMe, α/β-D-Gal*p*OMe, and α/β-D-Man*p*OMe, and compared with available solution-phase NMR experimental data, Table 2.9. The overall agreement between the GLYCAM06-derived and experimental data are quite

good. However, theoretical and experimental values for the $^3J_{H4H5}$ and $^3J_{H1H2}$ constants in α/β-D-GalpOMe and β-D-ManpOMe, respectively, were in weaker agreement. Superimpositions of the crystal structures with energy-minimized and average ring geometries of these pyranosides, computed from the MD simulations did not reveal any structural distortions (Figure 2.S2, supplementary materials). This error may be arising from electronic hyperconjugative effects between the lone pair of electrons on the ring oxygen atom and the antibonding orbital of the C5-H5 or the C1-H1 bonds when the anomeric center is in the β-configuration, which are not included in classical force fields. Using the B3LYP functional with the cc-pV5ZT and 5s2p1d basis sets for hydrogen and heavy atoms, respectively, QM NMR predictions from the B3LYP/aug-cc-pVTZ optimized neutron diffraction structures of α- and β-D-GalpOMe resulted in $^3J_{H4H5}$ values of 1.9 and 1.2Hz, respectively.

Another indication of the performance of a biomolecular force field is its ability to reproduce experimentally determined unit cell geometries. In general, unit cell dimensions are influenced by internal torsional rotations, as well as by van der Waals and electrostatic intermolecular interactions. However, it is recognized that in polar molecules such as carbohydrates, the treatment of electrostatics is crucial in reproducing experimental crystal geometries.[129] The suitably of the present treatment of electrostatics in GLYCAM06 was assessed by comparing the average difference between the unit cell geometries of a methyl α-D-Glcp crystal computed from a 1-ns MD simulation to those determined experimentally.[144] The average difference computed with GLYCAM06 was 1.2 Å, which showed better performance when compared with the average differences computed with GLYCAM_93,[114] CHARMM HGFB,[214] and GROMOS (93)[215] of 1.7 Å, 1.8 Å and 1.6 Å, respectively.

The results of the frequency calculations, assignments, and comparisons to the various experimental results are summarized in Figure 2.16. The character of the assignments generated by our program is illustrated in Table 2.10, where we have listed the primary motion assigned by DMODES to each vibrational frequency for the monomer calculated by NMODE. The notations are described in the caption. The assignments for the monomer were also checked by visually inspecting corresponding plots of the atom positions with energy-weighted motion vectors. For the comparisons with experiment, the DMODES-generated assignments were grouped into the broader classes used in Figure 2.16, Figure 2.17 and Table 2.11.

In general, the agreement between the calculated frequencies and the experimental results is good, with the only notable exceptions being the H—O stretches, which are about 5% blueshifted. This deviation is likely due to the use of the gas-phase H—O stretching frequency for methanol in the initial derivation of the AMBER HO—OH stretching constant.[216] Since the gas-phase H—O stretching frequency of methanol is about 360 cm^{-1} higher than the liquid phase value, it is reasonable to expect that a smaller force constant would produce better results in the condensed phases.

We have chosen to use the standard AMBER values for the H—O stretches, and also the H—C, for two reasons. First, we wish to minimize the introduction of new atom types, and associated complications, into the AMBER force field. Second, fitting force constants to quantum-derived separation energies for these bond stretches does not improve the function of the force field. Instead, the quantum derived force constants increase the normal mode frequency disagreement to 10–20%. However, although the current values work well, the AMBER force field is used primarily for condensed-phase simulations, so it may be valuable to reevaluate the use of the current HO—OH stretching constant.

Although the O–C stretches seem in Figure 2.16 to differ by ~50%, the issue is more complex, and is partly one of assignment. As with the experimental work, we also see O–C stretches in the region ~1000–1200 cm21, as is illustrated for the 64-mer in Figure 2.17. However, according to our molecular mechanics analysis of the monomer and the 64-mer, as well as the quantum analyses of the monomer (Tables 2.10 and 2.11 for the monomer data), the O–C stretches do not account for the major portion of the vibrational kinetic energy at these frequencies. In Figure 2.17, we present all frequencies calculated for the crystal that corresponded to a C–C or O–C stretch. The figure shows that the calculation yields many C–C and O–C stretches in the range of ''heavy atom stretches'' given by Wells and Atalla.[152] The more intense stretches at the lower frequencies are all parts of larger, more complex motions that would not typically be assigned as stretches, so frequencies in these ranges are not expected to correspond to experiment. On the other hand, the presence of the cluster of high-frequency C–C stretches, about 1400–1600, is puzzling. The cause of the discrepancy is not known. Kuttel *et al*.,[149] reported similar calculated frequencies, although theirs also included O–C stretches. They attributed the high frequencies to the influence of ring-atom bond and angle force constants that are not mitigated by the presence of cross terms in the force field. Although our outlying frequencies might have a similar cause, further investigation is needed to determine the precise cause of the discrepancy.

The agreement with experimental results for bending, distortion, and torsion modes, on the other hand, is very good (Figure 2.16). Our calculated ranges significantly overlap experimental ranges for these motion classes, though the experimental ranges do not always agree with each other. For example, consider the data for the H–O-C torsions. These show the least consensus among the other studies. This may in part be due to difficulties translating

between assignment conventions, but it may more likely represent a wide variability in those torsional frequencies, as is found for the 64-unit calculation. Similar arguments might be made for the ''other C6 & O6'' class. It should be noted that the Gregurick and Kafafi[151] study used a different assignment method than what was used in the other studies to which we compare our results, which might account for the difference between their results and those of the others within these classes.

Two features in Figure 2.16 that are specific to the results of this study require further discussion. The first is the presence of the highest frequency –C6– distortion modes. There were no assignments offered by the other studies that satisfactorily correlated with this motion, which is essentially the movement of the C6 atom along a line between the C5 and O6 atoms. There is accompanying motion in C5 and in the two hydrogens attached to C6, but the primary energy is in the motion of the C6. Note that this discussion only applies to the –C6– distortion modes above 1500 cm^{-1} (the lower frequency modes are other stretching motions of the C6 atom, for example, along the bond to one of the aliphatic hydrogens). Considering the strong stretching character of the motion, it is likely that the frequency reported here is only approximate. Normal mode analysis of the HF/6-31G* data for the monomer revealed a similar mode at 1635 cm^{-1} (Table 2.11). The density functional calculation generated a similar mode at 1458 cm^{-1}. This mode is likely to be real, but may be difficult to observe since the motion is highly symmetric. In all of the quantum calculations in which the vibrational kinetic energy was primarily coupled to the motion of C6 between C5 and O6, the IR intensity calculated for that mode was small.

The other feature of note is the small cluster of very high frequency H—O-C torsions noted with (*) in Figure 2.10. Note that the relative energy associated with the motion is small, but it was not small enough to be filtered out by the procedures described earlier. The motions

responsible for these anomalous assignments make physical sense, and are likely to occur in nature, but may be too weak to be observed. The cluster of torsions arises due to an accidental resonance. In the 64-unit crystal used for the calculation, the minimization resulted in a configuration such that a stretch of the O2-HO2 of a single molecule would necessarily excite torsion of the hydroxyl HO6 of another neighboring molecule. Since one of the molecules was situated at one a face of the crystal, and since NMODE does not employ periodic boundaries, the isolated behavior was not representative of the bulk properties.

2.4. Conclusions

This work describes the derivation of a parameter set for classical quadratic force fields that accurately models carbohydrates, but which can also be generalized to more diverse molecules by virtue of the non-carbohydrate specific nature of the parameters. QM calculations were employed to compute properties that are difficult or impossible to access experimentally, such as, bond and valence angle deformation force constants, dihedral angle rotational barriers, and electrostatic properties. The structures of the biomolecular building blocks (monosaccharides in the present case) were taken from experimental neutron diffraction data. Consistent with the AMBER force field, the PARM94 parameter set for van der Waals terms was employed. Particular attention was given to those properties that most impact the overall 3D structure and dynamics of biopolymers; namely, internal electrostatic interactions, solute-solvent interactions and internal rotational barriers. While we derived partial atomic charges by fitting to the QM molecular electrostatic potentials (so called ESP-fitting), we departed from the general practice of assigning partial atomic charges to every atom in the molecule adopted in our earlier work (GLYCAM_93, GLYCAM2000) by not fitting partial charges to aliphatic hydrogen atoms. In

addition, the current parameterization removes the need to treat 1-4 non-bonded interactions as a special case; that is, 1-4 scaling has been removed.

Although the parameter set is complete for carbohydrates, the methodology and the force field may be readily extended to lipids, proteins and nucleic acids. The generality of the parameters is exemplified by the utilization of a common set of terms for α- and β- carbohydrate anomers. When combined with appropriate charge sets, these common terms provide a means of predicting the relative energies of interconverting ring forms. Not surprisingly, for non-polar molecules the parameters derived by fitting to gas-phase B3LYP/6-31++G(2d,2p) data successfully reproduced the QM rotational energies. In addition, the force field was able to reproduce solid state ring puckering parameters, ring inter-proton torsion angles and distances, solution phase populations of the ω-angle, and scalar ^3J-coupling constants for the H_5-C_5-C_6-$H_{6R/S}$ atoms in representative carbohydrates. In highly polar molecules, however, the percentage errors in the torsional energies were occasionally high, and in the case of carboxylate ions the gas-phase QM-derived parameters failed to reproduce solution phase conformational properties. The expectation that condensed phase simulations, employing dihedral terms fit to gas-phase QM data, can reproduce solution rotamer populations is founded on the assumption that internal electrostatic polarization is relatively independent of internal rotation, either in the gas or condensed phases. This is frequently a reasonable assumption for relatively non-polar molecules, but we found that as molecular polarity increased the observed condensed phase rotamer populations began to deviate from experimental expectations. This is clearly the case for ionic systems, for which it was necessary, in the absence of an explicit treatment of polarizability, to adjust the dihedral terms empirically in order to achieve accurate behavior in the aqueous solution. However, the empirically-corrected rotational energy curves were shown to differ only

modestly from the gas-phase QM data. For such systems an iterative procedure, which involved explicit solvent MD simulations, was adopted to generate the torsion parameters. The inadequacy of the gas-phase QM-derived torsion terms when applied to highly polar molecules may be attributed to the absence of charge polarizability in the classical force field. The development of a polarizable version of GLYCAM is underway.

Table 2.1. Model compounds employed in the development of GLYCAM06

Molecular class	Parameter	Training set	Test set
Hydrocarbons	HC-CG-CG-HC	Ethane	
	CG-CG-CG-CG	Butane	2,3-dimethylbutane, 2-methylbutane and ethylcyclohexane
	HC-CG-CG-CG	Propane	2-methylpropane, 2,2-dimethylpropane and methyl cyclohexane (axial and equatorial)
Alcohols	HO-OH-CG-H1	Methanol	Propan-2-ol
	HO-OH-CG-CG	Ethanol, and propan-2-ol	
	OH-CG-CG-HC	Ethanol, and 2-methylpropan-2-ol	
	CG-CG-CG-OH	Propanol	1,1,2,2-tetramethylpropanol, 2,2-dimethylpropanol, hydroxymethylcyclohexane (equatorial) and butanol
	OH-CG-CG-OH	1,2-ethanediol and 1,2-propanediol	2-methyl-1,2-propanediol, 2,3-butanediol, 2-methyl-2,3-butanediol and 2,3-dimethyl-2,3-butanediol
Ethers	H1-CG-OS-CG	Methoxymethane	Methoxyethane, 2-methoxypropane and 2-methoxy-2-methylpropane
	CG-CG-OS-CG	Methoxyethane and 2-methoxypropane	2-methoxy-2-methylpropane, 2-methoxypropane, 2-*tert*-butoxy-2-methylpropane, methoxycyclohexane (axial and equatorial) and 2-isopropoxypropane
	CG-CG-CG-OS	Methoxypropane and 1-methoxy-2-methylpropane	1-methoxy-2,2-dimethylpropane
	OS-CG-CG-OS	1,2-dimethoxyethane	1,2-dimethoxypropane, 1,2-dimethoxy-2-methylpropane, 2,3-dimethoxybutane, 2,3-dimethoxy-2-methylbutane, and 2,3-dimethoxy-2,3-dimethylbutane
	OS-CG-OS-CG	1,1-dimethoxyethane, 2,2-dimethoxypropane and Tetrahydro-2-methoxy-2*H*-pyran (axial and equatorial)	Tetrahydro-2-methoxy-2*H*-pyran-3-ol (axial and equatorial) and Tetrahydro-2-(tetrahydro-2*H*-pyran-2-yloxy)- 2*H*-pyran

Table 2.1. (continued)

Molecular class	Parameter	Training set	Test set
Amides	N-CG-CG-CG	*N*-(2,33-trimethylbutan-2-yl) acetamide and *N*-propylacetamide	*N-tert*-pentylacetamide and *N-sec*-butylacetamide
	H-N-C-O, HC-CG-C-N	Acetamide	
	H1-CG-N-C/CG	*N, N*-dimethylacetamide, and *N*-isopropylacetamide	
	HC-CG-CG-N	*N-tert*-butylacetamide	
	CG-N-C-O CG-N-C-CG	*N, N*-dimethylacetamide	*N*-methylacetamide, *N*-ethylacetamide, *N*-isopropylacetamide and *N-tert*-butylacetamide
	CG-CG-N-C	*N-tert*-butylacetamide and *N*-isopropylacetamide, *N*-(tetrahydro-2*H*-pyran-3-yl) acetamide	*N*-ethylacetamide and *N*-(tetrahydro-2-methoxy-2*H*-pyran-3-yl) acetamide
Esters	H1-CG-OS-C	Methyl acetate	
	OS-C-CG-HC	Methyl acetate	
	CG-OS-C-O CG-C-OS-CG	Methyl acetate	
	CG-CG-OS-C	Ethyl acetate and tetrahydro-2*H*-pyran-3-yl acetate	tetrahydro-2-methoxy-2*H*-pyran-3-yl acetate and (tetrahydro-2*H*-pyran-2-yl)methyl acetate
Carboxylates	O2-C-CG-CG	2-methyl propanoate	
	OH-CG-CG-C	3-hydroxy propanoate	2-methyl-3-hydroxy propanoate
	HC-CG-C-O2	Acetate	

Table 2.1. (continued)
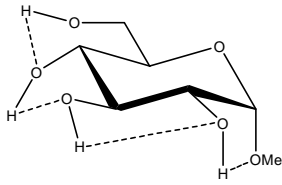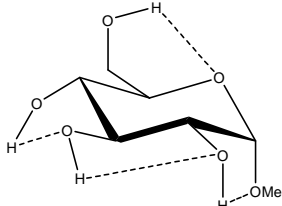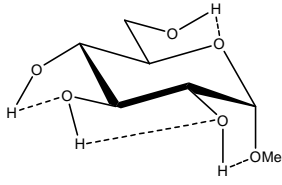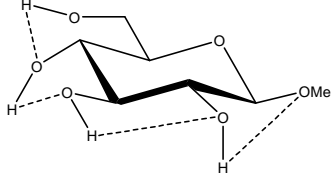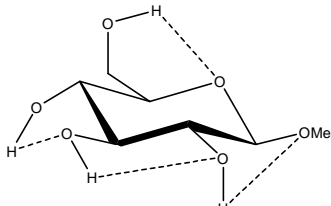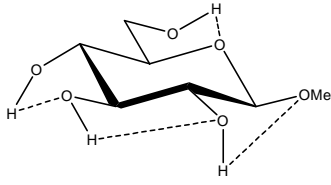
Molecular class	Parameter	Training set	Test set
Ether alcohols	OH-CG-CG-OS	Methoxyethanol and 2-methoxypropanol	2-methoxybutan-3-ol, 2-methoxy-2-methylpropanol, 2-methoxy-2-methylbutan-3-ol and 2-methoxy-2,3-dimethylbutan-3-ol
Ether amides	OS-CG-CG-N	*N*-(1,1-dimethyl-2-methoxyethyl) acetamide and *N*-(1-methoxypropan-2-yl)acetamide	*N*-(1-methoxy-2-methylpropan-2-yl)acetamide and *N*-(1-methoxypropan-2-yl)acetamide
	CG-OS-CG-N	*N*-(2-methoxyethyl)acetamide and *N*-(2-methoxypropyl)acetamide	
Alcohol amides	OH-CG-CG-N	*N*-ethanolacetamide	*N*-(2-hydroxypropyl)acetamide, *N*-(2-hydroxy-2-methylpropyl)acetamide and *N*-(3-hydroxy-2,3-dimethylbutan-2-yl)acetamide
Ether carboxylates	O2-C-CG-OS	2-methoxyacetate, 2-methoxypropanoate, 2,2-dimethoxyacetate, 2-methoxy-2-methylpropanoate and (*R*)-tetrahydro-2-methoxy-2*H*-pyran-2-carboxylic acid (ionized)	
	OS-CG-CG-C	3-methoxypropanoate	2-methyl-3-methoxypropanoate
	C-CG-OS-CG	2-methoxyacetate, 2-methoxypropanoate, 2-methoxy-2-methylpropanoate, (*R*)-tetrahydro-2-methoxy-2*H*-pyran-2-carboxylic acid (ionized) and (*R*)-tetrahydro-2-methoxy-2*H*-pyran-2-carboxylic acid (neutral)	α-5-*N*-acetylneuraminic acid

Table 2.2. Ensemble-averaged partial charge[a] sets for the methyl glycosides of D-Glc*p*, D-Man*p* and D-Gal*p*

Atom	D-Glc*p*OMe		D-Man*p*OMe		D-Gal*p*OMe	
	α	B	α	β	α	β
CMe	0.259 ± 0.01	0.267 ± 0.01	0.275 ± 0.01	0.262 ± 0.01	0.262 ± 0.01	0.264 ± 0.01
OMe	-0.470 ± 0.03	-0.454 ± 0.02	-0.532 ± 0.03	-0.406 ± 0.03	-0.478 ± 0.03	-0.455 ± 0.03
C1	0.527 ± 0.08	0.378 ± 0.08	0.496 ± 0.07	0.262 ± 0.05	0.488 ± 0.09	0.372 ± 0.08
C2	0.246 ± 0.09	0.310 ± 0.10	0.245 ± 0.07	0.241 ± 0.09	0.339 ± 0.10	0.357 ± 0.10
O2	-0.713 ± 0.03	-0.718 ± 0.02	-0.716 ± 0.03	-0.664 ± 0.03	-0.724 ± 0.02	-0.727 ± 0.03
H2O	0.437 ± 0.03	0.437 ± 0.02	0.446 ± 0.02	0.431 ± 0.02	0.431 ± 0.02	0.432 ± 0.02
C3	0.286 ± 0.10	0.284 ± 0.11	0.251 ± 0.11	0.263 ± 0.13	0.253 ± 0.09	0.259 ± 0.11
O3	-0.699 ± 0.03	-0.709 ± 0.03	-0.696 ± 0.03	-0.692 ± 0.04	-0.701 ± 0.02	-0.695 ± 0.02
H3O	0.427 ± 0.02	0.432 ± 0.01	0.431 ± 0.02	0.430 ± 0.02	0.433 ± 0.02	0.433 ± 0.02
C4	0.254 ± 0.12	0.276 ± 0.13	0.326 ± 0.16	0.349 ± 0.15	0.204 ± 0.08	0.203 ± 0.10
O4	-0.710 ± 0.03	-0.714 ± 0.03	-0.732 ± 0.04	-0.746 ± 0.03	-0.673 ± 0.04	-0.664 ± 0.04
H4O	0.436 ± 0.02	0.440 ± 0.02	0.438 ± 0.02	0.443 ± 0.02	0.436 ± 0.03	0.436 ± 0.02
C5	0.283 ± 0.11	0.225 ± 0.12	0.276 ± 0.15	0.206 ± 0.13	0.216 ± 0.09	0.140 ± 0.09
O5	-0.574 ± 0.07	-0.470 ± 0.06	-0.507 ± 0.07	-0.393 ± 0.06	-0.527 ± 0.07	-0.402 ± 0.07
C6	0.276 ± 0.04	0.282 ± 0.044	0.266 ± 0.05	0.273 ± 0.05	0.308 ± 0.05	0.319 ± 0.05
O6	-0.682 ± 0.02	-0.688 ± 0.02	-0.688 ± 0.02	-0.680 ± 0.02	-0.684 ± 0.03	-0.692 ± 0.02
H6O	0.418 ± 0.02	0.424 ± 0.02	0.422 ± 0.02	0.419 ± 0.02	0.418 ± 0.02	0.419 ± 0.02

[a]From 200 snapshots evenly extracted from a 50-ns MD simulation
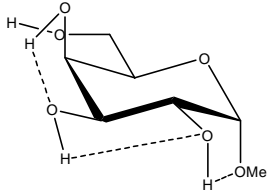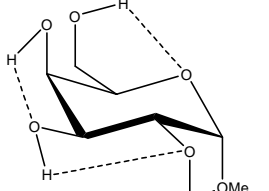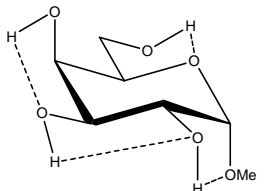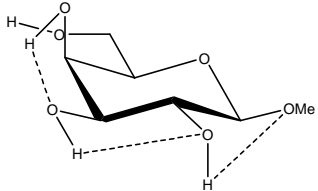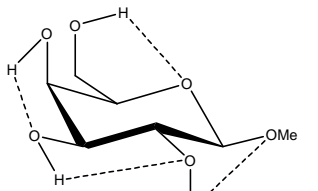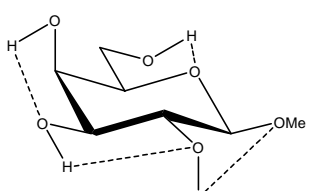
Table 2.3. Influence of Charge Protocol on Conformational Relative Energies (E_{rel}) computed with GLYCAM06 for the α- and β-anomers of D-Glc*p*OMe and D-Gal*p*OMe

| | | E_{rel} | | |
| | | Charge Protocol | | |
Glycoside conformation (D-Glc*p*OMe)	(HF/6-31G*//HF/6-31G*)[a]	EA-RESP[c] \<overall\>	EA-RESP[d] \< α, β\>	EA-RESP[e] \<α\>, \<β\>
α1	0.00[b]	0.00[b]	0.00[b]	4.45[b]
α2	0.11	0.69	0.58	4.97
α3	0.00	0.22	0.26	4.56
β1	1.30	1.12	1.29	0.00
β2	1.23	1.67	1.75	0.5
β3	1.46	1.24	1.44	0.29

[a] Relative energies taken from Ref. 156. Dotted lines indicate hydrogen bonds
[b] Relative to the lowest energy conformation for each glycoside. Units are kcal/mol. Charges computed by averaging the EA charges of: [c] the α- and β-anomers of all glycosides, [d] the α- and β-anomers of each glycoside, [e] each anomer on each glycan.

Table 2.3. (continued)

Glycoside conformation (D-Gal*p*OMe)	(HF/6-31G*//HF/6-31G*) [a]	E_{rel} Charge Protocol EA-RESP[c] \<overall\>	EA-RESP[d] \< α, β\>	EA-RESP[e] \<α\>, \<β\>
α1	1.62	1.76	1.89	13.65
α2	1.18	0.94	0.82	12.86
α3	0.00	0.00	0.00	11.71
β1	3.24	3.18	3.42	1.93
β2	2.90	2.99	2.65	0.77
β3	1.88	1.79	1.58	0.00

[a] Relative energies taken from Ref. 156. Dotted lines indicate hydrogen bonds
[b] Relative to the lowest energy conformation for each glycoside. Units are kcal/mol. Charges computed by averaging the EA charges of: [c] the α- and β-anomers of all glycosides, [d] the α- and β-anomers of each glycoside, [e] each anomer on each glycan

Table 2.4. Force field parameters in GLYCAM06

Bond	K_r[b]	r_{eq}[c]	Bond	K_r	r_{eq}
C –OS[a]	450.0	1.323	C –HC[a]	331.0	1.090
CG –CG	310.0	1.520	C –H1	410.0	1.092
CG –HC[a]	340.0	1.090	OH –HO[a]	553.0	0.960
CG –H1[a]	340.0	1.090	N –H[a]	434.0	1.010
CG –H2[a]	340.0	1.090	CY –OH[a]	320.0	1.410
CG –OH	320.0	1.430	CG –OS	285.0	1.460
CG –C	220.0	1.530	CG –OY	320.0	1.410
CG –N	337.0	1.450	CY –CG	310.0	1.520
C –O[a]	570.0	1.229	CY –OS[a]	320.0	1.410
C –N[a]	490.0	1.335	CY –OY[a]	320.0	1.410
C –O2[a]	656.0	1.250	CY –C	220.0	1.530

Angle	K_θ[d]	θ_{eq}[e]	Angle	K_θ	θ_{eq}
CG–OS–C	60.0	117.0	OS–CG–N	106.9	107.9
OS–C –O[a]	80.0	125.0	CG –CG–N[a]	80.0	109.7
OS–C –CG	95.0	110.8	H1–CG–N[a]	50.0	109.5
HC–CG–HC	40.0	109.5	H2–CG–N[a]	50.0	109.5
H1–CG–H1	45.0	109.5	C –CG–CG[a]	63.0	111.1
H2–CG–H2	45.0	109.5	C –CG–HC[a]	50.0	109.5
CG–CG–HC	45.0	112.0	C –CG–H1[a]	50.0	109.5
CG–CG–H1	45.0	111.0	C –CG–H2[a]	50.0	109.5
CG–CG–H2	45.0	111.0	C –N –CG[a]	50.0	120.0
CG–CG–CG	45.0	113.5	CG –N –CG[a]	50.0	118.0
OH–CG–CG	70.0	107.5	H –N –CG[a]	30.0	118.0
H1–CG–OH	60.0	110.0	C –N –H[a]	30.0	122.0
H2–CG–OH	60.0	110.0	H–N–H[a]	35.0	120.0
CG–OH–HO	55.0	109.5	H2–C –N	55.0	112.4
CY–OH–HO	55.0	109.5	H2–C –O	55.0	112.2
CG–OS–CG	50.0	111.6	OS–CG–C	63.0	112.36
CT–OS–CG	50.0	111.6	O2–C –OH[a]	80.0	126.0
OS–CG–CG	70.0	108.5	OS–CY–CG	70.0	108.5
OY–CG–CG	70.0	108.5	OH–CY–CG	70.0	108.5
H1–CG–OS	60.0	110.0	OY–CY–CG	70.0	108.5
H1–CG–OY	60.0	110.0	OH–CY–C	63.0	112.4
H2–CG–OS	60.0	110.0	OS–CY–C	63.0	112.4
H1–C –O	60.0	110.0	OY–CY–C	63.0	112.4
H1–C –N	55.0	112.4	OY–CY–OH	100.0	112.0
OS–CG–OS	100.0	112.0	OY–CY–OS	100.0	112.0
OH–CG–OS	100.0	112.0	CG–CY–C[a]	63.0	111.1
O2–C –O2	80.0	126.0	CG–OS–CY	50.0	111.6
O2–C –CG	70.0	115.0	CG–OY–CY	50.0	111.6
CG–C –N[a]	70.0	116.6	O2–C –CY	70.0	115.0
CG–C –O[a]	80.0	120.4	HC–CG–CY	45.0	112.6
N –C –O[a]	80.0	122.9	CG–CG–CY	45.0	113.5
N –CG–HC[a]	50.0	109.5			

Table 2.4. (continued)

Torsion	$V_n^f/2$	γ^g, n^h	Torsion	$V_n/2$	γ, n
OS–CG–CG–C	-0.10	0.0, 3	OS–CG–OS–CT	0.37	0.0, 3
	0.10	0.0, 2		1.27	0.0, 2
	-1.00	0.0, 1		0.30	0.0, 1
OH–CG–CG–C	0.10	0.0, 3	OH–CG–OS–CG	0.37	0.0, 3
	0.20	0.0, 2		1.27	0.0, 2
	-2.50	0.0, 1		0.30	0.0, 1
CG–CG–OS–C	-0.04	0.0, 3	H2–CG–OS–CG	0.10	0.0, 3
	0.47	0.0, 1		0.60	0.0, 2
CG–OS–C –O	-3.20	0.0, 2	H2–CG–OS–CT	0.10	0.0, 3
CG–C –OS–CG	3.00	0.0, 2		0.60	0.0, 2
H1–CG–OS–C	0.00	0.0, 3	HC–CG–CG–OS	0.05	0.0, 3
OS–C –CG–HC	0.00	0.0, 3	HC–CG–CG–OY	0.05	0.0, 3
HC–CG–CG–HC	0.13	0.0, 3	H1–CG–CG–OS	0.05	0.0, 3
H1–CG–CG–HC	0.17	0.0, 3	H2–CG–CG–OS	0.05	0.0, 3
H2–CG–CG–HC	0.17	0.0, 3	H –N –C –O	-2.50	0.0, 2
H1–CG–CG–H1	0.17	0.0, 3		2.00	0.0, 1
H1–CG–CG–H2	0.17	0.0, 3	H –N –C –H1	-2.50	0.0, 2
HC–CG–CG–CG	0.10	0.0, 3	H –N –C –CG	0.0	0.0, 1
H1–CG–CG–CG	0.15	0.0, 3	HC–CG–C –O	0.0	0.0, 3
H2–CG–CG–CG	0.15	0.0, 3	HC–CG–C –N	0.0	0.0, 3
HO–OH–CG–H1	0.18	0.0, 3	CG–OS–CG–N	-0.90	0.0, 1
HO–OH–CG–H2	0.18	0.0, 3	OS–CG–CG–N	-1.30	0.0, 1
HO–OH–CY–OY	0.18	0.0, 3	OY–CG–CG–N	-1.30	0.0, 1
HO–OH–CY–C	0.18	0.0, 3	OH–CG–CG–N	-1.50	0.0, 1
HO–OH–CY–CG	0.18	0.0, 3	CG–CG–CG–N	0.40	0.0, 1
HO–OH–CG–CG	0.18	0.0, 3	HC–CG–CG–N	0.10	0.0, 3
CG–CG–CG–CG	0.45	0.0, 1	H1–CG–CG–N	0.10	0.0, 3
OH–CG–CG–HC	0.05	0.0, 3	H2–CG–CG–N	0.10	0.0, 3
OH–CG–CG–H1	0.05	0.0, 3	N –CG–CG–N	0.00	0.0, 1
OH–CG–CG–H2	0.05	0.0, 3	HC–CG–CG–C	0.10	0.0, 3
OH–CG–CG–CG	0.10	0.0, 3	H1–CG–CG–C	0.10	0.0, 3
OS–CG–CG–CG	-0.27	0.0, 1	CG–CG–CG–C	0.45	0.0, 1
OS–CG–CG–CY	-0.27	0.0, 1	CG–CG–N –C	0.00	0.0, 1
OY–CG–CG–CG	-0.27	0.0, 1	OS–CG–N –C	2.03	0.0, 2
H1–CG–OS–CG	0.27	0.0, 3		1.00	0.0, 1
CG–OS–CG–CG	0.16	0.0, 3	OS–CG–N –H	-0.43	0.0, 2
CT–OS–CG–CG	0.16	0.0, 3		1.52	0.0, 1
OH–CG–CG–OH	0.55	0.0, 3	CG–CG–N –H	0.10	0.0, 3
	0.95	0.0, 2	H1–CG–N –CG	0.00	0.0, 3
	-0.10	0.0, 1	H2–CG–N –CG	0.00	0.0, 3
OH–CG–CG–OS	0.25	0.0, 2	H1–CG–N –H	1.00	0.0, 1
	-1.10	0.0, 1	H2–CG–N –H	1.00	0.0, 1
OH–CG–CG–OY	0.25	0.0, 2	H1–CG–N –C	-1.00	0.0, 1
	-1.10	0.0, 1	H2–CG–N –C	-1.00	0.0, 1
H1–CG–CG–OY	0.05	0.0, 3	CG–N –C –O	-2.80	0.0, 2
OS–CG–CG–OS	0.82	0.0, 2	CG–N –C –CG	-2.70	0.0, 2
OS–CG–CG–OY	0.82	0.0, 2	HC–CG–C –O2	0.00	0.0, 1
OS–CG–OH–HO	0.18	0.0, 3	CG–CG–C –O2	-1.50	0.0, 2
OS–CG–OS–CG	0.37	0.0, 3	H1–CG–C –O2	0.00	0.0, 1
	1.27	0.0, 2	H2–CG–C –O2	0.00	0.0, 1
	0.30	0.0, 1	OS–CG–C –O2	-2.25	0.0, 2
			CG–OS–CG–C	0.00	0.0, 2

Table 2.4. (continued)

Torsion	$V_n/2$	γ, n	Torsion	$V_n/2$	γ, n
H1–CT–OS–CG	0.27	0.0, 3	OS–CY–CG–HC	0.05	0.0, 3
CG–OS–CT–CT	0.16	0.0, 3	OY–CY–CG–HC	0.05	0.0, 3
CT–OS–CG–CG	0.16	0.0, 3	CG–OS–CY–CG	0.16	0.0, 3
OS–CG–OS–CT	0.37	0.0, 3	CG–OY–CY–CG	0.16	0.0, 3
	1.27	0.0, 2	OH–CY–C –O2	-2.25	0.0, 2
	0.30	0.0, 1	OS–CY–C –O2	-2.25	0.0, 2
CG–N –C –CT	0.00	0.0, 2	OY–CY–C –O2	-2.25	0.0, 2
OY–CY–OS–CG	0.37	0.0, 3	CG–CY–C –O2	-1.50	0.0, 2
	1.27	0.0, 2	CG–OS–CY–C	3.50	0.0, 1
	0.30	0.0, 1	CG–OY–CY–C	0.00	0.0, 1
OH–CY–OY–CG	0.37	0.0, 3	HC–CG–CY–C	0.10	0.0, 3
	1.27	0.0, 2	CG–CG–CY–C	0.45	0.0, 3
	0.30	0.0, 1	CY–OS–CG–CG	0.16	0.0, 3
OS–CY–OY–CG	0.37	0.0, 3	CY–OY–CG–CG	0.16	0.0, 3
	1.27	0.0, 2	CY–OS–CG–H1	0.27	0.0, 3
	0.30	0.0, 1	CY–OY–CG–H1	0.27	0.0, 3
OH–CY–CG–CG	-0.27	0.0, 1	CY–CG–CG–CG	0.45	0.0, 1
OS–CY–CG–CG	-0.27	0.0, 1	CY–CG–CG–HC	0.10	0.0, 3
OY–CY–CG–CG	-0.27	0.0, 1	CY–CG–CG–H1	0.15	0.0, 3
OH–CY–CG–HC	0.05	0.0, 3	CY–CG–CG–OH	0.10	0.0, 3

Improper torsions

X -X -N -H	1.0	180.0, 2
X -X -C -O	10.5	180.0, 2
X -X -N -CG	1.5	180.0, 2
X -O2 -C -O2	10.5	180.0, 2
X -O2 -C -OH	10.5	180.0, 2

van der Waals

	$R*^i$	ε^j
H	0.6000	0.0157
HO	0.0000	0.0000
HC	1.4870	0.0157
H1	1.3870	0.0157
H2	1.2870	0.0157
O	1.6612	0.2100
O2	1.6612	0.2100
OH	1.7210	0.2104
OS	1.6837	0.1700
OY	1.6837	0.1700
CG	1.9080	0.1094
CY	1.9080	0.1094
C	1.9080	0.0860

[a] Valence terms incorporated from the AMBER PARM94 parameter sets (CG corresponds to CT)
[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. [i] van der Waals radius (Å).
[j] Minimum energy in the Lennard-Jones 6-12 potential (kcal/mol).

Table 2.5. Average[a] overall errors in the GLYCAM06 torsional energies relative to B3LYP/6-31++G(2d,2p)//HF/6-31G* values for the molecules and specific atomic sequences in the training set

Atomic sequence	A	B	Relative Error		
			$\langle Error^c \rangle_{curve}$	% of maximum barrier [d]	$\langle Error^e \rangle_{minima}$
Hydrocarbons (Overall)			0.15	2.0	0.11
A-CG-CG-B	HC, CG	HC, CG	0.15	2.0	
Alcohols (Overall)			0.45	5.0	0.28
A-CG-CG-B	OH, H1	OH, H1, HC, CG	0.50	6.0	
A-OH-CG-B	HO	H1, CG	0.09	6.0	
Ethers (Overall)			0.56	4.0	0.35
A-CG-CG-B	OS, H1	OS, H1, HC, CG	0.30	2.0	
A-OS/OY-CG/CY-B	CG	OS, H1, H2	0.65	6.0	
Amides (Overall)			0.64	4.0	0.38
A-N-C-B	H, CG	CG, O	1.14	7.0	
A-CG-N-B	H1, CG	H, C	0.18	4.0	
A-CG-C-B	HC, CG	O, N	0.63	7.0	
A-CG-CG-B	N, H1	HC, H1, CG, N	0.41	4.0	
Esters (Overall)			0.44	3.0	0.23
A-CG-C-B	HC	OS	0.02	8.0	
A-CG-OS-B	H1,CG	C	0.46	4.0	
A-OS-C-B	CG	CG, O	0.40	3.0	
Carboxylates (Overall)			0.72	38.0	0.60
A-CG/CY[b]-C-B	HC, CG	O2	0.14	7.0	
Ether alcohols (Overall)			0.61	6.0	0.59
A-CG-CG-B	OS, H1	OH, H1	0.61	6.0	
Ether amides (Overall)			1.12	9.0	1.10
A-CG-N-B	OS, H2	H, C	1.93	15.0	
A-OS-CG-B	CG	N, H2	0.64	9.0	
A-CG-CG-B	N, H1	OS, H1	0.37	6.0	
Alcohol amides (Overall)			0.49	5.0	0.41
A-CG-CG-B	OH, H1	N, H1	0.49	5.0	
Ether carboxylates (Overall)			0.80	29.0	0.80
A-CG-C-B	OS, H1	O2	0.55	31.0	
A-OS-CG-B	CG	C	1.18	42.0	

[a]The polar moieties show the largest overall errors, which may not be surprising because the classical force field employed does not include atomic polarizabilities.
[b]CY = CG and OY = OS in sialic acids only. [c]Average relative error over entire curve. [d]($\langle Error \rangle_{curve}$/Maximum barrier height)*100. [e]Average relative error of mimima

Table 2.6. ω_8 (O7-C7-C8-O8) *trans* and *gauche* rotamer populations from a 10-ns explicit solvent MD simulation of α-Neu5Ac as a function of the V_1 coefficient

V_1 (kcal/mol)	*trans*	*gauche*
-1.0	30	70
-0.8	59	41
-0.6	32	68
-0.4	41	59
-0.2	88	12
0.0	84	16
0.2	67	33
0.4	83	17
0.6	86	14
0.8	88	12
1.0	98	2

Table 2.7. Scalar ^3J-couplings (Hz) and exocyclic hydroxymethyl rotamer populations computed from MD simulation (50 ns) versus experimental data for methyl α-D-Glc*p* and α-D-Gal*p*

	$^3J_{H5,H6R}$	$^3J_{H5,H6S}$	*gg* %	*gt* %	*tg* %
Methyl α-D-Glc*p*					
Experiment[a]	5.49	2.39	53	47	0
Experiment[b]	5.4	2.3	57	38	5
GLYCAM06	5.4 ± 1.7	2.9 ± 2.0	62	36	2
Methyl α-D-Gal*p*					
Experiment[a]	7.8	6.0	14	47	39
Experiment[b]	8.6	3.7	16	75	9
Experiment[c]	8.3	4.0	21	61	18[d]
			13	70	17[e]
			15	69	16[f]
GLYCAM06	7.9 ± 1.6	3.7 ± 1.8	8	75	18

[a]Ref. 48. [b]Ref. 50. [c]Ref. 49 (Employing equations [d]A, [e]B1, and [f]G (Table 3), with the $^3J_{HH}$ values derived from a Karplus equation that includes the effects of the electronegativities, and orientations of α-substituents)

Table 2.8. Distances and dihedral angles between the ring hydrogen atoms computed from MD simulation (50 ns) versus experimental data for methyl α-D-Gal$p$, α-D-Glc$p$ and α-D-Man$p$

	α-D-Gal$p$OMe		α-D-Glc$p$OMe		α-D-Man$p$OMe	
	Experiment[a]	GLYCAM06	Experiment[b]	GLYCAM06	Experiment[b]	GLYCAM06
H1-H2	2.396	2.40 ± 0.1	2.473	2.41 ± 0.1	2.562	2.52 ± 0.1
H1-H3	3.827	3.83 ± 0.1	3.795	3.86 ± 0.1	3.817	3.79 ± 0.1
H1-H4	4.874	4.95 ± 0.1	4.065	4.23 ± 0.1	4.109	4.21 ± 0.1
H1-H5	3.605	3.75 ± 0.1	3.634	3.74 ± 0.1	3.639	3.78 ± 0.1
H2-H3	3.053	3.05 ± 0.1	3.046	3.05 ± 0.1	2.440	2.42 ± 0.1
H2-H4	3.792	3.83 ± 0.1	2.500	2.78 ± 0.2	3.829	3.86 ± 0.1
H2-H5	3.907	3.99 ± 0.1	3.919	4.00 ± 0.1	4.057	4.17 ± 0.1
H3-H4	2.436	2.45 ± 0.1	3.033	3.05 ± 0.1	3.053	3.05 ± 0.1
H3-H5	2.485	2.53 ± 0.2	2.607	2.76 ± 0.2	2.592	2.66 ± 0.2
H4-H5	2.457	2.45 ± 0.1	3.050	3.04 ± 0.1	3.070	3.05 ± 0.1
H1C1C2H2	50.49	51.07 ± 7.5	56.86	54.02 ± 7.7	-69.03	-67.21 ± 6.9
H2C2C3H3	-176.08	-175.74 ± 7.3	-178.44	-171.18 ± 7.9	-53.69	-54.98 ± 6.9
H3C3C4H4	53.97	55.99 ± 6.9	174.53	166.95 ± 8.4	172.41	172.99 ± 7.7
H4C4C5H5	-59.23	-58.12 ± 6.8	-175.73	-171.72 ± 8.4	-173.52	-171.59 ± 8.3

[a] Ref. 228, [b] Ref. 229

Table 2.9. Ring $^3J_{HH}$ computed from a 50-ns MD simulation using GLYCAM06 compared to experimental values. (Experiment / GLYCAM06)

Spins	α-D-Glc$p$OMe	β-D-Glc$p$OMe	α-D-Gal$p$OMe	β-D-Gal$p$OMe	α-D-Man$p$OMe	β-D-Man$p$OMe
$^3J_{H1H2}$	3.8	8.0	4.0	7.9	1.8	0.9
	3.5 ± 0.8	9.9 ± 0.5	3.8 ± 0.8	9.8 ± 0.9	2.2 ± 0.5	3.0 ± 0.7
$^3J_{H2H3}$	9.8	9.4	10.3	9.9	3.5	3.2
	9.9 ± 0.4	9.7 ± 0.6	10.1 ± 0.2	10.0 ± 0.4	3.4 ± 0.8	3.2 ± 0.8
$^3J_{H3H4}$	9.1	9.2	3.4	3.4	9.5	9.6
	9.7 ± 0.6	9.7 ± 0.6	3.3 ± 0.7	3.2 ± 0.8	10.0 ± 0.3	10.0 ± 0.4
$^3J_{H4H5}$	10.1	9.7	1.0	1.1		9.7
	9.9 ± 0.4	10.0 ± 0.4	3.0 ± 0.7	2.9 ± 0.7	-	9.9 ± 0.6

Table 2.10. Primary assignments for monomer frequencies calculated by NMODE/GLYCAM.

Frequency, cm^{-1}	Primary Assignment
66	T[O6-C6-H61(1.000)]
91	T[O6-C6-C5(1.000)]
107	U(ETO)[O3(1.064),O4(0.904),O2(0.568),[&B,O3-C3-H3(1.000)]]
149	U(ETO)[O1(0.976),[&T,O1-C1-C2(1.000)]]
214	T[H4O-O4-C4(1.000)]
231	T[H4O-O4-C4(1.000),H6O-O6-C6(0.498),C6-O6-H6O(0.427),C6-C5-C4(0.422)]
253	U(ECO)[O1(1.027),[&S,O1-HO1(1.000)][&B,O1-C1-C2(1.000)]]
272	T[H6O-O6-C6(1.000)]
282	T[H2O-O2-C2(1.000)]
298	U(ECO)[O3(0.950),O4(0.907),O2(0.377),[&B,O4-C4-C3(1.000)]]
303	T[HO1-O1-C1(1.000)]
313	T[HO1-O1-C1(1.000)]
353	U(ERBO)[O3(1.076),[&S,O3-C3(1.000)]]
369	U(ERBO)[O2(0.982),[&B,O1-C1-H1(1.000)]]
382	T[H2O-O2-C2(1.000)]
410	T[H3O-O3-C3(1.000)]
422	T[H3O-O3-C3(1.000),HO1-O1-C1(0.742)]
458	T[H3O-O3-C3(1.000)]
481	U(RB)[O5(0.833),[&B,O5-C1-C2(1.000)]]
521	U(ROPD)[O5(0.735),C4(0.668),[&T,O5-C5-C4(1.000)]]
570	U(RC)[O5(0.657),C3(0.505),[&S,O5-C1(1.000)]]
584	U(RC)[C4(0.974),C5(0.734),[&B,C4-C3-O3(1.000)]]
661	U(RB)[C1(0.929),O5(0.386),[&T,C1-O1-HO1(1.000)]]
835	ST[H61(1.000),H62(0.620)]
881	U(RB)[O5(0.335),[&S,C1-H1(1.000)]]
951	T[C6-C5-O5(1.000)]
977	U(ROPD)[C4(0.870),[&T,C4-C3-O3(1.000)]]
983	U(RB)[C1(0.616),C4(0.356),[&T,C3-C4-O4(1.000)]]
995	U(RC)[C1(0.770),O5(0.755),[&S,C1-O5(1.000)]]
1028	U(ROPD)[C5(1.005),C2(0.388),[&S,C5-H5(1.000)]]
1054	U(RC)[C5(0.316),[&T,C3-C4-H4(1.000)]]
1073	U(ECH)[H3(0.989),[&T,H3-C3-O3(1.000)]]
1088	U(RB)[C2(0.768),C4(0.667),[&S,C2-O2(1.000)]]
1103	U(ECH)[H4(0.953),H5(0.877),[&T,H4-C4-O4(1.000)]]
1130	U(ETH)[H5(0.837),H4(0.529),[&T,H5-C5-C4(1.000)]]
1135	U(ECH)[H2(0.964),[&T,H2-C2-O2(1.000)]]
1236	B[HO1-O1-C1(1.000)]
1274	B[H3O-O3-C3(1.000)]
1285	B[H2O-O2-C2(1.000)]
1302	B[H6O-O6-C6(1.000)]
1303	B[H4O-O4-C4(1.000),H6O-O6-C6(0.588)]
1315	ST[H61(1.000),H62(0.900)]
1344	U(ECH)[H5(0.821),H4(0.802),[&T,H5-C5-O5(1.000)]]
1367	U(ECH)[H2(0.985),[&T,H2-C2-O2(1.000)]]
1393	U(ETH)[H4(1.005),H2(0.725),[&B,H3-C3-C4(1.000)]]
1433	U(ETH)[H2(0.984),H3(0.896),[&T,H2-C2-C1(1.000)]]
1466	U(ETH)[H5(0.923),H4(0.630),[&T,H5-C5-C4(1.000)]]

Table 2.10. (continued)

1513	U(ETH)[H1(0.999),[&T,H1-C1-C2(1.000)]]
1537	U(ECH)[H5(0.857),[&T,H5-C5-C6(1.000)]]
1545	U(RC)[C4(0.921),C1(0.456),[&T,C3-C2-H2(1.000)]]
1571	U(RB)[C3(0.580),[&S,C2-C1(1.000)]]
1599	U(RB)[C2(0.799),C4(0.558),[&T,C2-C1-O1(1.000)]]
1613	AT[H62(1.000),H61(0.999)]
1635	S[C6-O6(1.000),C6-C5(0.868)]
2912	SS[H61(1.000),H62(0.975)]
2944	U(EOPSH)[H3(0.993),H4(0.873),H2(0.434),[&S,H3-C3(1.000)]]
2947	U(ERBH)[H1(0.894),[&S,H1-C1(1.000)]]
2947	U(ERBH)[H1(0.893),[&S,H1-C1(1.000)]]
2951	U(EOPSH)[H5(1.000),H2(0.947),H3(0.422),[&S,H5-C5(1.000)]]
2956	U(EOPSH)[H4(0.997),H3(0.980),[&S,H4-C4(1.000)]]
2980	AS[H62(1.000),H61(0.954)]
3709	S[H2O-O2(1.000)]
3711	S[H6O-O6(1.000)]
3711	S[H4O-O4(1.000),H6O-O6(0.518)]
3714	S[H3O-O3(1.000),H4O-O4(0.420)]
3720	S[HO1-O1(1.000)]

Assignment notation details: Assignments begin with a class abbreviation (e.g., S, AT, ETH, etc.) and are followed by a list of atom sets significantly involved in that sort of motion. In any atom set, the first atom is the one whose motion is being considered. Stretches and bends are referenced to 1 and 2 other valence-connected atoms, respectively. Torsions are referenced to only two other atoms (and not three) because torsional motion is considered as being about a given bond. For toroidal and circumferential (ring) motions, each set includes only the atom(s) moving most significantly with respect to the ring. The weights in parentheses are a general indication of the significance of the motion. Weights for simple (non-ring) motions are all scaled so that the maximum significance is 1. Weights for ring-related motions are scaled accordingly, but after the simple motions are scaled, so their significances might be greater than one. *Assignment abbreviations:* Simple motions are noted without prefix or suffix. They are: S=stretch; B=bend; T=torsion. Complex motions typically have a prefix, a suffix, or both, in addition to an indication of motion type. The motion types, in addition to those employed in simple modes are: OPS=out of plane stretch (axial atoms); OPD=out of plane distortion; RB=ring "breathing" (expansion/contraction) motion; RC=ring circumferential motion; RT=ring toroidal motion. The prefix E indicates that the motion is of exocyclic atoms (O or H). In the combinations ET and EC, the T stands for toroidal and the C, for circumferential. The prefix S or A represents "symmetric" or "asymmetric." For example, SS indicates symmetric stretch of two equivalent atoms. An "O" or "H" suffix indicates oxygen or hydrogen. A lack of a suffix indicates motion of the main ring. Concerted ring-related motions that do not fall easily into one of the above classes are prefixed with "U."

Table 2.11. Comparison of vibrational analyses for the α-D-glucopyranose monomer using geometries computed at the molecular mechanics and quantum mechanics levels.

NMODE GLYCAM Frequency cm^{-1}	B3LYP / 6-31++g(2d,2p) // 6-31++g(2d,2p) Frequency cm^{-1}	IR Intensity $(km\cdot mol^{-1})$	HF / 6-31G* // 6-31G* Frequency cm^{-1}	IR Intensity $(km\cdot mol^{-1})$
	H-O Stretches			
3720	3857	39.1	4132	63.8
3714	3851	40.3	4119	56.3
3711	3826	45.9	4099	59
3711	3825	52	4095	95.4
3709	3811	49.4	4089	56.1
	H-C Stretches			
2980	3063	43.3	3283	78.6
2956	3057	15.7	3270	32.8
2951	3040	19.7	3266	34.3
2947	3031	17	3224	56.9
2947	3012	42.4	3207	58.6
2944	2992	17.9	3195	2.8
2912	2988	58.4	3189	49.8
	-C6- Distortions			
1635	1458	12.2	1635	0.6
	1065	52.3	1206	37.4
	421	3.2	463	42.7
	Other H-C (Bends, Distortions, Toroidal and Circumferential Motions)			
1613	1498	6.9	1667	7
1537	1441	19.1	1620	11
1513	1416	13.3	1593	6
1466	1410	3.8	1583	25.1
1433	1387	1.9	1567	27.5
1393	1382	12.6	1552	47.9
1367	1366	8.1	1536	30.5
1344	1357	4.8	1515	23.2
1315	1351	38.3	1502	17.7
1135	1290	7	1486	26
1130	1281	13.5	1416	6.4
1103	1265	14.7	1402	42.4
1073	1232	20.9	1368	91.3
835	1222	22.1	1300	102.5
	1194	25.7	944	24.5
	850	12.8		
	H–O-C Bends			
1303	1276	62.3	1437	72.5
1302			1339	51.6
1285			1321	57.5
1274				
1236				

Table 2.11. (continued)

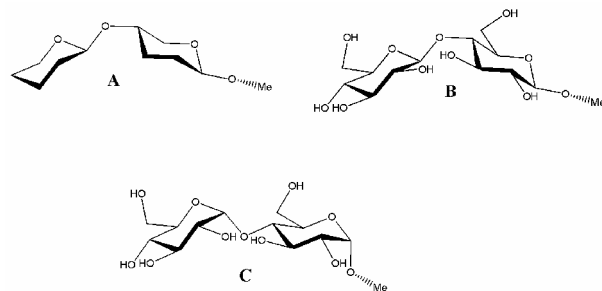| NMODE GLYCAM Frequency cm⁻¹	B3LYP / 6-31++g(2d,2p) // 6-31++g(2d,2p)		HF / 6-31G* // 6-31G*	
Frequency cm⁻¹	*Frequency cm⁻¹*	*IR Intensity (km·mol⁻¹)*	*Frequency cm⁻¹*	*IR Intensity (km·mol⁻¹)*
H–O-C Torsions				
458	384	65.1	444	84.8
422	366	58.7	424	51.9
410	360	44.9	419	105.2
382	332	93	386	98.9
313	226	79.7	306	20.1
303	221	50.8	262	123.2
282	171	94	252	55.4
272			207	90.3
231				
214				
Other Exocyclic (Non-Ring) O Motions				
881	1102	72.2	1298	69.2
570	1042	287.4	1274	190.4
521	637	28	696	25.6
481	555	23.7	602	18.3
369	534	18.9	581	51.4
353	436	25.7	486	65.9
298	400	38.5	399	17.2
253	269	13.6	293	2.8
149	251	8.6	274	2.9
107	247	9.9	155	1.8
	143	3.9		
	113	3		
Motions of the Ring				
1599	1464	12	1258	142.8
1571	1148	53.8	1255	29.1
1545	1129	13.3	1228	25.6
1088	1117	95.9	1215	207.9
1054	1082	71.9	1152	16.6
1028	1075	42.2	1126	32
995	1049	112.1	1099	16.5
983	1012	43	979	20.5
977	995	14.3	851	49.3
661	908	12.2	653	43.7
584	770	38.4		
	595	22.1		
Other Motions of C6 and O6				
951	306	14.3	1193	60.8
91	94	6.4	334	9.3
66	65	0.9	123	2.4
			106	4.7
			75	2.9

Figure 2.1. Tetrahydro-2-methoxy-5-(tetrahydro-2*H*-pyran-2-yloxy)-2*H*-pyran (A), Methyl β-D-Glucopyranosyl-(1→4)-β-D-Glucopyranoside (B), and Methyl α-D-Glucopyranosyl-(1→4)-α-D-Glucopyranoside (C)
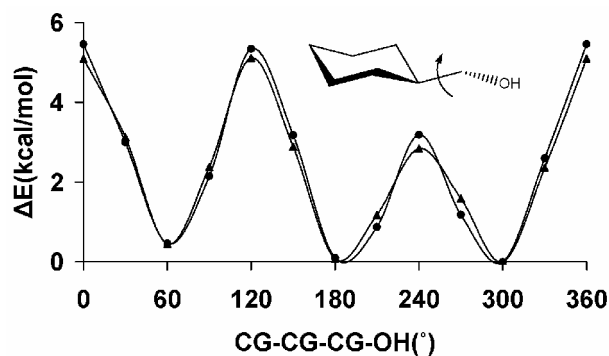

Figure 2.2. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves about the exocyclic CG-CG bond in hydroxymethylcyclohexane
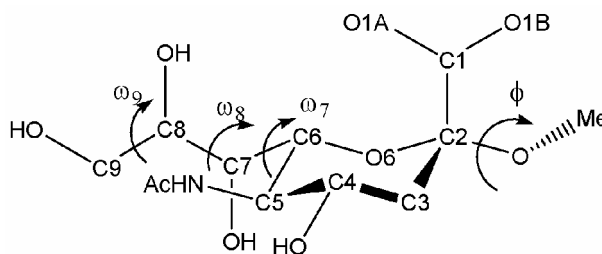

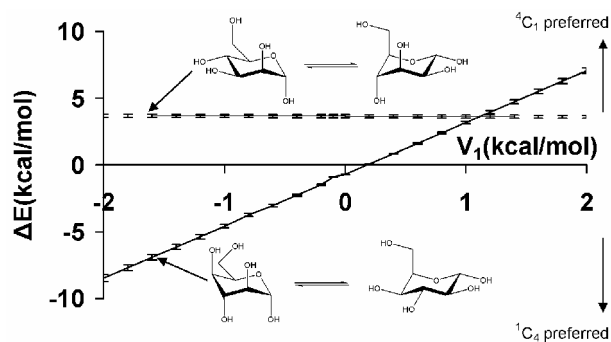**Figure 2.3**. Key exocyclic torsion angles in α-neuraminic acid

Figure 2.4. Differences in average molecular mechanical energies from MD simulation data (1 ns) for the $^4C_1$ and $^1C_4$ chair forms as a function of the V_1 coefficient for the OH-CG-CG-OH atomic sequence in α-D-mannopyranose (upper), and α-D-idopyranose (lower)
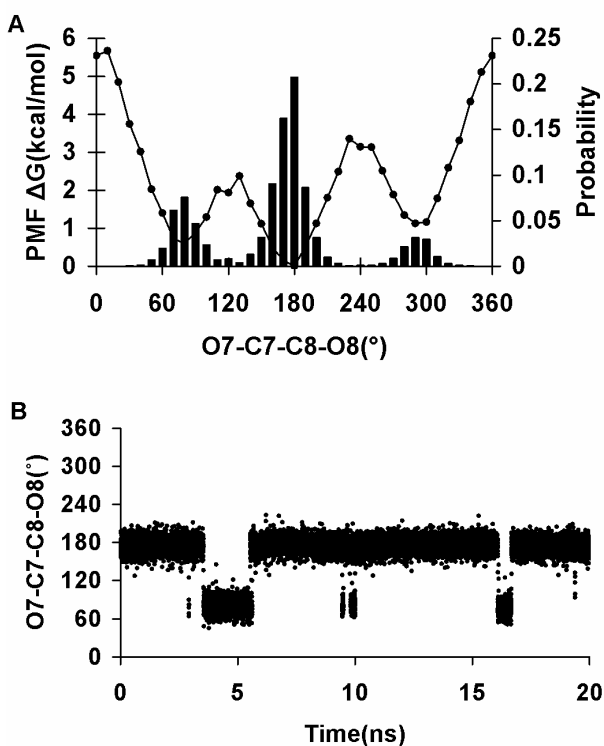

Figure 2.5. PMF-WHAM analysis for the ω_8 (O7-C7-C8-O8) torsion in α-5-*N*-acetyl-neuraminic acid (A). ω_8 population sampling over a 20-ns MD trajectory (86% *trans* and 14% *-gauche*) $V_1 =$ -0.1 kcal/mol (B)
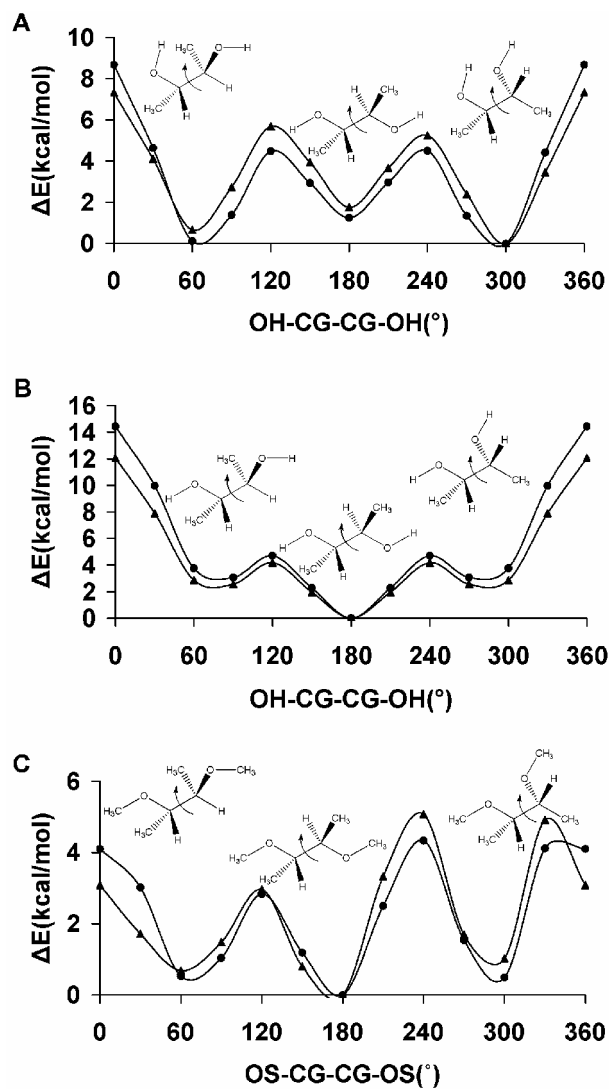
Figure 2.6. Rotational energy curves for 2,3-butanediol, with internal H-bonds (A), without internal H-bonds (B) and (C). 2,3-dimethoxybutane in which internal H-bonds are prohibited by the substitution of hydroxyl protons with methyl groups.   B3LYP/6-31++G(2d,2p) (▲); GLYCAM06 (●)
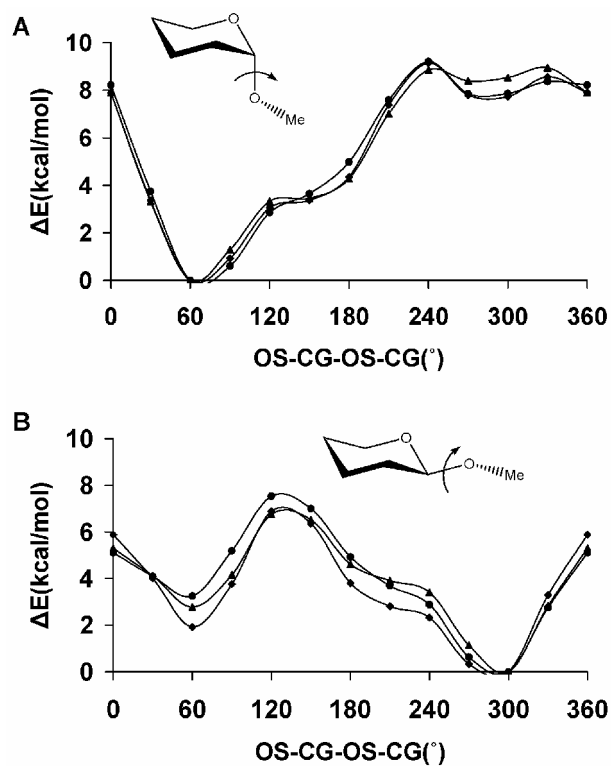
Figure 2.7. B3LYP/6-31++G(2d,2p) (▲), GLYCAM06 (●) and GLYCAM_93 (♦) rotational energy curves about the CG-OS bond in tetrahydro-2-methoxy-2*H*-pyran (axial), A, and tetrahydro-2-methoxy-2*H*-pyran (equatorial), B, for the united-atom (GLYCAM06) and all-atom (GLYCAM_93) charge models
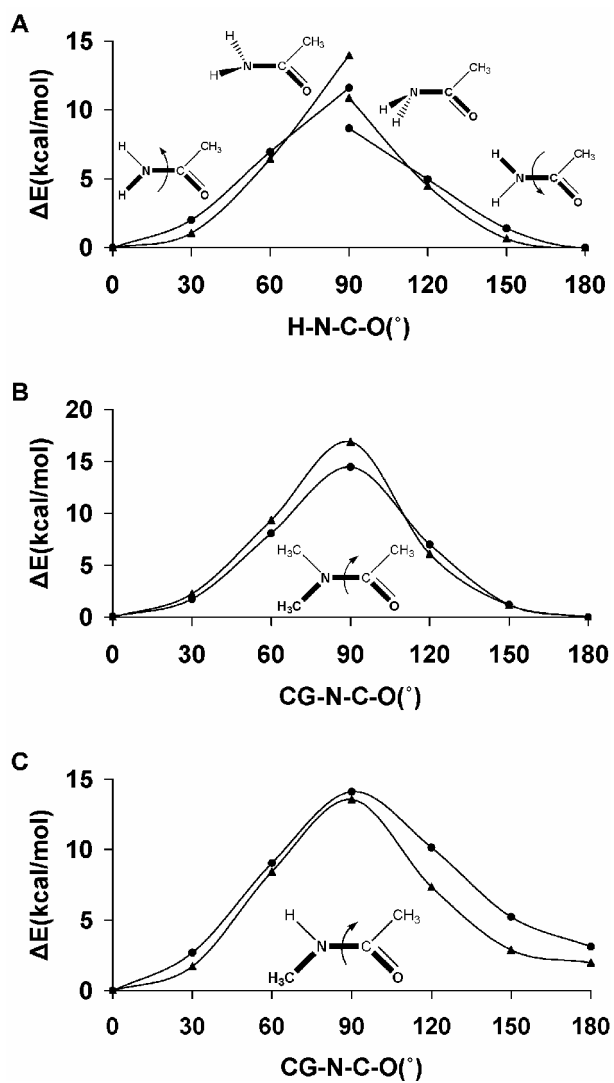
Figure 2.8. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves about the N-C bond in (A) acetamide, (B) *N,N*-dimethylacetamide and (C) *N*-methylacetamide
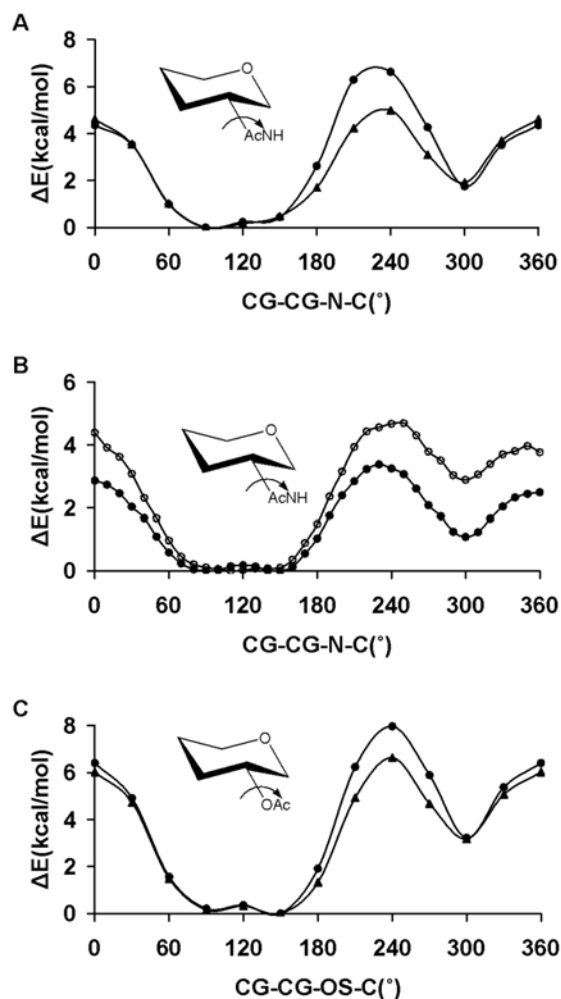
Figure 2.9. Rotational energy curves about the N-CG bond in *N*-(tetrahydro-2*H*-pyran-3-yl) acetamide (A), B3LYP/6-31++G(2d,2p) (▲), GLYCAM06 (●), H1-CG-N-H (V_1 = 0.0 kcal/mol) and H1-CG-N-C (V_1 = -0.17 kcal/mol), in vacuum. PMF-WHAM analysis for the CG-CG-N-C sequence in *N*-(tetrahydro-2*H*-pyran-3-yl) acetamide, H1-CG-N-H (V_1 = 1.0 kcal/mol) and H1-CG-N-C (V_1 = -1.0 kcal/mol) (○) and H1-CG-N-H (V_1 = 0.0 kcal/mol) and H1-CG-N-C (V_1 = -0.17 kcal/mol) (●), (B), in explicit solvent. Rotational energy curves about the OS-CG bond in tetrahydro-2*H*-pyran-3-yl acetate, B3LYP/6-31++G(2d,2p) (▲), GLYCAM06 (●), (C)
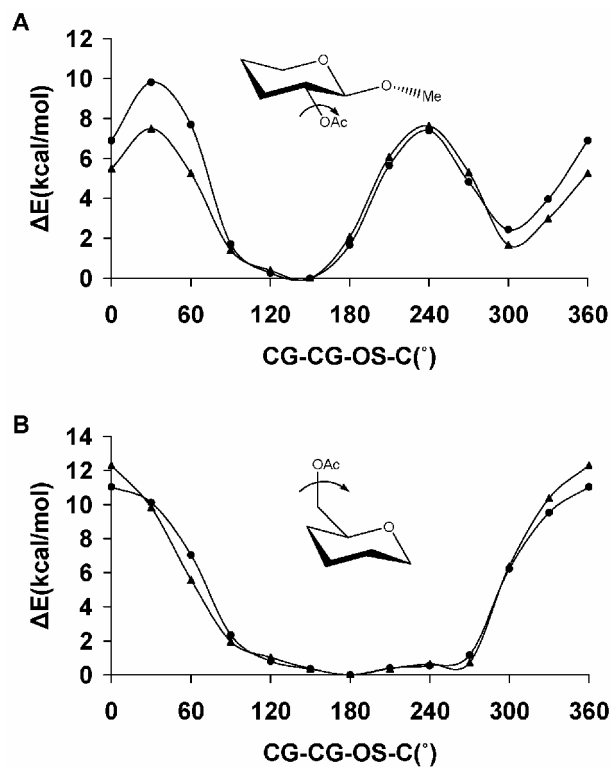
Figure 2.10. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves in (A)

tetrahydro-2-methoxy-2*H*-pyran-3-yl acetate and (B) (tetrahydro-2*H*-pyran-2-yl)methyl acetate
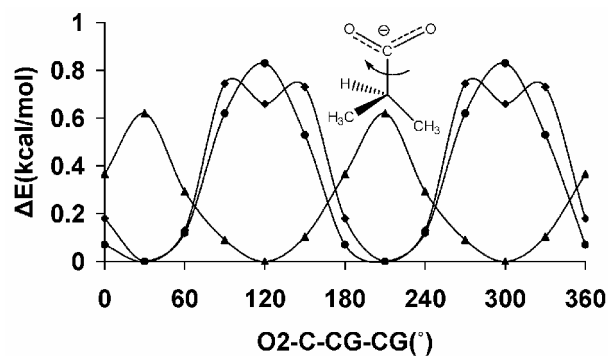
Figure 2.11. QM and GLYCAM06 rotational energy curves about the C-CG bond in 2-methylpropanoate MP2/cc-pVDZ (♦), B3LYP/6-31++G(2d,2p) (▲),GLYCAM06 (●)
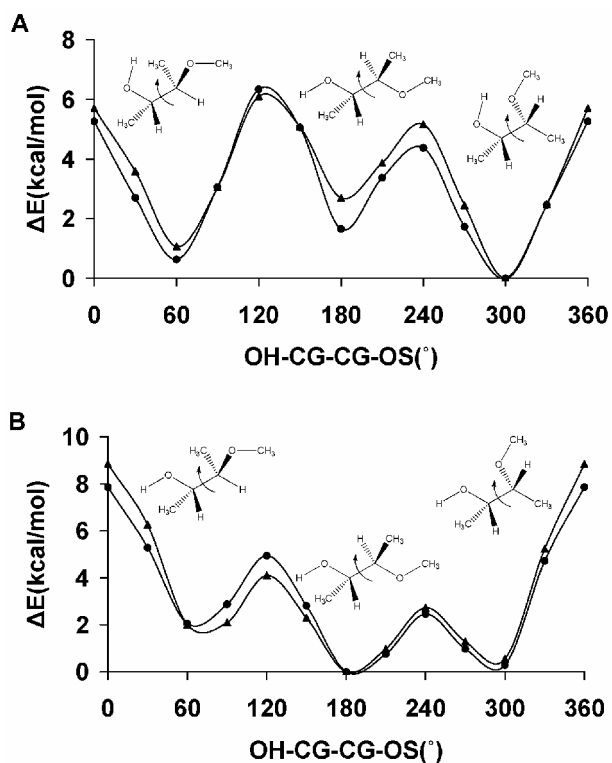
Figure 2.12. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves about the CG-CG bond in 2-methoxybutan-3-ol with internal hydrogen bonds (A) and without internal hydrogen bonds (B)
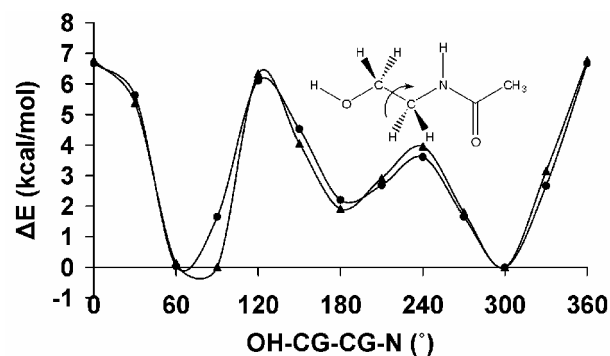
Figure 2.13. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves about

the CG-CG bond in *N*-ethanolacetamide
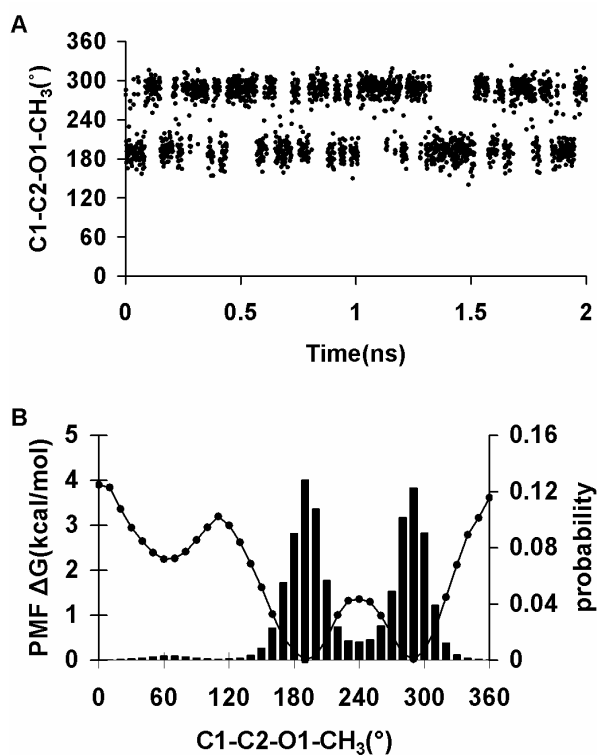
Figure 2.14. Population distributions from a 2-ns MD simulation trajectory $V_1 = 3.5$ kcal/ mol,

(A). Potential of Mean Force (PMF) and Statistical Weighted-Histogram Analysis Method

(WHAM) for φ (C_x-C_x-O_{x-1}-C_{x-1}) in methyl α-5-*N*-acetylneuraminic acid, (B)
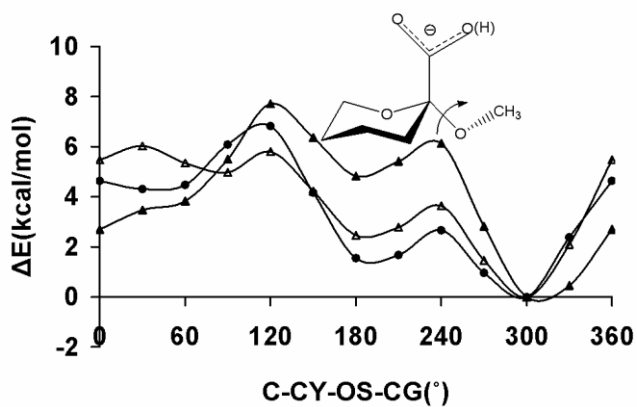
Figure 2.15. The rotational energy curves about the CG-OS bond in (*R*)-tetrahydro-2-methoxy-

2*H*-pyran-2-carboxylic acid computed with GLYCAM06 (●) employing the ionized species, and

at the B3LYP/6-31++G(2d,2p) level for both the ionized (▲) and neutral (Δ) species
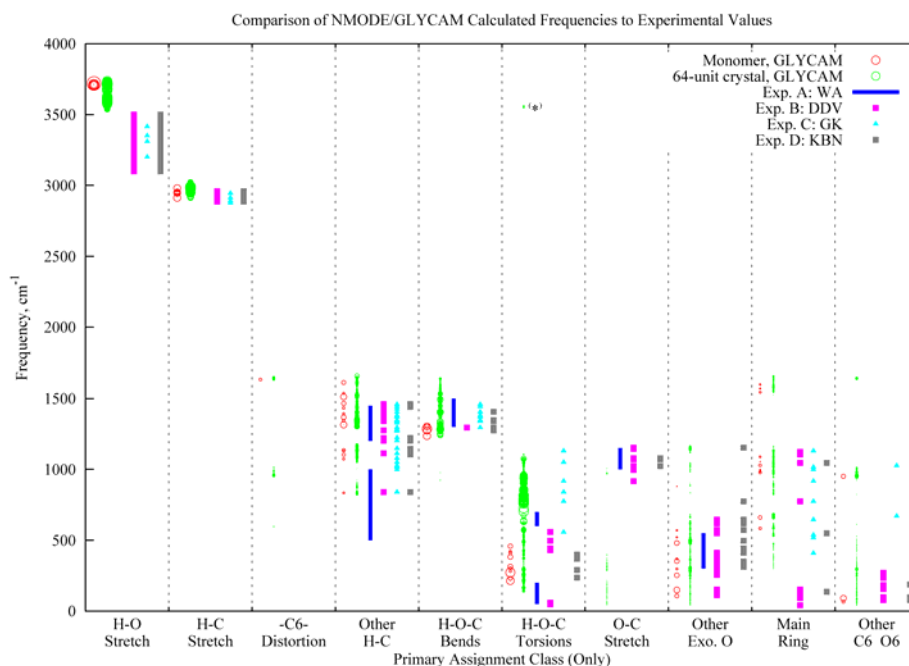
Figure 2.16. Comparison with experimental results of frequencies calculated by NMODE/GLYCAM for a monomer and a 64-unit crystal. The diameter of the circles for the calculated values corresponds to the relative kinetic energy associated with the motion used to determine the assignment, and not to photoabsorption intensity. For the monomer, the kinetic energy is relative to only the one molecule; for the crystal, it corresponds to a motion in a particular molecule, but relative to all 64 molecules. Some frequencies included here might not be observable by IR and/or Raman spectroscopy. Calculated crystal frequencies lower than 40 cm^{-1} are not included in the figure. In the list of assignment classes, "Exo. O" refers to exocyclic oxygens, oxygen atoms connected to the main ring but that are not part of the ring. (*) These frequencies occurred in a single molecule at the face of the crystal and are not expected to occur frequently in nature. Citations: (A) Wells and Atalla;[152] (B) Dauchez, Derremaux and Vergoten;[148] C, Gregurick and Kafafi;[151] D, Kuttel, Brady and Naidoo.[149] Please see the text for further discussion of the figure. The details of the individual assignments, as well as the frequencies, eigenvectors, etc., can be obtained by contacting the authors.
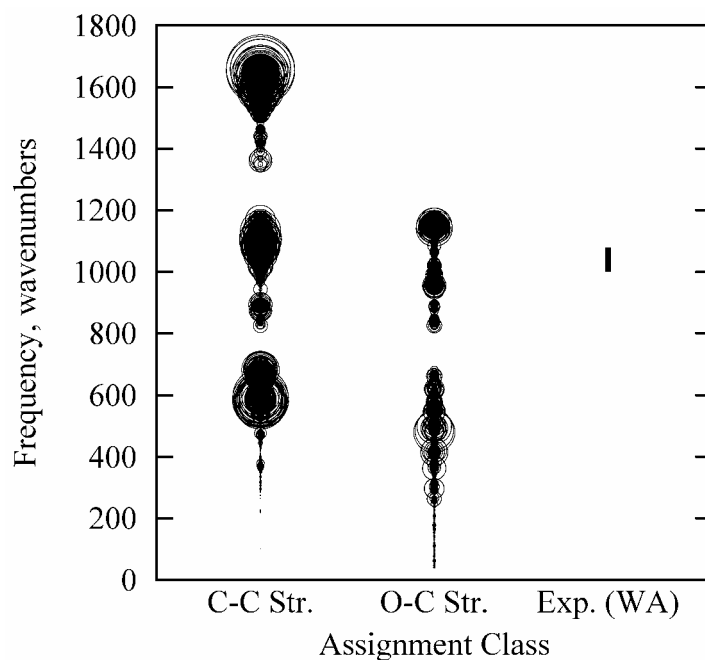
Figure 2.17. Range of frequencies associated with C-O and C-C stretches as calculated by NMODE/GLYCAM for the 64-unit monomer. The figure includes all C-O and C-C stretches in each molecule, regardless of relative importance to the mode. Circle diameters correspond to the relative kinetic energy associated with the stretching motion within the mode (see also discussion of Figure 16). The experimental range is taken from Wells and Atalla,[152] and corresponds to "heavy atom stretches."
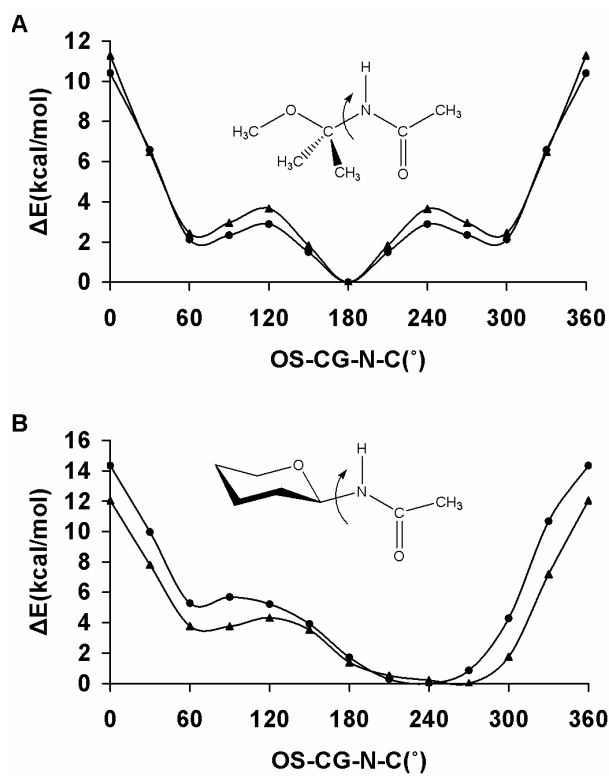
Figure 2.S1. B3LYP/6-31++G(2d,2p) (▲) and GLYCAM06 (●) rotational energy curves about

the CG-N bond in (A) *N*-(methoxyisopropyl)acetamide and (B) 2-*N*-acetyltetrahydropyran
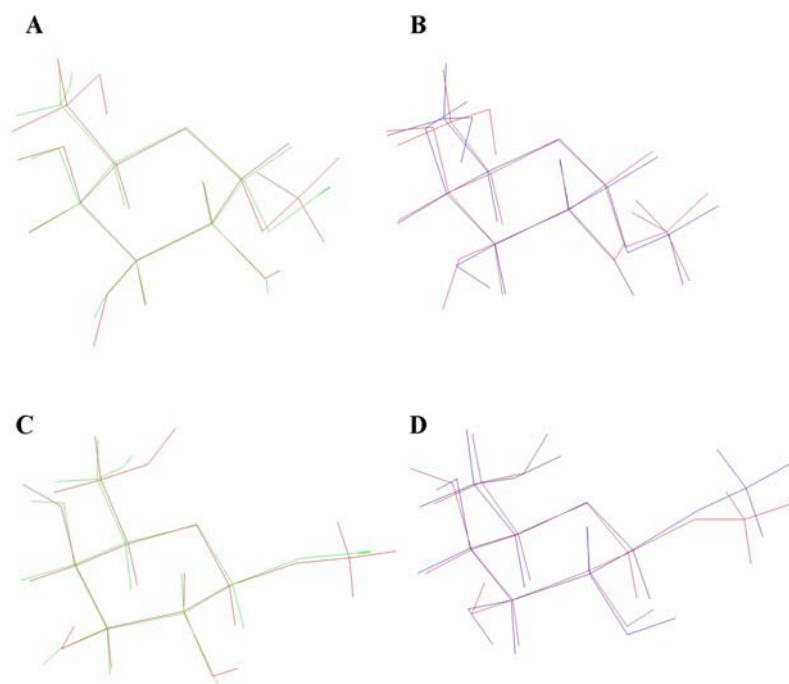
Figure 2.S2. Superpositions of the MD-derived average structures (green) and the energy minimized structures (blue) of methyl α-D-galactopyranoside (A and B), and methyl β-D-galactopyranoside (C and D) with their respective neutron diffraction structures (red)

CHAPTER 3

On achieving experimental accuracy from molecular dynamics simulations of flexible molecules:

aqueous glycerol[1]

Abstract

The rotational isomeric states (RIS) of glycerol at infinite dilution have been characterized in the aqueous phase via a 1-μs conventional molecular dynamics (MD) simulation, a 40-ns enhanced sampling replica exchange molecular dynamics (REMD) simulation, and a reevaluation of the experimental NMR data. The MD and REMD simulations employed the GLYCAM06/AMBER force field with explicit treatment of solvation. The much shorter time scale of the REMD sampling method gave rise to RIS and theoretical scalar $^3J_{HH}$ coupling constants that were comparable to those from the much longer traditional MD simulation. The $^3J_{HH}$ coupling constants computed from the MD methods were in excellent agreement with those observed experimentally. Despite the agreement between the computed and the experimental J-values, there were variations between the rotamer populations computed directly from the MD data and those derived from the experimental NMR data. The experimentally derived populations were determined utilizing limiting J-values from an analysis of NMR data from substituted ethane molecules, and may not be completely appropriate for application in more complex molecules, such as glycerol. Here, new limiting J-values have been derived via a combined MD and quantum mechanical approach, and were used to decompose the experimental $^3J_{HH}$ coupling constants into population distributions for the glycerol RIS.

113

3.1. Introduction

Glycerol has a complex conformational space due to its high flexibility and the presence of vicinal hydroxyl groups that are capable of stabilizing various rotamers through intramolecular hydrogen bonds (H-bonds), Figure 3.1. The physical properties of glycerol exhibit a peculiar dependence on variations in either temperature or pressure.[217,218] The number of H-bonds increases with pressure,[217] while the intermolecular distribution of pure liquid glycerol shows little temperature dependence from 193 K to 296 K.[218] Glycerol may also exist as a supercooled liquid, a property that makes its crystallization possible only through special techniques.[218] Decreases in temperature generally lead to the formation of a glass phase, at 185 K.[219] The resistance to crystallization has been exploited by nature, wherein glycerol is utilized alone[220] or in mixtures with trehalose, as a cryoprotectant.[221] *In vitro* glycerol helps preserve biomolecular structure,[222] as well as enhances the self-assembly of biomolecules.[223] These properties have stimulated extensive characterizations of the conformational equilibria of glycerol in different phases, employing either experimental[183,221,222,227-230] or theoretical methods.[1,9,13-20]

In glycerol, the dihedral angles defining rotations about the CC bonds are generally utilized to characterize the backbone rotamers. Employing Bastiansen's nomenclature:[181] α denotes a *gauche* torsion angle for the $O_1C_1C_2O_2$ sequence and an *anti* orientation for the $O_1C_1C_2C_3$ sequence; β corresponds to an *anti* $O_1C_1C_2O_2$ conformation; and γ is ascribed to a *gauche* and *anti* orientation for the $O_1C_1C_2O_2$ and $O_1C_1C_2H_2$ sequences, respectively, Figure 3.1. An unambiguous definition of all of the rotational isomeric states (RIS) of the backbone leads to six unique states: $\alpha\alpha$, $\alpha\beta$, $\alpha\gamma$, $\beta\beta$, $\beta\gamma$ and $\gamma\gamma$. Each backbone rotamer in turn displays different hydroxyl rotamers giving rise to a total of 27 H-O-C-C torsion angle RIS.

The exclusive presence of the αα backbone structure in the crystalline phase has been reported from neutron scattering experiments,[224] and is seen in short molecular dynamics (MD) simulations of neat crystalline glycerol.[225] For all the other phases, there is some disagreement between theoretical and experimental results, as well as between independent experimental studies, particularly in the gas phase. Electron diffraction gas-phase studies indicate a mixture of the αα and αγ rotamers,[181] while later experimental analysis from the microwave data identified γγ as the main rotamer, with a secondary presence of the αγ rotamer.[226] From a theoretical perspective, *ab initio* quantum mechanical (QM) calculations[246-248] of isolated rotamers of glycerol were in general agreement with the observations from the electron diffraction experiment. Gas-phase QM calculations[227-229] were generally consistent with gas-phase MD simulations,[230] which all indicated that αα and αγ were the major rotamers. The experimental conformational properties of glycerol in aqueous solution have been determined principally by NMR spectroscopy,[229,231] which indicated that the αγ rotamer was the most abundant, followed by αβ, αα, and βγ, in approximately comparable populations, while the γγ and ββ rotamers were the least populated. In contrast, in the pure liquid phase, X-ray[232] and neutron diffraction experiments[218,233] indicated the presence of only the αα rotamer. In addition to the αα rotamer, a combined QM and experimental approach that involved fitting the density functional theory (DFT) calculated IR-spectra of selected monomers of glycerol to the experimental spectra,[227] and MD simulation[225,230] studies, showed that the αγ rotamer was also a significant contributor in the neat liquid phase. MD simulations[225,230] have indicated that the αα and αγ rotamers are predominant in the glass phase, a conclusion that contrasts with neutron diffraction data that showed the presence of only the βγ rotamer.[219]

A number of empirical potential energy models for glycerol have been proposed, and applied to the gas, liquid, and crystalline phases.[233,244,251,253] And, a few MD simulations have been reported for glycerol, which seek to address mechanisms that occur under physiological conditions, such as the transportation of glycerol across cell membranes of micro-organisms.[234,235] In those MD studies, the conformational properties of glycerol in aqueous solution were not examined, despite the fact that it has been noted that a full understanding of the thermodynamics of initial binding would require including the dynamics of glycerol in the periplasm.[236] A model that is capable of reproducing the experimental aqueous phase conformational properties of glycerol would lend some confidence to the predictions of free energies of binding from more elaborate simulational studies. In a recent MD simulation investigation utilizing different concentrations of glycerol in the aqueous phase,[237] the rotamer populations were found to be insensitive to the composition of the mixtures. However, the very short MD time scale (500 ps) employed in that work[237] may limit the significance of the results. Here, we present a model, employing the recently reported GLYCAM06 force field,[45] for studying the conformational properties of glycerol in aqueous solutions at infinite dilution.[236,238] To the best of our knowledge, this study is the first attempt to characterize the conformational properties of dilute glycerol, via MD simulations on the physiologically relevant microsecond time scale.

3.2. Methods

3.2.1. Conventional MD simulations

The GLYCAM06 parameter set[45] was employed with the SANDER module of the AMBER8[136] software suite for all molecular mechanics (MM) and MD simulations. Ensemble-averaged partial atomic charges for glycerol were generated by restrained fitting (restraint weight

= 0.01) to the QM molecular electrostatic potentials (HF/6-31G*) of 100 rotamers, following the procedure described previously,[142,157] giving rise to the following values (in atomic units): Q_{O1} = Q_{O3} = -0.674, Q_{O2} = -0.663, Q_{C1} = Q_{C3} = 0.249, Q_{C2} = 0.302, Q_{HO1} = Q_{HO3} = 0.409, and Q_{OH2} = 0.393. Aliphatic protons carry zero net charge in GLYCAM06. A molecule of glycerol was immersed in a box of 235 pre-equilibrated TIP3P[117] water molecules, and the initial solvent configurations were subjected to energy minimization via 50 cycles of steepest descent followed by 950 cycles of conjugate gradient optimization. The entire system was then minimized via the same protocol. This was followed by a short simulated annealing of the system, in which it was heated from 5 to 300 K over 50 ps, then cooled to 5 K over another 50 ps. Initial atomic velocities were allocated from a Boltzmann distribution at 5 K. Prior to the production MD run the entire system was heated from 5 to 300 K over 70 ps, and maintained at that temperature for a further 80 ps. Production dynamics simulations were performed for 1 μs, under isobaric-isothermal (nPT) conditions with a 2-fs time step used to integrate the equations of motion. Long-range electrostatic interactions were treated using particle mesh Ewald summation.[239] To avoid potential imbalances in the internal energies of five- and six-membered intramolecular hydrogen bonds,[41] a unit scale factor was employed for all 1-4 nonbonded electrostatic and van der Waals interactions (SCEE = SCNB = 1.0). The SHAKE algorithm[138] was employed to constrain bonds containing hydrogen atoms to their equilibrium values.

3.2.2. Replica exchange MD simulations

The replica exchange MD (REMD) simulations were performed using the SANDER module of the AMBER9[143] software package. An exponential equation was fit to a range of sample temperatures available in the replica exchange section of the AMBER9 users' manual. An interpolation between the given temperatures was performed to obtain eight successive

approximate target temperatures (299.9, 308.0, 316.4, 324.9, 333.7, 342.8, 352.1 and 361.6 K). A molecule of glycerol was submersed in a pre-equilibrated octahedral water box of 1114 water molecules, and the system was equilibrated via the same protocol outlined in the conventional MD simulation section. During the simulation, long-range electrostatic interactions were treated using the particle-mesh Ewald summation, and scaling of 1-4 nonbonded electrostatic and van der Waals interactions were turned off. From this single equilibrated system, eight simulations were generated by heating to the eight approximate target temperatures noted above. A 10-ns production dynamics ensued under the constant temperature-volume ensemble. Gaussian distributions[52] of the potential energies from these simulations were plotted, in order to determine whether energy overlaps occurred within the temperature range, Figure 3.2. Utilizing the lowest target temperature (300 K) from the simulations, and a swap acceptance probability of 0.2, the final target temperatures (300, 303, 306, 309, 312, 315, 318 and 322 K) were obtained through an iterative process described previously.[52] Prior to the exchange dynamics, the systems were heated to their respective target temperatures. The hybrid solvent REMD approach was employed,[51] retaining 100 closest water molecules during each replica exchange process. A 2-fs time step was employed to integrate the equations of motion. An exchange was attempted every 250th step for a total of 250 x 80000 steps, resulting in an overall simulation time of 8 x 40 ns for the replicas.

Coordinates were collected every 5 and 20 ps from the REMD and conventional MD simulations, respectively, for subsequent analysis, which was performed with either the CARNAL or PTRAJ modules of AMBER8.

3.3. Results and Discussion

In the present study, the relative abundance of the backbone RIS was computed from a 1-µs explicit solvent MD simulation, and compared to available aqueous phase experimental NMR *J*-couplings and populations.[229,231] The MD data could be clustered into nine RIS, Figure 3.3A, which were subsequently grouped into six unique backbone RIS according to internal rotational symmetry. To determine whether the simulation had reached statistical equilibrium within the 1-µs time scale, the populations of the six unique backbone RIS were monitored as a function of simulation time, Figure 3.3B. During the initial equilibration stage (0 – 150 ps), only the αα and βγ rotamers were present with average populations of 42 and 58%, respectively. Throughout the first 200 ns the rotamer populations showed wide fluctuations that equilibrated only after approximately 300 ns and continued to display minor fluctuations in the population distribution up to 600 ns. The long simulational time required in the traditional MD simulation to achieve rotamer sampling equilibration raises the question as to whether similar results might not be achieved in a much shorter time, via the utilization of enhanced sampling methods. To this end, the REMD simulation approach described in the Methods section was employed. During the first 15 ns of the REMD simulations, the populations of the RIS varied significantly, showed less variation between the 15-30-ns interval, and became more stable during the last 10 ns of the simulation, Figure 3.3C.

3.3.1. Relative energies from a Boltzmann population analysis

The relative energies of the six backbone RIS, computed from a Boltzmann analysis of the experimental[231] and theoretical populations, are presented in Table 3.1. In the course of the MD and REMD simulations, each hydroxyl torsion angle displayed frequent transitions between all three staggered rotamers (data not shown). As such, the relative energies computed here are

averages over the staggered rotamers of the hydroxyl hydrogen atoms in the respective six unique backbone RIS, Figure 3.4. The most unstable rotameric state (ββ) was within 2 kcal/mol of the most stable (αγ). The relative energies of some backbone conformers of glycerol have also been reported at the QM SM5.42/HF/6-31G* and B3LYP/6-31+G**//SM5.42/HF/6-31G* levels of theory,[229] and as expected depend on the hydroxyl hydrogen torsion angles. The relative energies of each of the backbone RIS from the QM study[229] were calculated by averaging the relative energies of the hydroxyl rotamers and are shown in Table 3.1. The trends of the relative energies in decreasing order of stability are: αβ ≈ βγ < αα ≈ αγ < ββ ≈ γγ, and αα < αγ ≈ αβ < βγ < γγ < ββ, at the SM5.42/HF/6-31G* and B3LYP/6-31+G**//SM5.42/HF/6-31G* levels of theory, respectively. These trends in relative energies are at variance with those computed from the MD simulation, namely, αγ < αβ < αα ≈ βγ < γγ < ββ, and are disordered with respect to the relative energies derived from the experimental populations. However, it should be noted that with the exception of the ββ state, all of the relative energies computed from the NMR-derived populations are under 1 kcal/mol, placing extreme demands on any computational method.

3.3.2. Comparison with NMR data

In order to compare directly with the experimental NMR data, a generalized Karplus equation[240] was employed to compute the scalar $^3J_{HH}$-coupling constants of each terminal methylene hydrogen atom (H_B, $H_{B'}$, H_C and $H_{C'}$) to the central hydrogen atom (H_A), Figure 3.1A. These couplings were calculated for each snapshot, extracted at 5- and 20-ps intervals from the REMD and conventional MD trajectories, respectively, and then averaged, Table 3.2. In terms of the NMR spectrum, the coupling of each pair of prochiral methylene protons to the central proton is indistinguishable. As such, only two instead of four independent coupling constants are observed. To make a direct comparison with the experimental data, the MD-computed average

coupling constants between each methylene proton and the central proton were further averaged with those of its magnetically equivalent pair, to afford the final scalar $^3J_{HH}$-couplings, Table 3.2. Notably, the coupling constants computed from the 1-μs MD and 40-ns REMD simulation methods were essentially identical with each other and indistinguishable from the experimental data. This agreement suggested that the variations among the relative energies might be related more to the approximations imposed in the decomposition of the NMR J-values into rotamer populations, than to inaccurate MD data. To examine this possibility further, a detailed examination of rotamer populations was undertaken.

3.3.3. Rotamer populations

The rotamer populations computed from the traditional MD and REMD simulations are presented in Table 3.3. To quantify simulational convergence, as well as to judge the statistical significance of the results of the MD simulations, error estimates were computed by considering each rotamer population as a binomial random variable. That is, the population of each of the RIS ($\alpha\alpha$, $\alpha\beta$, etc.) within the simulation was incremented if any observed set of torsion angles occurred within the limits of the values employed to define the state. Employing the central limit theorem, which is appropriate given the large number of data points, the statistical properties of the RIS could then be readily characterized,[241] and are summarized here. The proportion, P, of the occurrence of a given state (X), in n trials is:

$$\mathbf{P} = \frac{\mathbf{X}}{n} \tag{1}$$

Since the relative population proportions are well established in the simulation, (see Figure 3.3), we can estimate the standard deviation of P (σ_P) by:

$$\sigma_P \approx \sqrt{\frac{P(1-P)}{n}} \qquad (2)$$

For example, for the αα state the observed P from the MD simulation is 0.17 and for the 1-μs simulation $n = 50{,}000$ (snapshots extracted at 20-ps intervals). Therefore, the standard deviation can be estimated as 0.0016, or approximately 0.2%.

The populations obtained from the much shorter REMD simulation were comparable with those determined from the 1-μs MD simulation. From the data in Table 3.3, the theoretical populations were in qualitative agreement with those derived from NMR experimental data.[229,231] Recent simulations of different compositions of the aqueous phase,[237] indicated the following trend in rotamer abundance: αα (40%), αγ (30%), αβ (20%), γγ (5%), βγ (5%) and ββ (0%), over all the concentrations examined. However, when compared to the present results, it appears probable that statistical equilibration was not achieved in that 500-ps study. In the present work, it is worth noting that the model predicted that rotamers that could form internal H-bonds were the least populated, namely, γγ (3 ± 0.1%) and ββ (2 ± 0.1%). These rotamers would be expected to be destabilized in solution primarily because of the breaking of internal H-bonds by solvent, as has been predicted to occur in carbohydrates in aqueous solution.[41]

The excellent agreement between the MD-derived coupling constants from this work and those determined experimentally suggested that there should be a corresponding agreement between the observed theoretical and experimentally derived rotamer populations; however, this was not the case. For example, the 1-μs MD populations of the αβ (28 ± 0.2%) and γγ (3 ± 0.1%) RIS were higher and lower, respectively, compared to the experimental values of 20-21% and 10-12%, respectively. It should be noted that the experimental rotamer populations were derived by utilizing the limiting $^3J_{HH}$-values determined from di-substituted ethane molecules.[231] The derivation of the populations of RIS from experimental J-coupling constants typically invokes a

linear combination of states weighted by associated state (or limiting) J-values. The final populations are therefore heavily dependent on these limiting J-values, which can rarely be determined directly from experimental data without invoking further approximations. The present results suggested a need to reevaluate the RIS populations from the experimental J-values.

Therefore, a combined MD-QM approach was utilized to compute limiting J-values for the *gauche* (J_G) and *anti* (J_T) coupling constants, Table 3.4. Additionally, by employing the states identified in the MD simulations, it was possible to take into consideration the effects of hydroxyl group rotations on the limiting J-values. Thus, the MD simulation was employed to identify the RIS, while QM methods, employing the Gaussian 03 software package,[242] were utilized to compute the J_G and J_T values for representative structures from each of the RIS. The 1-µs trajectory was clustered into the nine symmetry-related backbone RIS, identified during the MD simulation. On the basis of the staggered rotamers of the H-O-C-C dihedral angles, each of the RIS was further subdivided giving rise to 27 clusters for the single-weighted (αβ, αγ and βγ), or 15 clusters for the double-weighted (αα, ββ and γγ), backbone rotamers. An average structure was computed for each cluster, and a single structure that was the closest match to this average, based on root-mean-squared deviation in the atomic positions, was extracted from the MD trajectory and subjected to direct QM J-coupling calculations. Thus, a single "real" structure was employed to approximate the average "virtual" geometry of each RIS. Prior to the J-coupling calculations, each structure was optimized at the QM B3LYP/6-31G** level, while the backbone and hydroxyl torsion angles were frozen at their solution-preferred conformations. The J-couplings were computed with the B3LYP functional employing the HIIIsu3 basis set, as implemented recently for J-calculations.[206,243] For comparison, limiting J_G- and J_T-values were

123

also derived from the average RIS geometries employing a generalized Karplus equation.[240] To compute the J_G and J_T-values for each of the backbone RIS, a population-weighted average of the QM J-values was performed taking into consideration the population of each hydroxyl rotamer and the population of its symmetry-related cluster in the corresponding backbone cluster. Employing the QM-computed J_G and J_T-values computed in this work, the experimental NMR coupling constants[231] were decomposed into the populations of the individual RIS as described previously,[231] Table 3.3. The populations computed from these new J_G and J_T-values were comparable among the QM and MD simulation methods, and all suggested that the αα, αβ and αγ states account for approximately 75% of the conformational distribution of the aqueous phase, while the βγ, ββ and γγ states account for the remainder. However, the populations obtained using the QM-computed J_G and J_T-values (derived using average geometries) differed significantly from both the MD RIS populations and the experimentally derived RIS populations. These results indicate the significant influences that the choice of limiting J-value and model geometry may have when converting experimentally observable J-values into rotamer populations. Moreover, in decomposing the J-values into the populations of the experimental RIS, it was assumed that the J_G and J_T-values for all the RIS were identical,[231] suggesting that the OCCO atoms adopted ideal staggered conformations (±60°, 180°). Here, the J_G and J_T-values computed at the B3LYP/HIIIsu3//B3LYP/6-31G** level of theory, Table 3.5, are different for the symmetry-related protons as well as for protons on the same carbon atom, indicating that the aforementioned assumption may not be accurate.

3.4. Conclusion

The conformational properties of glycerol in the aqueous phase were examined through traditional MD and REMD simulation methods. Scalar $^3J_{HH}$ coupling constants computed from

both simulation methods were essentially identical, and were in excellent agreement with the available experimental data. That the explicit solvent MD simulations of glycerol with the GLYCAM06 force field reproduced the NMR J-couplings indicates that the force field together with the TIP3P water model can effectively compute the subtle balance between the intra-glycerol and glycerol-water nonbonded interactions.

The rotamer populations of the RIS from the MD and REMD simulations were in qualitative agreement with those derived from the NMR J-values; however, the approximations associated with deriving the experimental populations can be problematic. Employing QM-computed limiting J-values for state geometries derived from MD simulations resulted in an internally consistent set of populations. However, these populations differed from both the MD populations and from the experimentally derived values. The resulting rotamer populations suggest that the approximations employed in the experimental determination of these populations, which involved limiting J-values taken from substituted ethane molecules, may not be the most accurate approach. In this light, the J_G and J_T-values derived via the combined MD-QM approach may represent more consistent values to be employed in determining the solution conformational properties of glycerol. Ultimately, however, it is more accurate to compute and compare theoretical J-values with experimentally observable J-values, than to include the additional approximations necessary to decompose the experimental J-values into RIS populations. This problem is exacerbated in the case of glycerol, in which molecular symmetry reduces the number of experimental J-values.

In terms of the length of the simulations, it was shown that the use of an enhanced sampling method, such as REMD, gave rise to rotamer populations that were comparable to those from a 1-μs MD simulation, in a significantly shorter time. This fact is extremely

significant for larger flexible molecules, for which 1 μs MD simulations with explicit solvent are

presently unattainable.

Table 3.1. Relative energies[a] for glycerol RIS

Rotamer	NMR[224,b]	1 μs MD[b]	SM5.42 HF/6-31G*[229,c]	B3LYP/6-31+G**// SM5.42/HF/6-31G*[229,c]
αγ	0	0	0.22	0.47
αβ	0.20 – 0.21	0.13	0.00	0.53
αα	0.21 – 0.26	0.43	0.21	0.00
βγ	0.26 – 0.37	0.50	0.02	1.00
γγ	0.34 – 0.61	1.46	1.09	1.20
ββ	1.03	1.70	1.06	1.97

[a]In kcal/mol.
[b]From a Boltzmann analysis of the RIS populations.
[c]Determined from the reported QM data[229] by averaging the relative energies of all the hydroxyl group rotamers of each of the backbone RIS.

Table 3.2. Computed[a] and experimental scalar $^3J_{HH}$-couplings

			symmetry averaged		
	1 μs MD	40 ns REMD	1 μs MD	40 ns REMD	experimental[224]
$^3J_{AB}$	6.3 ± 1.0	6.4 ± 1.0	6.3 ± 1.0	6.3 ± 1.0	6.3 ± 0.2
$^3J_{AB'}$	6.3± 1.0	6.2 ± 1.0			
$^3J_{AC}$	4.0 ± 0.8	4.1 ± 0.8	4.0 ± 0.8	4.0 ± 0.8	4.4 ± 0.1
$^3J_{AC'}$	4.0 ± 0.8	4.1 ± 0.8			

[a]In Hz, computed using an empirical Karplus equation[240] and averaged over all structures extracted at 5 and 20 ps intervals for the REMD and conventional MD, respectively.

Table 3.3. Rotamer populations for glycerol from MD and REMD simulations, and derived from NMR data.

			NMR-derived			
rotamer	1 μs MD	40 ns REMD	QM limiting J-values[a]	empirical limiting J-values[b]	empirical limiting J-values[c]	experimental limiting J-values[224]
αγ	35 ± 0.2	33 ± 0.2	23	23	23	28-30
αβ	28 ± 0.2	31 ± 0.2	27	27	27	20-21
αα	17 ± 0.2	17 ± 0.1	27	28	26	18-21
βγ	15 ± 0.2	14 ± 0.1	11	11	12	15-17
γγ	3 ± 0.1	3 ± 0.1	5	5	5	10-12
ββ	2 ± 0.1	2 ± 0.0	7	7	7	5

[a]Limiting J-values computed at the B3LYP/HIIIsu3//B3LYP/6-31G** level, averaging over representative geometries for each of the RIS.
[b]Limiting J-values computed from a Karplus curve,[240] averaging over representative geometries for each of the RIS. [c]Limiting J-values computed from a Karplus curve,[240] averaging over all conformers in each of the RIS.

Table 3.4. Limiting J_G- and J_T-values computed from the 1-µs MD simulation data

	J_G	J_T
QM J-values[a] averaged over representative geometries for each of the RIS	2.46	9.86
empirical J-values[b] averaged over representative geometries for each of the RIS	2.35	10.04
empirical J-values[b] averaged over all conformers in the RIS	2.57	9.67

[a] Computed at the B3LYP/HIIIsu3//B3LYP/6-31G** level.
[b] Computed using an empirical Karplus equation in Hz. [240]

Table 3.5. Limiting J-values computed[a] for the coupled aliphatic protons in each RIS

Rotamer	J_{AB}	$J_{AB'}$	J_{AC}	$J_{AC'}$	$<J_G>$	$<J_T>$
αγ	8.88	1.69	1.29	2.35	1.77	8.88
αβ	9.01	4.61	1.49	11.21	3.05	10.11
αα	9.48	9.42	1.44	1.28	1.36	9.45
βγ	4.50	1.85	10.82	2.09	2.81	10.82
γγ	1.65	2.11	3.01	2.39	2.29	-
ββ	3.47	3.49	10.10	10.01	3.48	10.05

[a] B3LYP/HIIIsu3//B3LYP/6-31G**
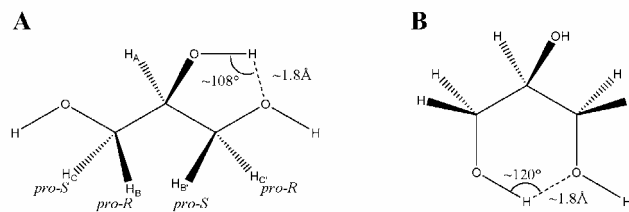
Figure 3.1. Schematic representations of glycerol indicating the atoms utilized to form potential five-membered (A) and six-membered (B) ring hydrogen bonds. The hydrogen atoms involved in scalar ^3J-couplings are also indicated (A).
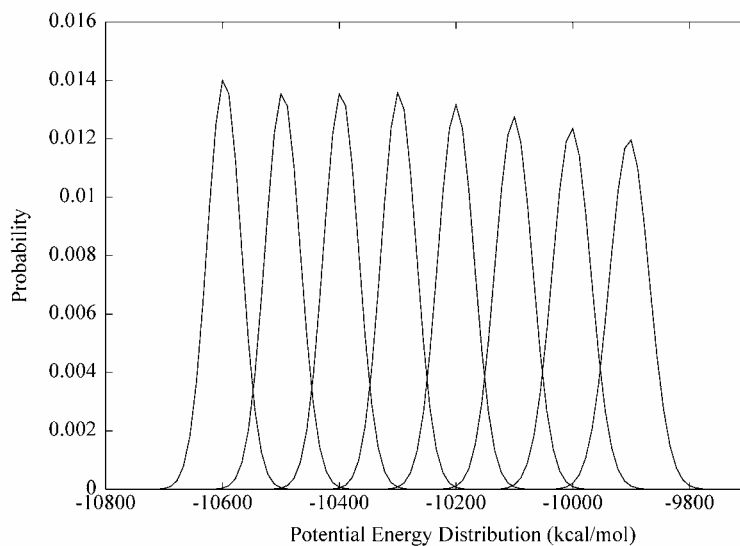
Figure 3.2. Gaussian potential energy distributions indicating the feasibility of temperatures employed in the replica exchange simulation
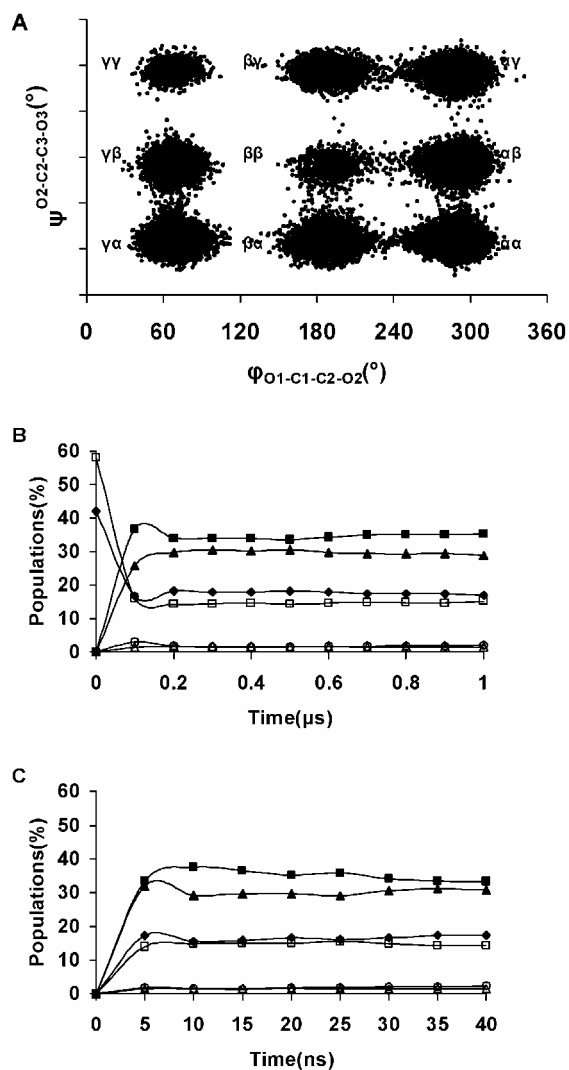
Figure 3.3. A scatter plot of two backbone dihedral angles, $\varphi(O_1\text{-}C_1\text{-}C_2\text{-}O_2)$ and $\psi(O_2\text{-}C_2\text{-}C_3\text{-}O_3)$, during the 1-$\mu$s MD simulation (A). Evolution of the six RIS of glycerol vs simulation time. Traditional MD (B) and REMD (C). $\alpha\gamma(\blacksquare)$, $\alpha\beta(\blacktriangle)$, $\alpha\alpha(\blacklozenge)$, $\beta\gamma(\square)$, $\gamma\gamma(\circ)$ and $\beta\beta(\triangle)$
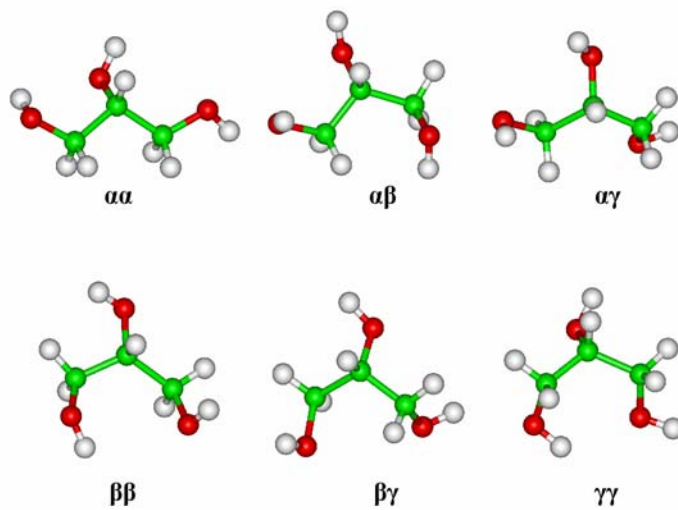
Figure 3.4. The six unique backbone RIS of glycerol extracted from the MD trajectory.

CHAPTER 4

The conformational properties of methyl α-(2,8)-di/trisialoside: Implications for anti-*Neisseria meningitidis* B vaccine design[1]

[1] Austin B. Yongye, Jorge Gonzalez-Outeiriño, John Glushka, Verena Schultheis, and Robert J. Woods. (Submitted to *Biochemistry*)

Abstract

The conformational properties of di- and trisaccharide fragments of the polysialic acid *O*-antigen capsular polysaccharide (CPS) of *Neisseria meningitidis* B (*Nm*B), have been investigated by a combination of solution phase NMR spectroscopy and explicit-solvent molecular dynamics (MD) simulations. Simulations employing 100 ns of conventional MD, as well as 160 ns of replica exchange MD (REMD), with the GLYCAM06 force field were shown to be in agreement with experimental NMR scalar *J*-coupling and nOe values. The presence of conformational families has been determined by monitoring inter-glycosidic torsion angles, by comparing structural superimpositions, as well as via a Bayesian statistical analysis of the torsional data. Attempts to augment the immunogenicity of *Nm*B CPS often involve chemical modifications of the *N*-acetyl moiety. Here the effects of these chemical group modifications on the conformational properties of the trisialoside have been probed via REMD simulations of the *N*-glycolyl, *N*-propionyl, *N*-propyl and *N*-butanoyl analogs. Although there were conformational families unique to each non-native analog, the chemical modifications resulted in largely equivalent overall conformational phase-spaces compared to the native trisialoside. On the basis of the conformational distributions, these shared conformational properties suggest that a recurrent global conformational epitope may be present in both the native and chemically modified CPS fragments. Explanations are therefore provided for monoclonal antibody cross-reactivity, in terms of recognition of a shared global CPS conformation, as well as for lack of cross-reactivity, in terms of fine structural differences associated with the *N*-acyl groups, which may be dominant in highly matured antibody responses.

Keywords: GLYCAM06, AMBER, molecular dynamics, *Neisseria meningitidis*, conformational properties, clustering, carbohydrates, cross-reactivities, capsular polysaccharides

4.1. Introduction

Neisseria meningitidis (*Nm*) is a gram negative bacterium that is responsible for septicemia and meningococcal disease in neonates.[1,244] Annually, about 500,000 cases of meningococcal meningitis are reported worldwide, with 2600 occurrences in the United States. Ten percent of infections result in deaths, while survivors may suffer from seizures, hearing impairments and brain damage.[245] Thirteen strains have been classified serologically, based on the identity of the polymeric repeat units and the types of glycosidic linkages in the *O*-antigen capsular polysaccharides (CPS) that envelop the surface of the bacteria.[7,246] Of these thirteen strains, five strains, A, B, C, W-135 and Y, have been identified as the main causative agents of virulent invasive meningococcal disease, Table 4.1.[8,267-270]

The lack of cross-reactivity between anti-CPS sera indicates that despite the fact that several *Nm* capsules bear striking carbohydrate sequence similarities, they must display unique antigenic determinants or epitopes. That is, each CPS must either have a unique overall 3D conformation, or must display its immunodominant regions in unique 3D configurations. For example, serogroups B (*Nm*B) and C (*Nm*C), which are both linear homopolymers of α-*N*-acetylneuraminic acid (Neu5Ac) containing α-(2,8) and α-(2,9) glycosidic linkages, respectively, elicit antibodies that do not cross-react with each other's CPS.[7,247] It should be noted that variations in the level and position of *O*-acetylation also exist between the serotypes, (Table 4.1).[8,248,249] It is remarkable that, despite the apparent sensitivity to carbohydrate sequence, cross-reactivity of a human monoclonal antibody (IgM^{NOV}) with the *Nm*B CPS and poly(Adenine) and other polynucleotides has been observed,[14] and attributed to similar conformational properties and spatial distribution of anionic charges between the carboxyl groups of the B polymer and the phosphate groups of the polynucleotide.[15] This result serves to illustrate a recurrent phenomenon,

namely that IgM class antibodies often show broader cross-reactivity than seen with more highly affinity matured IgG antibodies, and it is necessary to consider this factor when discussing CPS reactivities. Each *Nm*B CPS may also display differences in the apparent size of the epitope. For example, in binding inhibition assays a minimum of approximately ten residues of the *Nm*B CPS was required for effective inhibition of antisera binding to the intact polysaccharide antigen, while for strain C, maximal inhibition occurred with a more conventional hapten length of four or five residues.[9] In addition to antigenic variations arising from differences in glycosidic linkages types, epitope size may also be altered by chemical modifications. In the case of *Nm*B, inhibition studies indicated that reduction of the carboxylic acids to primary alcohols (sialitols) led to a decrease in the epitope size to approximately six residues.[16] The requirement in *Nm*B for an oligosaccharide length of at least ten residues, to inhibit antibody binding to the native CPS, has led to the postulation that this CPS adopts an extended conformational epitope.[9]

Currently, two CPS-based conjugate vaccines against *Nm* are available;[17,250] a bivalent form, composed of groups A and C, and a tetravalent form containing A, C, Y and W-135. A conjugate vaccine against *Nm*B is not currently available due to the poor immunogenicity of its CPS, which is structurally equivalent to the polysialic acid (PSA) autoantigen present in the gangliosides of human neural cell adhesion molecules, and is therefore not immunostimulatory.[18-21] The corollary to this structural similarity is that antibodies that bind to the *Nm*B CPS can also cross-react with PSA,[23] raising a question as to whether or not a CPS-based vaccine would have undesirable side effects. However, studies of human maternal and placental cord sera have found that "natural" anti-PSA antibodies were present in almost all matched pairs of sera;[251] suggesting that increased levels of anti-*Nm*B CPS antibodies elicited by vaccination would not be pathologic.[251] It has also been hypothesized that the weak

immunogenicity of *Nm*B CPS-based vaccines arises from the formation of transglycosidic lactone rings, between the hydroxyl group at C9 and the carboxylic acid in the preceding residue.[252,253] Lactonization, could alter the antigenicity and immunogenicity of the CPS by changing its conformation and charge distribution,[252,253] and indeed it has been observed that as little as 9% lactone content was sufficient to attenuate the antigenic properties of the CPS.[252] Recent NMR studies of the *Nm*B CPS displayed on *E. coli* K12 cell surfaces, under physiological conditions, showed no evidence of lactone formation, confirming that it is not a biologically common form of the CPS.[254]

Several attempts to develop immunogenic *Nm*B vaccines have focused on modifications of the amido group, for example by replacement of the *N*-acetyl with *N*-glycolyl (*N*-Gc), *N*-propionyl (*N*-Pr), *N*-propyl (*N*-Prop) or *N*-butanoyl (*N*-But) moieties.[16,22-24,26,277-281] Somewhat perplexing cross-reactivities, as well as varying levels of protection and bactericidal activity have emerged from immunizations with these conjugate vaccines. Several of the modified polysaccharides appear to present an extended epitope similar to that found in *Nm*B, as judged by cross-reactivity with monoclonal antibodies known to display specificity for the *Nm*B extended epitope. For example, the cross-reactivity of the *N*-Pr CPS with the well-characterized bactericidal *Nm*B monoclonal antibody 735 (IgG) indicates that both the *N*-Pr analog and the native CPS share a similar extended epitope of approximately ten residues.[26] However, some CPS analogs, such as the *N*-Prop derivative, bind poorly to mAb 735,[22] suggesting that the *N*-Prop analog differs significantly from the native CPS, either in overall conformation or in local contacts involving the *N*-acyl groups. Notably, the presence of an extended epitope is not sufficient to ensure cross-reactivity among the modified immunogens and or antisera. For example, although the *N*-Pr analog and the native CPS both bind to mAb 735,[26] immunization

with the *N*-Pr conjugate, while giving a strong antibody titre against the *N*-Pr CPS, gave little cross-reaction with *Nm*B CPS.[255] Similarly, a bactericidal monoclonal antibody, IgG 13D9, generated by immunization with the *N*-Pr conjugate vaccine showed no affinity for the native *Nm*B CPS,[26] although it did cross-react with the *N*-But CPS[256] Subsequent analysis confirmed that, mAb 13D9 does bind to the bacterial surface of *Nm*, leading to its bactericidal activity, but does not bind to *Nm*B CPS.[26] Thus, the bactericidal activity of mAb 13D9 can only be understood by accepting that it recognizes an epitope on the bacterial surface, other than that for which the vaccine was designed.[26]

As an alternative to modification of the *Nm*B CPS, attempts have been made to stimulate antibody production against both the *Nm*B and *Nm*C CPS by immunizing with vaccines made from the CPS of *E. coli* K92. The K92 CPS consists of alternating α-(2,8) and α-(2,9) linkages, and conjugate vaccines using this CPS could therefore be expected to produce by default anti-*Nm*B (2,8) or *Nm*C (2,9) responses.[11,251] Although vaccination with K92 CPS conjugates did yield anti-*Nm*C responses, they were poorer than direct immunization with *Nm*C conjugate.[251] Antisera from K92 vaccinations showed only a very modest if any affinity for *Nm*B CPS.[254,258] In an attempt to enhance the immune response to K92 conjugate, following on the results with chemically modified *Nm*B CPS, the *N*-Pr analog of K92 was generated. However immunization with *N*-Pr K92 conjugate did not lead to any improvement in cross-reactivity, but rather the antisera lost all affinity for both *Nm*B and *Nm*C CPS.[255] In contrast to the case of *Nm*B, the *Nm*C CPS does not display an extended epitope,[255] and the lack of affinity for *Nm*B displayed by the *E. coli* K92 antisera has been ascribed to the absence of an extended epitope in K92.[255] The origin of the deleterious effect of *N*-propionylation on the ability of K92 conjugate to elicit anti-

*Nm*C antibodies remains unclear, particularly since the *N*-Pr K92 antisera retained the ability to recognize native K92 CPS.

The potential relevance of an extended conformational epitope to the antigenicity, immunogenicity and bactericidal activity of *Nm*B CPS[9,283] has stimulated numerous investigations of the conformational properties of this CPS.[15,16,22,283-285] Understanding the structural features that determine the conformational properties that mediate the specificity and antigenicity of interactions between this CPS and bactericidal antibodies, could provide a basis for the interpretation of the antigenic properties of related structures and possibly for the rational development of more immunogenic and bactericidal vaccines against *Nm*B. Any chemical modification that alters the conformational epitope of a CPS-based immunogen may lead to attenuated antigenicity or unexpected cross-reactivity. By conserving the epitope conformation responsible for generating bactericidal antibodies (the bactericidal epitope) it should be possible to categorize putative immunogens not only in terms of their antigenicity, but also in terms of their potential to elicit bactericidal antibodies. However, the flexibility of *Nm*B in particular, as well as its homopolymeric sequence, currently prevents a direct resolution of the NMR data into definitive conformational families. The linkage regions between the residues of the *Nm*B CPS may be characterized by four conformationally significant rotatable bonds, Figure 4.1. The presence of these flexible dihedral angles (φ, ψ, ω_7, ω_8) renders the characterization of the 3-D structure and dynamics the *Nm*B CPS particularly challenging. When conformational variations are significant, such as in carbohdratees, the risk of generating virtual 3D conformations from the use of NMR restraints to guide the structural refinement are significant.[29] MD simulations can be employed to provide a structural basis for interpreting the NMR data or for *a priori* investigations of the effects of chemical modifications on the size and shape of the CPS epitopes.

For example, MD and NMR studies have been used to elucidate the role of sialic acid residues in determining the overall conformation of the CPS from type III group B *Streptococcus*.[33,102] For the specific case of the *Nm*B, molecular mechanics potential energy maps and MD simulations have been reported for the α-(2,8)-disialoside[15,258] and α-(2,8)-tetrasialoside,[257] respectively, in the gas phase. Explicit-solvent MD simulations employing simulation times of 3 ns[255] and 1-ns initiated from nine different low-energy conformations[258] of the α-(2,8)-disialoside, and 530 ps for the α-(2,8)-tetrasialoside[257] have also been reported. However, it is now well established that the life-times of rotamers in glycosidic linkage regions can span more than the 10-ns time frame.[33,41] In order to identify all of the significant conformational states, as well as to determine their populations, simulations on the order of 100 ns are typically required.[33,40,41]

Here, we characterize the conformational properties of methyl α-(2,8)-di- (**I**) and trisialoside (**II**) haptens from the CPS of the *Nm*B, employing data from 100-ns MD simulations performed with a force field (GLYCAM06)[259] that has been specifically developed for carbohydrate simulations. The theoretical conformations were then independently assessed for their experimental accuracy through comparison with new NMR nOe and J-coupling data for **I** and **II**.. Additionally, replica exchange MD (REMD) simulations were performed for **I** and **II** and the *N*-Gc, *N*-Pr, *N*-Prop and *N*-But analogs of **II**, Figure 4.1. The REMD protocol was recently shown to lead to excellent agreement with experimental populations for simulations of glycerol,[260] suggesting that this approach may be appropriate for determining the conformational properties of the glyceryl side chains in the Neu5Ac residues, which determine the conformational properties of *Nm*B. Our goal is to probe the effects of chemical modification on hapten conformation, and to provide a structural interpretation for the observed cross-reactivities or lack thereof between antibodies raised against the modified and native *Nm*B CPS.

4.2. Materials and methods

4.2.1. Computational methods

4.2.1.1. MD simulations

The strain B haptens (methyl α-(2,8)-di- and trisialosides) were built from the GLYCAM[259] library of monosaccharide residues using the XLEaP module of AMBER8.[136] The appropriate number of Na+ counter-ions was added to neutralize the haptens using XLEaP, and the complete systems were immersed in pre-equilibrated boxes of 670 and 743 TIP3P[261] water molecules for the di- and trisialoside, respectively. Nonbonded van der Waals and electrostatic scaling factors for 1-4 interactions were set to unity (SCEE = SCNB = 1). Long range electrostatic interactions were computed with particle-mesh Ewald summation,[239] with a nonbond cut-off distance of 8 Å. The constant pressure-temperature ensemble (nPT) was employed under periodic boundary conditions, with box dimensions of 30 x 33 x 35 Å and 35 x 31 x 36 Å for the di- and trisaccharide systems, respectively. Initially, the solvent configurations were energy minimized through 1000 cycles (50 steepest descent and 950 conjugate gradient minimization), while the CPS fragments were restrained. Subsequently, the entire system was energy minimized with the same protocol. Energy minimization was followed by a short simulated annealing stage, wherein the entire system was heated from 5 K to 300 K over 50 ps, then cooled to 5 K over another 50 ps. Initial atomic velocities were assigned from a Boltzmann distribution at 5 K. Prior to the MD production cycle, the temperature of the system was raised to 300 K over 70 ps and held at that temperature for a further 80 ps. The production cycle ensued for 100 ns. A 2-fs time step was utilized to integrate the equations of motion, with hydrogen-containing bonds constrained to their equilibrium lengths using the SHAKE algorithm.[138]

4.2.1.2 Replica exchange MD (REMD) simulations

In REMD simulations, multiple replicas of a system are simulated concurrently at progressively higher target temperatures that increase exponentially.[51,52] At specified intervals, the energies of adjacent replicas are swapped based on an acceptance probability.[52] Prior to performing the REMD simulations, the system was equilibrated following the same procedures outlined in the MD section. To derive the target temperatures employed in the present simulations, an exponential equation was fit to the range of sample temperatures given in the replica exchange section of the AMBER9 manual: 269.5, 300, 334, 371.8, 413.9, 460.7, 512.9 and 570.9 K. The exponential equation obtained from the fit was:

$$y = 242.1 \, e^{(0.1072 \, x)} \tag{1}$$

Eight approximate initial target temperatures (299.9, 308.0, 316.4, 324.9, 333.7, 342.8, 352.1 and 361.6 K) were generated by interpolating between the sample temperatures in incremental x-values of 0.25 starting from x = 2.0. From the initial single equilibration stage eight simulations were spawned by equilibrating to the eight approximate target temperatures noted above, followed by 10 ns of production dynamics under the constant volume-temperature (nVT) ensemble. Gaussian distributions[52] of the potential energies from these simulations were plotted, in order to confirm that energy overlaps occurred between successive replicas within the temperature range, Figure 4.2. Employing the lowest target temperature (300 K) from the simulations, and a swap acceptance probability of 0.2, the final target temperatures (300, 305, 310, 315, 320, 326, 331 and 337 K) were obtained through an iterative process described previously.[52] Prior to committing the replicas to exchange dynamics, each of the eight systems was heated to its target temperature. The hybrid solvent REMD method[51] was employed retaining 100 closest water molecules during each replica exchange process. An exchange was

attempted every 500th step for a total of 500 x 40000 steps, resulting in an overall simulation time of 8 x 20 ns for the eight replicas. A 1-fs time step was utilized to integrate the equations of motion, and a 10-kcal/mol restraint force constant was applied to the carbohydrate ring atoms to prevent the rings from distorting at high temperatures.

4.2.1.3. Bayesian cluster analysis

Cartesian coordinates from the MD simulations were collected every 5 ps for subsequent analysis with either the CARNAL or PTRAJ modules of AMBER8. In addition to typical analysis of torsion angle distributions,[40,41,262] global conformational states were also identified by determining transitions between Markov states, via a Bayesian statistical analysis of the glycosidic torsion angles.[262] A 25-component mixture of univariate Gaussians was used as a fuzzy partition of the data space. The centers and the widths of these Gaussians was optimized by a maximum likelihood approach.[263] Subsequently, the transition matrix for the 25 partition volumes was constructed. The two partition volumes connected by the fastest transitions were iteratively united. Thus, 25 Markov models and associated time-scales of the fastest transitions between the states of the respective model were obtained. In this approach, if the time scale (τ_l) of model l is much larger than the time scale of the subsequent model $l+1$ ($\tau_l \gg \tau_{l+1}$), then model $l+1$ contains rapidly mixing states, while model l contains slowly varying metastable states. In that case, model l represents a plausible conformational state of the system. This enables the definition of several relevant state models for the observed trajectory,[262] and identifies global relationships between glycosidic torsion angles that may not otherwise be readily determined.

4.2.2. NMR spectroscopy

Synthetic samples of the methyl α-2,8-di- and trisialosides were purchased from Dr. Nicolai Bovin (Lectinity Holdings, Lappeenranta, Finland). Samples were lyophilized from D_2O

then redissolved in 99.96% D_2O. Data were collected on Varian Inova 500, 800 and 900 MHz spectrometers at 25°C, except for some nOe data which were also collected at 10°C. All pulse programs were from the standard Varian library. Data were processed with Gaussian window functions, linear prediction, and zero-filling prior to Fourier transformation. For each sample, a COSY dataset was collected with 1582 x 1024 complex points and acquisition times of 0.40 and 0.258 seconds in F2 and F1, respectively. An HSQC dataset was collected with 789 x 400 complex points and acquisition times of 0.2 and 0.03 seconds in F2 and F1, respectively. Assignments are listed in Table 4.2. A second high-resolution HSQC spectrum was collected with the carbon decoupler turned off to allow a long acquisition time of 0.7 seconds during T2, in order to obtain proton-proton coupling data. 1D selective nOe experiments with mixing times of 100 to 800 ms were performed on the H3 axial signals in the Neu5Ac residues (H3ax) to determine the linear nOe buildup regime. Data was then collected, using a 300-ms mixing time, on those signals that had adequate spectral dispersion. Additional nOe data were obtained by acquiring a 2D NOESY spectrum utilizing a zero-quantum suppression element[264] and a mixing time of 300 ms. Data were collected with 420 x 2048 complex points and acquisition times of 0.459 and 0.09 s in F2 and F1. 1D data were processed with Varian VNMRJ software (Varian, Inc), and 2D data were processed and analyzed using NMRPipe software,[265] while volume integrations were performed with the NMRView software.[266]

4.2.2.1. Chemical shift assignments

Assignments were confirmed with COSY and HSQC data, Tables 4.2-4.3. Identification of the reducing, middle and terminal Neu5Ac residues was based on the existence of nOes between the H3ax proton and the OMe aglycon, and between the H3ax/eq and H8 protons of the preceding residue, as well as in the relative downfield chemical shift of C8 due to the glycosidic

linkage. *N*-Acetyl methyl groups were assigned based on nOes to protons H5 and H7. Chemical shift data were consistent between **I** and **II** as can be seen by comparing the reducing and terminal residues of the di- and trisialosides, Tables 4.2-4.3.

4.2.2.2. nOe measurements

The nOe data were collected at high field (800 and 900 MHz) and 10°C to promote a stronger negative nOe. The 2D NOESY spectra of the disialosyl and trisialosyl haptens are shown in Figures 4.3 and 4.4, respectively. Due to the high degree of spectral overlap, only a limited number of nOes could be confidently quantified from the many observed. A 1D selective NOESY spectrum of the trisaccharide, Figure 4.5, clearly indicated a long-range nOe between protons H9 of the middle residue and H7 of the reducing residue. Peaks in 2D NOESY datasets, processed using NMRView, were integrated by defining boundary boxes around the peaks and summing points within the box according to the NMRView analysis function. 1D nOe data were integrated using standard Varian functions. Integrated values were used for distance calculations based on the isolated spin-pair approximation, wherein the H3eq to H3ax distance was set to 1.79 Å.[15] Cross-peaks originating from the same source proton were then compared for their respective distances.

4.2.2.3. *J*-coupling measurements

Extracting accurate scalar coupling constants from strongly coupled protons frequently found in carbohydrates can be achieved by a combination of selective 1D TOCSY spectra, followed by a simulation of the resulting spin system. In the *Nm*B haptens, selective TOCSY methods were limited by the poor transfer from protons H6 to H7, and from overlapping resonances that prevented access to the important H7, H8 and H9 protons. Thus a 2D HSQC spectrum was acquired with high digital resolution in the proton dimension. The high-resolution

HSQC data were Fourier transformed with additional linear prediction and zero-filling in T2 and T1 in order to maximize digital resolution to 1.0 Hz per point. Traces through the carbon-proton peaks provided, in most cases, proton sub-spectra showing first-order coupling. *J*-values were measured directly from these traces.

4.3. Results and Discussion

4.3.1. NMR and conformational properties

The conformational properties of the α-(2,8)-linked Neu5Ac residues are largely determined by the rotatable bonds between the pyranose rings.[15,62,284] Previous MD simulations of **I** indicated some pyranose ring interconversions from the predominant $^2C_5$ chair conformation,[258] which were rationalized only via non-aqueous phase experimental and molecular mechanics studies. However, there were no indications that these ring fluctuations influenced the conformational properties of the dimer.[258] In the present work, the pyranosyl rings remained entirely in the $^2C_5$ conformation although the ring torsion angles displayed modest oscillations of approximately (±15º). Therefore, the conformational properties of the methyl α-(2,8)-di- and trisialosides were characterized here by monitoring the time-dependent evolution or populations of the φ, ψ, ω_8 and ω_7 angles (see Figure 4.1) during the MD and REMD simulations.

In order to ascertain whether the results from the present simulations were consistent with experiment, and so could be relied on to shed some light on the conformational properties of the strain B haptens, inter-proton distances and scalar $^3J_{HH}$-coupling constants were computed from the MD trajectories and compared with experimental data determined herein. The theoretical scalar *J*-coupling constants were computed employing a Karplus equation that takes into account the electronegativities and orientations of substituent groups.[240]

4.3.1.1. Methyl α-(2,8)-disialoside (a-b-OMe)

The experimental NMR coupling constants and those computed from the MD simulations of **I** are presented in Table 4.4. It can be seen that the values derived from the MD and REMD simulations were statistically indistinguishable, and the back-calculated $^3J_{HH}$ values showed good overall agreement with the experimental coupling constants. In terms of conformational properties, the small experimental $^3J_{bH6-bH7}$ and $^3J_{bH7-bH8}$ values of <1.0 and 1.5 Hz, respectively, for these sets of protons in the internal reducing Neu5Ac residue, were consistent with a ±*gauche* orientation, due to the symmetry of the Karplus relationship.[240] The precise rotamer assignment required the utilization of nOe data, Table 4.5. The strong nOes measured between the bH8-bH6 and bH8-bH7 protons, confirmed the –*gauche* and +*gauche* conformations for the bH6-bC6-bC7-bH7 and bH7-bC7-bC8-bH8 dihedral angles, respectively, which corresponded to a +*gauche* orientation for both the ω_7 (bO6-bC6-bC7-bO7) and ω_8 (bO7-bC7-bC8-bO8) angles. In the non-reducing terminal residue, the small magnitude of the $^3J_{aH6-aH7}$ coupling constant (1.4 Hz) for the glyceryl side chain indicated a +*gauche* conformation for the ω_7 (aO6-aC6-aC7-aO7) angle. In contrast, the large $^3J_{aH7-aH8}$ value of 9.5 Hz for this residue suggested an *anti* orientation for these protons, which corresponded to an *anti*-orientation for this ω_8 (aO7-aC7-aC8-aO8) torsion angle. These results are similar to those reported previously and indicate that the orientation of ω_8 depends heavily on whether O8 is glycosylated or free.[15,267] The MD and REMD simulations gave similar nOe values, which compared well with the experimentally-derived distances, Table 4.5. The inter-glycosidic nOe distances were largely determined by the linkage dihedral angles; therefore the fact that the MD simulations were sampling experimentally consistent torsion space was also supported by the good agreement between the experimental and the MD-derived inter-glycosidic nOe distances. For example the nOe distances measured

between the aH3ax–bH8 protons were 2.9 ± 0.7 and 3.0 ± 0.8 Å for the MD and REMD, respectively, which may be compared with the experimental value of 2.7 Å. The experimentally-derived nOe distance between the aH3eq–bH8 protons was 4.0 Å, compared to 3.9 ± 0.4 and 3.0 ± 0.4 Å for the MD and REMD simulations, respectively.

The conformational itineraries of the glyceryl exocyclic dihedral angles from the 100-ns MD simulation of **I** are presented in Figure 4.6. Despite the potential for the glycosidic linkage to populate a wide range of conformations, the data from the MD and REMD simulations suggested that well-defined subsets of conformations were preferentially adopted. The ω_7 angle predominantly sampled the +*gauche* rotamer in both Neu5Ac residues. In contrast, the ω_8 angle exhibited a preference for several staggered rotamers depending on whether the O8 atom was involved in a glycosidic bond. Consistent with the experimental vicinal $^3J_{\text{H7-H8}}$ coupling constants, the ω_8 angle was predicted to exist predominantly in the *anti* orientation when the O8 atom was not involved in a glycosidic bond, but preferred the *gauche* orientation otherwise. The predominance of the *anti* conformation when O8 is non-glycosidically linked, has been observed in X-ray crystal structures[59,180] and inferred from solution-phase NMR[15,57,59,62] data. Two intra-residue hydrogen bonds, between the oxygen atoms of the carboxylate group with the hydroxyl group at C8, and between the carbonyl group of the acetamido functionality and the hydroxyl group at C7, seem to be the origin of this rotamer's stability.[57] However, the magnitude of this stabilization is not clear, since a similar *anti* orientation for the ω_8 angle has been observed in the X-ray crystal structures of the β-anomer,[59] in which the formation of the intra-residue hydrogen bond to the carboxylate group is impossible.

The φ-angle in Neu5Ac residues is known to populate two conformations, -*gauche* and *anti*.[56,57] There are no conclusive experimental data on the solution conformational properties of

the ψ angle; however, potential energy calculations suggest that the ψ angle oscillates between 105-130°.[15] Over the course of the MD simulations, correlated motions were observed between the conformational itineraries of the ω_8, and φ and ψ dihedral angles of the glycosidic linkage region. During the first 65 ns, the ω_8 angle predominantly populated the +*gauche* rotamer (78 ± 13°), the φ angle displayed transitions between its two solution preferred conformations, (-161 ± 14°, -91 ± 17°), while the ψ angle adopted a value of 98 ± 18°. After 65 ns the ω_8 angle effected a transition to the *anti* rotamer. During the time interval that the ω_8-*anti* rotamer persisted, if the φ angle adopted either the -*gauche* (-91 ± 17°) or *anti* (-161 ± 14°) conformations then a +*gauche* ψ (98 ± 18°) would introduce unfavorable electrostatic repulsions due to the close proximity of the carboxyl groups and ring oxygen atoms on two adjacent Neu5Ac residues, Figure 4.7. Presumably to alleviate these interactions, the ψ angle flips to the *anti* rotamer, (142 ± 12°). In general, the preferred conformational preferences of the exocyclic dihedral angles observed during the MD simulations are in accordance with previous reports from experimental NMR studies of the respective angles.[15,56,267,268] The majority of the transitional correlations between the ω_8, and φ and ψ angles were also observed in the REMD simulation. Comparisons of the relative distributions of the rotamers of the exocyclic glyceryl side chain computed from experimental *J*-values, as well as the MD and REMD simulations are shown in Table 4.6. For the conformationally significant dihedral angles, particularly ω_7 and ω_8, the theoretical populations showed good agreement with experiment. Only a qualitative agreement was observed between the theoretical and experimental populations for the freely rotating ω_9 angle.

4.3.1.2. Methyl α-(2,8)-trisialoside (a-b-c-OMe)

Methyl α-(2,8)-trisialoside (**II**) was included in this study to investigate the effects of multiple Neu5Ac residues on the conformational properties of the 2,8-linkage. To the best of our

knowledge only two experimental studies have been carried out on the trisaccharide hapten.[15,270] In those studies the trisaccharide fragment terminated in the β-configuration, limiting the relevance of those data. The vicinal homonuclear $^3J_{HH}$-coupling constants in **II**, determined experimentally and those from our MD simulations, are presented in Table 4.7. The observed $^3J_{H6-H7}$ constants for all the Neu5Ac residues from the experiment were less than 1.5 Hz, which indicated a *gauche* relationship between protons H6 and H7. The average $^3J_{H6-H7}$ values computed from the MD-simulation were also of low magnitude, and were in excellent agreement with the experimental values. As in the case of the disaccharide, the large experimental $^3J_{aH7-aH8}$ value of 9.6 Hz indicates that these protons are predominantly in the *anti* orientation for the non-reducing terminal residue. The MD-computed value for $^3J_{aH7-aH8}$ (7.7 ± 0.6 Hz) also suggested a predominantly *anti* orientation, which upon further analysis was confirmed to be 78% *anti* and 22% *gauche*. In contrast to the large $^3J_{aH7-aH8}$ values observed in the non-reducing terminal residue, the $^3J_{H7-H8}$ values for the internal (b) and reducing-internal (c) residues, in which O8 is involved in glycosidic bonds, were less than 4.0 Hz, indicating a conformational change to the *gauche* orientation. That a *gauche* rotamer was preferred for the a-b and b-c linkages, respectively, was also demonstrated by the relatively small magnitude of the $^3J_{H7-H8}$ coupling constants observed in the MD data (2.1 ± 0.7 and 2.1 ± 0.8 for the MD and REMD simulations, respectively). In addition, the small-magnitude of the $^3J_{H7-H8}$ values for the residues b and c imply similarities between the conformational properties of the ω_8 angle in both the a-b and b-c linkages. In previous experimental NMR studies of the Neu5Ac trisaccharide[15,62] the anomeric center of the reducing-terminal Neu5Ac residues was in the β-configuration, which led to the conclusion that the conformational properties of the ω_8 angle in the two linkage regions differed.[267] Here, for the all α-configuration, we see that this is not the case. Good agreement was

also obtained between the experimental and MD-computed values for the $^3J_{\text{H8-H9R}}$ and $^3J_{\text{H8-H9S}}$ couplings; however, the rotameric preferences of the C8-C9 bond is not directly pertinent to the overall 3D conformational properties of the CPS. nOe distances were computed from the MD and REMD simulations of the trimer and compared with the available experimental data, Table 4.8. Unlike the case of the dimer, the agreement with the experimental values in the trisaccharide was only qualitative, suggesting a need for longer sampling time. The nOe values from the REMD simulation were in better agreement relative to the MD simulation, except for one case, bH3eq–cH7. This observation suggests that enhanced-sampling methods may be more efficient in studying the conformational properties of larger fragments of polysaccharides. Rotamer populations computed from the experimental *J*-values and directly from the MD simulations are shown in Table 4.9. In general, the agreement with experiment was excellent with the exception of the ω_8 populations of the internal linkages, for which the theoretical data, nevertheless, correctly ranked the rotameric preferences.

The time series for the rotational angles of the glyceryl side chain in **II** over the 100-ns simulation are presented in Figure 4.8. In all three Neu5Ac residues the ω_7 angle populated only the +*gauche* rotamer. The ω_8 angle predominantly adopted the *anti* conformation in the glyceryl side chain of the non-reducing terminal residue, while in the internal and reducing-terminal residues the +*gauche* rotamer was preferred. Over the course of the simulation, the ψ angle oscillated around $101 \pm 30°$ and $109 \pm 27°$ for the a-b and b-c linkages, respectively, while the φ angle frequently populated the *anti* and -*gauche* rotamers in both linkage regions. As in the case of the linkage properties of **I**, correlated motions were observed between the ω_8, and φ and ψ dihedral angles, which were strongly evident for the terminal (a-b) linkage in **II**. During the first 70 ns, the ω_8 angle predominantly populated the +*gauche* ($76 \pm 12°$) rotamer and the φ angle

150

displayed transitions between its two solution preferred conformations (-159 ± 14° and -84 ± 19°). After 70 ns the ω_8 dihedral angle effected a transition to the *anti* (174 ± 15°) rotamer. The presumed unfavorable electrostatic repulsions, discussed earlier in the disaccharide, again resulted in the ψ and φ angles effecting conformational transitions to their respective *-gauche* rotamers, -80 ± 8° and -104 ± 14°, respectively. However, unlike in the case of the disaccharide, these rotamers persisted for the remainder of the simulation. Despite the plausible explanation for the existence of these rotamers observed during the last 30 ns of the traditional MD simulation, we sought to determine whether they represented experimentally relevant conformations.

Therefore, *J*-values and inter-proton distances were computed pre- and post-conformational transition of the ω_8 angle, the first 70 and last 30 ns, respectively. A closer observation of the *J*-values computed from the traditional MD and REMD simulations indicated that the *J*-values were statistically similar except for the ${}^3J_{\text{bH7-bH8}}$ values, where the φ, ψ and ω_8 transitions occurred. The value computed from the traditional MD simulation was about a factor of two (3.6 ± 0.9 Hz) greater than that determined from the REMD simulation (1.9 ± 0.9 Hz). The higher value may be due to contributions from the *anti* rotamers of ω_8. The *J*-values computed before and after the conformational transition are presented in Table 4.10. These values are statistically similar and in good agreement with the experimental data except the ${}^3J_{\text{bH7-bH8}}$ values that are 1.4 ± 0.8 Hz and 8.7 ± 0.1 Hz for the first 70 and last 30 ns, respectively. The large value of 8.7 ± 0.1 Hz suggests that the last 30 ns may not be entirely experimentally relevant. However, the value of 1.4 ± 0.8 Hz is still small relative to the experimental value of < 4.0 Hz, suggesting that the conformational properties of the ω_8 angle may be described by

contributions from both the +*gauche* and *anti* rotamers, with a predominance of the +*gauche* rotamer.

The inter-proton distances prior to and after the conformational transitions are presented in Table 4.11. It should be noted that the inter-proton distances were statistically indistinguishable in the b-c linkage for both segments of the trajectory. However, notable differences were observed in the a-b linkage. For example, the aH3ax-bH8 distance during the first 70 ns was 3.1 ± 0.8 Å, which compared well with the experimental value of 3.0, relative to the distance of 4.1 ± 0.2 Å computed during the last 30 ns. The bH6-bH8 and aH3eq-bH8 were also in better agreement with the experimental data during the first 70 ns of the simulation compared to the last 30 ns. Coupled with the *J*-values, these inter-proton distances indicate that during the last 30 ns of the traditional MD simulation experimentally non-relevant conformations were sampled.

4.3.2. Conformational clustering

In order to identify unique global conformational states and correlated internal rotations, the trajectories of **I** were subjected to a clustering analysis.[262] Another key attribute of such an analysis is the accessibility of pertinent information about the interconversion pathways between conformational states, which cannot be easily discerned from traditional torsion angle scatter plots or from time dependent trajectory representations of dihedral angles. The conformer families were generated on the basis of inter-residue dihedral angles: φ, ψ, ω_8 and ω_7.

A logarithmic plot of the fastest transition time scales τ_l (relative to the time scale of the 25 dimensional model τ_{25}) for the hierarchy of Markov states obtained from the trajectory of **I** is shown in Figure 4.9A. The largest changes in time scale suggest a four-state model is adequate to describe the conformational space. The representative values of the glycosidic linkages of the

four clusters present in the four-state model, and the abundance of each state, are presented in Table 4.12. Two states differing in more one than dihedral angle will effect transitions through correlated or concerted changes, while states with differences in a single angle undergo interconversions through isolated transitions. Transition probabilities were generally low in the case of strongly coupled or correlated transitions between two or more torsion angles, which may be ascribed to the overall additive barrier to interconversion. Examination of the four-state transition matrix (Figure 4.9) indicates that state 1 (red) undergoes less frequent transitions to state 3 (blue) than to state 4 (yellow). The principal difference between states 1 and 3 lies in the ψ angles with values of 71 and 128°, respectively. States 1 and 4 differ in the (φ, ψ) domain with values of (-97°, 71) and (-167°, 106°), respectively. Given that there are two concerted transitions in the 1-4 pathway, it might be expected that the 1-3 pathway would be preferred. However, the transitions between states 1 and 4 occur with higher probability. This discrepancy is explained in terms of the low barrier for the transitions between the *-gauche* and *anti* rotamers of the φ angle involved in the 1-4 route. State 2 (green) is characterized by an *anti* (173°) orientation of the ω_8 angle, and by an increase in the ψ angle (143°). This is an isolated state, as indicated by the extremely low transition frequency (Figure 4.9), which is accessed essentially only from state 3 (blue). The reason for its isolation arises in the need for simultaneous changes in ω_8 and ψ, to compensate for the electrostatic repulsions that would otherwise result from a single transition in ω_8 (Figure 4.7). The 1-2 pathway involves three concerted transitions (φ, ψ and ω_8) and would be disfavored, compared to two concerted changes (φ and ω_8) and (ψ and ω_8) in the 2-3 and 2-4 routes, respectively. The 2-3 pathway is more probable compared to the 2-4 pathway, because the 2-3 route involves the φ angle oscillating frequently between the *-gauche* and *anti* rotamers. Pathways 3-4 and 1-4 both involve changes in the (φ, ψ) domain. However,

the 3-4 route occurs with a lower probability. It must be the case that the 3-4 pathway is strongly competing with another pathway. Further examination indicated that state 3 also frequently interchanges with state 1, the (1-3) route involving only the ψ angle. In terms of populations, which are indicated by the number of partition volumes in Table 4.12, the isolated state 2 accounted for 8% of the conformer distribution. States 1 and 3 accounted for 16 and 12%, respectively, while state 4 was the dominant state, at 64%.

Plots of the time evolution of the glycosidic angles of the 4-state model during the simulation, and scatter plots of the (φ, ψ) and (ω_8, ω_7) dihedral angles are presented in Figure 4.10. During the first 65 ns and the last 20 ns of the simulation interconversions occurred predominantly between states 1 and 4, with much lower sampling of states 2 and 3, respectively. However, within the 65-80 ns interval, states 2 and 3 were preferentially populated. It should be noted that the major pathway from state 2 was via state 3. The scatter plots, Figure 4.10, indicate that the (φ, ψ)-space is clearly separated into four states for the 4-state model, while the (ω_8, ω_7) space showed an overlap of states 1, 3 and 4.

4.3.3. Effects of chemical group modifications

The effects of chemical group modifications on the conformational properties of **II** were determined by comparing the (φ, ψ) and (ω_8, ω_7) scatter plots of the linkage dihedral angles with those of the N-Gc, N-Pr, N-Prop and N-But analogs. The multiple conformational distributions in the (φ, ψ) scatter plots indicate a significant degree of flexibility in the trisaccharide hapten and its analogs, Figure 4.11. For any given trisaccharide the φ and ψ angles for both the reducing and non-reducing terminal residues displayed similar distributions with the exception that the N-Pr and N-But analogs occupied additional (φ, ψ)-spaces in the non-reducing and reducing linkage regions, respectively. Despite these differences in the number of conformations between the

reducing and non-reducing linkage regions in each analog, subsets of the phase-space were common to both linkage regions, albeit to varying extents. The (φ, ψ)-space of the native trimer, compared to all the analogs, exhibited similar conformational properties in terms of the number and conformation of the states, except for the *N*-propionyl derivative, which displayed a significant population of a conformation with (φ, ψ)-values both between -120° and -60°. The (ω_8, ω_7)-space showed less flexibility in all cases, and considerable phase-space overlaps were seen in all the linkages of the trisaccharides, Figure 4.12. Amongst all the linkages the (ω_8, ω_7)-space showed more rigidity primarily due to the restricted flexibility of the ω_7 angle. However, in the *N*-Prop and *N*-But analogs the ω_7 angle populated isolated families with (ω_8, ω_7) values of about (-300, 180) and (60, 180), respectively. These additional conformational families were unexpected, given the preference for ω_7 to adopt a +*gauche* conformation in the native hapten.

In general the chemical group modifications resulted in similarities in the conformational families in the trisaccharides, although there were conformational states unique to some systems. To provide a structural interpretation for this observation, representative structures of conformational states identified via the Bayesian statistical method were generated. Plots of the relaxation time scales determined for each trajectory, Figure 4.13, indicated that the following number of states could adequately describe the conformational properties of each trisaccharide: **II** (10 states), *N*-Gc (16 states), *N*-Pr (16 states), *N*-prop (14 states) and *N*-But (13 states). The representative conformations of the most populated states of the trisaccharide fragments are presented in Table 4.13. Based on dihedral angle matches, conformer family A of the *N*-Ac trimer will have the highest similarity with conformer family A of the *N*-Gc, *N*-Pr, *N*-Prop and *N*-But analogs, respectively. It is also evident that conformer family B of **II** and *N*-But trimers are close structural matches. However, these angle-matching comparisons may fail to identify

certain global structural similarities. In particular, MD simulations have indicated that oligosaccharides may display similar overall topologies even though their internal linkages possess different dihedral angles,[269] which are due to correlated transitions between the dihedral angles in the linkages. Therefore, the conformational similarities between the representative structures of the most populated clusters of the analogs and the native trisaccharide were also determined by RMSD comparisons. This approach revealed further structural matches that were not identified initially based on angle-similarities, Table 4.14 and Figure 4.14. For example, the A and J conformers of **II** and the *N*-Gc analog, respectively, differ in the φ and ω_8 angles of the b-c linkage. Despite these differences, they share very similar overall structures, Figure 4.14. The relative abundance of the states, Table 4.14, was employed to determine the fraction of the states displaying similar overall topologies. In the case of **II** and the *N*-Gc trisialosides, 24% of the *N*-Gc conformers (A, C, J) exhibited similar backbone topologies as the A conformer of **II** (36%). Another 8% of the conformers of **II** (C), could be mapped to 12% of the *N*-Gc conformers (C), while 8% of the conformers of **II** (E), were similar to 8% of the *N*-Gc conformers (I). Such comparisons were be extended to all the other analogs (Table 4.14).

These topological similarities suggest that a common core conformational epitope is present in the *Nm*B CPS and its analogs. These observations may be utilized to rationalize the results from efforts to enhance the immunogenicity of the *Nm*B CPS by chemically modifying the *N*-acetyl moiety.[16,22-24,26,277-280] In those studies, both IgM[270] and IgG[22-24,274,275] classes of antibodies, respectively, were utilized. Attempts to characterize the epitopes of the *N*-propionylated CPS suggested that the IgG mAbs recognized both short and extended conformational epitopes,[23] and a subset of the mAbs cross-reacted with human PSA and encapsulated *Nm*B.[23,24] In another study[270] IgM mAbs cross-reacted equally with the native *Nm*B

CPS and its derivatives, while different reactivities were noted among other species. In that study *Nm*B CPS specific IgMs were bactericidal and safe when tested in mice. These studies indicate that low affinity IgM's and fully matured IgG's can show cross-reactivity between the native and modified *Nm*B CPS. Regardless of the immunoglobulin class, the observed cross-reactivity or lack thereof may be rationalized via the backbone conformations and the exocyclic group orientations, Figures 4.14 and 4.15. When cross-reactivity is observed the elicited mAbs are recognizing a common topology, Figure 4.14, whose fine specificities do not likely depend on the *N*-acyl groups or on strong interactions with them, Figure 4.16. On the other hand, the absence of cross-reactivity either arises from variations in the overall 3D shape (unlikely based on the data presented here) or from local group incompatibilities. An antibody that has evolved to recognize the non-native *N*-acyl derivatives, such as mAb 13D9, may not be able to make optimal contacts when the native CPS is encountered, despite the presence of a common conformational backbone. This is particularly evident in the comparison between the native structure and the synthetic *N*-Pr or *N*-But analogs, Figure 4.15. The smaller size of the *N*-acetyl group may not be able to interact effectively with a binding pocket deep enough to accommodate the larger *N*-Pr and *N*-But groups. This would explain the observation that although mAb 13D9 requires an extended epitope, as in the native CPS, it is unable to recognize the native CPS.[256] It was reported previously that absence of cross-reactivity between the native CPS and the synthetic *N*-Pr and *N*-But polysaccharides, could be due to the modifications inducing conformational changes in the CPS.[270] In contrast, the MD simulations performed in this work suggest there is significant overlap in the conformational space of the native and chemically modified saccharides, observations that are in accord with experimental NMR data.[16]

Two salient points emerge from all these considerations: cross-reactivity arises between the *N*-acyl analogs when overall backbone conformation is conserved and when the antibody has not matured to recognize the acyl moiety as an immunodominant epitope. Therefore, these simulations suggest that alternative chemical modifications may be sought, subject to the constraints that the integrity of the *N*-acetyl moiety and the backbone core conformations are not compromised. These modifications may involve the introduction of small hydrophobic groups at positions that were not previously derivatized. To the best of our knowledge, no such derivatizations have been performed with the *Nm*B CPS. However, within the accuracy limitations of the force field and MD simulations, it is now feasible to predict the effect of chemical modifications on the CPS structure and to identify the structural similarity between native and synthetic immunogens.

4.4. Conclusion

Because of the absence of a protective vaccine against *Nm*B, and the postulated importance of a conformational epitope in mediating the antigenicity and bactericidal activity related to this CPS, considerable attention has been given to characterizing its 3D shape and dynamics.[15,16,22,260,270,276,277] In addition, the weak immunogenicities of both the protein-conjugated and non-conjugated forms of the *Nm*B CPS have resulted in alternative vaccine design attempts that typically involve chemical group modifications of the acyl groups[16,22-24,26,255] the carboxylate group,[16] or less commonly the search for peptide mimics.[271]

As an initial step in quantifying the antigenicities of native and chemically modified *Nm*B CPS, conformational properties of di- and trisaccharide haptens have been determined by MD simulations and confirmed by NMR spectroscopy. That the MD simulations identified experimentally relevant conformations was demonstrated by the good agreement between the

experimental and theoretical NMR observables. In addition, a Bayesian statistical analysis[262] was employed to identify conformational states from the trajectories. To the best of our knowledge this is the first time such an analysis has been applied to investigate carbohydrate conformational properties. With the preliminary agreement demonstrated here between the MD data and the experimental data for the native haptens, the study was extended to shed light on the effects of chemical group modifications on epitope size and fine structure. These theoretical simulations present a unique opportunity to generate atomic level insight into the conformational space of native and synthetic CPS analogs. It is also suggested that efforts employed in the quest for more immunogenic carbohydrate vaccines that can mimic *Nm*B should consider chemical group modifications that do not perturb the *N*-acetyl moieties or the global conformation. MD simulations provide a much needed tool for the rational design of such synthetic vaccines, and provide a solid basis for the interpretation of the diverse and complex data from immunological studies.

Table 4.1. CPS structures of the five most virulent *N. meningitidis* serotypes

Serotype	Oligosaccharide repeat units	Acetylation
W-135	-6)-α-D-Gal-(1-4)-α-Neu5Ac-(2-	7 and/or 9 *O* partial [248]
Y	-6)-α-D-Glc-(1-4)-α-Neu5Ac-(2-	9 *O* partial [248]
C	-9)-α-Neu5Ac-(2-	7 and/or 8 *O* [8]
B	-8)-α-Neu5Ac-(2-	---- [8]
A	-6)-α-D-ManNAc-(1-OPO$_3$-	3 *O* [249]

Table 4.2. Proton chemical shifts of methyl α-(2,8)-di- and trisialoside, acquired at 25°C in D$_2$O, relative to internal DSS.

	H3ax	H3eq	H4	H5	H6	H7	H8	H9	NAc	OMe
					Methyl α-(2,8)-disialoside					
b	1.565	2.612	3.552	3.783	3.813	3.858	4.194	4.133, 3.693	2.059	3.310
a	1.726	2.759	3.666	3.825	3.610	3.581	3.905	3.872, 3.635	2.022	
					Methyl α-(2,8)-trisialoside					
c	1.572	2.607	3.531	3.801	3.838	3.890	4.144	4.141, 3.668	2.070	3.310
b	1.697	2.674	3.573	3.795	3.602	3.844	4.154	4.182, 3.674	2.062	
a	1.739	2.753	3.656	3.833	3.601	3.573	3.928	3.890, 3.634	2.022	

Table 4.3. Carbon chemical shifts of methyl α-(2,8)-di- and trisialoside, acquired at 25°C in D$_2$O. Referenced indirectly to DSS.

	C3	C4	C5	C6	C7	C8	C9	NAc	OMe
					Methyl α-(2,8)-disialoside				
b	43.06	70.48	55.09	76.81	72.17	81.19	64.26	25.08	54.49
a	43.25	71.15	54.40	75.30	70.81	73.36	65.26	24.83	
					Methyl α-(2,8)-trisialoside				
c	42.95	70.67	55.19	76.37	71.87	80.71	64.10	25.16	54.54
b	42.88	70.88	55.09	76.47	72.08	81.10	64.18	25.09	
a	43.24	71.30	54.44	75.34	70.94	74.45	65.28	24.81	

Table 4.4. NMR and computed $^3J_{HH}$ coupling constants (Hz) for inter-residue torsion angles in fragments of *Nm*B sialobioside, α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b).

Angle	Linkage	Spins	NMR	MD (100 ns)[a]	REMD (20 ns)[a,b]
ω₇	**Terminal**	**aH6-aH7**	**1.4 ± 0.1**	**1.0 ± 0.8**	**1.0 ± 0.8**
	Internal	**bH6-bH7**	**< 1.0**	**1.2 ± 0.9**	**1.2 ± 0.9**
ω₈	**Terminal**	**aH7-aH8**	**9.5 ± 0.9**	**7.9 ± 0.6**	**8.1 ± 0.6**
	Internal	**bH7-bH8**	**1.5 ± 0.2**	**2.0 ± 0.7**	**2.0 ± 0.7**
ω₉	Terminal (free rotor)	aH8-aH9*R*	6.1 ± 0.6	6.7 ± 1.5	5.9 ± 1.5
		aH8-aH9*S*	2.4 ± 0.2	3.7 ± 0.9	3.6 ± 0.9
	Internal (free rotor)	bH8-bH9*R*	6.1 ± 0.6	7.3 ± 1.6	7.4 ± 1.7
		bH8-bH9*S*	4.1 ± 0.4	2.9 ± 1.3	2.9 ± 1.3

[a] Conformation-defining intra-residue *J*-couplings are shown in bold.
[b] REMD simulations were performed at 8 temperatures for a total of 160 ns.

Table 4.5. Experimental (ISPA) NOE distances (Å) and average inter-proton distances from MD and REMD simulations for fragments of *Nm*B sialobioside, α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b).

Spins	NMR	MD 100 ns	REMD 20 ns	Spins	NMR	MD (100 ns)[a]	REMD (20 ns)[a,b]
aH3ax – aH3eq	1.7	1.7 ± 0.1	1.7 ± 0.1	**bH8 – bH6**	**2.6**	**2.5 ± 0.2**	**2.5 ± 0.2**
aH3ax – aH4	2.9	3.0 ± 0.1	3.0 ± 0.1	**bH8 – bH7**	**2.3**	**2.6 ± 0.2**	**2.6 ± 0.2**
aH3ax – aH5	2.6	2.7 ± 0.2	2.7 ± 0.2	**aH3ax – bH8**	**2.7**	**2.9 ± 0.7**	**3.0 ± 0.8**
aH3eq – aH4	2.5	2.5 ± 0.1	2.5 ± 0.1	**aH3eq – bH8**	**4.0**	**3.9 ± 0.4**	**3.4 ± 0.4**
bH3ax – bH4	3.1	3.0 ± 0.1	3.0 ± 0.1	bH5 – bNMe	4.3	4.5 ± 0.1	4.5 ± 0.1
bH3ax – bH5	2.6	2.7 ± 0.2	2.7 ± 0.2	bH9R – bH8	2.3	2.5 ± 0.2	2.9 ± 0.3

[a] Conformation-defining intra-residue nOe's are shown in bold.
[b] REMD simulations were performed at 8 temperatures for a total of 160 ns.

Table 4.6. NMR-based and computed populations (%) for inter-residue torsion angles in fragments of *Nm*B disialoside, α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b).

Angle	Linkage	NMR[a,b] *gauche/trans/-gauche*	MD (100 ns) *gauche/trans/-gauche*	REMD (20 ns)[d] *gauche/trans/-gauche*
ω₇	Terminal a	100/0/0	100/0/0	100/0/0
ω₇	Internal b	100/0/0	100/0/0	100/0/0
ω₈	Terminal a	0/100[c]/0	13/80/7	13/83/4
ω₈	Internal b	96 ± 3/3± 3/0	90/10/0	90/10/0
ω₉	Terminal a	54 ± 8 /0/50 ± 12	61/18/21	52/16/32
ω₉	Internal b	55 ± 8/24 ± 9/20 ± 15	73/2/25	75/4/21

[a] Range is consistent with the estimated error in the experimental *J*-values, only one symmetric solution is presented. [b] NMR populations derived by employing limiting *J*-values from rotational isomeric states computed from traditional MD (the REMD simulations did not give significantly different limiting *J*-values). [c] Estimated based on the magnitude of the *J*-value. [d] REMD simulations were performed at 8 temperatures for a total of 160 ns.

Table 4.7. Conformationally sensitive NMR and computed $^3J_{HH}$ coupling constants (Hz) for inter-residue torsion angles in fragments of *Nm*B trisialoside, α-Neu5Ac-(2-8)-α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b-c).

Angle	Linkage	Spins	NMR	MD (100 ns)[a]	REMD (20 ns)[a,b]
ω_7	Terminal	aH6-aH7	1.5 ± 0.2	1.0 ± 0.8	1.0 ± 0.8
	Internal	bH6-bH7	< 1.0	0.9 ± 0.8	0.9 ± 0.7
	Internal	cH6-cH7	< 1.0	1.1 ± 0.9	1.3 ± 1.0
ω_8	Terminal	aH7-aH8	9.6 ± 1.0	7.7 ± 0.6	8.1 ± 0.5
	Internal	bH7-bH8	< 4.0	3.6 ± 0.9	1.9 ± 0.9
	Internal	cH7-cH8	< 4.0	2.1 ± 0.7	2.1 ± 0.8

[a] Conformation-defining intra-residue *J*-couplings are shown in bold.
[b] REMD simulations were performed at 8 temperatures for a total of 160 ns.

Table 4.8. Experimental (ISPA) NOE distances (Å) and average inter-proton distances from MD and REMD simulations for fragments of *Nm*B trisialoside, α-Neu5Ac-(2-8)-α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b-c).

Spins	NMR	MD (100 ns)[a]	REMD (20 ns)[a,b]	Spins	NMR	MD (100 ns)[a]	REMD (20 ns)[a,b]
bH3ax – cH8	2.6	3.3 ± 0.8	2.6 ± 0.7	bH6 – bH8	2.3	2.6 ± 0.3	2.5 ± 0.1
bH3eq – cH8	3.5	4.1 ± 0.5	3.8 ± 0.5	aH3ax – bH8	3.0	3.4 ± 0.8	2.8 ± 0.7
bH3eq – cH7	3.9	4.7 ± 0.4	5.1 ± 0.3	aH3eq – bH8	3.4	4.2 ± 0.4	3.8 ± 0.4
bH6 – bH7	2.2	2.4 ± 0.1	2.4 ± 0.1				

[a] Conformationally insensitive nOe's for the trisialoside were similar to the disialoside values, and are omitted. [b] REMD simulations were performed at 8 temperatures for a total of 160 ns.

Table 4.9. NMR-based and computed populations (%) for inter-residue torsion angles in fragments of *Nm*B trisialoside, α-Neu5Ac-(2-8)-α-Neu5Ac-(2-8)-α-Neu5Ac-OMe (a-b-c).

Angle	Linkage	NMR[a,b] *gauche/trans/-gauche*	MD (100 ns) *gauche/trans/-gauche*	REMD (20 ns)[d] *gauche/trans/-gauche*
ω_7	Terminal a	100/0/0	100/0/0	100/0/0
ω_7	Internal b	100/0/0	100/0/0	100/0/0
ω_7	Internal c	100/0/0	100/0/0	100/0/0
ω_8	Terminal a	0/100[c]/0	15/74/7	14/82/4
ω_8	Internal b	60 ± 9 / 25 ± 3 / 0	70 / 30 / 0	92/8/0
ω_8	Internal c	66 ± 5 / 31 ± 9 / 0	89 / 11 / 0	82/10/8

[a] Range is consistent with the estimated error in the experimental *J*-values, only one symmetric solution is presented. [b] NMR populations derived by employing limiting *J*-values from rotational isomeric states computed from traditional MD (the REMD simulations did not give significantly different limiting *J*-values). [c] Estimated based on the magnitude of the *J*-value. [d] REMD simulations were performed at 8 temperatures for a total of 160 ns.

162

Table 4.10. NMR and computed $^3J_{HH}$ coupling constants (Hz) for inter-residue torsion angles in fragments of *Nm*B trisialoside during the first 70 ns and last 30 ns of the traditional MD simulation.

Pre-conformational transition (the first 70 ns)

Angle	Linkage	Spins	NMR	MD
ω_7	Terminal	aH6-aH7	1.5 ± 0.2	1.0 ± 0.8
	Internal	bH6-bH7	< 1.0	0.9 ± 0.7
	Internal	cH6-cH7	< 1.0	1.1 ± 0.9
ω_8	Terminal	aH7-aH8	9.6 ± 1.0	7.6 ± 0.6
	Internal	bH7-bH8	< 4.0	1.4 ± 0.8
	Internal	cH7-cH8	< 4.0	2.4 ± 0.7

Post-conformational transition (the last 30 ns)

Angle	Linkage	Spins	NMR	MD
ω_7	Terminal	aH6-aH7	1.5 ± 0.2	1.0 ± 0.8
	Internal	bH6-bH7	< 1.0	1.0 ± 0.8
	Internal	cH6-cH7	< 1.0	1.2 ± 0.9
ω_8	Terminal	aH7-aH8	9.6 ± 1.0	7.8 ± 0.6
	Internal	bH7-bH8	< 4.0	8.7 ± 0.1
	Internal	cH7-cH8	< 4.0	1.4 ± 0.7

Table 4.11. Experimental (ISPA) NOE distances (Å) and average inter-proton distances from MD and REMD simulations for fragments of *Nm*B trisialoside, during the first 70 ns and last 30 ns of the traditional MD simulation.

Pre-conformational transition (the first 70 ns)

Spins	NMR	MD	Spins	NMR	MD
bH3ax – cH8	2.6	3.3 ± 0.8	bH6 – bH8	2.3	2.4 ± 0.2
bH3eq – cH8	3.5	4.2 ± 0.5	aH3ax – bH8	3.0	3.1 ± 0.8
bH3eq – cH7	3.9	4.7 ± 0.4	aH3eq – bH8	3.4	4.0 ± 0.4
bH6 – bH7	2.2	2.4 ± 0.1			

Post-conformational transition (the last 30 ns)

Spins	NMR	MD	Spins	NMR	MD
bH3ax – cH8	2.6	3.1 ± 0.9	bH6 – bH8	2.3	3.0 ± 0.3
bH3eq – cH8	3.5	4.0 ± 0.4	aH3ax – bH8	3.0	4.1 ± 0.2
bH3eq – cH7	3.9	4.7 ± 0.4	aH3eq – bH8	3.4	4.5 ± 0.1
bH6 – bH7	2.2	2.4 ± 0.1			

Table 4.12. The conformations (°) of representative members of the four clusters in the four-state Markov model in the methyl α-(2-8)-disialoside

State	$\varphi_{(C1C2O8C8)}$	$\psi_{(C2O8C8C7)}$	$\omega_{8(O8C8C7O7)}$	$\omega_{7(O7C7C6O6)}$	Populations from partitions volumes (%)[a]
1 (red)	-97	71	73	65	16
2 (green)	-142	143	173	62	8
3 (blue)	-82	128	78	67	12
4 (yellow)	-167	106	74	69	64

[a] Provides an estimate of what fraction of the data set belongs to a particular conformation. A total of 25 partition volumes were employed.

Table 4.13. Linkage torsion angles of representative members of the predominant clusters populated during the REMD simulations of the native trimer and its *N*-acetyl substituted analogs (a-b-c-OMe)

Conformer	a-b	a-b	a-b	a-b	b-c	b-c	b-c	b-c
	φ	ψ	ω_8	ω_7	φ	ψ	ω_8	ω_7
	(C1C2O8C8)	(C2O8C8C7)	(O8C8C7O7)	(O7C7C6O6)	(C1C2O8C8)	(C2O8C8C7)	(O8C8C7O7)	(O7C7C6O6)
N-acetyl								
A (36%)	-170	106	78	70	-174	115	78	75
B (8%)	-73	132	77	62	-174	112	76	74
C (12%)	-159	103	76	65	-82	96	76	67
D (12%)	-173	108	73	67	-131	147	-176	59
E (8%)	61	121	143	62	-164	128	-87	65
F (8%)	-157	141	179	54	-178	100	87	75
N-glycolyl								
A (8%)	-175	109	65	59	178	115	-67	60
C (8%)	-177	90	92	67	-115	81	36	60
G (8%)	-64	139	-49	59	-178	109	-73	61
H (16%)	-179	112	-67	54	-76	125	-62	63
I (8%)	-175	109	65	59	178	115	-67	60
J (8%)	-171	106	73	65	-72	123	-67	61
K (8%)	-73	137	-93	59	-94	131	-93	60
N-propionyl								
A (12%)	-173	106	76	69	-172	114	74	72
L (8%)	-174	133	153	53	-175	109	-67	60
M (8%)	-169	104	81	71	-76	120	-70	59
N (20%)	-172	108	73	65	-179	121	-71	56
O (8%)	-177	113	-68	54	-73	126	-65	62
N-propyl								
A (12%)	-169	101	78	65	-176	112	74	72
P (12%)	-160	108	-8	64	-81	127	96	62
Q (8%)	-170	101	77	61	-179	117	-67	61
R (16%)	-170	101	75	65	-75	125	-64	62
S (12%)	-177	113	-74	47	-73	128	-61	64
T (8%)	-176	113	-75	47	-71	132	-68	-164
N-butanoyl								
A (32%)	-172	113	75	66	-173	113	77	76
B (8%)	-74	118	77	64	-171	112	76	76
U (12%)	-85	112	131	46	-176	96	77	160
V (8%)	-94	71	69	62	-174	110	75	74
W (8%)	-174	72	60	57	-158	109	53	74

Table 4.14. Backbone[a] RMSD (Å) comparison between the representative conformations of the native trimer and those of the analogs, populated during the REMD simulations

N-glycolyl	*N*-acetyl					
	A[b]	B	C	D	E	F
A	0.79	1.14	1.25	1.66	1.87	1.46
C	1.00	1.30	0.69	1.29	1.51	1.37
G	2.81	2.53	2.28	2.20	2.14	2.35
H	1.83	1.75	1.90	2.16	2.22	1.73
I	1.27	1.46	1.38	1.10	1.26	1.67
J	1.39	1.76	1.85	1.92	2.09	1.94
K	2.01	1.98	1.83	1.90	1.70	1.75
N-propionyl						
A	0.72	1.19	1.21	1.61	1.85	1.45
L	1.85	1.67	1.40	1.17	1.25	1.31
M	1.50	1.88	1.91	1.90	2.06	2.00
N	1.28	1.49	1.31	0.98	1.18	1.65
O	1.80	1.78	1.90	2.13	2.17	1.73
N-propyl						
A	0.49	1.29	1.61	1.91	2.03	1.58
P	1.86	1.78	1.32	1.22	1.67	2.15
Q	1.28	1.60	1.42	1.01	1.17	1.77
R	1.52	1.92	1.90	1.82	1.94	2.01
S	1.86	1.78	1.93	2.12	2.11	1.68
T	2.07	1.95	2.07	2.16	2.24	1.82
N-butanoyl						
A	0.48	1.00	1.18	1.60	1.83	1.40
B	1.17	0.35	1.23	1.67	2.02	1.23
U	1.22	1.10	1.63	1.80	2.04	1.37
V	0.87	0.92	1.39	1.75	2.13	1.59
W	0.84	1.40	1.55	1.86	2.08	1.84

[a] Pyranose ring and glycosidic linkage atoms. [b] Highlighted entries indicate the best structural matches (see also Figure 4.14).
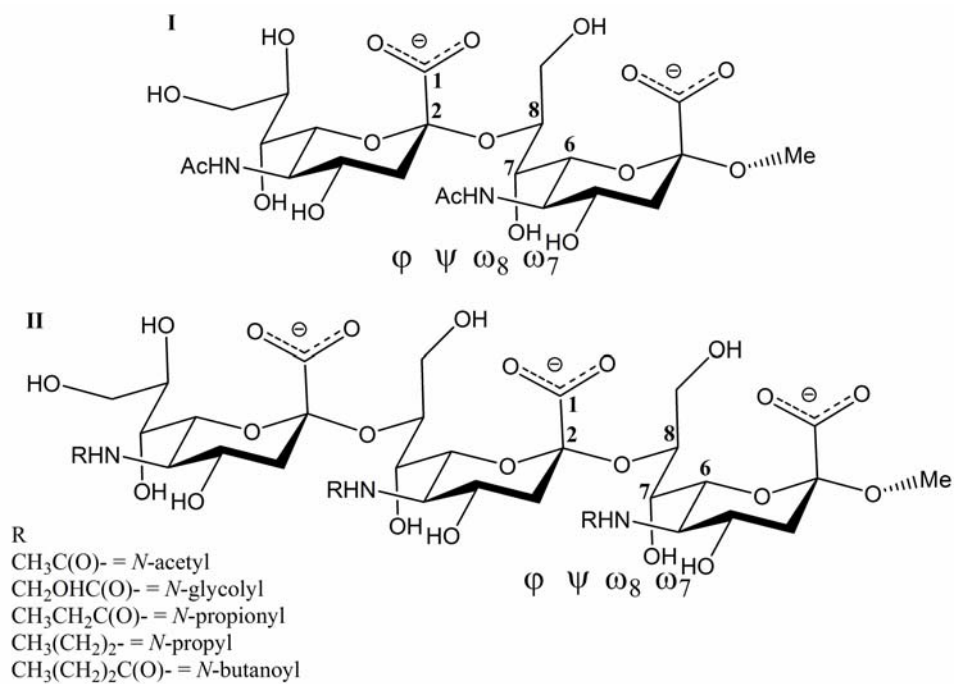
Figure 4.1. Schematic representation of the α-2,8-linked *N*-acetylneuraminic acid dimer (**I**) and trimer (**II**). The dihedral angles are defined as follows: φ = C1-C2-O8'-C8', ψ = C2-O8'-C8'-C7', ω_8 = O8-C8-C7-O7 and ω_7 = O7-C7-C6-O6
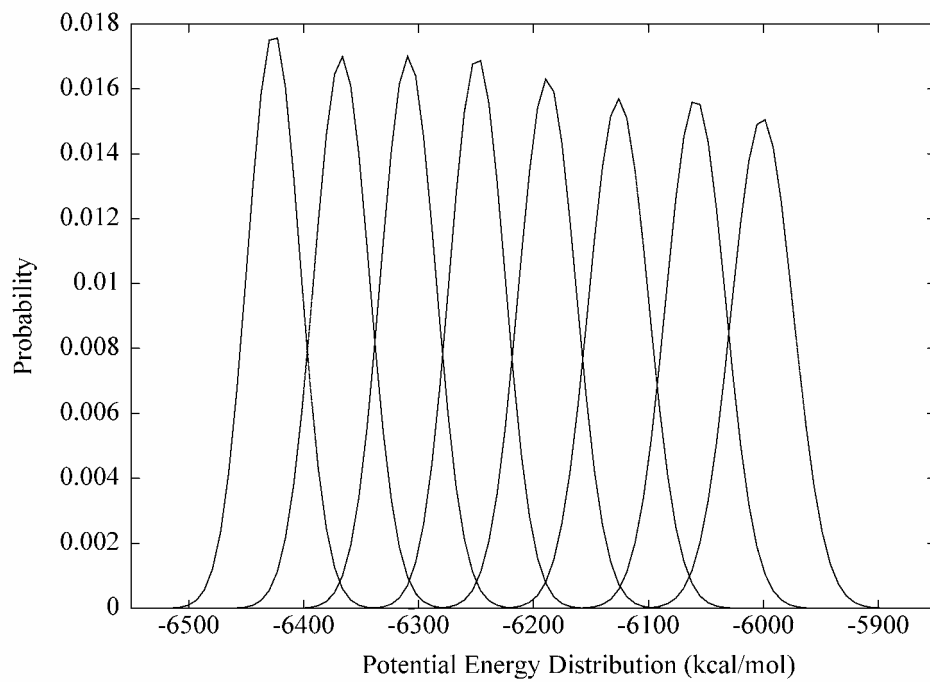
Figure 4.2. Gaussian potential energy distributions indicating the feasibility of the temperatures employed in the replica-exchange simulation
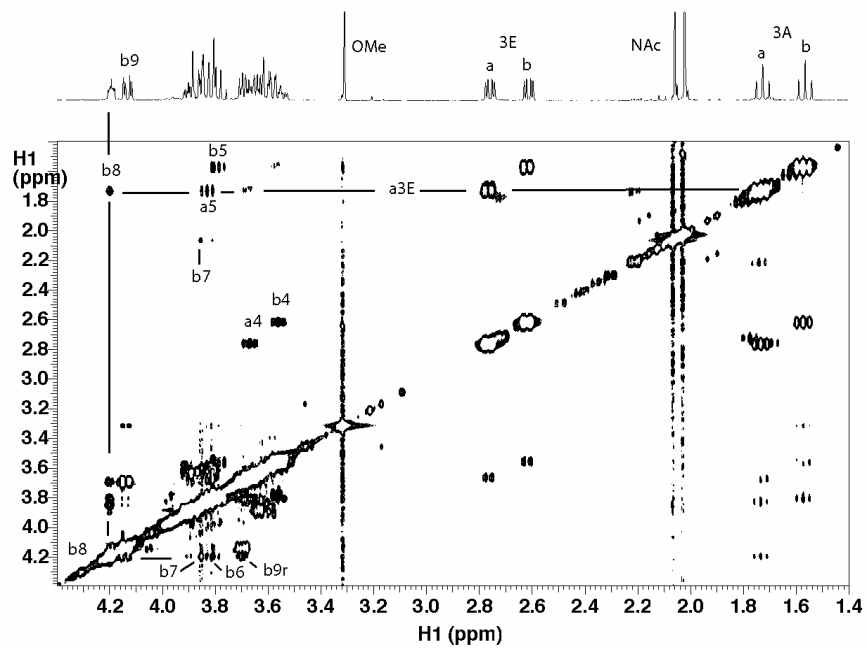
Figure 4.3. NOESY spectrum of disialoside collected at 10° C and 800 MHz with a mixing time of 300 ms. A few cross-peaks used in determining the linkage conformation are indicated
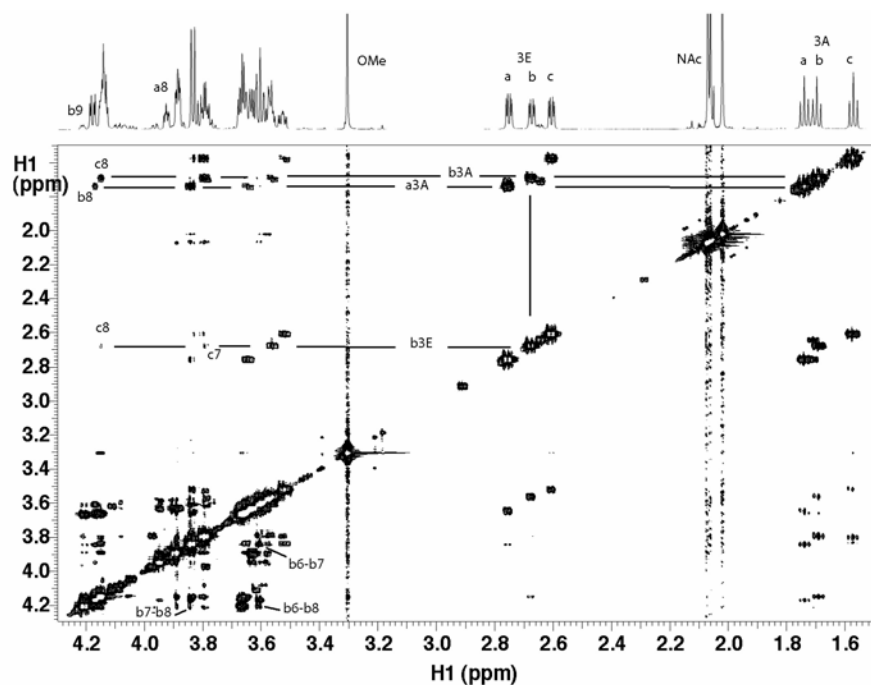
Figure 4.4. NOESY spectrum of trisialoside collected at 10° C and 800 MHz with a mixing time of 300 ms. A few key cross-peaks used in determining the linkage conformation are indicated. Notably, weak cross-peaks between the equatorial H3 proton of the middle residue (bH3eq) and H7 and H8 of the reducing residue (cH7 and cH8) can be observed. Despite substantial overlap of signals, distinct cross-peaks can be measured for some ring protons (e.g. bH6-bH8) of the middle residue
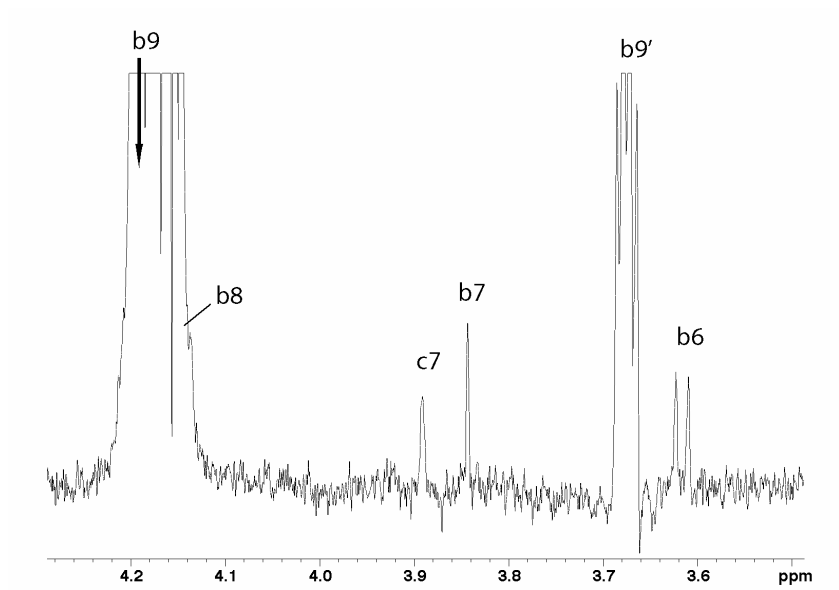
Figure 4.5. The 1D nOe from proton H9 of the middle residue (bH9) showing expected intra-residue nOes to adjacent protons (bH6, bH7, bH8, bH9') but also an unexpected long-range nOe to the H7 of the reducing residue (cH7)
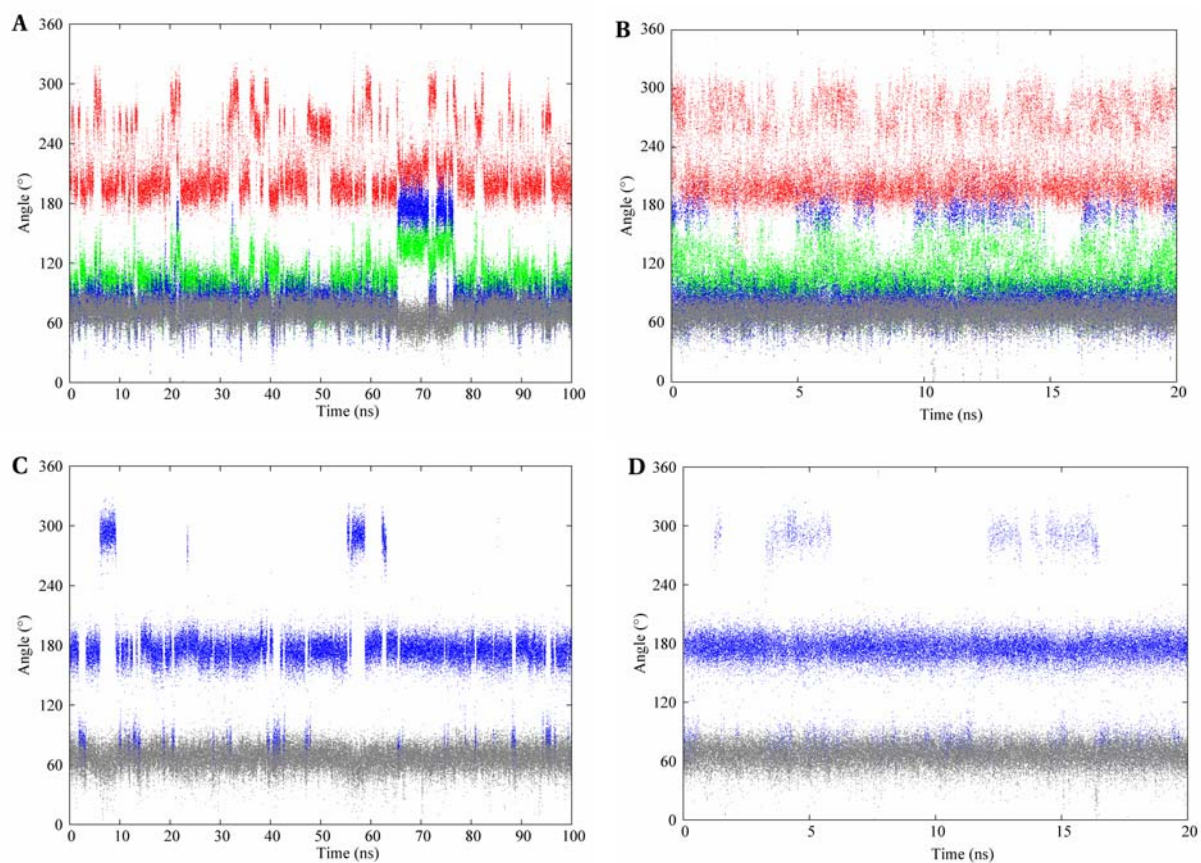
Figure 4.6. The conformational variations of the exocyclic glyceryl dihedral angles of the methyl α-(2,8)-disialoside. φ – red; ψ – green; ω_8 – blue; ω_7 – gray. The A–B and C–D panels depict dihedral angle variations of the glyceryl region involved and not involved in glycosidic bond formation, respectively. The left and right columns represent trajectories from the traditional MD and REMD simulations, respectively
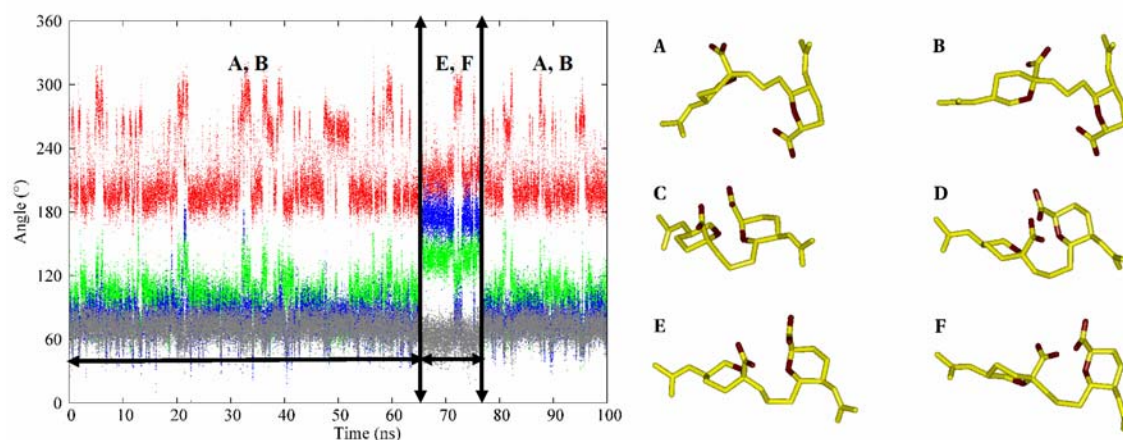
Figure 4.7. Possible repulsive interactions between the charged carboxyl groups and the O6 ring oxygen atoms on two adjacent residues. $(\varphi, \psi, \omega_8) = (-161, 98, 78)$ and $(-91, 98, 78)$ in A and B, respectively, denoting the solution-preferred structures during the first 65 ns of the simulation. $(\varphi, \psi, \omega_8) = (-161, 98, 175)$ and $(-91, 98, 175)$ in C and D, respectively, represent unfavorable interactions that would occur if the ω_8 angle adopted the *anti* rotamer while the ψ angle remained in the +*gauche* (98°) state. To alleviate the repulsive interactions, the ψ angle makes a transition to the *anti* rotamer, E and F with $(\varphi, \psi, \omega_8) = (-161, 142, 175)$ and $(-91, 142, 175)$, respectively
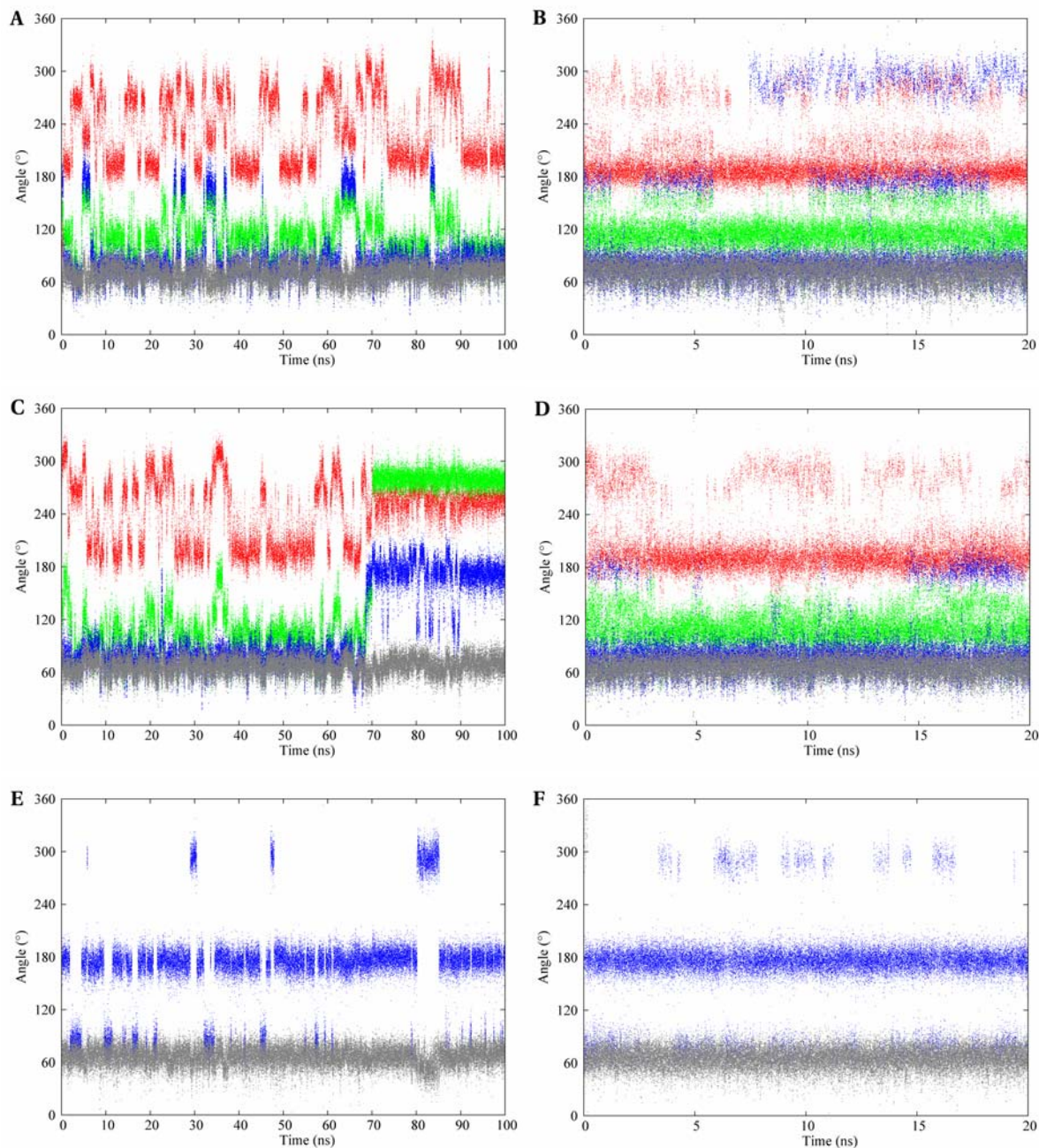
Figure 4.8. The conformational variations of the exocyclic glyceryl dihedral angles of the methyl α-(2,8)-trisialoside. φ – red; ψ – green; ω_8 – blue; ω_7 – gray. The top, middle and bottom panels represent trajectories from the reducing, internal, and terminal glyceryl side chains, respectively. The left and right columns represent the traditional MD and REMD simulations, respectively
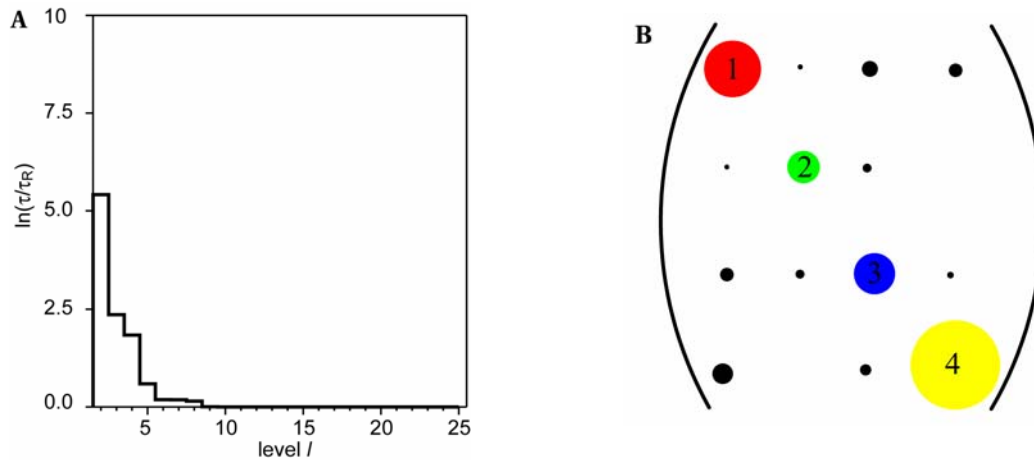
Figure 4.9. Plot of the relaxation time-scale, $\ln\tau/\tau_R$, between the various hierarchic Markov states identified from the MD-generated trajectories of the methyl α-(2,8)-disialoside, A. A graphical representation of the transition matrix between the Markov states when a four-state model is employed to describe the molecular conformational distribution of the dimer, B. The diameters of the diagonal peaks represent the populations of each state and cross-peaks denote populations of states and transition probabilities, respectively. The sizes of the cross-peaks are related to the additive energy barrier to interconversion
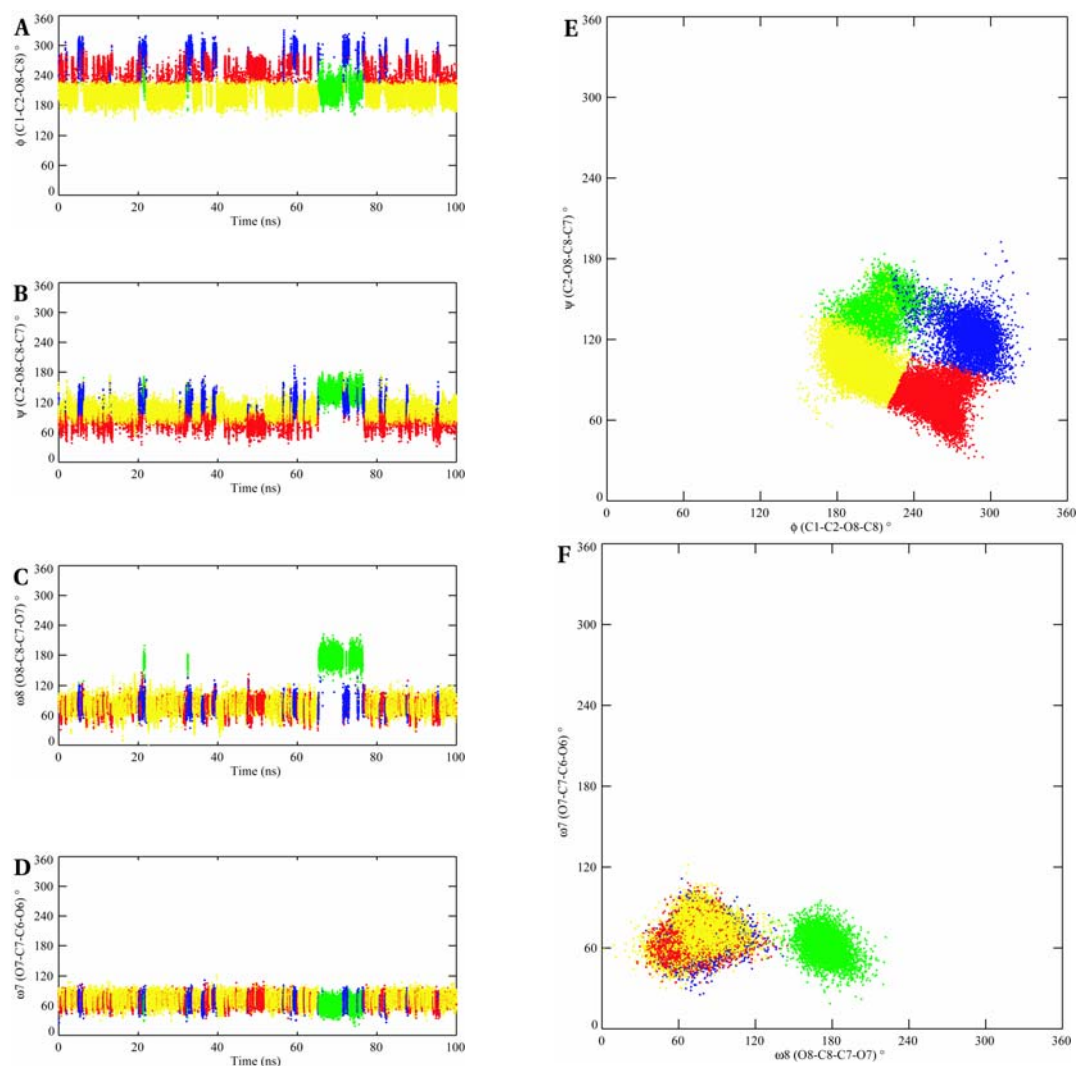
Figure 4.10. The conformational transition pathway of the four plausible clusters of methyl α-(2,8)-disialoside over the course of the MD simulation, left column. A, B, C and D represent the pathways of the φ, ψ, ω_8 and ω_7 angles, respectively. The scatter plots of the exocyclic dihedral angles, right column: E − (φ, ψ) and F − (ω_8, ω_7). The figures employ the color-coding of Figure 9 to indicate the various conformations
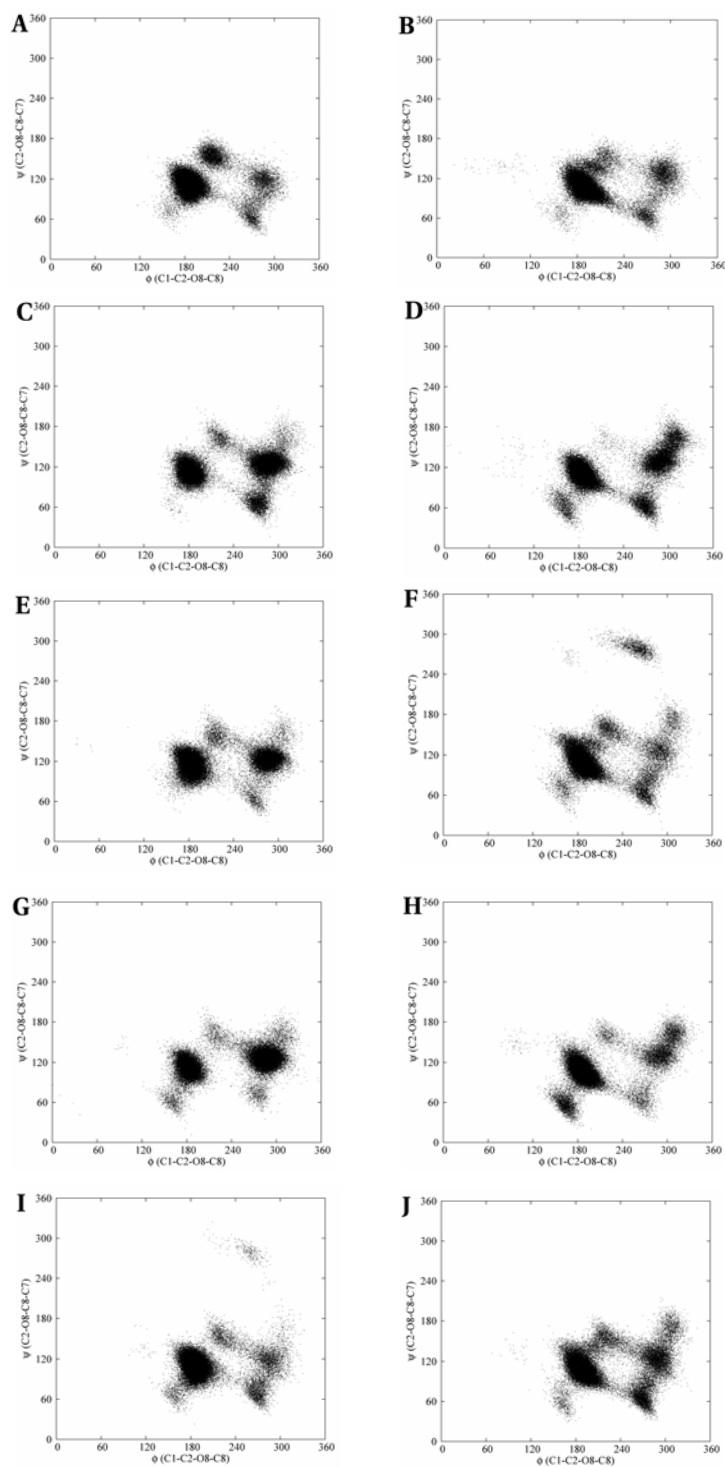
176

Figure 4.11. The (φ, ψ) conformational space of the native *N*-acetyl α-(2,8)-trisialoside and its analogs. The reducing and non-reducing linkages are on the left and right columns, respectively. *N*-acetyl (A, B), *N*-glycolyl (C, D), *N*-propionyl (E, F), *N*-propyl (G, H) and *N*-butanoyl (I, J)

Figure 4.12. (ω_8, ω_7) scatter plots of the conformational space of the native *N*-acetyl α-(2,8)-trisialoside and its analogs. The reducing and non-reducing linkages are on the left and right columns, respectively. *N*-acetyl (A, B), *N*-glycolyl (C, D), *N*-propionyl (E, F), *N*-propyl (G, H) and *N*-butanoyl (I, J)
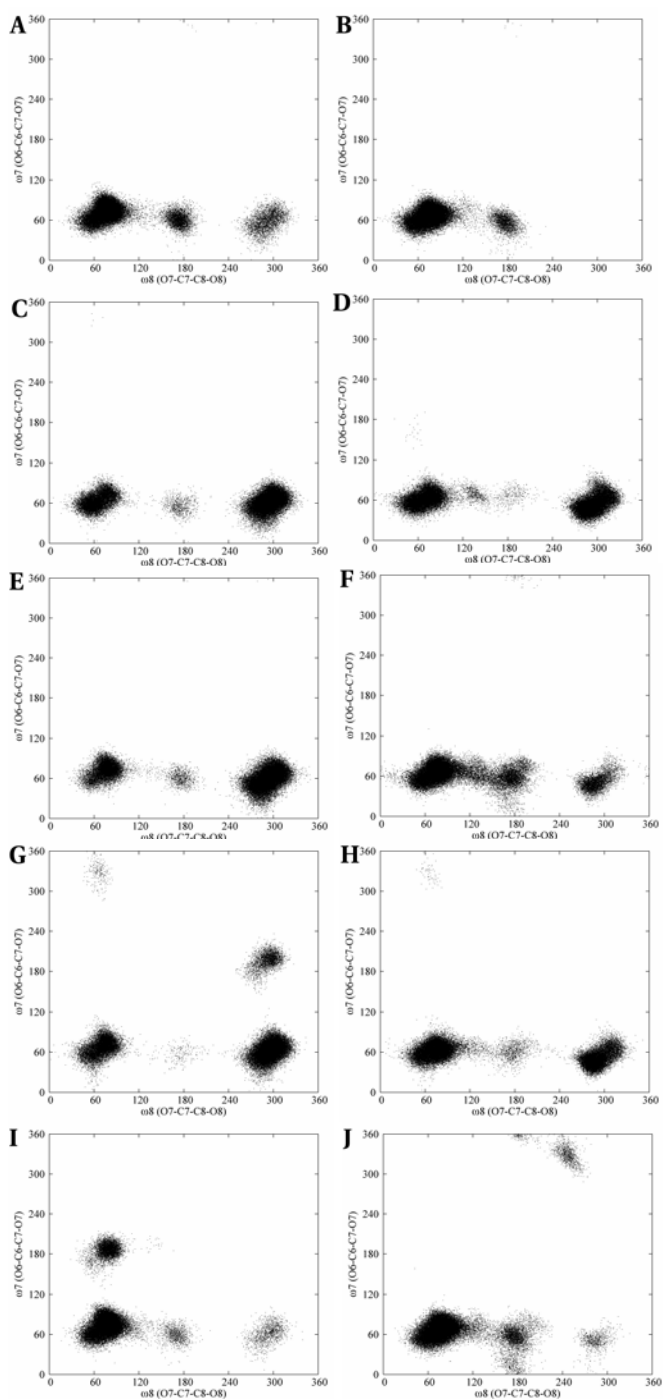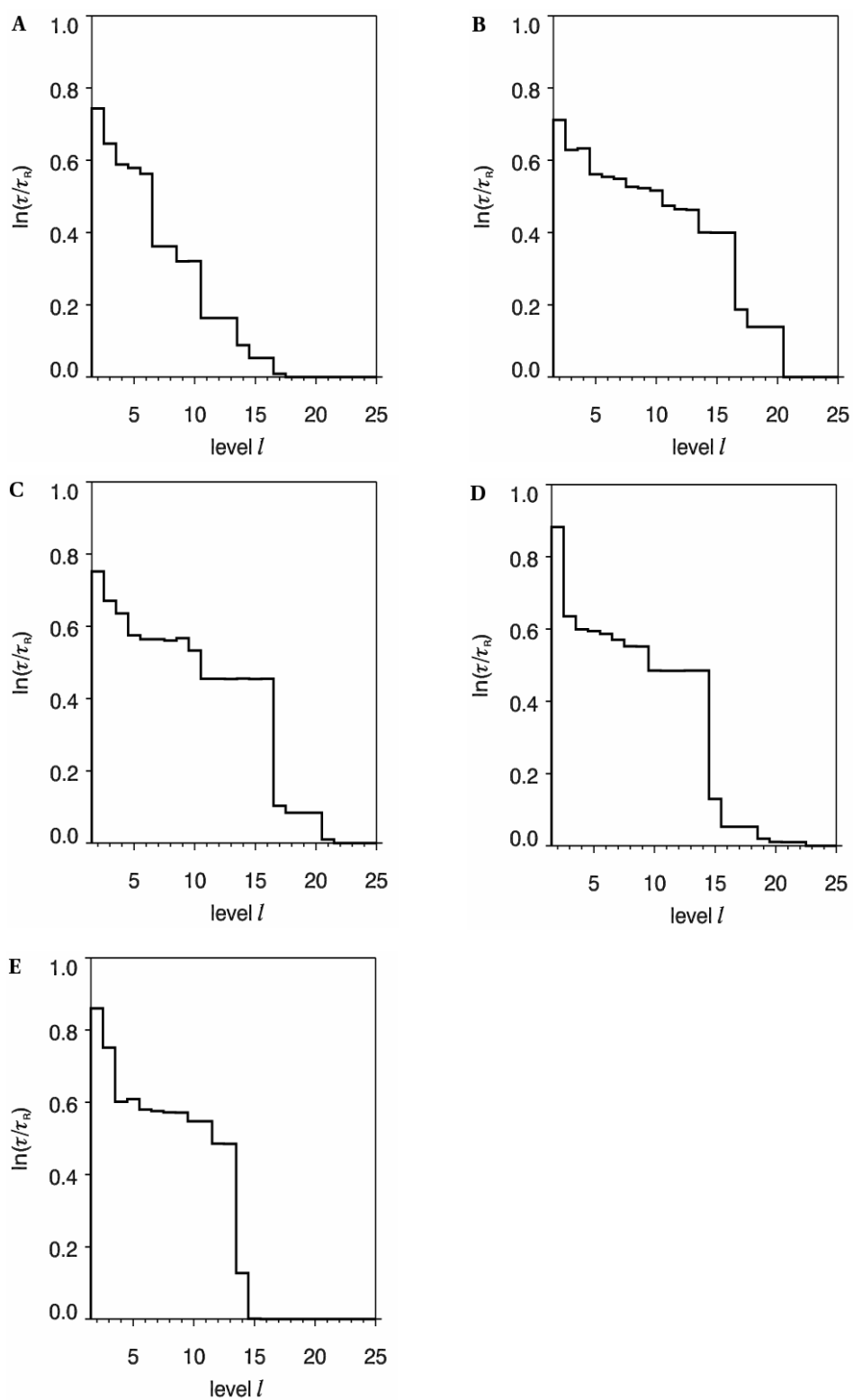
Figure 4.13. Plot of the relaxation time scale, $\ln(\tau/\tau_R)$ between the hierarchic Markov states identified from the REMD trajectories of **II** (A), N-Gc (B), N-Prop (C), N-Pr (D) and N-But (E)
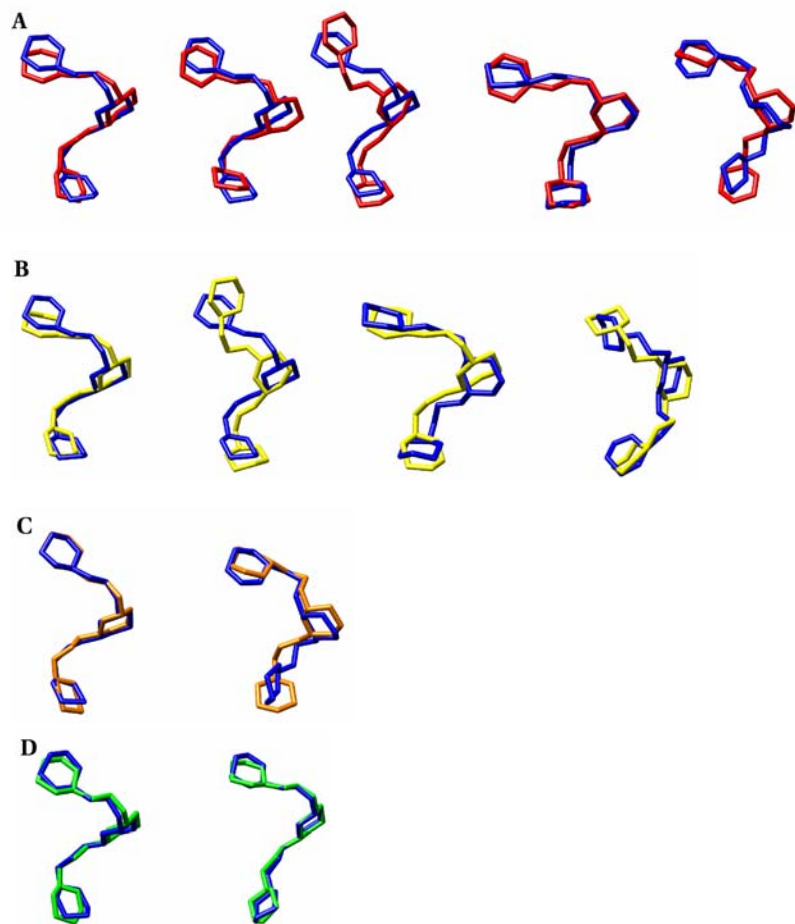
Figure 4.14. Backbone (ring and glycosidic linkage atoms) superposition of the representative members from the most populated clusters during the REMD simulations of the native and *N*-acetyl analogs of the methyl α-(2,8)-trisialoside

A. *N*-acetyl (blue) and *N*-glycolyl (red): A-A, A-C, A-J, C-C, E-I
B. *N*-acetyl (blue) and *N*-propionyl (yellow): A-A, A-M, C-A, E-N
C. *N*-acetyl (blue) and *N*-propyl (orange): A-A, E-Q
D. *N*-acetyl (blue) and *N*-butanoyl (green): A-A, B-B
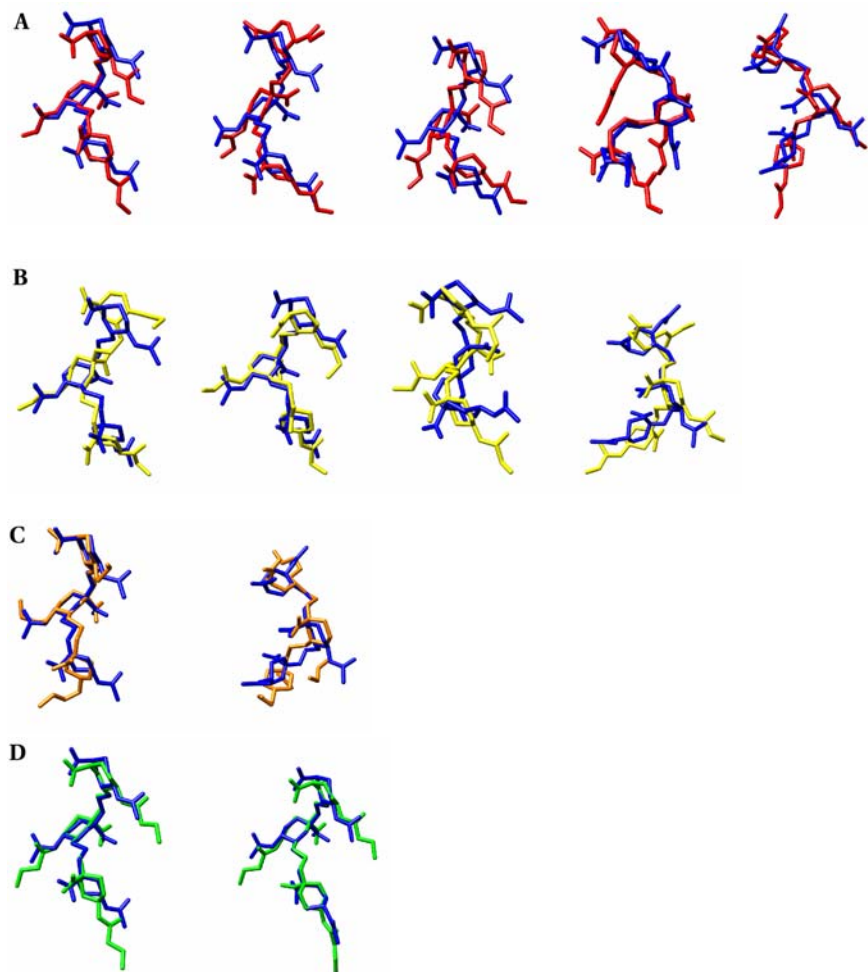
Figure 4.15. The display of key exocyclic groups relative to the backbone (ring and glycosidic linkage atoms) conformations of the representative cluster members of the native methyl α-(2,8)-trisialoside and its analogs

A. *N*-acetyl (blue) and *N*-glycolyl (red): A-A, A-C, A-J, C-C, E-I
B. *N*-acetyl (blue) and *N*-propionyl (yellow): A-A, A-M, C-A, E-N
C. *N*-acetyl (blue) and *N*-propyl (orange): A-A, E-Q
D. *N*-acetyl (blue) and *N*-butanoyl (green): A-A, B-B
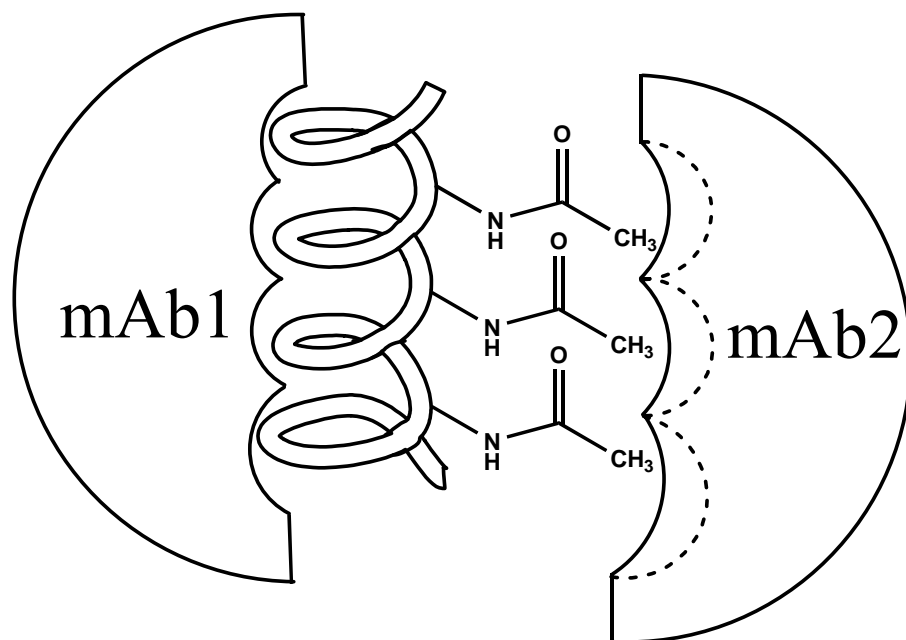
Figure 4.16. A schematic representation of the interactions between hypothetical monoclonal antibodies (mAb1 and mAb2) and synthetic derivatives of the conformational epitope of the *Nm*B CPS. mAb1 would cross-react because its epitope is common amongst the intact and derivatized CPS, while mAb2 would not display cross-reactivity due to the differences in fine specificities involving the *N*-acyl moiety. The dashed curves of mAb2 represent the binding surface of a monoclonal antibody that was optimized to recognize the longer *N*-acyl derivatives of the CPS

CHAPTER 5

Computational modeling of antigenic carbohydrate-antibody interactions

Abstract

An assessment of the ability of computational protocols to model protein-carbohydrate binding interactions, utilizing the GLYCAM06 carbohydrate and AMBER protein force fields, has been undertaken. Two criteria are employed to judge the performance of these protocols: the ability of reproduce experimental crystal complexes, and the ability to predict the relative affinities of a protein for a series of ligands. Trisaccharide and pentasaccharide fragments from the capsule of *Shigella flexneri* variant Y and a monoclonal antibody raised against these oligosaccharides are employed as test cases. Automated docking approaches are employed to generate complexes of the pentasaccharide-antibody systems by docking to variable fragment (Fv) domains derived from the co-crystallized and unbound crystal structures, as well as a comparative model of the antibody. It is noted that the theoretical complexes are increasingly poor compared to the co-crystallized crystal structures. However, refinements via molecular dynamics (MD) simulations are able to improve the qualities of the theoretical complexes. Thermodynamic integration (TI) MD simulations are employed to compute the relative binding affinities of the antibody for a series of trisaccharide analogs. Two sets of partial atomic charges, generated by fitting to the HF/6-31G* and B3LYP/cc-pVTZ computed electrostatic static potentials of the monosaccharide residues, are implemented. It is shown that the utilization of charges derived from the larger basis set, results in relative binding affinities that are more consistent with the available experimental data.

5.1. Introduction

The roles of carbohydrates in mediating important biological processes have been extensively documented.[27,28,113,297] Some of these include cell-cell interactions,[106] ovulation,[272] neuronal development,[104] and host-pathogen interactions.[297,299] Carbohydrates are also prevalent on the surfaces of bacteria and viruses masking vulnerable peptide epitopes and conferring antiphagocytic properties.[8,300] However, the surface exposure of these carbohydrates also makes them potential targets of the hosts' humoral immune system, primarily antibodies. The efficient recognition of the carbohydrate antigens by antibodies in a host's immune system is crucial for the successful elimination of pathogens. Because of the crucial role antibodies play in the immune response, investigating the key structural features in these antigens, which are responsible for determining the antibody specificities, can provide insights that are potentially applicable in the design better antigenic vaccines.

Despite the prevalence of carbohydrates in most undesirable host-pathogen interactions, the modes of interaction between antibodies and antigenic carbohydrates are still not well defined. Evidence for this is reflected in the low availability of experimental structural and thermodynamic data reported for both unbound carbohydrates and antibody-bound carbohydrates. Primarily, the availability of experimental structures of antigenic carbohydrate-antibody complexes will be important in ascribing the nature of key intermolecular interactions as well as the origins of specificity. For example, in the case of the mannose-binding C-type lectin, stereochemistry has been shown to be the main structural determinant for its selective recognition of α-D-mannose over α-D-galactose,[273] while a balance between solvation, enthalpy and entropy drive the thermodynamic aspects of the interaction. The presence of water-mediated hydrogen bonds is also evident in carbohydrate-protein interactions.[117,302,303] In spite of the

polyhydroxy character of carbohydrates, it is well known that hydrophobic interactions are involved in driving or enhancing the formation or stability, respectively, of carbohydrate-antibody complexes.[281-284] The scarcity of structural data can be attributed to the microheterogeneity and multi-conformational nature of carbohydrates, which make their structural characterization a challenge to either NMR spectroscopy or X-ray crystallography. Surface plasmon resonance and titration microcalorimetry are the typical sources of thermodynamic data. However, studies that involve both structural and thermodynamic characterizations of carbohydrate-antibody complexes are very rare.

An alternative approach to generating and characterizing antigen-antibody complexes is via computational modeling techniques, which entail the correct determination of both the structures of complexes and their relative or absolute binding free energies. The availability of computational protocols that provide accurate structures of carbohydrate-antibody complexes as well as a quantitative breakdown of the interaction energies into contributions from different energetic components would be valuable in providing useful insights about what structural and thermodynamic factors drive recognition and specificity at the molecular level. In the absence of experimental structures of proteins, comparative modeling can be utilized to generate low-resolution structures from the available protein sequences.[74,75] Automated docking approaches are the most widely employed methods to generate the initial *in silico* complexes of biomolecules,[81,88,97,274-276] followed by refinements employing explicit-solvent molecular dynamics (MD) simulations to account for induced-fit mechanisms.[102] Intermolecular interaction energies are most commonly determined via free energy perturbations (FEP), thermodynamic integrations (TI) and molecular mechanics generalized Born surface area (MM-GBSA) methods, which are all extensions of conventional MD simulations.[99,101,102] The utilization of these

thermodynamic methods provide molecular level resolutions of the effects of functional group modifications on receptor-ligand binding affinity.[99,277-279] In one example of FEP calculations applied to oligosaccharide-antibody interactions, Pathiaseril *et al.*[99] were able to reproduce the binding free energies between oligosaccharide haptens and an anti-*Salmonella* monoclonal antibody, Se155-4. In that study, the experimental relative free energies were within 1 kcal/mol placing high demands on computational methods.

In this study, to assess the accuracy of employing computational methods to predict the binding modes of carbohydrate-antibody complexes, the *O*-antigen of *Shigella flexneri* (*S. flexneri*) variant Y was selected. The *S. flexneri O*-antigen was chosen because it has been extensively characterized experimentally and is one of the few examples of an oligosaccharide co-crystallized with the antigen-binding fragment (Fab) of an antibody.[280] In addition, experimental X-ray structures have been solved for the free Fab (pdbid: 1M71), and for the Fab in complex with a fragment of the natural *O*-antigen pentasaccharide (pdbid: 1M7I) and a synthetic trisaccharide analog (pdbid: 1M7D). The synthetic ligand differs from the natural ligand by the presence of 2-deoxy α-L-rhamnopyranose (α-L-Rha*p*) at residue two of the former, instead of α-L-Rha*p* in the latter, Figure 5.1. Moreover, thermodynamic binding data are available for several deoxy derivatives of the trisaccharide.[280] In this light, the *S. flexneri* system provides an opportunity to test the ability of computational methods to reasonably predict the binding modes of antigen-antibody complexes. The primary aim of this work is to determine whether the available computational protocols can correctly model the experimental ligand affinities and the crystal structures of the *S. flexneri* capsular polysaccharide-antibody complexes.

5.2. Methods

All energy minimizations and MD simulations were performed using the AMBER8 software package.[136] The GLYCAM06[259] force field was employed to model carbohydrate properties, while the PARM94[79] protein force field available in the AMBER8 suite of programs was utilized to model the properties of the antibodies.

5.2.1. Comparative modeling

A comparative model of the Fv domain of the anti-*S.flexneri* variant Y antibody was generated in order to establish whether our current protocol would perform reasonably well when no experimental structure was available. To identify template structures for the model, a BLAST[76] search implementing the BLOSUM62 matrix was performed with the sequences of the variable light (V_L) and heavy (V_H) domains of the antibody. Only structures from immunoglobulin molecules with high sequences identities were selected, and the backbone canonical conformations[61,62] of the complementary determining regions (CDR) were further analyzed. The initial selection was further streamlined to include only structures that possessed CDR canonical conformations identical to those of the anti-*S. flexneri* antibody. The comparative model was generated by employing the homology module of the InsightII software package.[281] For aligned and identical residues, the coordinates of the primary template were transferred to the comparative model. For aligned non-identical residues, only the coordinates of the backbone atoms of the primary template were transferred to the model, while coordinates of the side chain atoms were assigned from a standard rotamer library available in InsightII. In cases whereby the lengths of one of the CDR's of the primary template and the target were not equal, another search was performed employing only the CDR of the target with four additional residues flanking both sides of the CDR. The CDR of the target was then constructed by grafting the

coordinates of this secondary template unto the growing comparative model. Prior to the grafting stage, the first and last four residues adjacent to the CDR's of the primary and secondary templates were superimposed. Particular attention was given to the CDRH3 loop because it is located at the center of the CDR and is known to be crucial for antigen specificity.[63] To generate the Fv domain of the target, the modeled V_L and V_H domains were oriented relative to each by superimposing the framework residues on the respective domains of another antibody.

The resultant model was refined via explicit-solvent MD simulations. A spherical cap of TIP3P water[117] was placed around the CDR's. First the solvent configurations were optimized in 100 steps steepest descent followed by 9900 steps conjugate gradient method, while keeping the solute fixed. Next, the CDR's and water molecules were subjected to energy minimization cycles as before, followed by an annealing stage in which the entire system was heated from 5 K to 300 K over 50 ps and cooled to 5 K over another 50 ps. A short simulation was performed next whereby the system was heated to 300 K over 70 ps and maintained at this temperature for 80 ps. A production dynamics stage ensued for 10 ns with a time step of 2 fs employed to integrate the equations of motion. At the end of the production stage, the system was cooled from 300 K to 5 K over 1 ns, and subjected to another cycle of minimization comprising of 5000 steps of steepest descent and 45000 steps of conjugate gradient. During the entire refinement, a restraining force of 10 kcal/mol was applied to keep the water molecules in the droplet, while another restraining force of 50 kcal/mol was employed to constrain motion in the framework residues. A nonbonded cutoff distance of 20.0 Å was employed, while 1-4 electrostatic and van der Waals interactions were scaled by 1.2 and 2.0, respectively. The quality of the final structure was determined by performing statistical analysis with the PROCHECK[282] program.

5.2.2. Ligand docking

The Autodock 3.0.5[82] software package was employed for all docking simulations. In Autodock, the ligand is treated with full atomistic detail, while the receptor is modeled rigidly as a three-dimensional affinity grid, consisting of regularly spaced points that store the electrostatic and van der Waals interaction energies between each atom-type in the ligand and all the atoms in the receptor. Three docking simulations, employing a pentasaccharide repeat unit of the *S. flexneri* capsular polysaccharide (CPS), were performed: 1). A positive control in which the pentasaccharide co-crystallized with the Fab (1M7I), was extracted and re-docked to the Fv domain. This represented an ideal case whereby the antibody's surface was already optimally shaped to recognize the pentasaccharide; 2). The pentasaccharide was docked to the free Fv domain (1M71), representing the scenario of an antibody making initial contacts with an antigen; 3). The pentasaccharide was finally docked to a comparative model of the anti-*S. flexneri* monoclonal antibody. This depicted a situation whereby the amino acid sequence of an antibody had been determined, but no experimental structural model was available.

An affinity grid box with 80 x 90 x 80 grid points in the x-, y- and z-dimensions, and a grid point spacing 0.375Å was centered on the binding groove or CDR of the Fv domain. At the beginning of the docking experiments, the pentasaccharide was positioned at the center of the grid. The Lamarckian genetic algorithm (LGA), and the Solis and Wets[98] local search methods were employed for searching the ligand's conformational phase-space in the binding pocket of the Fv domain. The total number of docking runs was set to 150, while Autodock default values were utilized for all other parameters.

5.2.3. MD simulations of the pentasaccharide -Fv complexes

Because Autodock 3.0.5 treats the receptor as a rigid entity, it is more appropriate to determine the structure of the complex only after induced-fit effects have been accounted for. Moreover, the presence of multiple conformations of glycosidic bonds and exocyclic groups in carbohydrates may limit the ability of the search algorithm employed in Autodock to sample an appreciable surface of the phase-space in the course of the docking simulation. Based on these considerations, the docked complexes were subjected to MD simulations in order to allow for further sampling of exocyclic dihedral angles and also account for induced-fit effects in the Fv domain due to the presence of the pentasaccharide, and vice versa.

A 25-Å radius droplet of TIP3P[117] water molecules, centered on the C2 atom of the β-D-GlcNAc residue of the pentasaccharide, was placed around the binding site of the Fv domain, and the appropriate number of counter-ions was added to achieve overall charge neutrality. During all stages of the refinement, a force constant of 10.0 kcal/mol was applied to restrain the water molecules around the vicinity of the binding pocket. A 50.0 kcal/mol restraining potential was applied on the backbone atoms, while the belly option was utilized to allow motion in both the solvent and only solute atoms that were enclosed in the droplet. First the coordinates of the solvent were optimized while keeping the solute fixed through 10000 cycles of which the first 100 cycles employed the steepest-descent method. Next the entire system was minimized through the same number of steps. An annealing stage followed wherein the system was heated from 5-300 K in 50 ps, and cooled to 5 K in another 50 ps, with initial velocities assigned from a Boltzmann distribution. The entire system was then heated from 5-300 K in 70 ps, and maintained at this temperature for 80 ps. Finally, a 20-ns production dynamics stage ensued during which the temperature was maintained at 300 K. A 20-Å cut-off was applied to

nonbonded van der Waals interactions, while 1-4 nonbonded interactions were not scaled (SCEE=SCNB=1.0).

5.2.4. Thermodynamic integration: free energy calculations

The scoring function employed in Autodock is a generalized ligand scoring function that was not specifically optimized for carbohydrates. In recognition of this deficiency, Laederach *et al*.[97] developed a carbohydrate-specific scoring function and independently implemented it in Autodock. The energy function was able to predict the free energy of the formation of complexes to within 1.1 kcal/mol, but did not perform as well in complexes that contained tightly bound water molecules.[97,283] Given the importance in distinguishing the affinities of ligands for a common receptor as the basis for drug or vaccine candidate selection, it is necessary to employ a more robust approach to computing ligand binding affinities via free energies of complexes. To this end, the TI approach was selected. Despite the computationally intensive nature of TI simulations, their robustness in providing accurate values of free energy changes has been noted.[284] Due to the requirement for statistical equilibration in configurational sampling in free energy simulation methods only ligands with modest structural differences can be employed.[100,283] One advantage of TI calculations is that explicit solvent molecules are included, which is especially necessary for crystallographic water molecules involved in carbohydrate-antibody complexes. The specific application of TI methods in carbohydrate-antibody systems is still in the fledgling stages. It should be noted that TI can be utilized to compute only the relative and not the absolute free energy changes for the binding of different ligands to a common receptor and often requires the availability of high resolution structures of the complexes involved. The task here was to determine whether TI employed in MD simulations will be able to distinguish between the antigenicities of an antibody for different carbohydrate antigens.

In TI simulations a coupling parameter, λ, is employed to transform one ligand into another through a series of simulations employing successive values of λ.[100] In the course of a TI perturbation, the total energy of the system is expressed as a linear combination of the full properties of the two ligands:

$$H_\lambda = \lambda^n H_L + (1 - \lambda^n)H_{L*}$$

wherein L and L* are the two ligands of interest.

The contribution of the properties of each end state to any intermediate state depends on the value of λ. The intermediate states in the simulation are alchemical states that do not represent any real physical properties of the two ligands. Nonetheless, they define a path through which one ligand can be transformed into the other. Given that free energy is a state function, the path taken from the initial to the final states is irrelevant, and only the free energies of the endpoints are important. The successive states are simply employed to ensure maximum overlaps between the Hamiltonians of the endpoints.[100]

The relative affinity of a receptor R, for the two ligands can be computed following the scheme illustrated in Figure 5.2. In the cycle, ΔG_1 and ΔG_2 are the experimental free energy changes for the binding of R to L and L*, respectively. ΔG_3 and ΔG_4 represent the alchemical processes of transforming L to L* unbound in solution and bound on the surface of the receptor, respectively. They are determined computationally and take into account the differences in solvation (ΔG_3) and binding free energies (ΔG_4) of the two ligands. From the thermodynamic cycle, the difference in the changes in free energies can be expressed as:

$$\Delta\Delta G = (\Delta G_4 - \Delta G_3) = (\Delta G_2 - \Delta G_1)$$

if $\Delta\Delta G < 0$, then the "mutant", L*, is a higher affinity ligand or a better inhibitor.

In this work, the coordinates of the Fv domain and trisaccharide ligand were obtained from a 2.3-Å resolution crystal structure (pdbid: 1M7D) of a synthetic trisaccharide moiety of the *O*-antigen of *S. flexneri* variant Y complexed with the antigen-binding fragment (Fab) of a monoclonal antibody. The natural ligand was generated from the synthetic ligand in the presence and absence of the Fv domain, by incorporating α-L-Rha*p* at residue two of the crystal structure of the synthetic ligand. For computational efficiency only the coordinates of the Fv domain were utilized. Nine closest crystallographic water molecules were retained in the co-crystallized complex. The unbound ligand and Fv-ligand complex were immersed in a periodic box of 648 and 7219 TIP3P water molecules, respectively, and the appropriate number of counter ions was added to achieve overall charge neutrality. Both systems were minimized under constant volume conditions via 15000 cycles of steepest-descent, followed by 5000 cycles of conjugate gradient methods without constraining bonds involving hydrogen atoms at their equilibrium values with the SHAKE algorithm.[138] The free ligand and complex in solution were separately heated from 5-300 K in 50 ps, and maintained at this temperature for 450 ps under the isothermal-isobaric ensemble, while applying the SHAKE algorithm. As for the system containing the complex, restraints were applied to the ligand, counter ions, crystallographic water molecules and the protein backbone atoms during this stage. Next, the temperature of the complex in solution was maintained at 300 K for another 500 ps during which the entire system was allowed to relax fully. A 10.0Å cutoff was applied for nonbonded interactions, while long range interactions were treated via the particle-mesh Ewald summation method.[239] A 1-fs time-step was used to integrate the equations of motion, and scaling of 1-4 nonbonded interactions was turned off (SCEE=SCNB=1.0).

In selecting which perturbations to perform the following considerations were taken into account, based on experimental data: A mutation that resulted in 1) increased affinity for the mutant, 2) decreased affinity for the mutant, and 3) no effect. These mutations should also involve the probing of hydrophobicity, hydrophilicity, electrostatics, or van der Waals (vdW) effects. Only thermodynamic data generated in liquid-phase assays were included, in order to be internally consistent.

The perturbations involving the ligand, both free and bound, were performed sequentially in 12 stages with successive clambda values of: 0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206 and 0.99078.[136] klambda values of 1 and 6 were utilized for the electrostatic and vdW perturbations, respectively. At each clambda value, a constant pressure equilibration was performed for 20 ps, followed by 100 ps constant volume production dynamics, during which running averages of changes in free energy were collected every 5 ps. The perturbations were performed in two stages by decoupling the electrostatic and vdW perturbations, for a total of 5.76 ns for each mutation cycle. For instance, in the charge perturbation stage, only the charges were perturbed while the vdW properties were unchanged, while for the vdW perturbation stage only the vdW properties were perturbed while the partial charges derived at the end of charge perturbation stage were fixed. This method has the advantage that the change in free energy can be decomposed into specific contributions from the vdW and Coulombic components. Finally, the value of the TI integral was determined numerically through the Gaussian quadrature method.

5.3. Results and discussion

5.3.1. Comparative modeling

The pdbid's of the template structures that were selected to generate the target structure of the anti-*S. flexneri* antibody are shown in Table 5.1. The V_L domain from an antibody with pdbid 1MNU[285] was selected as the template for assigning the coordinates of the V_L chain of the target structure. The CDRL1 and CDRL2 loops of 1MNU and the target structure both belonged to the canonical structure class 4 and 1, respectively. However, the CDRL3 loop of 1MNU belonged to the class 1 canonical structure and differed from that of the target, which belonged to class 3. In addition, the CDRL3 loop of 1MNU was one residue longer than that of the target. To avoid errors introduced during loop splicing or deletions, the coordinates of the CDRL3 loop were assigned from the CDRL3 loop of another antibody, 1KEG[286], which had the exact loop length and required canonical structure. Prior to incorporating the coordinates of the CDRL3 loop of 1KEG into the model, the trace atoms (Cα) of the first and last four residues adjacent to the CDRL3 loop of 1KEG and 1MNU were structurally aligned, with a root-mean-squared deviation (RMSD) of 0.10 Å.

The V_H domain of the model was derived from an immunoglobulin with pdbid 1AXT.[287] The canonical structures of the CDRH1 and CDRH2 loops of 1AXT and the target belonged to classes 1 and 4, respectively. The CDRH3 loop differed in length as well as in structural classification. The template CDRH3 loop was two residues shorter and adopted the "torso-non bulge"[64,65] conformation, while the CDRH3 loop of the model existed in the "torso-bulge" conformation. The coordinates of CDRH3 loop were subsequently assigned from the immunoglobulin with pdbid 1JHL,[288] that had the "torso-bulge" conformation and same residue length as the target. Again the Cα atoms of the first and last four residues adjoining the CDRH3

loops of 1AXT and 1JHL were superimposed prior to the grafting of coordinates. The RMSD of the superposition was 0.30 Å.

Given that the V_L and V_H structures of the target were derived from the structures of different immunoglobulin templates, the Fv domain was generated by superimposing the structurally conserved regions (SCR) of both domains on those of 1MNU. The Cα superpositions of the SCR's were 0.43 Å and 1.25 Å for the V_L and V_H domains, respectively. The goodness of the model was evaluated before and after structural refinement by comparing RMSD superposition with the crystal structure of the anti-*S. flexneri* antibody, pdbid 1M71[280] and via Ramachandran diagrams, Figure 5.3. The Cα RMSD of the entire Fv domain before the refinement was 0.99 Å. As may be expected, the most pronounced differences occurred in the CDRL3 and CDRH3 loops, which were generated by transferring coordinates from other immunoglobulin molecules. Nonetheless, the MD refinements were able to align these loops in close proximity to those of the 1M71 structure. The RMSD between the refined target structure and 1M71 was 1.03 Å. Despite the improvement with the CDRL3 and CDRH3 loops, the CDRH2 loop was slightly displaced relative to the 1M71 structure. However, this displacement was not as a result of any internal conformational changes, as the Cα RMSD between the CDRH2 of the model and 1M71 was only 0.65 Å. This difference can be attributed to atomic fluctuations that occurred during the MD refinement. The Ramachandran analysis before the MD refinement, identified three amino acids, VAL56, HIS98 and ALA215, with unusual (φ, ψ) backbone values of (50.7, -27.6), (67.2, -93.7) and (101.4, -12.9), respectively. Closer examination showed that these residues were in "turn" regions of the model and were adopted from the template structures. These backbone outliers are very common in the loop regions of immunoglobulin structures. The VAL56 and HIS98 residues occurred in classic γ-turns that were

stabilized by hydrogen bonds. In the case of VAL56 the γ-turn in the 1MNU template was stabilized by a 2.60-Å hydrogen bond between LYS L55 (CO) and SER L57 (NH). For HIS98, whose corresponding amino acid in the 1KEG template was LEU L93, a 2.80-Å hydrogen bond between SER L92 (CO) and VAL L94 (NH) was the stabilizing factor. The exact stabilization of the "turn" occurring at ALA215, or ASN H101 in the template, was not evident. It should be noted that the 1KEG and 1JHL structures were co-crystallized in the presence of antigens. It is possible that the unusual backbone conformations were adopted in order to accommodate the ligand. For example in 1KEG the backbone NH group of LEU L93 was within hydrogen-bonding distance (2.47 Å) from a phosphate group in the ligand, while in 1JHL some hydrophobic interactions could be seen between TYR102 of the immunoglobulin and TYR23 of the antigen. Nonetheless, the MD refinement eliminated these "turns" except in VAL56, which was derived from the unbound structure of 1MNU, Figure 5.3.

5.3.2. Docking the pentasaccharide to the Fv domain

The docking results from the three different scenarios are portrayed in Figure 5.4. In the first panel employing the Fv domain from the co-crystallized Fab, the positive control (**I**), Autodock was able to accurately align the pentasaccharide in the Fv binding pocket when compared with the experimental structure. In the second panel, wherein the oligosaccharide was docked to the free Fv (**II**), deviations in the structural alignments between the theoretical and experimental structures were observed. These are not major differences and may be ascribed to differences in protein side chain orientations between the Fv domains extracted from the free and co-crystallized Fabs. In the last panel, docking to the comparative model (**III**), the pentasaccharide showed the worst alignment amongst all the three docking experiments. These differences may be due to different glycosidic angles adopted by the pentasaccharide during the

docking simulations. Consequently, the glycosidic angles of the ligand in all three complexes were closely examined.

A comparison of the glycosidic (φ, ψ) angles of the pentasaccharide in the experimental complex and those found in the docked ligand is shown in Table 5.2. Overall, the (φ, ψ) angles in the experimental and the docked pentasaccharide in **I** compared quite well, which is reflected in the good structural alignment observed between these structures, Figure 5.4. In this case, the side chains and backbone folds of the Fv domain are already optimally predisposed to recognize the ligand, which explains the observed excellent alignment.

In **II**, the φ- and ψ-angles in the linkages between residues 1-2 (34°, 41°) of the docked complex compared quite well with those of the ligand in the crystal structure (58°, 29°). In the linkage between residues 2-3, the dihedral angle difference was in the sign of the ψ-angle, with values of 2° and -11° in the experimental structure and in **II**, respectively. This difference is manifest in the poor alignment of the penultimate α-L-Rha*p* residue, Figure 5.4. The alignment became increasingly better in the last two residues, because of the agreement in the (φ, ψ)-angles of the linkage between residues 3-4 and 4-5; (6°, 29°) and (34°, -62°), respectively, in the crystal structure and (28°, 39°) and (46°, -13°), respectively, docked pentasaccharide.

Despite the similarities of the last three (φ, ψ)-angles of the ligand in **II** and **III**, the structural alignment between the ligand in **III** and the experimental structure was considerably worse. There was an evident translational shift in the structure of the docked ligand, in addition to a remarkable difference in the (φ, ψ)-angles between residues 1-2, (58°, 29°) and (-16°, -8°) for the experimental structure and the ligand in **III**, respectively. The translational shift was primarily due to a non-optimal spatial offset in backbone conformation of the CDRH3 loop of the comparative model with respect to the experimental structure. The cascading effect was an

induced-fit in the ligand, which resulted in the re-orienting of the (φ, ψ)-angles between residues 1-2.

5.3.3. Refining the complexes through MD simulations

Given the rigid treatment of the receptor by Autodock 3.0.5, induced-fit effects were accounted for by subjecting the complexes to structural refinements via MD simulations. An assessment of whether the MD simulations could improve the orientations of the ligand in the docked complexes was gauged by the ability of simulations of the theoretical models to reproduce the intermolecular interactions, primarily hydrogen bond distances, observed in the crystal structure. Because X-ray structures are essentially static representations of inherently dynamic systems, the intermolecular hydrogen bonds present during the MD simulations of the docked complexes were compared to those observed during an MD simulation initiated from the co-crystallized complex,[280] Table 5.3. In the course of the MD simulation of the experimentally derived complex, all the hydrogen bond distances reported in the crystal structure were well reproduced. However, a false positive interaction (Thr92 O – Rha5 HO3) with a 27% occupancy was introduced in the course of the refinement. Despite the presence of this false positive, all the hydrogen bond distances were reproduced in high occupancies, except the Thr91 Oγ1 – Rha3 HO2 hydrogen bond with an occupancy of 5%. The low occupancy of the Thr91 Oγ1 – Rha3 HO2 hydrogen bond interaction occurred because in the experimental structure the Rha3 HO2 forms two competing hydrogen bonds with the Fab: Thr91 Oγ1 – Rha3 HO2 and GlyB99 O – Rha3 HO2 with O---O distances of 2.92 and 3.15 Å, respectively. In the course of the MD simulation, the Rha3 HO2 atom preferentially interacted with the backbone carbonyl oxygen atom of GlyB99 (GlyB99 O – Rha3 HO2) for 99% of the simulation time. As a result of the loss

of the Thr91 Oγ1 – Rha3 HO2 hydrogen bond, the Thr91 side chain interacted with the Rha3 O5 atom (Thr91 Oγ1 – Rha3 O5) with an occupancy of 33%.

In system **II**, prior to the MD refinement, only two out of eight hydrogen bonds present in the experimental complex were reproduced, in addition to two false positive interactions, Table 5.3. After the MD simulation, seven out of eight hydrogen bond interactions were correctly reproduced. When compared to the MD simulation initiated from the experimental complex, the interactions were well reproduced except the Thr91 Oγ1 – Rha3 HO2 interaction. As in the case of the MD simulation of the co-crystallized complex, the Rha3 HO2 preferentially formed a hydrogen bond with the backbone carbonyl oxygen atom of GlyB99 (GlyB99 O – Rha3 HO2), which persisted 100% of the time, while the Thr91 Oγ1 – Rha3 O5 interaction was present for only 7% of the time. With the exception of the Thr91 Oγ1 – Rha3 HO2 and the false positive Asn31 O – Rha1 HO4 interactions, the MD simulations of the experimental complex and **II** essentially converged to the same structure.

In the case of the complex in **III**, only three out of eight hydrogen bond interactions were correctly predicted, with the presence of one false positive (Val98 O – Rha3 HO2) pre-refinement. However, post-refinement, six out of eight experimental hydrogen bond interactions were reproduced, in addition to the elimination of the false positive. The Thr91 Oγ1 – Rha3 HO2 and Trp33 Hε1 – Rha1 O4 interactions that were observed in the experimental complex were still not reproduced. In **III**, the Rha3 HO2 preferentially formed a hydrogen bond with the backbone carbonyl oxygen atom of GlyB99 (GlyB99 O – Rha3 HO2), with a 56% occupancy, while the Thr91 Oγ1 – Rha3 O5 interaction was present for only 10% of the time. With regard to the Trp33 Hε1 – Rha1 O4 interaction, the Rha1 O4 atom instead interacted with the SerB52C Oγ

atom for 7% of the time. Now prohibited from interacting with the Rha1 O4 atom, the Trp33 Hε1 interacted with the Rha1 O2 atom for 42% of the time.

From the data in Table 5.3, it was demonstrated that the MD simulations were able to reasonably reproduce the experimental intermolecular interactions compared to the initial docking simulations. It is also possible that the side chain orientations of some of these residues intrinsically prevent them from making contacts with sugar residues.

The glycosidic angles, φ (H_1-C_1-O_x-C_x) and ψ (C_1-O_x-C_x-H_x), of the pentasaccharide in the complexes were monitored over the course of the simulation, in order to determine differences between the experimental and docked complexes. In addition, the dihedral angles were also monitored for the unbound pentasaccharide, to highlight conformational changes that resulted from the binding of the ligand to the Fab of the antibody, Figure 5.5. The average values of the φ and ψ are presented in Table 5.4.

In the first linkage region (between residues one and two) the φ-angle of the unbound pentasaccharide existed primarily (99%) in the *+gauche* conformation, 44 ± 12°, and a negligible 1% of the population in the *-gauche* conformation, -19 ± 12°. The ψ-angle fluctuated between the *+gauche* and *-gauche* rotamers with an average value of -16 ± 28°, which on average is the *syn* rotamer. The predominance of the *+gauche* and *syn* rotamers of the φ- and ψ-angles, respectively, was observed in all the other linkages of the unbound pentasaccharide, Figure 5.5, and is also consistent with the known experimental solution conformational properties of carbohydrates typically involved in (1,3)- or (1,4)-linkages.[46]

For the complexes, the rotameric preferences of the (φ, ψ) angles for each linkage in **I** were remarkably similar to those in **II**, Figure 5.5. For instance, in the 1-2 linkage the φ and ψ-angles were exclusively in the *+gauche* conformation with values of (50 ± 9°, 35 ± 9°) and (72 ±

8°, 56 ± 9°) in **I** and **II**, respectively. While the values of the φ-angles were similar to that of the predominant (99%) rotamer in the unbound ligand, 44 ± 12°, in contrast, the ψ-angles preferred only the +*gauche* rotamer. This was the case, because rotating the ψ-angle in the static complex of either **I** or **II** resulted in steric clashes with the Fab (data not shown). In simulation **III**, the 1-2 linkage was relatively the most flexible, with two φ- rotamers with values of 34 ± 14° (60%) and -34 ± 14° (40%), as well as two ψ-rotamers at -41 ± 19° (93%) and -161 ± 18° (7%). The preference for the -*gauche* rotamer in the comparative model, as opposed to the +*gauche* rotamer observed in simulations **I** and **II**, could be attributed to the poor initial fit of the complex pre-refinement, Table 5.2. In addition, residue 1 (Rha1) was more solvent exposed in **III** compared to either **I** or **II**, which may account for the greater flexibility observed in this linkage.

The conformational properties of the linkages between residues 2-3 and 3-4 were similar in simulations **I**, **II** and **III**. However, while the φ-angles adopted the solution-preferred conformation, 52 ± 7°, 53 ± 7° and 45 ± 9° in **I**, **II** and **III**, respectively, the ψ-angle primarily existed only in the -*gauche* rotamer with values of -28 ± 8°, -25 ± 8° and -34 ± 10° in **I**, **II** and **III**, respectively. These linkages were the least conformationally flexible amongst all the linkages of the pentasaccharide and sampled only one rotamer, Table 5.4. The residues adjacent to these linkages are the central residues of the pentasaccharide and are located in a deep groove on the surface of the antibody.

The glycosidic angles between residues four (GlcNAc) and five (Rha5) in simulations **I** and **II**, with (φ, ψ) values of (16 ± 12°, -42 ± 13°) and (9 ± 11°, -49 ± 10°), respectively. In simulation **III**, while the +*gauche* rotamer dominated (48 ± 14°, 92%) the rotamer preferences, a smaller percentage of the *anti*-rotamer was also observed, (154 ± 19°, 8%). The ψ-angle oscillated around the *syn* rotamer, (-17 ± 22°, 100%), though the -*gauche* rotamer (-26 ± 16°,

78%) was preferred over the +*gauche* rotamer (14 ± 10°, 22%). The relative flexibility of this linkage in simulation **III** compared to **I** and **II**, may be rationalized in terms of the position of Rha5 in the binding pocket of the complexes. Similar to the linkage between residues one and two in **III**, the high flexibility could be directly related to the greater solvent exposure of Rha5 compared to **I** and **II**, Figure 5.4, which might be indirectly related to the poor initial conformation of the CDRH3 loop of the model.

5.3.4. The ability of the TI calculations to reproduce experimental relative free energies of binding (ΔΔG)

The structural differences between the native and synthetic trisaccharide ligands represented comparatively modest chemical mutations of the native structure, which involved the generation of deoxy analogs. In this study, the trisaccharides analogs were chosen to represent three possible scenarios; an analog that displayed 1) comparable affinities with the wild-type, 2) higher affinity and 3) lower affinity. A schematic representation of these mutations is shown in Figure 5.6.

For the comparable binder, the 4-OH group of the terminal α-L-Rha*p* residue (Rha1) was converted to a hydrogen atom. This Rha1 4-OH group is solvent-exposed and makes indirect contact with the antibody through a non-crystallographic water-mediated hydrogen bond.[280] Therefore, it is expected that the chemical modification of the Rha1 4-OH would have little effects on the antibody-antigen binding energy.

An analog that resulted in enhanced affinity for the antibody involved the (Rha2 2-OH → 2-H) mutation. In this perturbation, the internal α-L-Rha*p* residue was deoxygenated at the C2 position. In the experimental studies,[280] this mutation unexpectedly resulted in an increased affinity for the antibody. The basis for the better affinity towards the synthetic trisaccharide was

revealed in the co-crystallized complex of the antibody and the 2-deoxy analog. The axial proton of Rha2 was oriented towards a hydrophobic pocket comprised of Y32A, L33A, the hydrophobic portions of K50A and T91A, and V98B. In the native trimer, the axial Rha2 2-OH group would point into this hydrophobic pocket. This suggested that a less electronegative group at the C2 position of the internal α-L-Rha*p* would also result in a more favorable binding free energy. In the experimental studies, when the slightly less electronegative chlorine atom, relative to oxygen, was placed at the C2 position of this residue (Rha2 2-OH → 2-Cl), an increase in the affinity for the 2-deoxy-2-chloro mutant was observed. Besides the enhanced binding observed in this analog, its inclusion in the computational studies presented a possible test of how well the force field would perform with a second-row element.

For a worse binder, the 4-OH group of the 2-deoxy-*N*-acetyl-β-D-glucopyranose residue (β-D-GlcNAc) was mutated to a hydrogen atom, (GlcNAc3 4-OH → 4-H). In the crystal structure,[280] the 4-OH group makes a direct hydrogen bond interaction with the carboxyl group of E50B. Intuitively, the elimination of this interaction would lead to a lower affinity for the antibody.

Initially, ensemble-averaged partial charge sets developed in the GLYCAM force field by reproducing HF/6-31G* electrostatic potentials (ESP) with a restraint weight of 0.01 were employed.[45] The results of the computed ΔΔG for the four structurally-related trisaccharides are presented in Table 5.5. The ΔΔG values can be decomposed into energetic contributions from electrostatic interactions and vdW contacts. Overall, the vdW component was consistently worse for all the mutants, while the electrostatic component preferentially favored the mutants, except for the (GlcNAc3 4-OH → 4-H) and (Rha1 4-OH → 4-H) mutations whereby it was worse and insignificant, respectively. The HF/6-31G* ESP-derived charges correctly ranked the affinities of

the antibody for three of the mutants relative to the native trisaccharide. The predicted ΔΔG for the solvent exposed Rha1 4-OH → 4-H perturbation was -0.36 ± 0.5 kcal/mol, while the experimental value was 0.6 kcal/mol.[280] Despite the negative trend of the predicted ΔΔG, the value is insignificant within the statistical error margin. The experimental trends for the (Rha2 2-OH → 2-H) and (GlcNAc3 4-OH → 4-H) mutations were also correctly predicted by the HF/6-31G* ESP-derived charge set. However, when (Rha2 2-OH → 2-Cl) mutation was performed the predicted ΔΔG value of was -0.20 ± 0.6, which suggested that the affinities for both the native and the synthetic ligands were similar. This result contrasted the experimental ΔΔG value that indicated that the 2-deoxy-2-chloro mutant was preferred by about 2.1 kcal/mol.[280] This inconsistency may suggest that the HF/6-31G* ESP was not ideal for generating partial charges of second-row elements that have more diffuse electron shells.

To this end, another perturbation cycle was performed in which ensemble-averaged partial charges of the central α-L-Rha*p* and 2-deoxy-2chloro-α-L-Rha*p* residues were derived by reproducing the QM ESP computed at the B3LYP/cc-pVTZ level, employing a restraint weight of 0.001. These level of theory and restraint weight value are also currently employed in the AMBER force field.[289] First, the charges derived from the DFT ESP were applied only to the central α-L-Rha*p* and 2-deoxy-2-chloro-α-L-Rha*p* residues being perturbed, while in the second cycle the DFT-derived charges were applied to all the residues of both the native and mutant trisaccharides. The utilization of the more diffuse basis functions in the new perturbation cycle resulted in a marked improvement in the calculated ΔΔG values compared to those obtained with HF ESP-derived charges, Table 5.5. Next, it was determined whether B3LYP/cc-pVTZ ESP derived partial atomic charges would also perform satisfactorily in all the other perturbations. In all the perturbation cycles, the charges derived from the DFT-computed ESP out-performed

205

those derived from the HF-derived ESP. It should be noted that the magnitudes of the electrostatic components for the (GlcNAc3 4-OH → 4-H) mutations in both the HF and DFT ESP-derived charges were much smaller compared to those of all the other mutations, and even more pronounced in the DFT ESP-derived charges. The origin of this low magnitude is not entirely known. A comparison of the charges derived from the two wavefunctions showed a linear correlation, with an R^2 value of 0.9951 and a slope of 1.0855, Figure 5.7. Hence, the charge derivation scheme may be ruled out as a possible source of the low magnitude of the electrostatic component. It is worth noting that the populations of the rotational isomeric states around the C5-C6 bond are known to be influenced by the stereochemistry of the hydroxyl group at the C4 position.[41] It is conceivable that the perturbation of the C4 hydroxyl group to a hydrogen atom resulted in different rotamer populations that were not statistically sampled in the course of the simulation. Nonetheless, the data presented in Table 5.5 demonstrated the ability of the TI calculations to reasonably reproduce the experimental binding affinities of the antibody for these closely related ligands.

5.4. Conclusions

The ability of computational protocols to reproduce the binding modes of antibody-carbohydrate antigen systems has been demonstrated. First, by utilizing a combination of knowledge-based and *ab initio* methods, accurate comparative models of the Fv domains of antibodies were generated. The canonical structure approach was employed to identify potential template structures, while MD simulations were utilized to refine the model. Compared to the crystal structure, this model was a low resolution structure as ascertained from the lower density of residues in the highly favored regions of the Ramachandran plot, Figure 5.3. However, the

MD simulations were able to correct the poor orientation of the CDRH3 loop relative to the other CDR loops.

The docking simulations underscored the importance of utilizing high quality structures of receptors when docking ligands to receptors. When the binding pocket was well disposed to bind to the ligand, the docked complex compared favorably with the experimental complex. However, the reproduction of the experimentally co-crystallized complex, became increasingly poor when side chains were not pre-disposed to recognize the ligand or with different backbone structures. Nonetheless, the inclusion of induced-fit effects via MD simulations significantly improved the quality of the complexes.

The ability of computational protocols to compute the relative affinities of an antibody for closely related carbohydrates was demonstrated via MD TI simulations. Noting the modest nature of the mutations in this series of ligands, the TI calculations were able to reasonably reproduce the experimental binding affinities. In the case of a row two element (chlorine) it was shown that the utilization partial atomic charges derived from an electrostatic potential computed by utilizing a small basis set, no diffuse functions and no electron correlation (HF/6-31G*//HF/6-31G*) was inadequate. The inclusion of diffuse function, electron correlation and larger basis sets (B3LYP/cc-pVTZ//HF/6-31G*) in computing the electrostatic potential, resulted in partial atomic charges that were superior in terms of reproducing the relative binding affinities of the ligands. These results are encouraging and suggest that the AMBER force field augmented with the GLYCAM06 force field for carbohydrates can be employed to compute the relative free energies of binding in carbohydrate-protein interactions.

Table 5.1. PDB structures employed as templates to generate the comparative models

PDBID	Identity[a] (%)	Identity (%) with positives included
V$_L$		
1MNU	91	98
1KEG	89	95
V$_H$		
1AXT	80	86
1JHL	50	74

[a] Relative to the Fv domain of the anti-*Shigella flexneri variant* Y antibody

```
VL                     L1                         L2                                         L3
SYA/J6  DVVLTQTPLSLPVRLGDQASISCRSSQSLLHSDGNTYLHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYFCSQTTHVP-TFGGGTKLEIKR
1MNU    DIVMTQTPLSLPVSLGDKASISCRSSQALVHSNGNTYLHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVFFCSQSTHVPRTFGGGTKLEIKR
1KEG    DVLMTQTPLSLPVSLGDQASISCRSSQSIVHSNGNTYLEWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGSLVP-TFGGGTKLEIKR


VH                      H1                         H2                                                         H3
SYA/J6  EVKVEESGGGLVQPGGSMKLSCVASGFTFSNYWMEWVRQSPEKGLEWVAEIRLKSNNYATHYAESVKGRFTISRDDSKSSVYLQMNNLRAEDTGIYYC---TRGGAVGAMDYWGQGTSVTVSSA
1AXT    EVKLEESGGGLVQPGGSMKLSCVVSGLTFSRFWMSWVRQSPEKGLEWVAEIRLKSDNYATHYAESVKGKFTISRDDSKSRLYLQMNSLRTEDTGIYYCKI-----YFYSFSYWGQGTLVTVSAA
1JHL    QVQLQQSGAELVRPGASVKLSCKASGYTFISYWINWVKQRPGQGLEWIGNI-YPSDSY-TNYNQKFKDKATLTVDKSSSTAYMQLSSPTSEDSAVYYC---TRDDNYGAMDYWGQGTTVTV
```

Table 5.2. Glycosidic angles $(\varphi/\psi)^a$ between the pentasaccharide residues[b] in the bound, control, and free Fab crystal structures, and comparative model

Linkage[c]	Bound crystal	Control	Free Fab	Model
1-2	**58,29**	41,2	34,41	**-16,-8**
2-3	40,2	45,12	66,-11	38,-24
3-4	6,29	6,20	28,39	26,9
4-5	34,-62	21,-23	46,-13	20,-10

[a] $\varphi = H_1\text{-}C_1\text{-}O_x\text{-}C_x$, $\psi = C_1\text{-}O_x\text{-}C_x\text{-}H_x$ [b] α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-GlcNAc-(1→2)-α-L-Rha-OMe. [c] Residue numbering begins from the non-reducing end.

Table 5.3. Hydrogen bonds (Å) in complexes generated from docking of the pentasaccharide to the crystal structure of the free Fv or to the comparative model, followed by MD refinement.

Protein – Carbohydrate Hydrogen bond	Experimental, (20 ns MD of Experimental Structure)		Docked to free Fab, 20 ns MD		Docked to homology model of Fab, 20 ns MD	
His27 Hε2 – Rha5 O3	3.21	(2.9 ± 0.1, 94%)	----[a]	(2.9 ± 0.1, 98%)	----	(3.2 ± 0.2, 12%)
Tyr32 OH – Rha2 HO4	2.52	(3.1 ± 0.2, 61%)	----	(3.1 ± 0.2, 64%)	2.5	(3.2 ± 0.2, 3%)
Trp33 Hε1 – Rha1 O4	2.70	(3.0 ± 0.1, 96%)	----	(3.2 ± 0.2, 51%)	----	(----)
Glu50 Oε1 – GlcNAc4 HO4	2.81	(3.1 ± 0.2, 32%)	----	(3.2 ± 0.2, 11%)	3.3	(2.7 ± 0.2, 83%)
Thr91 O – GlcNAc4 HN	2.68	(2.8 ± 0.1, 100%)	3.0	(3.0 ± 0.1, 100%)	----	(2.9 ± 0.2, 88%)
Thr91 Oγ1 – Rha3 HO2	2.92	(3.4 ± 0.1, 5%)	----	(----)	----	(----)
Ala97 O – Rha2 HO3	3.02	(2.7 ± 0.1, 100%)	----	(2.7 ± 0.1, 100%)	2.6	(3.1 ± 0.2, 12%)
Gly99 O – Rha3 HO2	3.15	(2.9 ± 0.2, 99%)	2.8	(2.9 ± 0.2, 100%)	----	(3.2 ± 0.2, 56%)
H-bonds correctly predicted	----	**(8/8)**	**2/8**	**(7/8)**	**3/8**	**(6/8)**
False Positives						
Asn31 O – Rha1 HO4			3.0	(3.0 ± 0.2, 67%)	----	(----)
Thr92 O – Rha5 HO3	----	(2.8 ± 0.2, 27%)	2.4	(3.0 ± 0.3, 4%)	----	(----)
Val98 O – Rha3 HO2		(----)	----	(----)	2.9	(----)
False Positives	**0**	**(0)**	**2**	**(0)**	**1**	**(0)**
Cα RMSD (Å)	0	(0)[b]	0	(0)[b]	1.0	(1.1 ± 0.1)
Oligosaccharide RMSD (Å)	0	(0.7 ± 0.1)	0.	(0.6 ± 0.1)	1.2	(1.7 ± 0.3)

[a] Dashes: Signify the hydrogen bond was not observed. [b] Restraints applied to the backbone atoms during the MD simulation.

Table 5.4. Average glycosidic angles (°)[a] and populations (%) from the MD simulations

Linkage	Angle	Unbound	In co-crystal structure	Docked to free Fab	Docked to comparative model
			Environment of the ligand		
1-2	φ_1	(44 ± 12, 99%)	(50 ± 9, 100%)	(72 ± 8, 100%)	(34 ± 14, 60%)
	φ_3	(-19 ± 12, 1%)	----	----	(-34 ± 14, 40%)
	ψ_1	----	(35 ± 9, 100%)	(56 ± 9, 100%)	----
	ψ_2	----	----	----	(-161 ± 18, 7%)
	ψ_3	----	----	----	(-41 ± 19, 93%)
	ψ_4	(-16 ± 28, 100%)	----	----	----
2-3	φ_1	(47 ± 12, 99%)	(52 ± 7, 100%)	(53 ± 7, 100%)	(45 ± 9, 100%)
	φ_3	(-24 ± 16, 1%)	----	----	----
	ψ_1	----	----	----	----
	ψ_2	(160 ± 14, 1%)	----	----	----
	ψ_3	----	(-28 ± 8, 100%)	(-25 ± 8, 100%)	(-34 ± 10, 100%)
	ψ_4	(-5 ± 25, 99%)	----	----	----
3-4	φ_1	(37 ± 14, 98%)	(24 ± 9, 100%)	(25 ± 9, 100%)	(28 ± 10, 100%)
	φ_3	(-21 ± 21, 2%)	----	----	----
	ψ_1	----	(27 ± 8, 100%)	(31 ± 8, 100%)	(21 ± 10, 100%)
	ψ_2	----	----	----	----
	ψ_3	----	----	----	----
	ψ_4	(3 ± 25, 100%)	----	----	----
4-5	φ_1	(35 ± 15, 94%)	(16 ± 12, 100%)	(9 ± 11, 100%)	(48 ± 14, 92%)
	φ_2	(178 ± 15, 3%)	----	----	(154 ± 19, 8%)
	φ_3	(-13 ± 12, 3%)	----	----	----
	ψ_1	----	----	----	----
	ψ_2	----	----	----	----
	ψ_3	----	(-42 ± 13, 100%)	(-49 ± 10, 100%)	----
	ψ_4	(-4 ± 33, 100%)	----	----	(-17 ± 22, 100%)

[a] φ (H_1-C_1-O_x-C_x) and ψ (C_1-O_x-C_x-H_x). φ_1, φ_2 and φ_3 denote the +*gauche*, *anti* and -*gauche* rotamers, respectively. ψ_1, ψ_2, ψ_3 and ψ_4 denote the +*gauche*, *anti*, -*gauche*, and *syn* rotamers, respectively.

Table 5.5. Employing thermodynamic integration calculations in the prediction of relative binding free energies (kcal/mol) of antigenic oligosaccharides[a] from *S. flexneri* to Fv (1m7d)

	Ligand in water		Ligand in complex		ΔE_{coul}	ΔE_{vdW}	$\Delta\Delta G$
A: 4-OH → 4-H	Coulomb	vdWaals	Coulomb	vdWaals			
HF/6-31G*	-33.45 ± 0.3	0.03 ± 0.1	-33.53 ± 0.2	-0.25 ± 0.3	-0.08 ± 0.4	-0.28 ± 0.3	-0.36 ± 0.5
B3LYP/cc-pVTZ	-31.06 ± 0.2	0.11 ± 0.1	-30.75 ± 0.3	0.20 ± 0.5	0.31 ± 0.4	0.09 ± 0.5	0.40 ± 0.6
						Experiment	0.6
B: 2-OH → 2-H							
HF/6-31G*	-46.78 ± 0.4	3.47 ± 0.2	-48.66 ± 0.3	4.48 ± 0.2	-1.88 ± 0.5	1.01 ± 0.3	-0.87 ± 0.6
B3LYP/cc-pVTZ	-45.17 ± 0.2	3.51 ± 0.2	-46.97 ± 0.2	3.87 ± 0.2	-1.80 ± 0.3	0.36 ± 0.3	-1.44 ± 0.4
						Experiment	-1.6
B: 2-OH → 2-Cl							
HF/6-31G*	-39.69 ± 0.4	2.48 ± 0.2	-41.33 ± 0.3	3.92 ± 0.3	-1.64 ± 0.5	1.44 ± 0.4	-0.20 ± 0.6
B3LYP/cc-pVTZ[b]	-39.51 ± 0.3	2.66 ± 0.2	-41.88 ± 0.2	3.24 ± 0.2	-2.37 ± 0.4	0.58 ± 0.3	-1.79 ± 0.5
B3LYP/cc-pVTZ	-39.17 ± 0.3	2.86 ± 0.2	-41.95 ± 0.3	3.97 ± 0.2	-2.78 ± 0.4	1.11 ± 0.3	-1.67 ± 0.5
						Experiment	-2.1
C: 4-OH → 4-H							
HF/6-31G*	-18.30 ± 0.3	2.00 ± 0.3	-16.40 ± 0.4	2.94 ± 0.3	1.90 ± 0.5	0.94 ± 0.4	2.84 ± 0.6
B3LYP/cc-pVTZ	-5.51 ± 0.3	2.26 ± 0.2	-5.02 ± 0.3	3.57 ± 0.1	0.49 ± 0.4	1.31 ± 0.2	1.80 ± 0.4
						Experiment	inactive

[a] Employing α-L-Rha-(1,3)-α-L-Rha-(1,3)-β-D-GlcNAc-OMe (A-B-C-OMe).
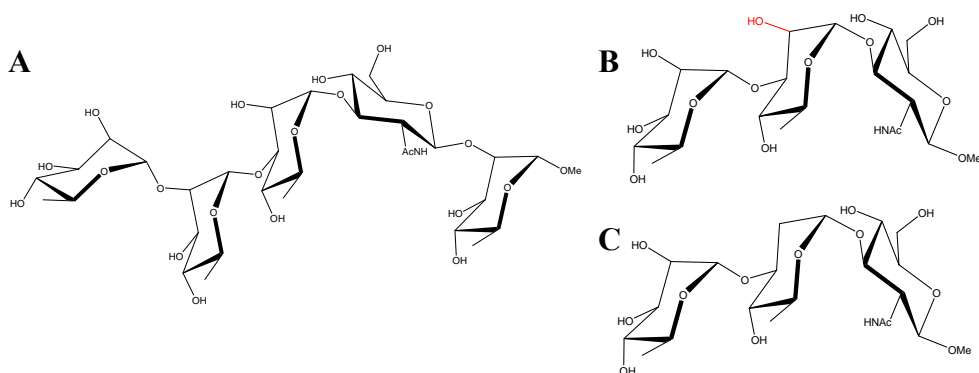[b] Charges only on perturbed residue

Figure 5.1. Schematic representations of the natural pentasaccharide (A) and trisaccharide (B), and the synthetic trisaccharide (C) fragments from the *S. flexneri* variant Y *O*-antigen
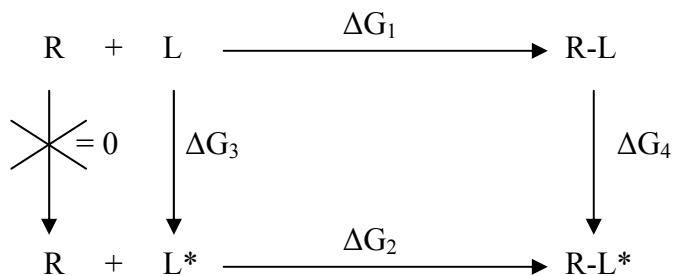

Figure 5.2. Schematic representation of a thermodynamic closed-cycle employed to determine the relative affinities of an antibody (R) for two closely related ligands L and L*. ΔG_1 and ΔG_2 are the changes in free energy of R binding L and L*, respectively. ΔG_3 and ΔG_4 are the changes in solvation and binding free energies, respectively, between L and L*
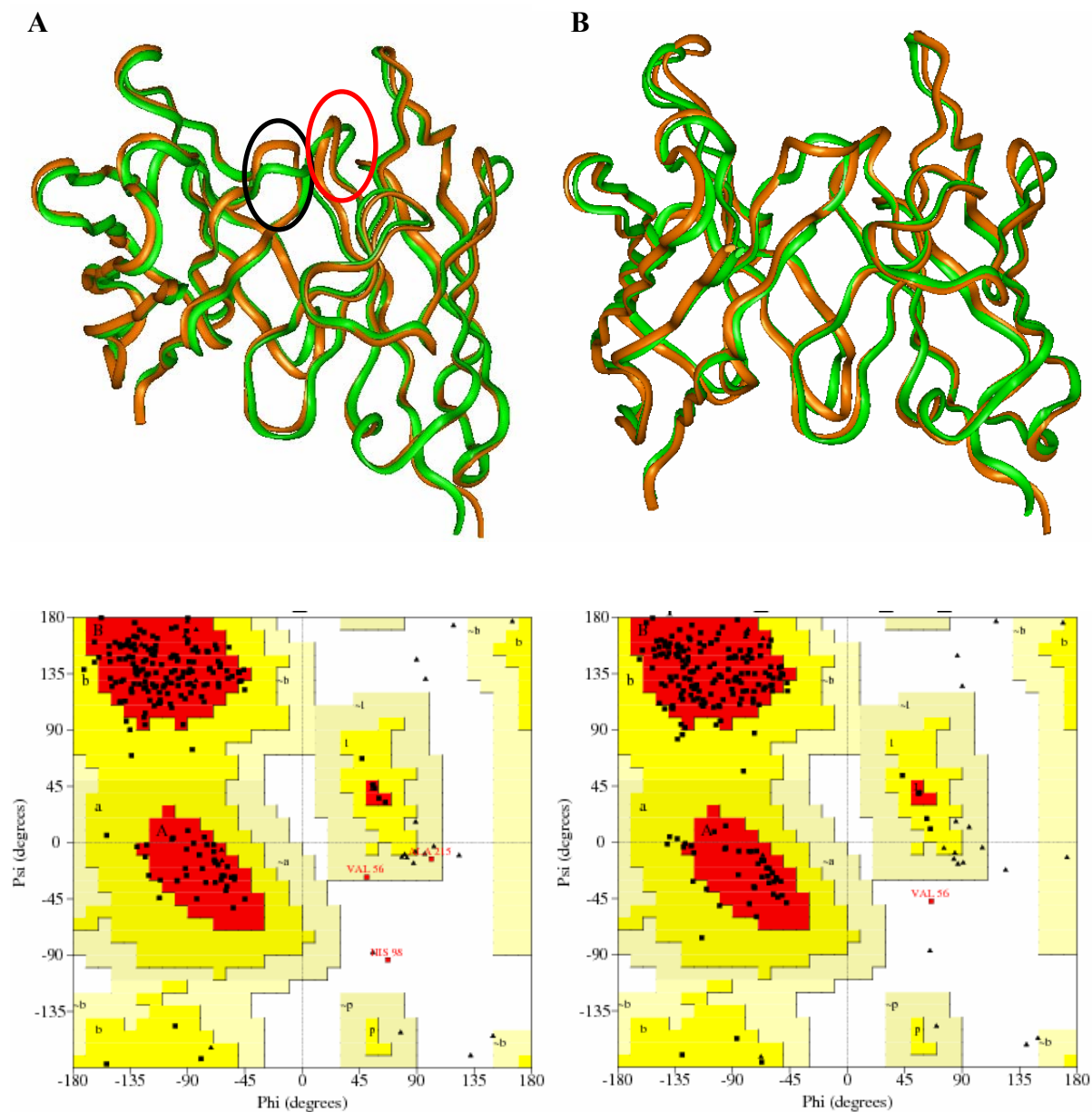
Figure 5.3. Backbone-atom superposition between the comparative model (orange) and crystal structure (green), before (A) and after (B) refinements. The CDRL3 and CDRH3 loops are highlighted in the red and black ovals, respectively. C and D represent Ramachandran plots of the comparative model before and after refinements, respectively
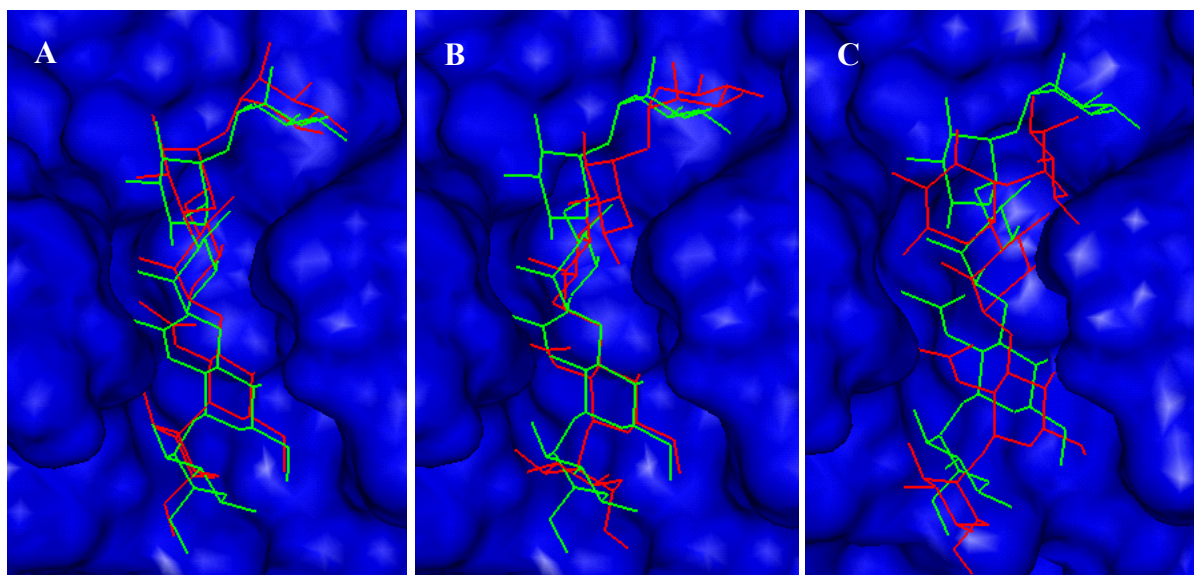
Figure 5.4. Docking results from three different schemes: A) utilizing the Fv from the Fab-oligosaccharide X-ray structure; B) Fv from X-tal structure of free Fab; C) Comparative modeled Fv. The experimental structure of the oligosaccharide is depicted in green, the docked in red. *S. flexneri* variant Y pentasaccharide: α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-GlcNAc-(1→2)-α-L-Rha-OMe
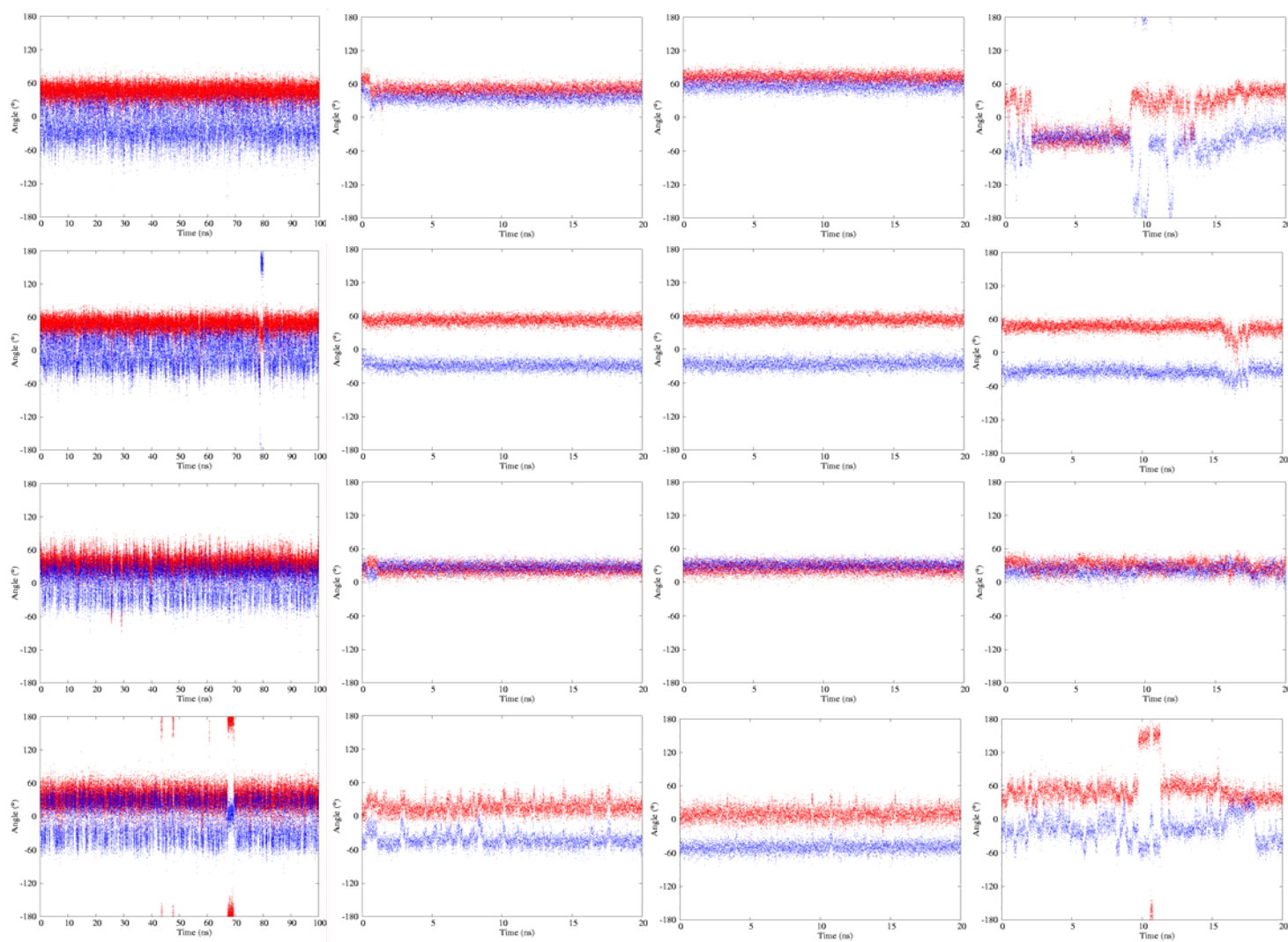
Figure 5.5. Trajectories of the φ (H$_1$-C$_1$-O$_x$-C$_x$) and ψ (C$_1$-O$_x$-C$_x$-H$_x$) angles of the pentasaccharide during the MD simulations. Left to right: the free ligand in solution; simulation initiated from the co-crystallized Fab; complex generated by docking to the free Fab, and the comparative model. Top to bottom: linkages between residues 1-2, 2-3, 3-4, and 4-5, respectively. φ – red, ψ – blue
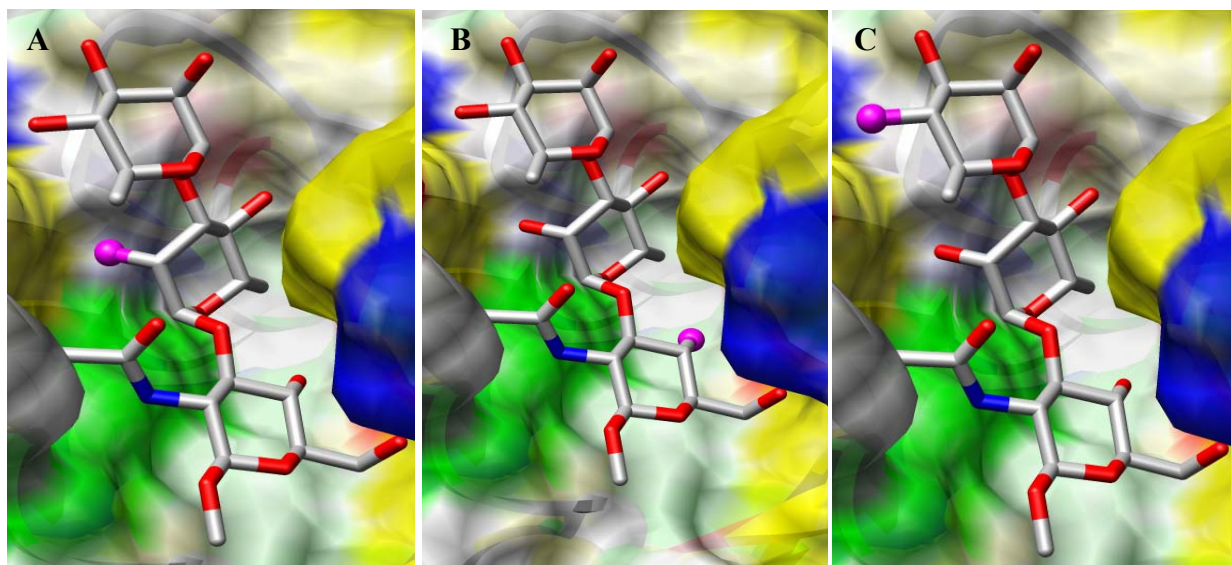
Figure 5.6. Schematic representation of a 2.3-Å resolution complex of the trisaccharide and the Fab utilized in the thermodynamic integration calculations. The synthetic ligand was converted to the natural ligand by substituting the non-terminal glycan residue with α-L-Rha*p*. The panels indicate the following mutations of the natural ligand: A; 2-hydroxyl group to both the 2-deoxy and 2-deoxychloro glycans. B and C represent the 4-hydroxy to the 4-deoxy mutations.

+ve = blue; -ve = red; polar = green; nonpolar = gray; aromatic = yellow
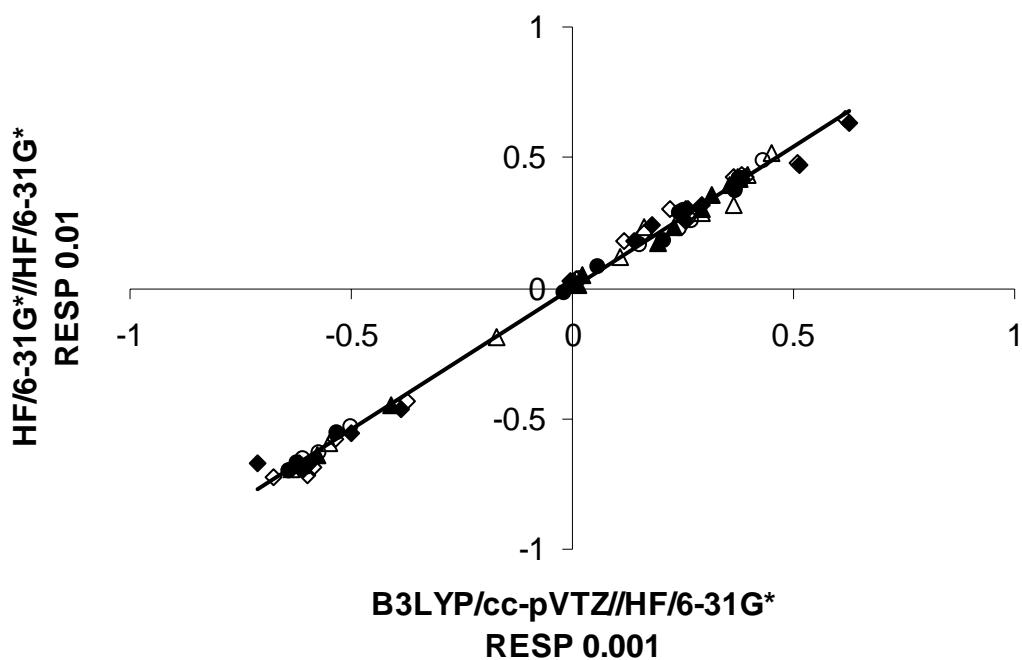
Figure 5.7. A comparison of the atom-centered charges derived by employing restraint weights of 0.001 and 0.01 for the B3LYP/cc-pVTZ//HF/6-31G* and HF/6-31G*//HF/6-31G* electrostatic potentials, respectively. β-D-GlcNAc (◊); 4-deoxy-β-D-GlcNAc (♦); α-L-Rha (○); 2-deoxy-α-L-Rha (●);2-deoxy-2-chloro-α-L-Rha (Δ); 2-deoxy-α-L-Rha (▲)

CHAPTER 6

Conclusions and future directions

The derivation of a generalizable MM parameter set (GLYCAM06) for modeling a broad range of biomolecules, with emphasis on carbohydrates, has been presented. Atom-centered partial charges of molecules were derived by reproducing their respective QM molecular electrostatic potentials. The vdW parameters implemented are those currently employed in the PARM94 protein force field. Valence bond and angle equilibrium values were obtained by empirical adjustments to reproduce experimental values obtained from neutron diffraction or X-ray crystallography studies, while force constants and torsion angle parameters were derived by minimizing the error between MM and QM distortion and rotational energy curves, respectively.

Unlike in previous versions of GLYCAM, aliphatic hydrogen atoms bear no partial charges and the need to scale 1-4 nonbonded interactions has been eliminated. In addition, a common set of terms can now be employed for both the α- and β-anomers of carbohydrates, which makes it feasible to determine the relative stabilities of common carbohydrate ring forms, when combined with the appropriate charge sets. The parameters are able to reproduce the experimental and QM gas phase rotational energy profiles and barriers of small molecules. Moreover, explicit-solvent MD simulations employing the parameter sets also reproduce solution phase experimental properties, such as the populations of rotational isomeric states, scalar ^3J-coupling constants and nuclear Overhauser effect distances of carbohydrates. For some highly charged carbohydrates, such as ulosonic acids with a negatively charged carboxyl group at the anomeric center, it is noted that utilizing torsion angle parameters developed by reproducing the

QM generated rotational energy curves of small molecules does not result in experimentally consistent rotamer populations. In this specific case, dihedral angle parameters have been generated by empirically adjusting torsion terms and performing MD simulations until the correct experimental rotamer populations are obtained. For such highly polar systems, the inability of the QM-derived torsion terms to reproduce the experimental populations may be attributed to the lack of charge polarizabilities in this classical force field. It is conceivable that the presence of a negatively charged group at the anomeric center, introduces stereoelectronic effects that cannot be included in this classical force field. The inclusion of a polarizable potential in future versions of GLYCAM may eliminate the necessity of employing explicit-solvent MD simulations when developing the torsion terms for highly polar carbohydrate residues.

Except for a few elaborate studies, the majority of explicit solvent MD simulations aimed at investigating the conformational properties of carbohydrates are typically less than 10 ns. It is now well established that the lifetimes of carbohydrate rotamers span the 10-ns regimes. Therefore, longer simulations times or enhanced sampling methods are needed in order to provide structural interpretations that are consistent with experimental studies. Here, the conformational properties of carbohydrates have been investigated through explicit solvent traditional MD and REMD simulation methods in the 100-ns regime. For small carbohydrate fragments, results from both simulations are statistically indistinguishable. However, for larger size oligosaccharides enhanced-sampling methods are superior, which suggests that simulations employing enhanced-sampling methods will be more efficient in attaining experimentally relevant conformational distributions in systems involving larger oligosaccharides.

GLYCAM06 has been employed to provide a structural interpretation to the experimentally observed immunological properties of the capsular polysaccharides from *Nm*B. The current lack of a highly immunogenic vaccine against *Nm*B, and the hypothesized key role of an extended conformational epitope in determining the antigenic properties and bactericidal activity of this CPS have served as an impetus in studying its dynamics and conformational properties. Efforts to design more immunogenic vaccines have involved chemical group modifications of the acyl groups. As an initial step in quantifying the antigenicities of native and chemically modified *Nm*B CPS, the conformational properties of di- and trisaccharide haptens have been characterized via MD simulations and confirmed by NMR spectroscopy. Results from the simulations suggest that the search for more immunogenic carbohydrate vaccines that can mimic the *Nm*B CPS should consider chemical group modifications that do not alter the *N*-acetyl moieties or the global conformation. The studies presented here, have focused on small size oligosaccharides for which there are reliable experimental NMR nOe and scalar *J*-couplings data to validate the MD simulations. While these simulations have provided initial structural insights about the possible origins of the observed antibody cross-reactivity, future studies will include the utilization of oligosaccharides of at least ten residues, which is the length required to form an immunologically relevant epitope of the type B CPS. The availability of experimental NMR data for the larger size CPS fragments will facilitate the validation of NMR observables predicted from MD simulations. These MD simulations will also aid in elucidating the conformational distribution of states of the CPS, by serving a complementary role with the experimental data. In the future, a full characterization of the antigenicities of the native and synthetic analogs will include larger oligosaccharide fragments. Complexes of oligosaccharide-antibody structures will

be generated, subjected to MD simulations to include induced-fit effects and relative binding energies will be determined via MM-GBSA methods.

The ability of computational protocols to reproduce the binding modes (structures of complexes and relative affinities) of antibody-carbohydrate antigen systems has been determined by employing docking simulations and thermodynamic integration MD simulations. In addition to the utilization of experimental structures, the Fv domain of an antibody was generated via a combination of knowledge-based and *ab initio* methods. Despite the low quality of the initial model, MD simulations were able to improve the quality of the model by correcting loop alignments relative to the experimental structure. The docking simulations highlighted the significance of employing high quality structures of receptors for generating ligand-receptor complexes. When the binding pocket was optimally folded to recognize the ligand, the alignment of the ligand in the docked complex compared well with its alignment in the experimental complex. The reproduction of the experimentally complex became increasingly poor when the ligand was docked to the Fv domain derived from the free Fab, due to the side chains not pre-disposed to recognize the ligand or with different backbone structures. However, MD simulations corrected these discrepancies.

In the case of predicting ligand-receptor affinities, the TI calculations reasonably reproduced the experimental relative binding affinities. It is noted that the perturbations in this series of ligands are relatively modest. Nonetheless, partial charge sets derived by reproducing the HF/6-31G*//HF/6-31G* molecular electrostatic potential did not accurately reproduce the relative affinities of the antibody for the series of ligands. The inclusion of diffuse functions, electron correlation and larger basis sets (B3LYP/cc-pVTZ//HF/6-31G*) in computing the molecular electrostatic potential, resulted in partial atomic charges that were superior in terms of

221

reproducing the relative binding affinities of the ligands. These results are encouraging and suggest that the AMBER force field augmented with the GLYCAM06 force field for carbohydrates can be employed to compute the relative free energies of binding in carbohydrate-protein interactions.

GLOSSARY

Antigen	A foreign substance that is recognized by an immune response
Antigenicity	The ability of an antibody to bind to an antigen with a certain affinity
Complementarity Determining Region	The region of an antibody's surface that is responsible for binding to an antigen
Epitope	The region of an antigen that is specifically recognized by an antibody
Immunogen	A foreign substance that is capable of stimulating an immune response
Immunogenicity	The ability of a foreign material to stimulate an immune response
Structurally Conserved Region	The structural regions of an antibody molecule, which are widely conserved among all their structures
Tolerance	Unresponsiveness of an immune system to an antigen or immunogen

REFERENCES

(1)   Ferreiros, C. M.; Gomez, J. A.; Criado, M. T. *Rev. Med. Microbiol.* **1998**, *9*, 29-37.

(2)   Anonymous *Weekly Epidemiological Record* **2002**, *77*, 331-339.

(3)   Jones, C. *An. Acad. Bras. Cienc.* **2005**, *77*, 293-324.

(4)   Michon, F.; Brisson, J. R.; Roy, R.; Ashton, F. E.; Jennings, H. J. *Biochemistry* **1985**, *24*, 5592-5598.

(5)   Spinosa, M. R.; Progida, C.; Tala, A.; Cogli, L.; Alifano, P.; Bucci, C. *Infect. Immun.* **2007**, *75*, 3594-3603.

(6)   Jarvis, G. A.; Vedros, N. A. *Infect. Immun.* **1987**, *55*, 174-180.

(7)   Tsang, R. S. W.; Zollinger, W. D. *Clin. Diagn. Lab. Immunol.* **2005**, *12*, 152-156.

(8)   Bhattacharjee, A. K.; Jennings, H. J.; Kenny, C. P.; Martin, A.; Smith, I. C. P. *J. Biol. Chem.* **1975**, *250*, 1926-1932.

(9)   Jennings, H. J.; Roy, R.; Michon, F. *J. Immunol.* **1985**, *134*, 2651-2657.

(10)   Kabat, E. A. *Structural concepts in immunology and immunochemistry*; 2nd ed.; Holt, Rinehart and Winston: New York, 1976.

(11)   Lifely, M. R.; Lindon, J. C.; Williams, J. M.; Moreno, C. *Carbohydr. Res.* **1985**, *143*, 191-205.

(12)   Glode, M. P.; Lewin, E. B.; Sutton, A.; Le, C. T.; Gotschlich, E. C.; Robbins, J. B. *J. Infect. Dis.* **1979**, *139*, 52-59.

(13)   Glode, M. P.; Robbins, J. B.; Liu, T. Y.; Gotschlich, E. C.; Orskov, I.; Orskov, F. *J. Infect. Dis.* **1977**, *135*, 94-102.

(14)   Kabat, E. A.; Liao, J.; Osserman, E. f.; Gamian, A.; Michon, F.; Jennings, H. J. *J. Exp. Med.* **1988**, *168*, 699-711.

(15)   Brisson, J. R.; Baumann, H.; Imberty, A.; Perez, S.; Jennings, H. J. *Biochemistry* **1992**, *31*, 4996-5004.

(16)   Baumann, H.; Brisson, J. R.; Michon, F.; Pon, R.; Jennings, H. J. *Biochemistry* **1993**, *32*, 4007-4013.

(17)   Weintraub, A. *Carbohydr. Res.* **2003**, *338*, 2539-2547.

(18)   Finne, J.; Leinonen, M.; Makela, P. H. *Lancet* **1983**, *2*, 355-357.

(19)   Finne, J.; Makela, P. H. *J. Biol. Chem.* **1985**, *260*, 1265-1270.

(20)   Hayrinen, J.; Jennings, H.; Raff, H. V.; Rougon, G.; Hanai, N.; Gerardy-Schahn, R.; Finne, J. *J. Infect. Dis.* **1995**, *171*, 1481-1490.

(21)   Muhlenhoff, M.; Eckhardt, M.; Gerardy-Schahn, R. *Curr. Opin. Struct. Biol.* **1998**, *8*, 558-564.

(22)   Evans, S. V.; Sigurskjold, B. W.; Jennings, H. J.; Brisson, J. R.; To, R.; Tse, W. C.; Altman, E.; Frosch, M.; Weisgerber, C.; Kratzin, H. D.; Klebert, S.; Vaesen, M.; Bittersuermann, D.; Rose, D. R.; Young, N. M.; Bundle, D. R. *Biochemistry* **1995**, *34*, 6737-6744.

(23)   Granoff, D. M.; Bartoloni, A.; Ricci, S.; Gallo, E.; Rosa, D.; Ravenscroft, N.; Guarnieri, V.; Seid, R. C.; Shan, A.; Usinger, W. R.; Tan, S.; McHugh, Y. E.; Moe, G. R. *J. Immunol.* **1998**, *160*, 5028-5036.

(24) Moe, G. R.; Dave, A.; Granoff, D. M. *Mol. Immunol.* **2006**, *43*, 1424-1431.
(25) Moe, G. R.; Dave, A.; Granoff, D. M. *Infect. Immun.* **2005**, *73*, 2123-2128.
(26) Pon, R. A.; Lussier, M.; Yang, Q. L.; Jennings, H. J. *J. Exp. Med.* **1997**, *185*, 1929-1938.
(27) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683-720.
(28) Varki, A. *Glycobiology* **1993**, *3*, 97-130.
(29) Cumming, D. A.; Carver, J. P. *Biochemistry* **1987**, *26*, 6664-6676.
(30) Imberty, A.; Perez, S. *Chem. Rev.* **2000**, *100*, 4567-4588.
(31) Woods, R. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 591-598.
(32) Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371-386.
(33) Gonzalez-Outeirino, J.; Kadirvelraj, R.; Woods, R. J. *Carbohydr. Res.* **2005**, *340*, 1007-1018.
(34) Almond, A.; Sheehan, J. K. *Glycobiology* **2003**, *13*, 255-264.
(35) Sayers, E. W.; Prestegard, J. H. *Biophys. J.* **2002**, *82*, 2683-2699.
(36) *NMR spectroscopy and computer modeling of carbohydrates: Recent advances*; Vliegenthart, J. F. G.; Woods, R. J., Eds., 2006.
(37) Lemieux, R. U.; Koto, S.; Voisin, D. *In anomeric effect, Origin and Consequences* Washington, DC, 1979.
(38) Perez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J.-F.; van Eyek, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenusi, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141-155.
(39) Almond, A.; Sheehan, J. K. *Glycobiology* **2000**, *10*, 329-338.
(40) Gonzalez-Outeirino, J.; Kirschner, K. N.; Thobhani, S.; Woods, R. J. *Can. J. Chem.* **2006**, *84*, 569-579.
(41) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10541-10545.
(42) Hoffmann, M.; Rychlewski, J. *J. Am. Chem. Soc.* **2001**, *123*, 2308-2316.
(43) Angyal, S. J. *Angewandte Chemie-International Edition English* **1969**, *8*, 157-166.
(44) Lii, J. H.; Chen, K.-H.; Durkin, K. A.; Allinger, N. L. *J. Comput. Chem.* **2003**, *24*, 1473-1489.
(45) Kirschner, K. N.; Yongye, A. B.; Gonzalez-Outeirino, J.; Tschampel, S. M.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2007**, *In press*.
(46) Dabrowski, J.; Kozar, T.; Grosskurth, H.; Nifantev, N. E. *J. Am. Chem. Soc.* **1995**, *117*, 5534-5539.
(47) Rockwell, G. D.; Grindley, T. B. *J. Am. Chem. Soc.* **1998**, *120*, 10953-10963.
(48) Bock, K.; Duus, J. O. *J. Carbohydr. Chem.* **1994**, *13*, 513-543.
(49) Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. *J. Carbohydr. Chem.* **1988**, *7*, 239-250.
(50) Nishida, Y.; Ohrui, H.; Meguro, H. *Tetrahedron Lett.* **1984**, *25*, 1575-1578.
(51) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420-433.
(52) Rathore, N.; Chopra, M.; J. de Pablo, J. *J. Chem. Phys.* **2005**, *122*, 024111.
(53) Naidoo, K. J.; Chen, J. Y. J. *Mol. Phys.* **2003**, *101*, 2687-2694.
(54) Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244-2252.
(55) Schauer, R. *Zoology* **2004**, *107*, 49-64.
(56) Breg, J.; Kroon-Batenburg, M. J. L.; Strecker, G.; Montreuil, J.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1989**, *178*, 727-739.

(57)   Poppe, L.; Stuike-Prill, S.; Meyer, B.; van Halbeek, H. *J. Biomol. NMR* **1992**, *2*, 109-136.

(58)   Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327-335.

(59)   Kooijman, H.; Kroonbatenburg, L. M. J.; Kroon, J.; Breg, J. N.; Deboer, J. L. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1990**, *46*, 407-410.

(60)   Ginalski, K.; Grishin, N. V.; Godzik, A.; Rychlewski, L. *Nucleic Acids Res.* **2005**, *33*, 1874-1891.

(61)   Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901-917.

(62)   Al-Lazikani, B.; Lesk, A. M.; Chothia, C. *J. Mol. Biol* **1997**, *273*, 927-948.

(63)   Shirai, H.; Kidera, A.; Nakamura, H. *FEBS Lett.* **1996**, *399*, 1-8.

(64)   Morea, V.; Tramontano, A.; Rustici, M.; Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1998**, *275*, 269-294.

(65)   Morea, V.; Tramontano, A.; Rustici, M.; Chothia, C.; Lesk, A. M. *Biophys. Chem.* **1997**, *68*, 9-16.

(66)   Morea, V.; Lesk, A. M.; Tramontano, A. *Methods* **2000**, *20*, 267-279.

(67)   Chothia, C.; Gelfand, I.; Kister, A. *J. Mol. Biol* **1998**, *278*, 457-479.

(68)   Martin, A. C. R.; Cheetham, J. C.; Rees, A. R. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9268-9272.

(69)   Moult, J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 285-289.

(70)   Higo, J. I.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33-43.

(71)   Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137-168.

(72)   R. M. Fine, H. W. P. S. S. D. L. Y. C. L. *Proteins: Structure, Function, and Genetics* **1986**, *1*, 342-362.

(73)   Fernandez-Fuentes, N.; Oliva, B.; Fiser, A. *Nucleic Acids Res.* **2006**, *34*, 2085-2097.

(74)   Ginalski, K. *Curr. Opin. Struct. Biol.* **2006**, *16*, 172-177.

(75)   Sadowski, M. I.; Jones, D. T. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 476-485.

(76)   Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389-3402.

(77)   Pearson, W. R.; Lipman, D. J. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2444-2448.

(78)   Fernandez-Fuentes, N.; Rai, B. K.; Madrid-Aliste, C. J.; Fajardo, J. E.; Fiser, A. *Bioinformatics* **2007**, *23*, 2558-2565.

(79)   Weiner, J. S.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 230-252.

(80)   Qian, B.; Ortiz, A. R.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15346-15351.

(81)   Sotriffer, C. A.; Flader, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varga, J. M. *Methods* **2000**, *20*, 280-291.

(82)   Morris, G. M.; Goodsell, D. S.; Huey, R.; Hart, W. E.; Halliday, R. S.; Belew, R. K.; Olson, A. J.; The Scripps Research Institute, CA, USA: 2001.

(83)   Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*.

(84)   Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175-1189.

(85)   Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-489.

(86)   Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151-166.

(87)   Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Structure, Function and Genetics* **2002**, *47*, 409-443.

(88)   Wang, J.; Kollman, P. A.; Kuntz, I. D. *Proteins: Structure Function and Genetics* **1999**, *36*, 1-19.

(89)   Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213-218.

(90)   Stanfield, R. L.; Fieser, T. M.; Lerner, R. A.; Wilson, I. A. *Science* **1990**, *248*, 712-719.

(91)   Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A. *Science* **1992**, *255*, 959-965.

(92)   Guddat, L. W.; Shan, L.; Anchin, J. M.; Linthicum, D. S.; Edmundson, A. B. *J. Mol. Biol* **1994**, *236*, 247-274.

(93)   Alonso, H.; Bliznyuk, A. A.; Gready, J. E. *Med. Res. Rev.* **2006**, *26*, 531-568.

(94)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(95)   Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.

(96)   Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791-804.

(97)   Laederach, A.; Reilly, P. J. *J. Comput. Chem.* **2003**, *24*, 1748-1757.

(98)   Solis, F. J.; Wets, R. J. B. *Mathematics of Operations Research* **1981**, *6*, 19-30.

(99)   Pathiaseril, A.; Woods, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 331-338.

(100)   Leach, A. R. *Molecular Modelling. Principles and applications*; Second Edition ed.; Pearson Education Limited, 2001.

(101)   Ford, M. G.; Weimar, T.; Kohli, T.; Woods, R. J. *Proteins: Structure, Function and Genetics* **2003**, *53*, 229-240.

(102)   Kadirvelraj, R.; Gonzalez-Outeirino, J.; Foley, B. L.; Beckham, M. L.; Jennings, H. J.; Foote, S.; Ford, M. G.; Woods, R. J. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8149-8154.

(103)   Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14330-14335.

(104)   Weinhold, B.; Seidenfaden, R.; Rockle, I.; Muhlenhoff, M.; Schertzinger, F.; Conzelmann, S.; Marth, J. D.; Gerardy-Schann, R.; Hildebrandt, H. *J. Biol. Chem.* **2005**, *280*, 42971-42977.

(105)   Jin, L.; Abrahams, J. P.; Skinner, R.; Petitou, M.; Pike, R.; Carrel, R. W. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14683-14688.

(106)   Haltiwanger, R. S.; Lowe, J. B. *Annu. Rev. Biochem.* **2004**, *73*, 491-537.

(107)   Sanders, R. W.; Venturi, M.; Schiffner, L.; Kalyanaraman, R.; Katinger, H.; Lloyd, K. O.; Kwong, P. D.; Moore, J. P. *J. Virol.* **2002**, *76*, 7293-7305.

(108)   Karaveg, K.; Siriwardena, A.; Tempel, W.; Liu, Z. J.; Glushka, J.; Wang, B. C.; Moremen, K. W. *J. Biol. Chem.* **2005**, *280*, 16197-16207.

(109)   Kogelberg, H.; Solis, D.; Jimenez-Barbero, J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 646-653.

(110)   Landersjo, C.; Widmalm, G. *Biopolymers* **2002**, *64*, 283-291.

(111)   Rice, K. G.; Wu, P.; Brand, L.; Lee, Y. C. *Biochemistry* **1991**, *30*, 6646-6655.

(112)   Wooten, E. W.; Edge, C. J.; Bazzo, R.; Dwek, R. A.; Rademacher, T. W. *Carbohydr. Res.* **1990**, *203*, 13-17.

(113)   Cano, F. H.; Foces-Foces, C.; Jimenez-Barbero, J.; Alemany, A.; Bernabe, M.; Martin-Lomas, M. *J. Org. Chem.* **1987**, *52*, 3367-3372.

(114)   Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832-3846.

(115)   Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydr. Res.* **2004**, *339*, 937-948.

(116)   Corzana, F.; Motawia, M. S.; Du Penhoat, C. H.; Perez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. *J. Comput. Chem.* **2004**, *25*, 573-586.

(117) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(118) Jeffrey, G. A. *J. Mol. Struct.* **1994**, *322*, 21-25.

(119) Gruza, J.; Koca, J.; Perez, S.; Imberty, A. *J. Mol. Struct. (Theochem)* **1998**, *424*, 269-280.

(120) Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421-8425.

(121) Chen, S.-Y.; Lin, T.-H. *J. Phys. Chem. B* **2005**, *109*, 9764-9772.

(122) Chevalier, F.; Lopez-Prados, J.; Perez, S.; Martin-Lomas, M.; Nieto, P. M. *Eur. J. Org. Chem.* **2005**, *16*, 3489-3498.

(123) Nguyen, D. H.; Colvin, M. E.; Yeh, Y.; Feeney, R. E.; Fink, W. H. *Biophys. J.* **2002**, *82*, 2892-2905.

(124) Perera, L.; Darden, T. A.; Pedersen, L. G. *J. Comput. Chem.* **2002**, *23*, 35-47.

(125) Shroll, R. M.; Straatsma, T. P. *Biophys. J.* **2003**, *84*, 1765-1772.

(126) Tempel, W.; Tschampel, S.; Woods, R. J. *J. Biol. Chem.* **2002**, *277*, 6615-6621.

(127) Umemura, M.; Yuguchi, Y.; Hirotsu, T. *J. Mol. Struct. (Theochem)* **2005**, *130*, 1-8.

(128) Vishnyakov, A.; Widmalm, G.; Kowalewski, J.; Laaksonen, A. *J. Am. Chem. Soc.* **1999**, *121*, 5403-5412.

(129) Woods, R. J.; Chappelle, R. *J. Mol. Struct. (Theochem)* **2000**, *527*, 149-156.

(130) Kirschner, K. N.; Woods, R. J. *J. Phys. Chem. A* **2001**, *105*, 4150-4155.

(131) Lii, J. H.; Ma, B. Y.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593-1603.

(132) Momany, F. A.; Willet, J. L. *Carbohydr. Res.* **2000**, *326*, 210-226.

(133) Momany, F. A.; Willet, J. L. *Carbohydr. Res.* **2000**, *326*, 194-209.

(134) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, Jr., R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A.G.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; 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.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E.S. and Pople, J. A.; Revision A.11.3 ed.; Gaussian 98. Gaussian, Inc.: Pittsburgh, PA, 2002.

(135) Case, D. A. P., D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Weiner, P. K.; Kollman, P. A.; AMBER 7. University of California: San Francisco, CA: 2002.

(136) Case, D. A. D., T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Lou, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, P.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.; AMBER 8. University of California: San Francisco, CA: 2004.

(137) Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.

(138) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311-1327.

(139) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

(140) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275-282.

(141) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

(142) Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* **2001**, *22*, 1125-1137.

(143) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Lou, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, 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.; AMBER 9. University of California: San Francisco, CA: 2006.

(144) Brown, G. M.; Levy, H. A. *Acta Crystallogr., Sect. B: Struct. Sci* **1979**, *35*, 656-659.

(145) Korolevich, M. *J. Mol. Struct.* **1994**, *306*, 261-268.

(146) Korolevich, M.; Sivchik, V.; Zhbankov, R.; Marchenko, G.; Zabelin, L. *J. Appl. Spectrosc.* **1992**, *56*, 217-223.

(147) Korolevich, M.; Zhbankov, R.; Sivchik, V. *J. Mol. Struct.* **1990**, *220*, 301-313.

(148) Dauchez, M.; Derremaus, P.; Vergoten, G. *J. Comput. Chem.* **1992**, *14*, 263-277.

(149) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236-1243.

(150) Huvenne, J.; Vergoten, G.; Fleury, G.; Legrand, P. *J. Mol. Struct.* **1981**, *74*, 169-180.

(151) Gregurick, S.; Kafafi, S. *J. Carbohydr. Chem.* **1999**, *18*, 867-890.

(152) Wells, H.; Atalla, R. *J. Mol. Struct.* **1990**, *224* 385-424.

(153) Prima, A.; Zhbankov, R.; Marupov, R. *J. Struct. Chem.* **1967**, *5*, 783-788.

(154) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith Jr, V. H. *J. Comput. Chem.* **1990**, *11*, 297-310.

(155) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361-373.

(156) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910-8922.

(157) Francl, M. M.; Carey, C.; Chirlian, L. E. *J. Comput. Chem.* **1996**, *17*, 367-383.

(158) *Conformational behavior of six-membered rings: analysis, dynamics and stereochemical effects*; Juaristi, E., Ed., 1995.

(159) Appell, M.; Strati, G.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2004**, *339*, 537-551.

(160) Jeffrey, G. A.; Pople, J. A.; Radom, L. *Carbohydr. Res.* **1972**, *25*, 117-131.

(161) Jeffrey, G. A.; Pople, J. A.; Binkley, J. S.; Vishveshwara, S. *J. Am. Chem. Soc.* **1978**, *100*, 373-379.

(162) Schleifer, L.; Senderowitz, H.; Aped, P.; Tartakovsky, E.; Fuchs, B. *Carbohydr. Res.* **1990**, *206*, 21-39.

(163) Dinur, U.; Hagler, A. T. *In Reviews in Computational Chemistry: Chapter 4*; VCH Publishers: New York, 1991.

(164) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, P. M. *Biopolymers* **1984**, *23*, 1513-1518.

(165) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher,

W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(166) Mackerell Jr, A. D. *J. Comput. Chem.* **2004**, *25*, 1584-1604.

(167) Potenzone, J. R.; Hopfinger, A. J. *Carbohydr. Res.* **1976**, *46*, 67-73.

(168) Arnott, S.; Scott, W. E. *J. Chem. Soc., Perkin Trans. 2* **1972**, 324-335.

(169) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(170) Compton, D. A. C.; Montero, S.; Murphy, W. F. *J. Phys. Chem.* **1980**, *84*, 3587-3591.

(171) Durig, J. R.; Bucy, W. E.; Wurrey, C. J.; Carreira, L. A. *J. Phys. Chem.* **1975**, *79*, 988-993.

(172) van Alsenoy, C.; Scarsdale, J. N.; Williams, J. O.; Schafer, L. *J. Mol. Struct.* **1982**, *86*, 365-376.

(173) Schafer, L.; van Alsenoy, C.; Scarsdale, J. N. *J. Mol. Struct.* **1982**, *86*, 349-364.

(174) Barrow, G. M. *J. Chem. Phys.* **1952**, *20*, 1739-1744.

(175) Sasada, Y. *J. Mol. Struct.* **1988**, *190*, 93-97.

(176) Weldon, A. J.; Vickrey, T. L.; Tschumper, G. S. *J. Phys. Chem. A* **2005**, *109*, 11073-11079.

(177) Houk, K. N.; Eksterowicz, J. E.; Wu, Y.-D.; Fuglesang, C. D.; Mitchell, D., B *J. Am. Chem. Soc.* **1993**, *115*, 4170-4177.

(178) Flippen, J. L. *Acta Crystallogr., Sect. B: Struct. Sci* **1973**, *29*, 1881-1886.

(179) Alagona, G.; Ghio, C. *J. Mol. Struct. (Theochem)* **1992**, *254*, 287-300.

(180) Bako, I.; Grosz, T.; Palinkas, G. *J. Chem. Phys.* **2003**, *118*, 3215-3221.

(181) Bastiansen, O. *Acta Chem. Scand.* **1949**, *3*, 415-421.

(182) Chidichimo, G.; Imbardelli, D.; Longeri, M.; Saupe, A. *Mol. Phys.* **1988**, *65*, 1143-1152.

(183) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1994**, *116*, 3892-3900.

(184) Erdem, S. S.; Varnali, T.; Aviyente, V. *J. Phys. Org. Chem.* **1997**, *10*, 196-206.

(185) Gubskaya, A. V.; Kusalik, P. G. *J. Phys. Chem. A* **2004**, *108*, 7165-7178.

(186) Hommel, E. L.; John, M. K.; Ma, G.; Hadad, C. M.; Allen, H. C. *J. Phys. Chem. B* **2005**, *109*, 811-818.

(187) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*, 3639-3646.

(188) Nagy, P. I.; Dunn III, W. J.; Alagona, G.; Ghio, C. *J. Am. Chem. Soc.* **1992**, *113*, 6719-6729.

(189) Pachler, K. G. R.; Wessels, P. L. *J. Mol. Struct.* **1970**, *6*, 471-478.

(190) Saiz, L.; Padro, J. A.; Guardia, E. *J. Chem. Phys.* **2001**, *114*, 3187-3199.

(191) Fateley, W. G.; Miller, F. A. *Spectrochim. Acta* **1962**, *18*, 977-993.

(192) Bedrov, D.; Borodin, O.; Smith, G. D. *J. Phys. Chem. B* **1998**, *102*, 5683-5690.

(193) Liu, H.; Muller-Plathe, F.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *1042*, 1722-1730.

(194) Smith, G. D.; Jaffe, R. L.; Yoon, D. Y. *J. Am. Chem. Soc.* **1995**, *117*, 530-531.

(195) Goutev, N.; Ohno, K.; Matsuura, H. *J. Phys. Chem. A* **2000**, *104*, 9226-9232.

(196) Begum, R.; Matsuura, H. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 3839-3848.

(197) Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 2078-2086.

(198) Fogarasi, G.; Szalay, P. G. *J. Phys. Chem. A* **1997**, *101*, 1400-1408.

(199) Langley, C. H.; Allinger, N. L. *J. Phys. Chem. A* **2002**, *106*, 5638-5652.

(200) Ross, B. D.; True, N. S.; Matson, G. B. *J. Phys. Chem.* **1984**, *88*, 2675-2678.

(201) Shen, Y.; Kohla, G.; Lrhorfi, A. L.; Sipos, B.; Kalthoff, H.; Gerwig, G. J.; Kamerling, J. P.; Schauer, R.; Tiralongo, J. *Eur. J. Biochem.* **2004**, *271*, 281-290.

(202) Sheridan, J.; Bossert, W.; Bauder, A. *J. Mol. Spectrosc.* **1980**, *80*, 1-11.

(203) Blom, C. E.; Gunthard, H. H. *Chem. Phys. Lett.* **1981**, *84*, 267-172.

(204) O'Connell, A. M. *Acta Crystallogr., Sect. B: Struct. Sci* **1973**, *29*, 2320-2328.

(205) Davis, J. T.; Hirani, S.; Bartlett, C.; Reid, B. R. *J. Biol. Chem.* **1994**, *269*, 3331-3338.

(206) Mobli, M.; Almond, A. *Org. Biomol. Chem.* **2007**, *4*, 2243-2251.

(207) Smondyrev, A. M.; Berkowitz, M. L. *J. Comput. Chem.* **1999**, *25*, 531-545.

(208) Momany, F. A.; Willet, J. L. *J. Comput. Chem.* **2000**, *21*, 1204-1219.

(209) Stenutz, R.; Carmichael, I.; Widmalm, G.; Serianni, A. S. *J. Org. Chem.* **2002**, *67*, 949-958.

(210) Haasnoot, C. A. G.; Deleeuw, F.; Deleeuw, H. P. M.; Altona, C. *Biopolymers* **1981**, *20*, 1211-1245.

(211) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354-1358.

(212) Jeffrey, G. A.; McMullan, R. K.; Takagi, S. *Acta Crystallogr., Sect. B: Struct. Sci* **1977**, *33*, 728-737.

(213) Takagi, S.; Jeffrey, G. A. *Acta Crystallogr., Sect. B: Struct. Sci* **1979**, *35*, 902-906.

(214) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydrate Research* **1988**, *180*, 207-221.

(215) Kouwijzer, M. L. C. E.; Vaneijck, B. P.; Kroes, S. J.; Kroon, J. *J. Comput. Chem.* **1993**, *14*, 1281-1289.

(216) Weiner, J. S.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. J.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

(217) Root, L. J.; Berne, B. J. *J. Chem. Phys.* **1997**, *107*, 4350-4357.

(218) Champeney, D. C.; Joadar, R. N.; Dore, J. C. *Mol. Phys.* **1986**, *58*, 337-347.

(219) Dawidowski, J.; Bermejo, F. J.; Fayos, R.; Fernandez Perea, R.; Bennington, S. M.; Criado, A. *Phys. Rev. E* **1996**, *53*, 5079-5088.

(220) Lee, R. E.; Chen, C.-P.; Denlinger, D. L. *Science* **1987**, *238*, 1415-1417.

(221) Dirama, T. E.; Carri, G. A.; Sokolov, A. P. *J. Chem. Phys.* **2005**, *122*, 244910.

(222) Dirama, T. E.; Carri, G. A.; Sokolov, A. P. *J. Chem. Phys.* **2005**, *122*, 114505.

(223) Na, G. C.; Timasheff, S. N. *J. Mol. Biol* **1981**, *151*, 165-178.

(224) van Koningsveld, H. *Recl. Trav. Chim* **1968**, *87*, 243-254.

(225) Chelli, R.; Procacci, P.; Cardini, G.; Della Valle, R. G.; Califano, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 871-877.

(226) Maccaferri, G.; Caminati, W.; Favero, P. G. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 4115-4117.

(227) Chelli, R.; Gervasio, F. L.; Gellini, C.; Procacci, P.; Cardini, G.; Schettino, V. *J. Phys. Chem. A* **2000**, *104*, 5351-5357.

(228) Chelli, R.; Gervasio, F. L.; Gellini, C.; Procacci, P.; Cardini, G.; Schettino, V. *J. Phys. Chem. A* **2000**, *104*, 11220-11222.

(229) Callam, C. S.; Singer, S. J.; Lowary, T. L.; Hadad, C. M. *J. Am. Chem. Soc.* **2001**, *123*, 11743-11754.

(230) Chelli, R.; Procacci, P.; Cardini, G.; Califano, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 879-885.

(231) van Koningsveld, H. *Recueil* **1970**, *89*, 801-812.

(232) Soltwisch, M.; Steffen, B. *Z. Naturforsch. A: Phys. Sci.* **1981**, *36*, 1045-1051.

(233) Garawi, M.; Dore, J. C.; Champeney, D. C. *Mol. Phys.* **1987**, *62*, 475-487.

(234) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6731-6736.

(235) Grayson, P.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 36-48.

(236) Lu, D.; Grayson, P.; Schulten, K. *Biophys. J.* **2003**, *85*, 2977-2987.

(237) Dashnau, J. L.; Nucci, N. V.; Sharp, K. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2006**, *110*, 13670-13677.

(238) Kosztin, I.; Schulten, K. *Phys. Rev. Lett.* **2004**, *93*, 238102(4).

(239) Darden, T.; York, D.; Pederson, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(240) Haasnoot, C. A. G.; Deleeuw, F.; Altona, C. *Tetrahedron* **1981**, *36*, 2783-2792.

(241) *Statistical Methods for Business and Economics, Revised Edition*; Pfaffenberger, R. C.; Patterson, J. H., Eds.; Richard D. Irwin: Homewood, Illinois, 1981.

(242) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., 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, J.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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, A.; 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.; Pople, J. A.; Revision C.02 ed.; Gaussian 03. Gaussian, Inc.: Wallingford, CT, 2004.

(243) Lutnaes, O. B.; Ruden, T. A.; Helgaker, T. *Magn. Reson. Chem.* **2004**, *42*, S117-S127.

(244) Riedo, F. X.; Plikaytis, B. D.; Broome, C. V. *Pediatr. Infect. Dis. J.* **1995**, *14*, 643-657.

(245) Tikhomirov, E.; Santamaria, M.; Esteves, K. *World Health Stat. Q.* **1997**, *50*, 170-177.

(246) Liu, T.-Y.; Gotschlich, E. C.; Jonssen, E. K.; Wysocki, J. R. *J. Biol. Chem.* **1971**, *246*, 2849-2858.

(247) Nato, F.; Mazie, J. C.; Fournier, J. M.; Slizewicz, B.; Sagot, N.; Guibourdenche, M.; Postic, D.; Riou, J. Y. *J. Clin. Microbiol.* **1991**, *29*, 1447-1452.

(248) Bhattacharjee, A. K.; Jennings, H. J.; Kenny, C. P.; Martin, A.; Smith, I. C. P. *Can. J. Biochem.* **1976**, *54*, 1-8.

(249) Bundle, D. R.; Smith, I. C. P.; Jennings, H. J. *J. Biol. Chem.* **1974**, *249*, 2275-2281.

(250) Jodar, L.; Griffiths, E.; Feavers, I. *Vaccine* **2004**, *22*, 1047-1053.

(251) Devi, S. J. N.; Robbins, J. B.; Schneerson, R. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7175-7179.

(252) Lifely, M. R.; Gilbert, A. S.; Moreno, C. *Carbohydr. Res.* **1981**, *94*, 193-203.

(253) Kielczynski, W.; Bartholomeusz, R. K.; Harrison, L. C. *Glycobiology* **1994**, *4*, 791-796.

(254) Azurmendi, H. F.; Vionnet, J.; Wrightson, L.; Trinh, L. B.; Shiloach, J.; Freedberg, D. I. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11557-11561.

(255) Pon, R. A.; Khieu, N. H.; Yang, Q. L.; Brisson, J. R.; Jennings, H. J. *Can. J. Chem.* **2002**, *80*, 1055-1063.

(256) Pon, R. A.; Biggs, N. J.; Jennings, H. J. *Glycobiology* **2007**, *17*, 249 - 260.

(257) Venable, R. M.; Bizik, F.; Henderson, T. J.; Egan, W. *J. Mol. Struct. (Theochem)* **1997**, *395*, 375-388.

(258) Vasudevan, S. V.; Balaji, P. V. *Biopolymers* **2002**, *63*, 168-180.

(259) Kirschner, K. N.; Yongye, A. B.; Gonzalez-Outeirino, J.; Tschampel, S. M.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622-655.

(260) Yongye, A. B.; Foley, B. L.; Woods, R. J. *J. Phys. Chem. A* **2008**, *112*, 2634-2639.

(261) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569-578.

(262) Schultheis, V.; Hirschberger, T.; Carstens, H.; Tavan, P. *J. Chem. Theory Comput.* **2005**, *1*, 515-526.

(263) Kloppenburg, M.; Tavan, P. *Phys. Rev. E* **1997**, *55*, R2089 - R2092.

(264) Thrippleton, M. J.; Keeler, J. *Angew. Chem. Int. Ed.* **2003**, *42*, 3938-3941.

(265) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

(266) NMRView Software; One Moon, Scientific, Inc.

(267) Michon, F.; Brisson, J. R.; Jennings, H. J. *Biochemistry* **1987**, *26*, 8399-8405.

(268) Rutherford, T. J.; Spackman, D. G.; Simpson, P. J.; Homans, S. W. *Glycobiology* **1994**, *4*, 59-68.

(269) Woods, R. J.; Pathiaseril, A.; Wormald, M. R.; Edge, C. J.; Dwek, R. A. *Eur. J. Biochem.* **1998**, *158*, 372-386.

(270) Lifely, M. R.; Esdaile, J. *Immunology* **1991**, *74*, 490-496.

(271) Park, I.; Choi, I.-H.; Kim, S. J.; Shin, J.-S. *Yonsei Medical Journal* **2004**, *45*, 755-758.

(272) Blundell, C. D.; Mahoney, D. J.; Almond, A.; DeAngelis, P. L.; Kahmann, J. D.; Teriete, P.; Pickford, A. R.; Campbell, I. D.; Day, A. J. *J. Biol. Chem.* **2003**, *278*, 49261-49270.

(273) Liang, G.; Schimdt, R.; Yu, H.-A.; Cumming, D.; Brady, J. *J. Phys. Chem.* **1996**, *100*, 2528-2534.

(274) Rockey, W. M.; Laederach, A.; Reilly, P. J. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 299-309.

(275) Laederach, A.; Dowd, M. K.; Coutinho, P. M.; Reilly, P. J. *Proteins: Struct., Funct., Genet.* **1999**, *37*.

(276) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks III, C. L. *J. Comput. Chem.* **1998**, *19*, 1612-1622.

(277) Simonson, T. *J. Phys. Chem. B* **2000**, *104*, 6509-6513.

(278) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430-437.

(279) Neumann, D.; Lehr, C.-M.; Lenhof, H.-P.; Kohlbacher, O. *Adv. Drug Del. Rev.* **2004**, *56*, 437-457.

(280) Vyas, N. K.; Vyas, M. N.; Chervenak, M. C.; Johnson, M. A.; Pinto, B. M.; Bundle, D. R.; Quiocho, F. A. *Biochemistry* **2002**, *41*, 13575-13586.

(281) Biosym/MSI San Diego, 2000.

(282) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.

(283) Laederach, A.; Reilly, P. J. *Proteins: Struct. Funct. Bioinf.* **2005**, *60*, 591-597.

(284) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. *J. Chem. Phys.* **2006**, *125*, 184114.

(285) van den Elsen, J.; Vandeputte-Rutten, L.; Kroon, J.; Gros, P. *J. Biol. Chem.* **1999**, *274*, 1495-1501.

(286) Yokoyama, H.; Mizutani, R.; Satow, Y.; Komatsu, Y.; Ohtsuka, E.; Nikaido, O. *J. Mol. Biol.* **2000**, *299*, 711-723.

(287) Barbas, C. F.; Heine, A.; Zhong, G. F.; Hoffmann, T.; Gramatikova, S.; Bjornestedt, R.; List, B.; Anderson, J.; Stura, E. A.; Wilson, I. A.; Lerner, R. A. *Science* **1997**, *278*, 2085-2092.

(288) Chitarra, V.; Alzari, P. M.; Bentley, G. A.; Bhat, T. N.; Eisele, J. L.; Houdusse, A.; Lescar, J.; Souchon, H.; Poljak, R. J. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7711-7715.

(289) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.