DERIVATION AND APPLICATION OF A TIP5P-CONSISTENT FORCE FIELD

by

SARAH MARGARET TSCHAMPEL WITTKOPP

(Under the Direction of Robert J. Woods)

ABSTRACT

The inclusion of zero-mass point charges (commonly called lone-pairs or extended points) around electronegative atoms, such as oxygen, within force fields is known to improve hydrogen bonding directionality. In parallel, inclusion of lone-pairs in the TIP5P water model increased its ability to reproduce both gas-phase and condensed-phase properties over its non-LP predecessor, TIP3P. In general, force field parameters relevant for modeling solvent are developed independently of those for biomolecules, generally focusing on the ability of the parameters to reproduce bulk properties of the solvent. Such an approach is not generally extendible to solutes, which themselves never exist in the pure liquid phase. Therefore, a different ansatz has to be implemented to derive the most optimal lone-pair description for each type of solute, which is not always consistent with the solvent model.

Currently, the partial charges for most biomolecular parameter sets are computed via fitting of the classical molecular electrostatic potential to the quantum molecular electrostatic potential. Application of this methodology to optimize lone-pair description is therefore consistent with current approaches of modeling electrostatics and is straightforward to implement. Here, we present an atom-type specific lone-pair model, which leads to the most optimal lone-pair placement for each atom-type, and notably, results in reproduction of the lone-pair description in TIP5P. Carbohydrates are rich in hydroxyl groups and development of a lone-pair inclusive carbohydrate parameter set for use with a lone-pair containing water model, such as TIP5P, ensures compatibility between these two models. Implementation of this lone-pair model into the GLYCAM series of parameter sets improves the geometry of a series of hydrogen bonded clusters over the non–lone-pair containing parameter set and is shown to reproduce crystalline and aqueous-phase properties for mono- and disaccharides. In addition, this methodology can easily be implemented to derive a general biomolecular parameter set and a preliminary investigation of this application to the protein segment, in conjunction with the carbohydrate segment, is examined.

INDEX WORDS: force field, lone-pair, GLYCAM, carbohydrate, conformational properties, parameter development

DERIVATION AND APPLICATION OF A TIP5P-CONSISTENT FORCE FIELD

by

SARAH MARGARET TSCHAMPEL WITTKOPP

B.A. Hamilton College, 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
2005

© 2005

Sarah Margaret Tschampel Wittkopp

All Rights Reserved

DERIVATION AND APPLICATION OF A TIP5P-CONSISTENT FORCE FIELD

by

SARAH MARGARET TSCHAMPEL WITTKOPP

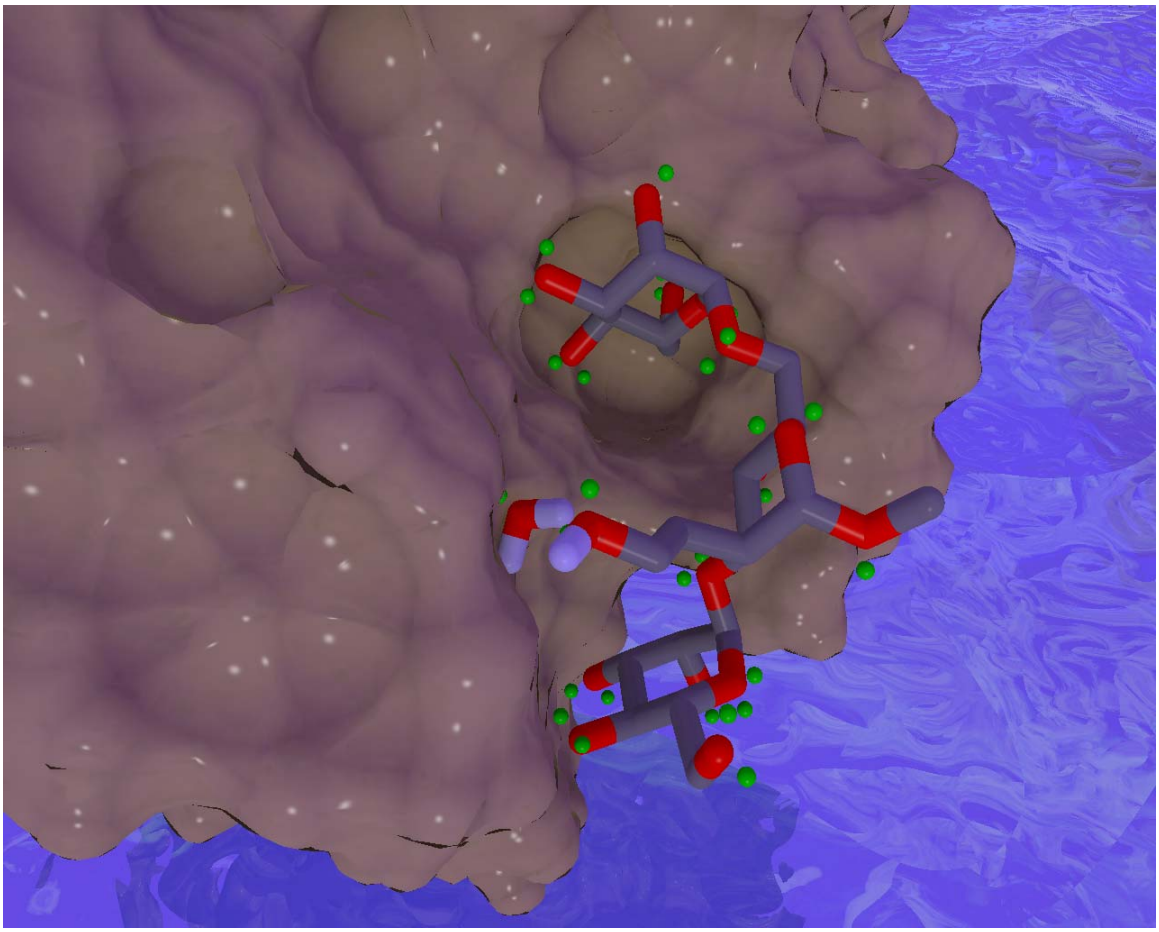
Major Professor:	Robert J. Woods
Committee:	Lionel A. Carreira
	Will S. York

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2005

DEDICATION

To my family with much love and thanks



"Yeah, well, you know, that's just, like, your opinion, man." - The Dude

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Rob Woods, for guidance and support over the last five years. I am deeply grateful for his demonstrated faith in my abilities, providing me with the necessary confidence all researchers require.

I am also grateful for the help given from all group members, both past (K. Kirschner, M. Basma, and M. Kennerty) and present (R. Kadirvelraj, L. Foley, N. Seyfried, A. Pathiaseril, J. Dyekjaer, S. Kawatkar, D. Elking, A. Yongye, M. Nagarajan), members of the CCRC (A. Kishore, C. Bougalt, J. Glushka, and C. Bergmann), as well as my advisory committee (Professors Butch Carreira, Will York, and Bob Scott) for their help.

In addition, I would like to thank the general CCRC members and staff for providing such an amazing work environment, especially the co-directors, Professor Peter Albersheim and Professor Alan Darvill. All of which would not be possible without the help from the wonderful women upstairs: including Sheilah Dixon, Dr. K. P., Melissa Akins, Lynn Berryman, Lane Guyer, Diane Hermosillo, Karen Howard, Tasashia McCormick, Beverly Chalk, and Lydia Shearer.

Much thanks to my undergraduate advisor, Professor George Shields, without whose guidance, both scientific and as a mentor, I would most likely not have chosen this path. He has exemplified to me the profound effect one individual can truly have on another person.

Most importantly, I would like to thank my parents for their never-ending support and love over the last 27 years, you can't imagine how thankful I truly am.

In addition, I would like to thank Nana and Papa for their gentle support; Maggie, Sierra, and Austin for the reminder of what life is really about; Tim for the comical reality check; and Kathy for the late night phone conversations about everything except what was wrong with my computations;

And Alexander, for among many other things bringing breakfast to my door.
You have made the last three years the best and hardest of my life -
Thanks for everything.

TABLE OF CONTENTS

Page

DEDICATION……iv

ACKNOWLEDGEMENTS……v

1 INTRODUCTION……...1

Conformational Properties of Oligosaccharides…..…….2

Challenges in Carbohydrate Force Field Development…….……..7

Empirical Non-Polarizable Water Models……...9

2 BASIC PRINCIPLES OF CLASSICAL MOLECULAR DYNAMICS SIMULATIONS IN AMBER…...…….14

3 DERIVATION OF THE OPTIMAL LONE-PAIR PLACEMENT FOR OXYGEN ATOMS...…….19

Introduction……19

Methods……...……24

Compatibility with the TIP5P Water Model……..25

A Lone-Pair Model for Alcohols and Ethers…….28

A Lone-Pair Model for Ketones...…….31

An Assymetric Lone-Pair Model for Amides……32

A Lone-Pair Model for Carboxylate Groups?…..…….....33

Conclusions……37

4 MODEL SYSTEMS TO INVESTIGATE THE ROLE OF WATER IN PROTEIN-CARBOHYDRATE AND CARBOHYDRATE-WATER INTERACTIONS..……38

Introduction……38

Methods……..41

Water Clusters……44

Uncharged Clusters……....47

Charged Clusters…..……...51

Protein-Carbohydrate Interactions…….56

Conclusions……59

5 VALIDATION AND TESTING OF THE LONE-PAIR MODEL..…….61

Introduction……....61

Methods……..63

Analysis of Hydrogen-Bonded Gas-Phase Clusters……..71

MD Simulations of Monosaccharide Crystals……...75

Conformational Properties of Monosaccharides……84

Solution Properties of Disaccharides…….86

Conclusions……..104

6 CARBOHYDRATE-LECTIN COMPLEXES MEDIATED BY BOUND WATER……..……...…..106

Introduction……..106

Methods……..……..110

Results……..…….....114

Conclusions……..118

7 PRELIMINARY EXTENSION OF THE GLYCAM LONE-PAIR MODEL TO DEVELOP A MODIFIED AMBER PROTEIN FORCE FIELD……..119

Introduction……..119

Methods……125

Results: Parameters for the φ and ψ Torsions of the Peptide Backbone……..134

Results: Parameters for the χ_1 and χ_2 Torsions……137

Crystalline-Phase MD Simulations of the Alanine Dipeptide Zwitterion…….147

Conclusions……..150

8 CONCLUSIONS AND FUTURE DIRECTIONS …….153

9 WORKS CITED..……...155

10 ABBREVIATIONS…….173

11 APPENDIX...……...177

Perl Program: AddLP.pl ……...…….177

Modifications to AMBER7 source for GLYCAM04-LP……179

1. INTRODUCTION

Carbohydrates are involved in many biological recognition processes, including the immune response to bacterial pathogens, adhesion of bacteria and virus particles to host cells, in the hyperacute rejection of tissue transplants from nonhuman sources (xenograft transplantation) and in inflammatory response.[1] Further, the covalent attachment of carbohydrates (glycosylation) is one of the most prevalent post-translational modifications of proteins. The heightened awareness of carbohydrate function has increased the desire and need to understand their three-dimensional structures. Traditionally, experimental techniques, such as X-ray diffraction and nuclear magnetic resonance (NMR), have been used to study the conformations of carbohydrates and their interaction with proteins. However, these techniques yield static, or average, pictures of inherently dynamic systems.

Computational chemistry has proven to be a valuable tool in studying the dynamics of protein and nucleic acid systems, and it has already been shown to be successful in studying systems involving carbohydrates.[2] Molecular modeling tools enable the exploration of sequence-structure relationships in oligosaccharides, which are experimentally difficult to establish due to the challenges associated with generating modified carbohydrate sequences in sufficient quantity and purity. In addition, computational chemistry can provide accurate models for the interpretation of experimental data that are only indirectly related to structure, such as spectroscopic data, but also from biological and immunological studies. As the interest in glycobiology grows, so grows the demand for improved theoretical methodologies.

Quantum mechanical methods are powerful tools for examining the conformational stabilities of molecules in the gas phase. The applicability of *ab initio* methods to systems under biologically relevant conditions (*i.e.* in water) is limited to relatively small systems, whereas methods based on Newtonian physics, such as molecular mechanics and dynamics, alleviate this limitation. As the accuracy of *ab initio* methods depends heavily on the choice of basis set, the accuracy of the force field depends on the rigorous derivation of force field parameters. With a properly tuned force field one can model entire carbohydrate and glycoprotein systems with a remarkably high degree of confidence in the obtained structures. Protein parameter sets have been proposed and refined for the past several decades with many research groups contributing to their development. The growing interest in structural glycobiology has led to the need to develop protein-consistent carbohydrate parameter sets. The application of existing classical methods, such as molecular dynamics (MD), in combination with the more recently developed protein-consistent carbohydrate parameter sets can potentially become a key structural tool in the field of glycobiology.

Conformational Properties of Oligosaccharides

The torsion angle associated with the rotation around the C5-C6 bond is generally defined by two letters (*gg*, *gt*, or *tg*); in which *g* and *t* refer to the *gauche* and *trans* orientation of the O6-C6-C5-O5 angle (deemed the ω-angle) and the O6-C6-C5-C4 angle, respectively. Experimentally, these three low energy conformations are found to be populated in varying percentages for the hydroxymethyl group of hexopyranoses in solution. Experimentally, the NMR derived $^{3}J_{\text{H5-H6R}}$ and $^{3}J_{\text{H5-H6S}}$ scalar couplings can be related to their respective torsion angles through a Karplus-type equation, Figure 1.1.[3]

Each ω-angle conformation corresponds to a value for the ^3J scalar couplings, which has been determined both experimentally and theoretically.[4-7] The measured ^3J couplings can be decomposed into the populations corresponding to each conformation, Equation 1.1-1.2.[4]

$$^3J_{\text{observed}} = x_{gg}(^3J_{gg}) + x_{gt}(^3J_{gt}) + x_{tg}(^3J_{tg}) \quad [1.1]$$

$$1 = x_{gg} + x_{gt} + x_{tg} \quad [1.2]$$

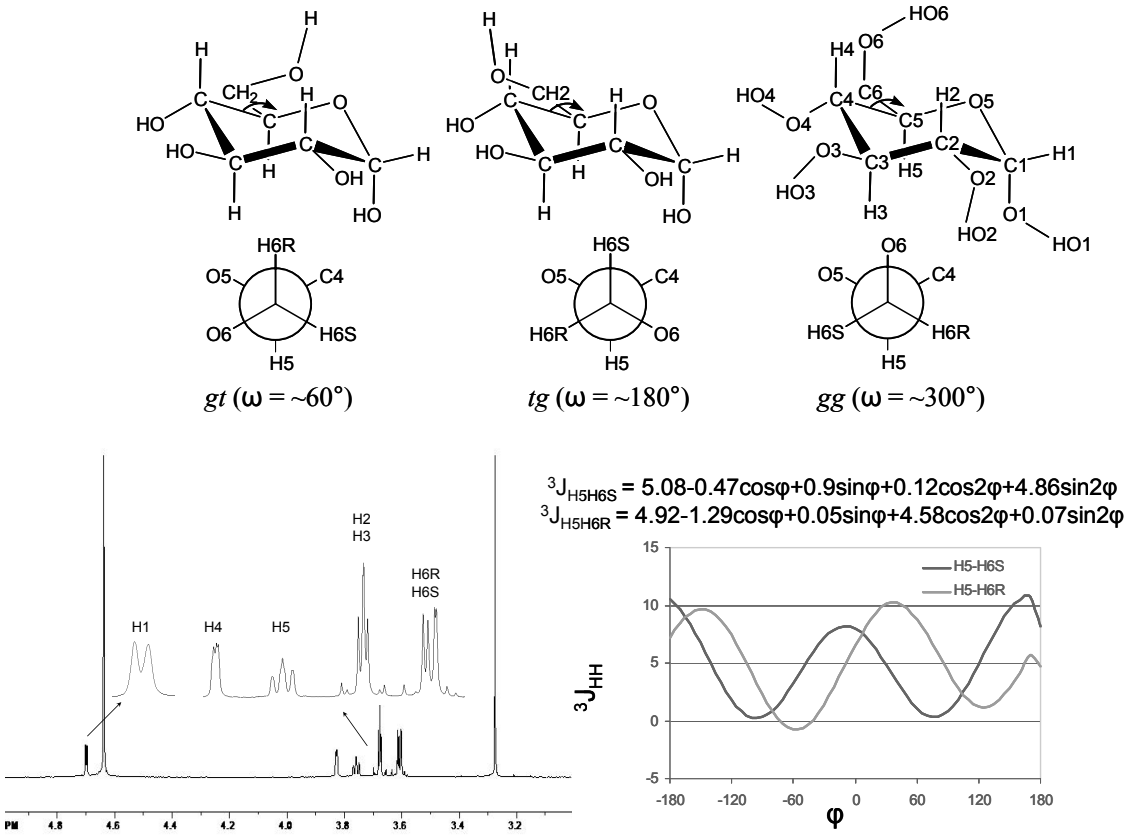


Figure 1.1 Two different representations depicting the nomenclature for the hydroxymethyl group conformation in hexopyranoses (top). A 1D ^{1}H NMR spectrum obtained for α-D-Glc*p*-OMe in D_2O (lower left) illustrates the splitting between each peak ($^3J_{\text{H-H}}$). The Karplus-type equation for the two representative $^3J_{\text{H-H}}$ coupling constants relevant for the description of the ω-angle (lower right).[8] The ^{1}H NMR spectrum is reproduced with the kind permission of Dr. J. González Outeiriño.

All three conformers may be populated in solution, with the different population ratios showing a dependence on the polarity of the solvent as well as a strong sensitivity to the configuration at C4.[7] For example, in D_2O the *gt:tg:gg* ratio is *41:7:52*[9] for α-D-Glc*p*-OMe and *56:32:12*[10] for its C4 epimer, α-D-Gal*p*-OMe. The population ratio for methyl 2,3,4-trimethyl-α-D-Glc*p*-OMe shifts from *40:0:60* in D_2O to *28:12:60* in $CDCl_3$, and to *18:9:73* in C_6D_{12}, which have dielectric constants of 80.2, 4.8, and 2.0, respectively.[7]

In the gas phase, the global minimum energy conformation exhibits the maximum number of intramolecular hydrogen bonds.[11] In this low energy conformation, the O6 group is *gauche* to the ring oxygen (*gt*), in accord with the *gauche* effect for Glc*p*, that is, the preference for two electronegative groups that are separated by three bonds to adopt a *gauche* orientation.[12] Of the three possible staggered conformations obtained from rotation of the C5-C6 bond, the maximum benefit from hyperconjugation is obtained in the *gauche* orientation, Figure 1.1. Therefore, due to the difficulty in separating the energetic contribution resulting from the *gauche* effect versus hydrogen bonding, the magnitude of their role in stabilization of the low energy conformation is not quantitatively known. In contrast, the lowest energy conformation for Gal*p* (*tg*), the C4 epimer of Glc*p*, maximizes the number of internal hydrogen bonds and does not coincide with the *gauche* effect. In conclusion, the *gauche* effect plays a smaller, secondary role to hydrogen bonding in carbohydrates, especially when solvent is introduced. In solution, water molecules compete with the intramolecular hydrogen bonds to form intermolecular hydrogen bonds, lowering the population of the global minimum found in the gas phase.[13]

In aldoses, the hydroxyl group at the C1 carbon may exist as a mixture of axial and equatorial configurations (called anomers) in solution, the distribution for Glc*p* at 300 K is 65% equatorial (β) and 35% axial (α).[14] Both solvent and stereoelectronic effects associated with the C-O-C-O epitope play a role in the stability of each anomer. The stabilizing effect in the axial configuration is due to electron density from the highest occupied molecular orbital (HOMO) on the ring oxygen, n_p, being donated into the lowest unoccupied molecular orbital (LUMO), the σ* antibonding orbital, for the C1-OMe bond, and is termed the anomeric effect, Figure 1.2.[12] In contrast, in the equatorial configuration the $n_p \rightarrow \sigma^*$ transition is disallowed due to the absence of the symmetry required for the overlap. Quantum mechanical computations utilizing second order Møller Plesset[15] (MP2) perturbation theory with a standard Pople basis set[16] (6-311++G(2d,2p)) determined the axial conformer of 2-methoxytetrahydropyran (2-MTHP) to be favored over the equatorial configuration by 1.66 kcal·mol^{-1} in the gas phase.[17]

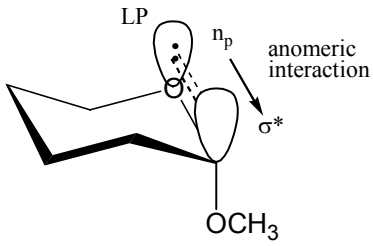


Figure 1.2 Donation of electron density from the HOMO into the LUMO in an α-hexopyranoside.

Polymerization of carbohydrates leads to additional degrees of conformational complexity, characterized by the φ and ψ linkage angles, which determine the orientation of the two monosaccharides relative to each other, Figure 1.3. The orientation of the φ angle is governed by the *exo*-anomeric effect, which is governed by the same principles

as the anomeric effect, φ = -60° for α-D- and β-L-hexoses and φ = +60° for β-D- and α-L-hexoses.[12] In contrast, the ψ angle has a wider range of motion from -50° to +50°. Commonly, the *syn*, ψ = 0°, conformation is observed, characterized by the H1' and HX atoms residing on the same face of the molecule.[12,18] Infrequently, although still observed, the *anti* conformation may also be populated, in which ψ = 180°.[12,18] In addition to the flexibility arising from the primary alcohol as well as the linkages between rings, the conformation of the ring is also variable. Typically, the $^4C_1$ conformation is more stable for D-hexopyranosides, while the $^1C_4$ is more stable for L-hexopyranosides, Figure 1.4.[19]

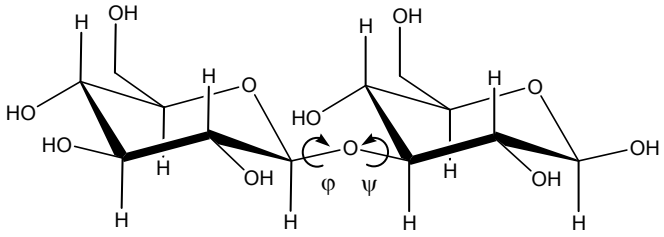


Figure 1.3 The φ (H1'-C1'-OX-CX) and ψ (C1'-OX-CX-HX) torsion angles for β-D-Glc*p*-(1→3)-β-D-Glc*p*-OMe, in which X represents the linkage position, X=3 here.

Differences in conformational properties of the glycosidic linkage in the gas phase versus condensed phase may be observed, and this difference is exemplified in the φ, ψ potential energy surface of α-D-Xyl*p*(1→4)-α-D-Xyl*p*.[20] The structure at the free energy global minimum in the condensed phase (φ = -37, ψ = 52°) is not even a local minimum in the gas phase. In addition, the two local minima, C (φ = 40, ψ ≈ 30°) and A (φ ≈ -61, ψ ≈ 40°), which exist in both phases, are energetically reversed going from gas to condensed phase. Specifically, in the gas phase C and A are 0.10 and 0.89 kcal·mol^{-1} higher than the global minimum, respectively, while in the condensed phase, A and C are 0.55 and 1.37 kcal·mol^{-1} higher in energy than the global minimum, respectively.[20]

Challenges in Carbohydrate Force Field Development

Development of a carbohydrate parameter set is a challenging endeavor due to the inherent flexibility of oligosaccharides. At a minimum, a carbohydrate force field must address the stability and structural properties of glycosyl rings, rotational properties of exocyclic bonds (*i.e.* the C5-C6 ω-angle in pyranoses), and the conformational properties associated with the O-C-O moiety (*i.e.* anomeric and *exo*-anomeric effects).

Amino acids polymerize in a single, linear fashion to form polypeptides. In contrast, carbohydrates polymerize by linking through any of several different hydroxyl positions to form both linear and branched structures, Figure 1.4. While only one peptide can be formed from two alanyl residues, up to eleven chemically unique disaccharides can arise from combining two glucopyranosyl residues through permutations of linkage position and anomeric configuration. Each of these disaccharides may adopt distinct conformational families about their linkage depending on the shape of the local potential energy surface.

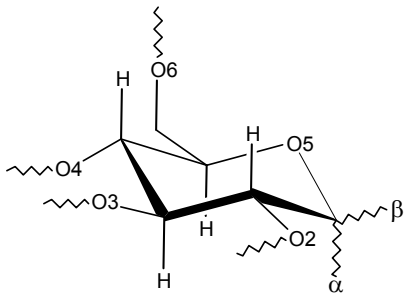


Figure 1.4 Possible linkage positions (2,3,4,6) and anomeric configuration (α,β) involved in a central gluco-residue of an oligosaccharide.

Due to the large number of hydroxyl groups, carbohydrates have the ability to form several inter-residue and solute-solvent hydrogen bonds. Water can compete with internal hydrogen bonds, attenuating the extent of inter-residue interactions and increasing molecular flexibility.[13,21,22] For these reasons, carbohydrates present a unique challenge in the development of a robust parameter set for modeling the physics in oligosaccharides.

In addition to purely carbohydrate systems, the ability to integrate the carbohydrate parameter set with a protein parameter set allows the former to be of broader scientific use.[23] Typical quadratic biomolecular force field equations implemented in OPLS, AMBER, GROMOS and CHARMM, include contributions from bond stretching, angle bending, and non-bonded interactions (Coulomb and Lennard-Jones terms).[24,25,26,27-29] In order to account for non-classical behavior, such as the hyperconjugation involved in the rotation around the C-O-C-O torsion angle in carbohydrates, the addition of a torsion term is necessary in the potential energy function. It should be noted that the inclusion of torsion terms can be considered a correction that partially compensates for the deficiencies in the classical model, such as the omission of molecular polarizability and other quantum effects. The derivation of parameters that influence the overall three-dimensional structure of oligosaccharides, such as the partial atomic charges and torsion terms associated with the glycosidic linkages, is frequently the most crucial task in the parameterization of a carbohydrate parameter set.

In order to achieve compatibility between carbohydrate and protein parameter sets, both models need to be derived for use with the same potential energy function, the same solvent model, and must employ compatible partial atomic charges (q). In short,

mixing parameters developed for use in different force fields or from different derivation methods can lead to a critical loss of internal consistency. In some cases, protein force fields employ a scaling factor to dampen 1,4 van der Waals (S_{vdw}) and electrostatic interactions (S_q).[24,25] This treatment is problematic when deriving carbohydrate parameters and is of particular relevance to the conformational distribution of the ω-angle, Figure 1.5. The bond distances between O6···O4 and O6···O5 are essentially equivalent in the *gg* rotamer. However, the former is a 1,5 interaction, while the latter is a 1,4 and is artificially weakened by the application of the 1,4 scaling factor. Removal of the scaling factors or introduction of additional scaling factors has been shown to be necessary to obtain the best fit to rotational properties for this linkage, but may result in loss of compatibility with the protein force field.[13,30]

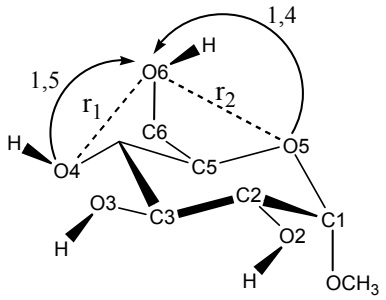


Figure 1.5 The *gg* rotamer of α-D-Glc*p*-OMe, in which $r_1 \approx r_2$.

Empirical Non-Polarizable Water Models

Compatibility between carbohydrate and protein parameter sets is integral for investigation of protein – ligand complexes and glycosylated proteins. In contrast, due to the aforementioned necessity of including solvent in MD simulations of carbohydrates, compatibility between the carbohydrate parameter set and solvation model is integral for

all MD simulations of sugars. In some cases, such as the AMBER* carbohydrate force field, the non-bonded interactions have been specifically developed for use with an implicit solvent model, but the majority of carbohydrate force fields have been developed with the intention of implementing an explicit solvent model. There are a variety of water models readily available for use in explicitly solvated MD simulations. Two commonly utilized and closely related models are TIP3P[31] (Transferable Intermolecular Potential, 3 Point) and SPC/E[32] (Extended Simple Point Charge). The TIP3P model was the first water model in the TIP series to be widely implemented in MD simulations (the TIP series also includes TIP4P[33] and TIP5P[34]), while the SPC/E model is a modification of the SPC[35] water model. The TIP3P, SPC, and SPC/E models are each three point models, in which a partial charge is placed at the nuclei of the oxygen and hydrogen atoms, Figure 1.6. The TIP4P model includes an additional partial charge, which is located at the midpoint between the bisector of the HOH angle 0.15 Å from the oxygen atom. The TIP5P water model has two additional partial charge sites over TIP3P, which are located in a tetrahedral arrangement 0.7 Å from the oxygen atom, which itself has a partial charge of zero.

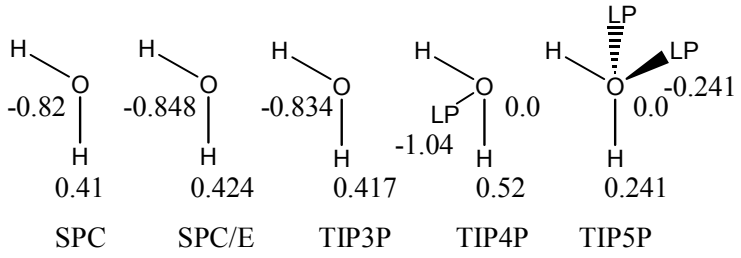


Figure 1.6 Common water models with their corresponding atomic partial charges. The hydrogen nuclei and, in the case for TIP5P, both LPs have the same charge.[31-35]

Each aforementioned water model was developed to reproduce as many bulk liquid properties of water as possible, such as the heat of vaporization (ΔH_{vap}) and density (ρ). Therefore, each model has a dipole moment (μ) that is greater in magnitude than the gas phase dipole moment. Typically, the three-dimensional structure of bulk water is assessed via computing the radial distribution functions (RDFs) around a given atomic center. For example, the first peak for the O···O RDF has the highest intensity revealing the first shell to be the most ordered with a maximum value at 2.8 Å.[36] The area under each peak corresponds to the average number of oxygen atoms in each solvent shell. Therefore, reproduction of both structural and thermodynamic properties of bulk water is required for a general-use water model, Table 1.1. Of the five water models discussed, TIP5P performs the best with respect to density, temperature at maximum density (TMD), and the location of the maximum O···O distance in the O···O RDF, while SPC/E performs better with respect to ΔH_{vap} (and the directly related interaction, or average non-bonded, energy of liquid water) and the self-diffusion coefficient (*D*). Overall, the TIP5P model reproduces the majority of the properties in Table 1.1 with the highest degree of accuracy. For the individual properties where TIP5P is not the most accurate model, the water model with the highest degree of accuracy performs only marginally better than TIP5P, while TIP5P is the only model yielding a reasonable value for the TMD.

Clearly, for the purpose of biomolecular MD simulations, reproduction of the bulk liquid properties of water is desirable, but the small cluster properties of water are also important. Specifically, when water is separated from the bulk, such as in a water-mediated lectin – carbohydrate complex, as well as in surface interactions between the

solute and the solvent, the water is not necessarily well described by a bulk-water model. The main success of the three point models has been in the bulk phase, but possible failures of these models can occur when non-bulk properties are exhibited. Poor non-bulk behavior correlates with the inability of the TIP3P and SPC/E water models to reproduce the general orientation of the water monomers in the water dimer, Table 1.1. TIP5P does substantially better at reproducing the hydrogen-bond directionality in the water dimer, 24° closer to experiment relative to models that do not contain any out-of-plane interaction sites. The established behavior of TIP3P combined with the lack of thorough testing of the TIP5P model, increased computational expense, as well as the only recent ability of AMBER to correctly analytically model the additional zero-mass partial charge sites, lone-pairs (LPs), are the main reasons for the infrequent use of the TIP5P model.

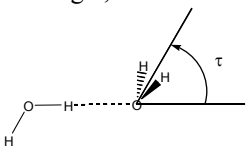
In some cases, the biomolecular parameter set has been developed explicitly for use with a single water model, *i.e.* CHARMM with TIP3P and OPLS-AA with TIP4P, or validation and testing have all been performed with a single water model, *i.e.* AMBER with TIP3P.[28,37,38] In contrast, no biomolecular parameter sets have been developed in conjunction with TIP5P. Several parameter sets have been derived that employ LPs, but the more recent developments also include polarization and are not explicitly developed to be compatible with a lone-pair inclusive water model. Inclusion of lone-pairs in a tetrahedral arrangement around the sp^3 type oxygen atom in water allows the model to reproduce both bulk and small cluster properties. Therefore, it can be hypothesized that a similar implementation of LPs in the biomolecular-solute will also be advantageous.

Table 1.1 Experimental and theoretically computed properties for water.

	SPC[35]	SPC/E[32]	TIP3P[31]	TIP4P[33]	TIP4P-EW[d,39]	TIP5P[34,40]	TIP5P-EW[41]	Exp
				Water Monomer				
q_H (*e*)	0.41	0.4238	0.417	0.520	0.52422	0.241	0.241	
q_O (*e*)	-0.82	-0.8476	-0.834	0.0	0.0	0.0	0.0	
q_{LP} (*e*)	-	-	-	-1.040	-1.04844	-0.241	-0.241	
ϵ_{OO} (kcal·mol^{-1})	0.1553	0.1553	0.1521	0.1550	0.162750	0.160	0.178	
σ_{OO} (Å)	3.166	3.166	3.15061	3.15365	3.16435	3.120	3.097	
R_{O-H}	1.0	1.0	0.9572	0.9572	0.9572	0.9572	0.9572	0.9572
R_{O-LP}	-	-	-	0.15	0.1250	0.70	0.70	-
θ_{H-O-H}	109.47	109.47	104.52	104.52	104.52	104.52	104.52	104.52
$\theta_{LP-O-LP}$	-	-	-	-	-	109.47	109.47	-
μ (10^{-18} esu-cm)	2.274	2.351	2.35	2.18	2.321	2.29	2.29	1.85[42]
				Properties of the Water Dimer				
R_{O-O} (Å)	2.75	-	2.74	2.75	-	2.68	-	2.98±0.02[43]
τ[b]	26	-	27	46	-	51	-	57±10[43]
$-\Delta E$[c] (kcal·mol^{-1})	6.59	-	6.50	6.24	-	6.78	-	5.4±0.5[43]
				Bulk Water Properties				
ρ[a] (g/cm^3)	0.970	0.998	1.002±0.001[34]	1.001±0.001[34]	0.9954	0.999±0.001	1.0000	0.997
μ[a] (10^{-18} esu-cm)	2.274	2.351	2.35	2.18	2.321	2.29	2.29	2.65 (3.0)
D[a] (10^{-5} cm^2·s^{-1})	4.3 (3.85)	2.5	5.19	3.29	2.4±0.06	2.62	2.8	2.4 (2.3[44])
τ_c[a] (ps)						1.58	1.55	2.46[44]
TMD[a] (°C)	-	-38	-13	-13[45]/-23[46]/25	1	4	0-12.5	3.984[47]
$-E_{liq}$[a] (kcal·mol^{-1})	10.18	-	9.82±0.01[34]	10.06±0.01[34]	-	9.87±0.01	-9.78	9.92
ΔH_{vap} (kcal·mol^{-1})	10.77	-	10.41±0.01[34]	10.65±0.01[34]	10.583	10.46±0.01	10.377	10.51

[a] Determined at 298 K and 1 atm, where ρ is the density, μ is the dipole moment, D is the self-diffusion coefficient, τ_c is the correlation time, and TMD is the temperature at maximum density of bulk liquid, E_{liq} is the intermolecular energy per molecule in the liquid phase.

[b] τ is the tilt angle, defined in the illustration below.



[c] ΔE is the dimerization energy.

[d] EW refers to the adaptation of the water model for use in MD simulations that employ EWALD summation for long-range electrostatics.

2. BASIC PRINCIPLES OF CLASSICAL MOLECULAR DYNAMICS SIMULATIONS IN AMBER

In classical mechanics the potential energy, V, for a molecule or group of molecules is computed from the force field Equation 2.1, which is composed of parameters that have been adapted from experiment or derived quantum mechanically to reproduce experiment or theory.[26,48] The AMBER force field equation,

$$V = \sum_{j=1}^{N} \frac{K_{i,r}}{2}\left((r_i - r_{i,o})^2\right) + \sum_{j=1}^{N} \frac{K_{i,j}}{2}\left((\theta_i - \theta_{i,o})^2\right)$$
$$+ \sum_{j=1}^{N}\sum_{n=1}^{N_n} \frac{V_{n,\varphi}}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left[S_{vdw}\left(\frac{A_{ij}}{R_{ij}^{*12}} - \frac{B_{ij}}{R_{ij}^{*6}}\right) + S_q\left(\frac{q_i q_j}{\varepsilon R_{ij}}\right)\right] \qquad [2.1]$$

was utilized in all MD simulations presented in this work.[49] The first two terms represent bond stretching and angle bending, respectively, and employ a harmonic approximation. The third term is a cosine series expansion, which represents the barrier around a given torsion angle, ϕ. The number of terms in the expansion varies from force field to force field, but in the force fields specified here the expansion is at most taken to be third order, in which the individual Fourier coefficients are referred to as V_1, V_2 and V_3. The last term includes the Lennard-Jones and Coulomb potentials to describe the non-bonded interactions, in which S_{vdw} and S_q are scaling factors employed to dampen the 1,4 van der Waals (S_{vdw}) and electrostatic interactions (S_q), respectively. The Lennard-Jones potential has essentially two adjustable parameters, the well depth, ϵ_i, and the internuclear separation at the minimum energy position, R_{ij}^*, Equation 2.2.[48]

$$A_{ij} = \epsilon_{ij} R_{ij}^{*12},\ B_{ij} = 2\epsilon_{ij} R_{ij}^{*6},\ \epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2},\ R_{ij}^* = R_i^* + R_j^* \qquad [2.2]$$

The Coulomb term requires atomic partial charges, q_i, which are a convenient approximation, although physically unobservable, and specification of the dielectric constant, ε_r, which is typically set to the experimental value, *i.e.* 80 for water.

Previously, it was common practice to specify a cut-off distance for these non-bonded interactions. Ewald summation was developed to provide a better description of the electrostatic interactions and allows a charge to see the mirror image of the solvent/solute box in all directions. Unfortunately, this method is very time consuming and a quicker alternative known as Particle Mesh Ewald (PME) is generally utilized. PME is an interpolative approximation of the Ewald summation method and lowers the time dependence to $N \cdot \log N$ from N^2, where N is the number of atoms.[48,50] Due to the success of the PME method, it is now the default for MD simulations in AMBER.

Newton's second law of motion (F = ma),[48] has been integrated to correlate to the potential energy, *V,* which is obtained from the force field Equation 2.1, to the change in atomic position *r*, Equation 2.3.

$$\mathbf{F} = m\mathbf{a} = m\frac{d^2r}{dt^2} = -\frac{dV}{dr} \qquad [2.3]$$

Therefore, since Equation 2.1 is utilized to determine the value for the numerator, the change in position during time, *t*, needs to be determined. Utilizing the change in position from the previous step, $\mathbf{r}(t - \delta t)$ to the future positions at time $t + \delta t$, the velocities (**v**), and accelerations (**a**) can be approximated as a Taylor series expansion, illustrated in Equations 2.4 and 2.5, which yield Equation 2.6 when combined.[48]

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \tfrac{1}{2}\delta t^2\mathbf{a}(t) + \ldots \qquad [2.4]$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t\mathbf{v}(t) + \tfrac{1}{2}\delta t^2\mathbf{a}(t) + \ldots \qquad [2.5]$$

$$0 = \mathbf{r}(t) - \delta t\mathbf{v}(t) + \tfrac{1}{2}\delta t^2\mathbf{a}(t) - \mathbf{r}(t - \delta t) \quad [2.6]$$

The original Verlet algorithm[51] combines Equations 2.4 and 2.6 to yield Equation 2.7,

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2\mathbf{a}(t) \quad [2.7]$$

which has the drawback of not explicitly determining the velocity.[48] The variation of the Verlet algorithm implemented within the simulations in this work was the leap-frog algorithm, which determines the velocity at $t + \frac{1}{2}\delta t$, Equation 2.8,

$$\mathbf{v}(t + \tfrac{1}{2}\delta t) = \mathbf{v}(t - \tfrac{1}{2}\delta t) + \delta t\mathbf{a}(t) \quad [2.8]$$

which is then utilized to calculate the position at $t + \delta t$, Equation 2.9. [48,52]

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t + \tfrac{1}{2}\delta t) \quad [2.9]$$

The leap-frog algorithm has the obvious disadvantage that the positions and velocities are not determined for the same point in time, but due to the small integration time step, typically 2 fs, the velocity at time t can be approximated by the average velocity, over the time step $t - \frac{1}{2}\delta t$ to $t + \frac{1}{2}\delta t$.

At the start of each simulation, a set of velocities must be assigned. Typically, the starting structure for the MD simulation is not a global minimum, but a local minimum on the potential energy surface. Since there may be large forces associated with the non-global minimum structure, utilizing the forces would lead directly to high initial velocities, which would increase the amount of time necessary to reach an energetic equilibrium. Therefore, the initial velocities are usually derived from a Maxwell-Boltzman distribution at 5 K, Equation 2.10.[48]

$$p(v_{ix}) = \left(\frac{m_i}{2k_BT}\right)^{\frac{1}{2}} \exp\left[-\frac{m_i v_{ix}^2}{2k_BT}\right] \quad [2.10]$$

In general, the MD simulations presented within this work are equilibrated to a temperature of 300 K. In order to maintain this temperature, a coupling constant to an external thermal bath is specified.[53] Depending on the size of this coupling constant, τ_T, the deviation from the desired temperature can be controlled, Equation 2.11.[48] A small value for τ_T results in tight coupling to the heat reservoir, while a large value corresponds to a weak coupling. Since the velocities are scaled accordingly, ensembles with different values for τ_T will not necessarily converge to the same results over the MD simulation.

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{bath} - T(t)), \qquad [2.11]$$

All the MD simulations presented here were run under constant pressure, which is maintained in a similar fashion to the temperature. A coupling constant for the pressure, τ_P, is presented in Equation 2.12.[48]

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{bath} - P(t)) \qquad [2.12]$$

Since the pressure and volume are inversely related, τ_P can be used to determine a positional scaling factor λ for the solvent box size, in which κ is the isothermal compressibility, Equation 2.13.[48]

$$\lambda = 1 - \kappa\frac{\Delta t}{\tau_P}(P - P_{bath}) \qquad [2.13]$$

The new positions, $\mathbf{r}_i'$, are then calculated using this scaling factor. Each box dimension is scaled by an equivalent factor when isotropic pressure scaling is employed, while each dimension is scaled independently during the implementation of anisotropic pressure scaling, using Equation 2.14 and 2.15, respectively.[48]

$$\mathbf{r}_i' = \lambda^{\frac{1}{3}}\mathbf{r}_i \qquad [2.14]$$

$$\mathbf{r}_{i,x}^{'} = \lambda_x \mathbf{r}_{i,x}, \mathbf{r}_{i,y}^{'} = \lambda_y \mathbf{r}_{i,y}, \mathbf{r}_{i,z}^{'} = \lambda_z \mathbf{r}_{i,z}, \lambda = \lambda_x \lambda_y \lambda_z \quad [2.15]$$

At specified intervals, typically every ps, the total energy and its components, as well as the coordinates (referred to as a snapshot) are saved for analysis.

3. DERIVATION OF THE OPTIMAL LONE-PAIR PLACEMENT FOR OXYGEN ATOMS

Introduction

Partial charges are non-physical entities frequently employed in the computation of the non-bonded Coloumb interaction in molecular mechanics force fields, Equation 3.1. The frequent use of atom-centered partial charges (monopoles) is based on the assumption that the contribution from the higher order terms (*i.e.* dipole and quadrupole) can be ignored due to the much faster rate at which the higher order energetic contributions diminish with respect to distance. A variety of protocols have been developed to obtain partial atomic charges for biomolecules. The first wide-spread method utilized the charges derived from a Mulliken population analysis of the QM wavefunction derived from either *ab initio* or semi-empirical methods.[54] During the Mulliken population analysis, the electron density is systematically localized to each atom center, in which the localization process is independent of chemical environment, and the resultant partial charges do not steadily reproduce molecular properties, such as molecular dipole moments and molecular electrostatic potentials (MEPs).[55] Due to the limited applicability of the results from this method, several groups pioneered a more generalized methodology that utilized the computed quantum mechanical MEP, Equation 3.2, to derive partial charges.[55-60] Partial charges are assigned to the atomic centers in a given molecule so as to optimize the agreement between the classical MEP arising from these partial charges and quantum mechanical MEP (Equation 3.3) using a least squares fitting algorithm, minimizing χ (Equation 3.4).[58]

$$V_{electrostatic} = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \frac{q_i q_j}{\varepsilon R_{ij}} \quad [3.1]$$

$$MEP_{\mathrm{QM}}(r) = \sum_{A} \frac{Z_A}{r - R_A} - \sum_{\mu,\upsilon} P_{\mu,\upsilon} \int \frac{\varphi_\mu \varphi_\upsilon}{|r - r'|} dr' \quad [3.2]$$

$$MEP_{classical}(i) = \sum_{j=1}^{n} \frac{q_j}{R_{ij}} \quad [3.3]$$

$$\chi^2 = \sum_{i} (MEP_{QM,i} - MEP_{Classical,i})^2 \quad [3.4]$$

Due to the more polarized environment found in condensed phase biomolecular systems, the dipole moment is larger in aqueous solution, relative to that in the gas phase. Therefore, in order to utilize partial charges derived from gas phase computations in a non-polarizable force field, the dipole moment must be scaled by a factor greater than one. The Hartree-Fock (HF) level of theory combined with a Pople basis set, which includes 'd' functions on second row elements (6-31G(d)), results in a 10-20% overestimate of gas-phase dipole moments, which makes this method applicable without inclusion of a scaling term.[61]

The Merz-Kollman (MK[57,60]) electrostatic fitting algorithm employs a low density of points located on shells around the nuclei and is the current methodology implemented in the derivation of additive AMBER force fields.[49] Altering the point selection to a grid based method, which put a 0.3 - 0.8 Å spaced grid of points around a molecule extending 2.8 Å from each atom center led to the CHarges from ELectostatic Potentials (CHELP) using a grid method (CHelpG).[58,59] Due to the high sampling of the MEP, the CHelpG protocol displays a decreased change in partial charges with respect to molecular conformation, resulting in a more robust charge set. Derivation of partial atomic charges utilizing the CHelpG protocol is the current methodology implemented in the

development of the Glycosides and Glycoproteins with AMBER (GLYCAM[13,62]) carbohydrate force field.

The second point of variation between the partial charge derivation implemented in the AMBER and GLYCAM force fields occurs at the fitting stage. The PARM94, 96, 98, 99, 02, and 02EP AMBER protein and nucleic acid force field utilized the Restrained Electrostatic Potential (RESP[63]) fitting methodology, Equation 3.5.[25,64,65] The RESP fitting process allowed a hyberbolic restraint, Equation 3.6, to be included in the fitting. Sequentially performing two RESP fittings was implemented as an improvement over the single stage fitting to derive the partial charges for the Cornell *et al.* series of protein parameter sets in AMBER. The first RESP fitting, stage 1, fits all atoms with a weak restraint of 0.0005. During the second stage, any non-polar atoms affected by symmetry restraints are refit with a RESP weighting of 0.001, *i.e*. freely rotating (in solution) methyl hydrogens.[25,63,64] As a result, the non-polar atoms that are chemically equivalent in solution can be constrained during the fit as opposed to averaging the charges *posterior*, which resulted in a large decrease in the dipole moment. In monosaccharides, there are only a few chemically equivalent hydrogen atoms, which eliminates the necessity for a two-stage fitting. Therefore, GLYCAM implements a one-stage fitting for all atoms with a RESP weighting of 0.010. This RESP value was determined through reproduction of crystal data.[13,66,67] In addition, it has been shown previously that the inclusion of charges on aliphatic hydrogen atoms in carbohydrates leads to over determination to the MEP, resulting in small, variant charges on the aliphatic hydrogens.[59,68,69] Therefore, in GLYCAM04 the partial charges on aliphatic hydrogen atoms are set to zero.[70]

$$\chi^2_{RESP} = \chi^2 + \chi^2_{restraint} \quad [3.5]$$

$$\chi^2_{restraint} = k_{restraint} \sum_i (\sqrt{(q_j^2 - b^2)} - b) \text{, where b = 0.1 } e \quad [3.6]$$

Despite subtle differences in the derivation of the partial charges, most of the current MEP-based techniques have been shown to perform well in practice and are widely employed. In contrast, the necessity to include point charges around oxygen, nitrogen and sulfur to mimic lone-pair electrons continues to be controversial. The option to include LPs in MD simulations has been available for over a decade in the AMBER force field.[26] However, version 7 of the AMBER package allows for an analytical treatment of LPs with zero-mass. Previously, LPs in the AMBER force field were treated as pseudo-atoms with both a mass (3 Dalton) and a partial charge.[49] Along with the conversion to an analytical treatment of lone-pairs (extended points (EP) in the AMBER nomenclature), the distance of LPs from their corresponding oxygen nucleus changed only marginally from the earlier PARM81 parameter set (0.20 Å) to the PARM02EP parameter set (0.35 Å)[71,72] The method implemented to determine the lone-pair oxygen (LP-O) distance in PARM02EP was based on the location of critical points in the charge density.[73] The critical points were obtained for a test set of 21 compounds, consisting of sp, sp^2, sp^3, and aromatic nitrogen atom types and sp^2, sp^3, and aromatic oxygen atom types, which revealed similar critical point positions for all of these atom types. Due to the insensitivity of critical point position to chemical environment, it was inferred that the average value of 0.35 Å was transferable to all oxygen and nitrogen atom types. However, the critical point location in the electron density is dependent on the level of theory and basis set in the quantum mechanical computation.[60] Furthermore, the charge density has not been shown to lead to good reproduction of the MEP.[67] Therefore,

since fitting partial charges to the quantum mechanical MEP has been shown to be appropriate for condensed-phase MD simulations, a lone pair placement derived from this approach was examined for application with the AMBER/GLYCAM parameter sets.

Empirically adjusting partial charges to reproduce experimentally known condensed phase properties, such as density and the enthalpy of vaporization, from liquid state MD simulations is an alternative charge derivation method utilized for pure liquids. This derivation was utilized to develop the partial charges and LP placement in the TIP5P water model.[31,34] In the case of TIP5P, the LP placement is specific for the hydroxyl-type oxygen present in the water molecule. Typically, this empirical approach cannot be applied to amino acids, nucleic acids or carbohydrates since they do not exist as pure liquids. This technique was applied to the derivation of the LP-O distance in methanol, which was utilized as a representative oxygen atom containing compound. A distance of 0.47 Å was determined via this method and applied to all oxygen atom types in a polarizable version of the OPLS force field.[74]

Due to the large number of hydroxyl groups, as well as the presence of a ring oxygen atom, carbohydrates inherently possess a large number of potential LP sites; for example, in a typical hexopyranoside, such as α-Glc*p*-OMe, there are 12 potential LP placements. Therefore, the effects seen for the addition of LPs to electronegative atoms may be amplified in carbohydrates.[34,71,74] Due to the number of hydroxyl groups, carbohydrates have the ability to form multiple inter- and intra-residue, as well as solute-solvent, hydrogen bonds. As a result, inclusion of LPs may have a substantial, beneficial effect on the hydrogen-bond directionality in carbohydrate-containing systems, such as glycoproteins and carbohydrate – protein complexes.

Derivation of an approach that can predict the LP description in TIP5P, and yet is applicable to biomolecular solutes, would result in a TIP5P-consistent approach to developing a LP-containing biomolecular parameter set. Currently, the GLYCAM[13,62,70] parameter set is consistent with the general AMBER methodology for partial charge derivation, which involves determination of partial charges that produce a classical MEP with the best fit to the quantum mechanical MEP. Extension of this technique to determine the LP-O distance, as well as the LP partial charges, will ensure consistency with the current methodology, as long as the quantum mechanical MEP is computed from the same HF/6-31G(d) wavefunction. The effect level of theory and basis set utilized to obtain the quantum mechanical MEP has on the determination of the LP-O separation will be determined. Analysis of a variety of test compounds, which include sp^2 and sp^3 type oxygen atoms, show the correlation between the LP description (LP-O distance and partial charge) and chemical environment.

Methods

All quantum mechanical computations were performed with the Gaussian 98 suite of programs, version A.11.3.[75] Initially, a water molecule was optimized with Hartree Fock (HF[76]) theory and density functional theory (DFT), in which the Becke-3[77] Lee, Yang, and Parr[78] correlation and exchange functionals (B3LYP) were implemented in the DFT computation, with the 6-31G(d) basis set. The MEPs were determined at each of these levels and geometries. For each test compound, both the geometry and quantum mechanical MEP was determined at the HF/6-31G(d) level and the B3LYP level with Dunning's augmented correlation consistent basis set, aug-cc-pVTZ.[79] The use of this unusually large basis set with B3LYP was employed to maintain consistency with

previously developed charge models in the AMBER parameter sets. Additionally, the MEPs and geometries of select test compounds were also computed at the MP2/aug-cc-pVTZ level[15]. The MEPs were computed with the CHelpG[58] algorithm as implemented in Gaussian 98, with the density of points varied from the default, $1/Å^2$, to $10/Å^2$. Partial charges were derived with the RESP[64] scheme, implementing a weighting factor of 0.010.[67] For all sp^3 type oxygen atoms, the LP geometry was assumed to prefer a tetrahedral placement, while for all sp^2 oxygen atoms the LPs were constrained to be in the plane of the carbonyl group and its two substituents, forming a 120° angle with respect to the carbon-oxygen carbonyl bond. The perl program AddLP.pl was developed to facilitate this process (Appendix).

Compatibility with the TIP5P Water Model

Initially, the optimal placement of LPs around sp^3 type oxygen atoms was assessed via determining the quality of the fit to the quantum MEP as the LP-O distance was varied from 0.01 Å to 1.00 Å in a water molecule with the hypothesis that this approach would yield a LP placement similar to that in TIP5P. The ability to reproduce the LP position currently employed in a LP containing water model would affirm the validity of this method. Two different approaches were taken during the RESP stage. The first scheme (A) did not constrain the partial charge on oxygen to zero, $q(O) \neq 0$, while the second scheme (B) fixed the oxygen partial charge on oxygen to zero, $q(O) = 0$ and all the negative charge was associated with the lone-pairs, as in TIP5P. The first scheme resulted in a minimum at a short LP-O distance, but possessed unintuitive charges, Figure 3.1. Specifically, as the LP-O distance was increased from 0.01 Å to the minimum in the fit between the quantum and classical MEP, the partial charge on oxygen

became increasingly negative when the charge on oxygen was not set to zero, and was accompanied by positively charged LPs. Application of the second scheme led to a minimum at a longer LP-O distance of approximately 0.75 Å, which had an intuitive charge scheme ($q_{LP} \approx -0.2$) and an LP-O separation close to that present in the TIP5P water model (0.70 Å).

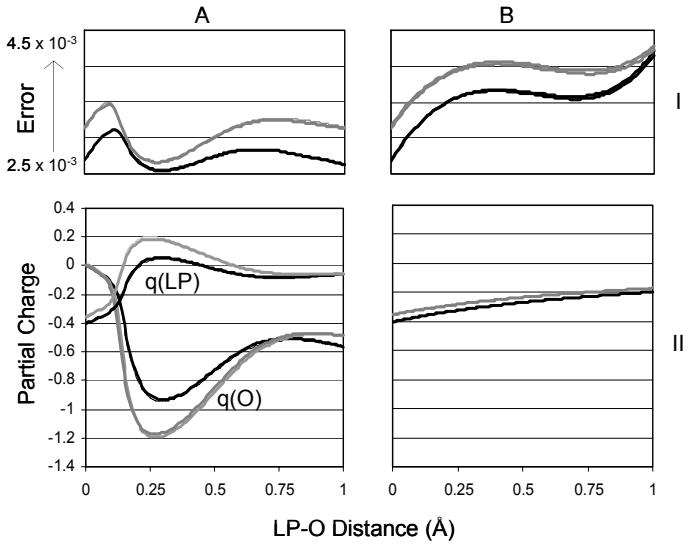


Figure 3.1 The quality of the fit (I) to the QM MEP and the corresponding partial charges for H_2O (II). Charges were calculated from the B3LYP/6-31G(d)// B3LYP/6-31G(d) and B3LYP/6-31G(d)//HF/6-31G(d) MEPs (gray) as well as the HF/6-31G(d)//HF/6-31G(d) and HF/6-31G(d)//B3LYP/6-31G(d) MEPs (black).

In order to ensure the LP-O placement was not dependent on the level of theory or basis set, the results from RESP-fitting to the MEPs determined at the B3LYP/aug-cc-pVTZ and HF/6-31G(d) optimized geometries for water were compared. In addition, during the computation of the quantum mechanical MEP, the density of points was varied from $1/Å^2$ to $10/Å^2$. For all fits determined with the charge on oxygen constrained to zero, similar fits to the error function were obtained, with a long LP-O distance resulting in a best fit. Different RESP weightings were also implemented in the

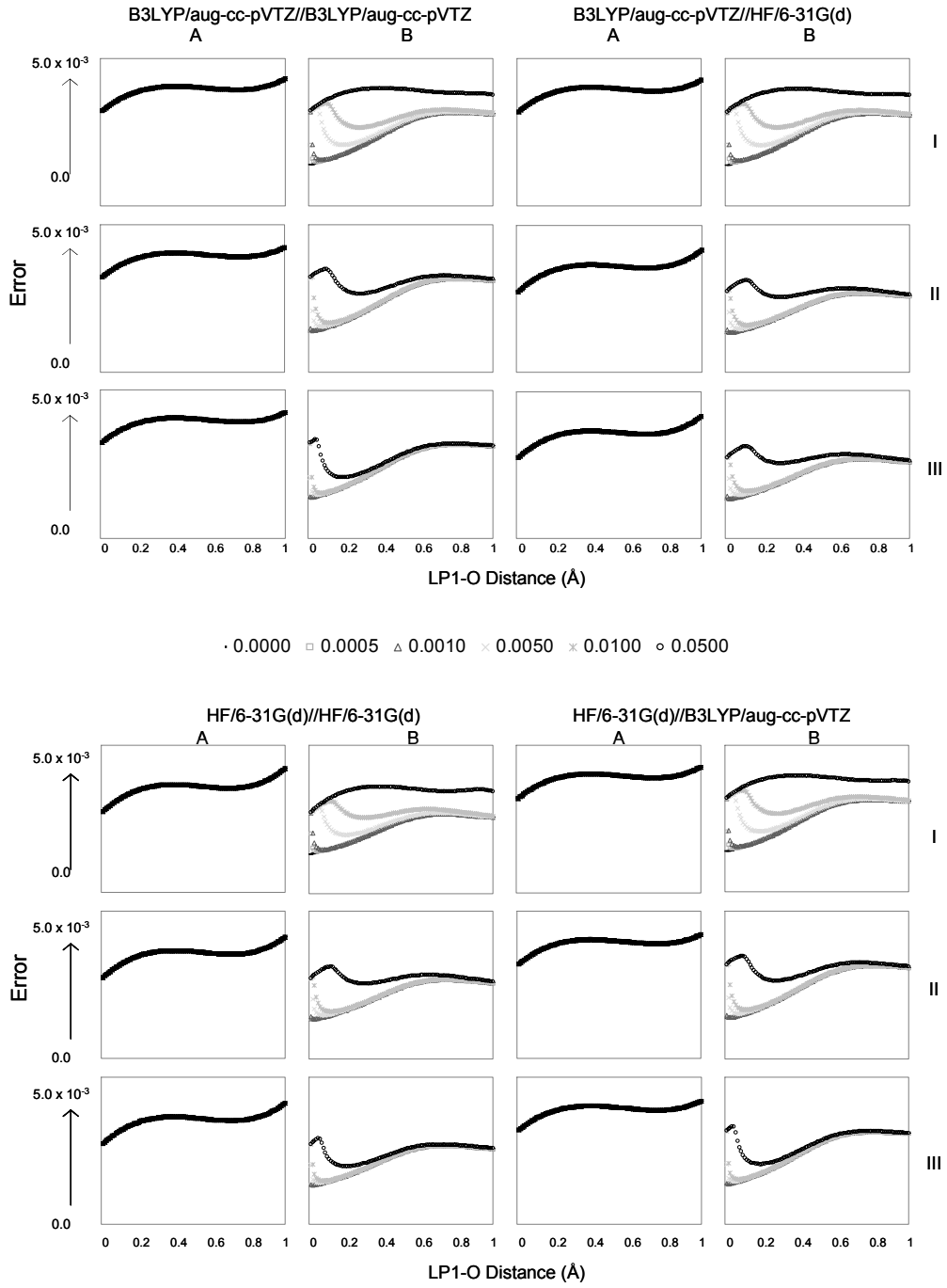


Figure 3.2 Correlation of the quality of the fit between the QM and classical MEP relative to the density of points, I = 1/Å^2, II = 5/Å^2, III = 10/Å^2, and the different levels of theory/basis sets when q(O) = 0 (A) and when q(O) is not constrained (B) for water. The center legend describes the different weightings implemented in the RESP fitting.

case where the charge on oxygen was not constrained, which resulted again in a short LP-O distance at the best fit, with similar unintuitive partial charges, Figure 3.2

Overall, the LP-O distance at which the error function is a minimum is constant over a variety of different levels and basis sets. Therefore, fitting to the MEP to obtain optimal LP distances is applicable for deriving partial charges for both non-polarizable and polarizable force fields, due to the different level of theory implemented to compute the MEP for varying parameter sets. The longer, optimal LP-O distance obtained is consistent with that of the TIP5P water model. Different force fields that employ the RESP-fitting algorithm to derive partial charges have chosen different RESP weightings. As a result, it was integral to illustrate that regardless of RESP weighting, an intuitively charged LP model cannot be obtained if $q(O) \neq 0$ to confirm that the partial charge on oxygen should be zero. In conclusion, determination of the most optimal LP-O distance via minimizing the error function to the MEP results in the first derivation of the partial charge placement for an empirically derived water model that has already been shown reproduce both bulk liquid and small cluster properties of water without implementing an iterative partial charge fitting scheme to reproduce experimentally observable properties.

A Lone-Pair Model for Alcohols and Ethers

Both charge fitting schemes, $q(O) = 0$ and $q(O) \neq 0$, were applied to a series of test molecules, which included methanol, ethanol, 2-butanol, 2-propanol, 3-pentanol, *t*-butanol, dimethylether, ethyl methylether, methyl isopropylether, and diisopropylether, to determine the optimal LP description for sp^3 type oxygen atoms. In the $q(O) \neq 0$ model, the average LP-O distances at the best fit were 0.69 Å, 0.72 Å, and 0.71 Å for the HF/6-31g(d), B3LYP/aug-cc-pVTZ, and MP2/aug-cc-pVTZ levels of theory,

respectively, Table 3.1 and Figure 3.3. At the same levels of theory, in the q(O) = 0 model, the optimal LP-O distances were 0.71 Å, 0.75 Å, and 0.71 Å. In contrast to the case of water, the partial charges determined for the alcohols and ethers in the test set revealed that even when the charge on the oxygen atom was not constrained to zero, only a small partial charge is localized on the oxygen atom in alcohols and ethers.

Table 3.1 The LP-O distance at the best fit to the ESP and the corresponding q(O) and q(LP) for each molecule in the sp^3 type oxygen atom test set at the HF/6-31G(d) (I), B3LYP/ aug-cc-pVTZ (II), and MP2/aug-cc-pVTZ (III) levels.

		q(O)≠0			q(O)=0	
Compound		LP-O	q(O)	q(LP)	LP-O	q(LP)
Methanol	I	*0.70*[a]	-0.029	-0.178	0.80	-0.177
	II	*0.70*	0.017	-0.169	0.88	-0.149
	III	0.89	-0.046	-0.146	0.84	-0.163
Ethanol	I	0.65	-0.047	-0.195	0.57	-0.221
	II	0.74	-0.043	-0.168	0.68	-0.188
	III	0.71	-0.048	-0.179	0.64	-0.202
2-butanol	I	0.53	0.008	-0.240	0.53	-0.237
	II	0.58	-0.005	-0.205	0.56	-0.209
	III	0.53	0.001	-0.228	0.52	-0.229
2-propanol	I	*0.70*	-0.097	-0.169	0.70	-0.199
	II	*0.70*	-0.074	-0.157	0.77	-0.173
	III	*0.70*	-0.094	-0.163	0.72	-0.190
3-pentanol	I	0.59	0.031	-0.232	0.61	-0.218
	II	0.65	0.013	-0.197	0.66	-0.191
	III	0.60	0.014	-0.216	0.60	-0.213
t-butanol	I	*0.70*	-0.080	-0.173	0.76	-0.192
	II	*0.70*	-0.076	-0.160	0.80	-0.174
	III	*0.70*	-0.093	-0.170	0.77	-0.192
dimethylether	I	0.92	-0.011	-0.105	0.91	-0.115
	II	0.90	0.110	-0.129	0.70	-0.106
	III	0.81	0.183	-0.168	0.70	-0.116
ethyl methylether	I	0.70	0.052	-0.150	0.80	-0.126
	II	0.79	0.158	-0.177	0.88	-0.109
	III	0.74	0.188	-0.188	0.91	-0.114
methyl isopropylether	I	0.65	0.047	-0.163	0.67	-0.145
	II	0.72	0.097	-0.167	0.79	-0.131
diisopropylether	I	0.73	0.041	-0.160	0.74	-0.147
	II	0.75	0.049	-0.158	0.76	-0.143

[a] The italic entries represent systems where the resolution was not fine enough to detect an exact minimum, so the partial charges at an LP-O distance of 0.70 Å are shown.

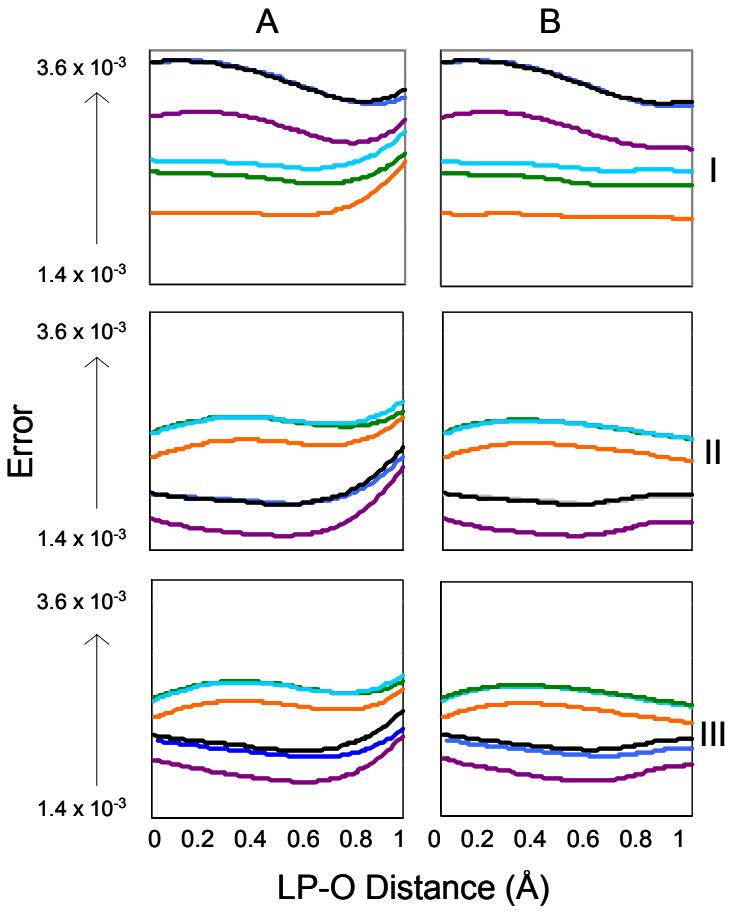


Figure 3.3 Correlation of the quality of the fit computed for methanol (purple), ethanol (blue), 2-butanol (black), 2-propanol (orange), 3-pentanol (green), and *t*-butanol (light blue) when q(O) = 0 (A) and when q(O) is not constrained (B) at the HF/6-31G(d) (I), B3LYP/aug-cc-pVTZ (II), MP2/aug-cc-pVTZ (III) levels.

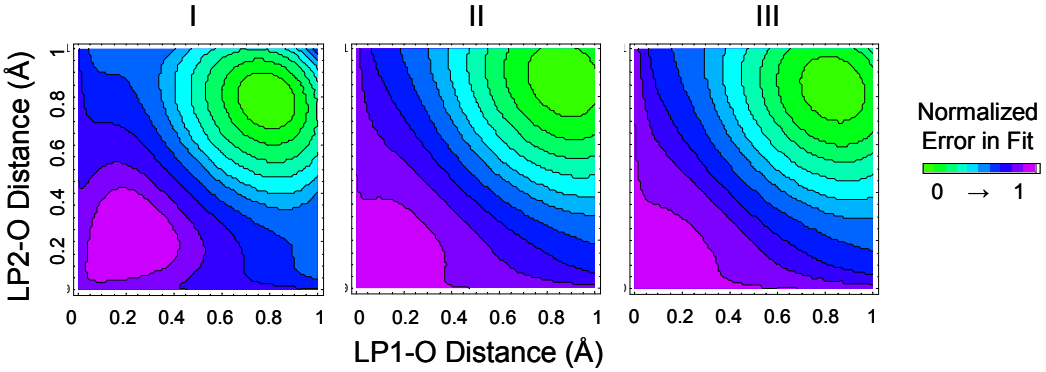


Figure 3.4 Variation of χ with respect to LP–O distance at the HF/6-31G(d) (I), B3LYP/aug-cc-pVTZ (II), MP2/aug-cc-pVTZ (III) levels with the partial charge on oxygen set to zero for methanol.

In conclusion, the results for the sp^3 type oxygen atoms in water, the alcohols, and ethers show that the partial charge on oxygen should be zero and the LPs should have equivalent negative charges, which are determined directly from the RESP-fitting. Reduction of the number of partial charge sites by one ($q(O) = 0$) is consistent with previous studies that have shown the number of centers used in the fitting of carbohydrates should be minimized.[69] For each LP-O distance derived thus far, the LP1-O and LP2-O distances were constrained to be equivalent, which is an approximation consistent with previous studies. When both LP-O distances were permitted to vary independently for each of the alcohol and ether type model compounds, the results converged to the aforementioned symmetric results, Figure 3.4, verifying the validity of the approximation.

A Lone-Pair Model for Ketones

In order to extend this model to sp^2 type oxygen atoms, a test set consisting of ketones was devised, which included 2-butanone, 3-pentanone, 2,4-hexadione, 2,5-hexadione, and 3,5-heptadione. For the ketones, when the charge on oxygen was not constrained, only at an LP-O distance of less than 0.5 Å, was the fit not degraded. In the region less than 0.5 Å, the charge on oxygen was approximately zero, while as the distance increased from 0.5 to 1.0 Å, the charge on oxygen became increasingly negative and the LP partial charges became less negative, Figure 3.5. Therefore, the LP-O distance should be less than 0.5 Å and the partial charge on oxygen should be set to zero. An LP-O distance of 0.3 Å was chosen due to the agreement with the previously determined LP-O distances in AMBER.[71,72]

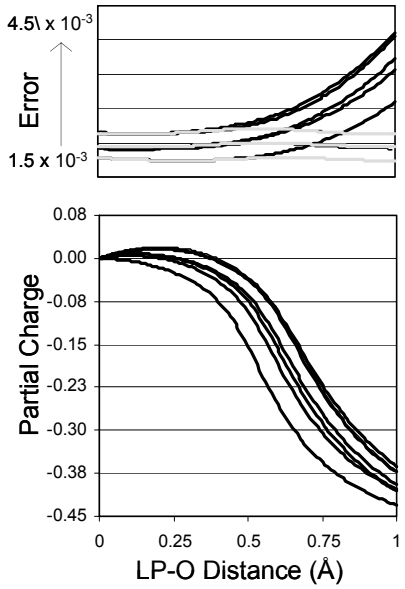


Figure 3.5 Error in the fit between the molecular and quantum mechanical MEP (top) and the magnitude of q(O) (bottom) for the ketone type test compounds. The fit performed with q(O) = 0 is shown in grey, while the results from the fit in which no constraint was placed on the magnitude of q(O) are shown in black.

An Asymmetric Lone-Pair Model for Amides

For each LP-O distance derived thus far, the optimal LP1-O and LP2-O distances were found to be equivalent.[71,74] In contrast to the results for sp^3 type oxygen atoms and sp^2 type oxygen atoms in ketones, constraining the LP-O distances to be equivalent for the carbonyl oxygen in amides led to a worsening in the fit to the MEP. Allowing each LP to adopt a unique distance from the oxygen atom and partial charge led to a better fit for the amides, with the best fit at a combination of longer sp^3 type, 0.7 Å, and shorter sp^2 type, 0.3 Å, LP-O distances, Figure 3.6.

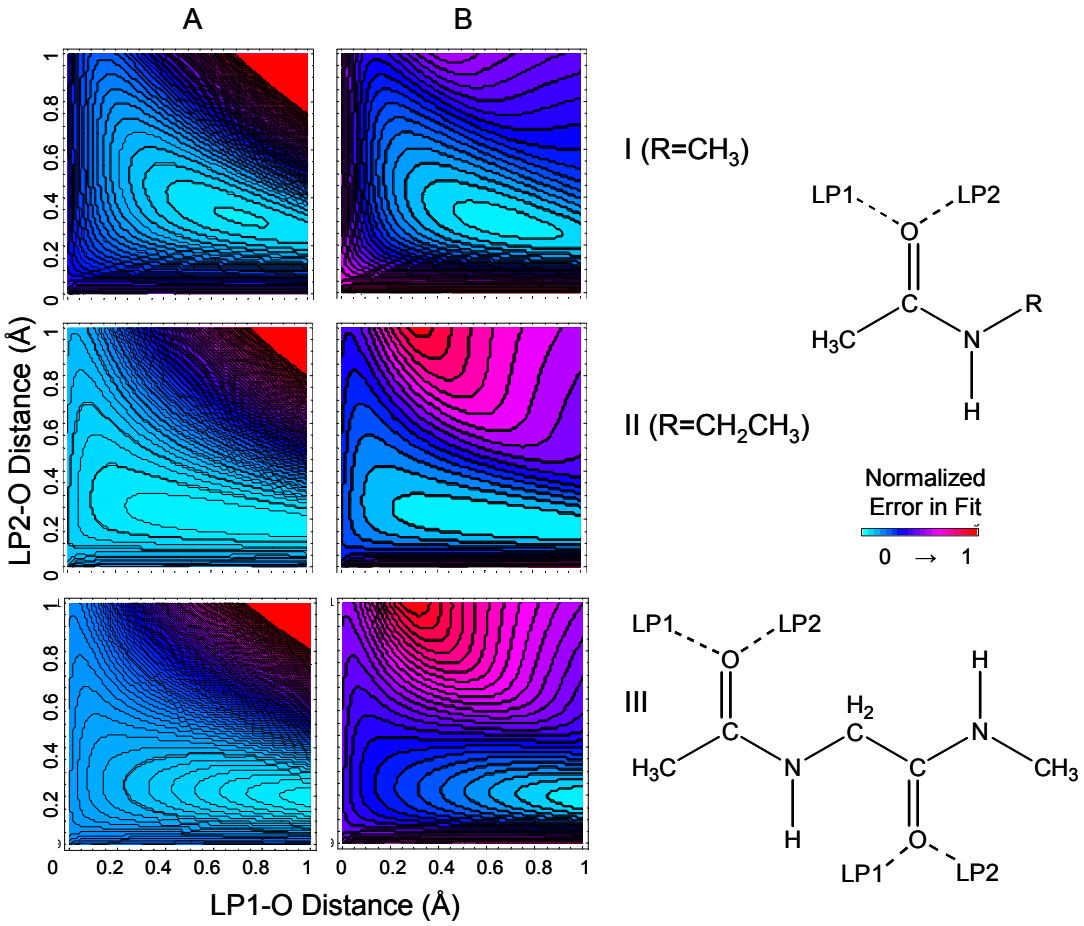


Figure 3.6 Variation of χ with respect to LP-O distance (Å) in *N*-methyl acetamide (I), *N*-ethyl acetamide (II), and the glycine dipeptide (III). In A, the partial charge on oxygen was freely fit, while q(O) was constrained to zero during the RESP fit in B.

A Lone-Pair Model for Carboxylate Groups?

In the PARM02EP AMBER parameter set, there were no LPs included for carboxylate groups.[71] Therefore, here the initial guess for LP placement consisted of treating each carbonyl oxygen atom as a ketone-type oxygen with the standard sp^2 type oxygen LP-O-LP angle of 120°, as in Figure 3.7(I). In order to be consistent with the molecular symmetry, the two lone-pairs on the outer side of the carboxylate group (LP_{outer}) were restrained to have equivalent charges and LP-O distances. In addition, the lone-pairs inside the O-C-O angle in acetate were also restrained to have equivalent

electrostatic parameters. Upon fitting to the quantum mechanical MEP, only models with short LP-O distances did not substantially degrade the fit, regardless of the location of the inner or outer lone-pairs, Figure 3.8. Reducing the number of lone-pair sites to three for the carboxylate moiety, Figure 3.7(II), did not lead to an improved fit over the initial lone-pair placement (I), either.

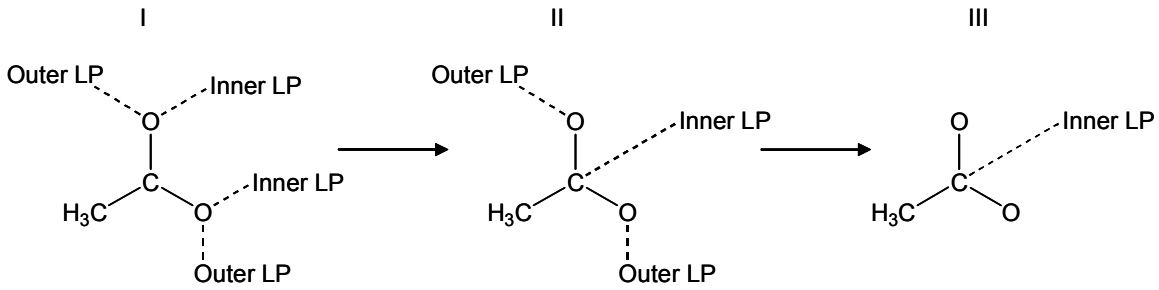


Figure 3.7. Three LP models tested for carboxylate groups for which the outer LP forms a 120° angle with the C-O bond for all. In model I, the LP_{inner} location is referred to with respect to the oxygen atom it is closest to, while in II and III, the location of LP_{inner} is with respect to the carbon atom in the carboxylate group.

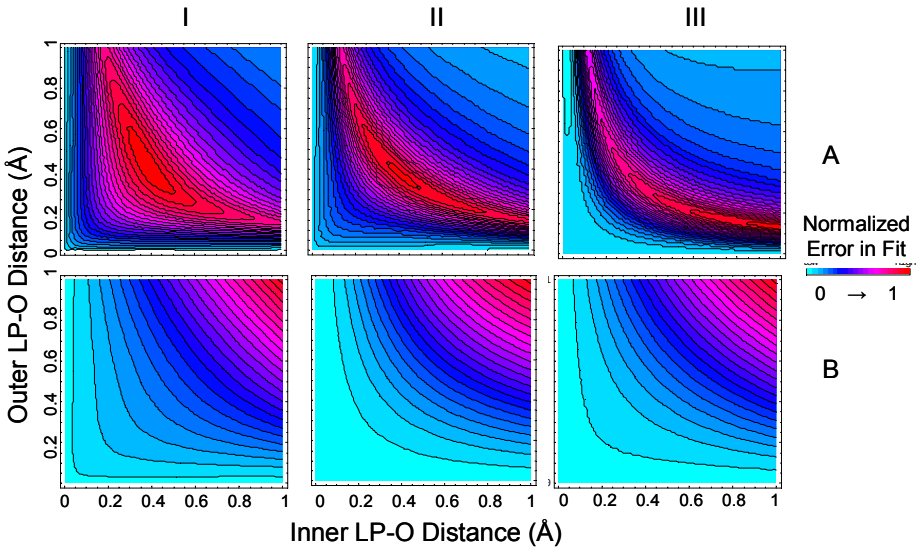


Figure 3.8. Variation of χ with respect to LP–O distance at HF/6-31G(d) (I), B3LYP/aug-cc-pVTZ (II), MP2/aug-cc-pVTZ (III) levels/basis sets with q(O) not constrained (A) and q(O) = 0 (B) for acetate.

Lastly, a single LP was included that bisected the O-C-O angle of the carboxylate moiety, Figure 3.7(III), which is the model that directly corresponds to the deficiency of the non-LP containing carboxylate model in MD simulations. Specifically, a bifurcated hydrogen-bond between a hydrogen-bond donor and the carboxylate group may be observed in the X-ray diffraction structure. Introduction of the crystal structure to biologically relevant conditions, 298 K and explicit solvent, increases the energy in the system. The additional energy may disrupt the bifurcated hydrogen bond and the water molecule may become associated with only a single oxygen atom in the carboxylate moiety, as opposed to dynamically associating with both oxygen atoms and partially retaining its bifurcated character. This additional negative point charge, LP_{inner}, located between the two negatively charged oxygen atoms may help stabilize bifurcated hydrogen bonds. The inclusion of this single lone-pair had a negligible effect on the error incurred upon fitting the classical MEP to the QM MEP, Figure 3.9.

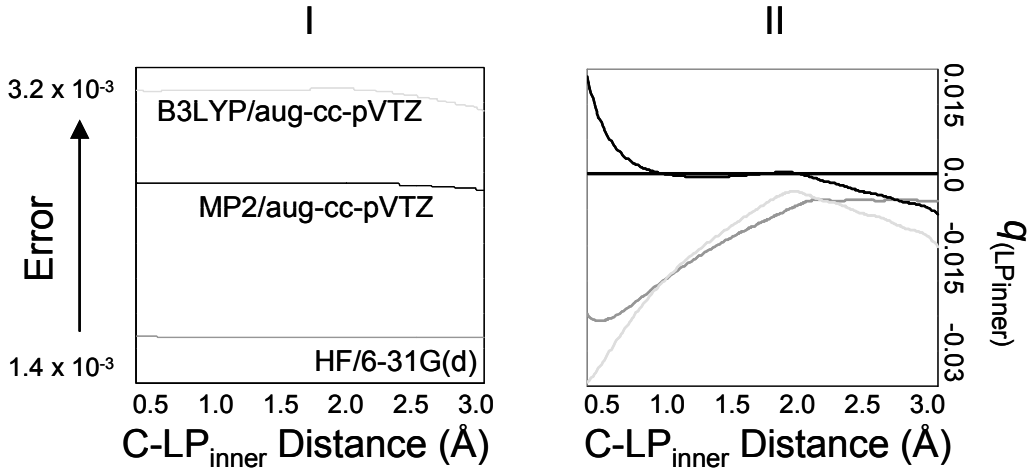


Figure 3.9 Variation of χ (I) and charge on LP_{inner} (II) with respect to LP_{inner}–C distance when the QM level was HF (gray), B3LYP (light gray), or MP2 (black) for acetate.

Due to the insensitivity of the fit to the electrostatic potential with respect to variation in the LP position for carboxylate groups, additional data was needed to determine if the carboxylate model should or should not have lone-pairs. Accordingly, the charge on the lone-pair should be negative to account for the deficiency in the carboxylate model observed during MD simulations. Overall, the majority of the negative charge is located on the two oxygen atoms, while the charge on LP_{inner} is only slightly negative, ranging only to -0.03 *e* for the HF and B3LYP levels and becomes positive for all LP_{inner}-C distances less than 2 Å at the MP2 level, Figure 3.9. This is in agreement with the decrease in both the MEP and electron density for acetate between the oxygen atoms, Figure 3.10. Similar results were obtained for propanoic acid anion, 2-methyl butanoic acid anion, and 2-methyl propanoic anion.

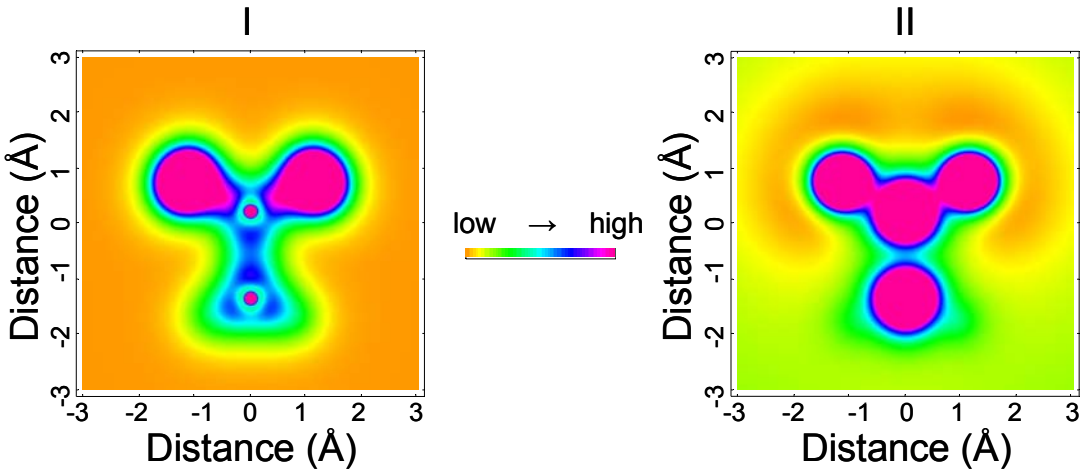


Figure 3.10 Electron density around acetate at the HF/6-31G(d) (I) and MP2/6-31+G(d,p) levels (II).

Conclusions

Utilizing the MEP to determine the distance of the LPs from the respective oxygen atom leads to results which reproduce the LP description in the TIP5P water model. Reproduction of the lone-pair description in TIP5P, which was originally obtained via fitting to reproduce bulk water properties, reveals that the MEP fitting methodology is appropriate for deriving LP locations in molecules for use in solvated (TIP5P) MD simulations. Constraining the partial charge on the aliphatic hydrogen atoms, and in some cases the partial charge on the oxygen atom, to zero leads to a robust partial charge set with very similar charges for the various atom types in similar environments.

4. MODEL SYSTEMS TO INVESTIGATE THE ROLE OF WATER IN PROTEIN-CARBOHYDRATE AND CARBOHYDRATE-WATER INTERACTIONS

Reproduced in part with permission from Tschampel, S.M. & Woods, R.J., "Quantifying the Role Water of Water in Protein-Carbohydrate Interactions." *J. Phys. Chem. A* 107(43): 9175-9181 **2003**. Copyright 2003 American Chemical Society

Introduction

Accurately simulating the conformational and energetic properties of solvated polar molecules requires the development of parameters that reproduce the delicate balance of electrostatic forces between the solvent and the solute. Generally, the electrostatics of a system are represented by partial charges located on each atomic center. The most commonly used solvent models have been developed independently of the models for biomolecular systems (i.e. proteins, nucleic acids, and carbohydrates). Therefore, due to the different methodologies employed, it is not guaranteed that the interactions between the solvent and the biomolecular solute will be balanced. Development of a series of test systems to assess these interactions will both facilitate solute parameter development, as well as provide insight into the individual interactions.

Considerable experimental and theoretical data are available pertaining to the quantification of intermolecular interactions in water.[80-83] The focus of a substantial portion of this research has been the water dimer, which, because of its size, has facilitated treatment with high-level *ab initio* quantum mechanical methods.[83] One of the most well-studied heterodimers is the methanol-water system, which has a theoretical binding energy of -5.9 kcal·mol^{-1} at the complete basis set (CBS) limit at second order Møller Plesset (MP2) pertubation theory,[84] which may be compared to -5.0 kcal·mol^{-1} for

the water dimer at the CBS limit of the coupled-cluster level with single and double substitutions, with triple excitations included perturbatively (CCSD(T)).[85] The binding energies for both the $(H_2O)_2$ and H2O-MeOH complexes are enhanced upon addition of a water molecule to form $(H_2O)_3$ and $(H_2O)_2$-MeOH, respectively. The binding energy corresponding to the formation of the three-bodied complexes from the three individual monomers is -10.8 kcal·mol^{-1} at the MP4 level, which lowers to -15.8 kcal·mol^{-1} at the CBS limit[86], for the water trimer and -11.5 kcal·mol^{-1} at the Gaussian-2 (G2)(MP2) level for the $(H_2O)_2$-CH_3OH trimer.[87,88]

Studies of the water trimer by Xantheas *et al.* revealed that the total two-body interaction energy is -11.6 kcal·mol^{-1}, in which 21% (-2.4 kcal·mol^{-1}) results from three-body (cooperative) interactions.[89] Notably, the magnitude of the three-body interaction depended on the orientation of the monomers in the trimer. A similar study on water-mediated base-pairing in RNA revealed that 9-13% of the interaction energy was due to cooperativity (at the MP2 and HF levels of theory).[90] The dimers investigated here each contain one water molecule, as well as a representative fragment of side chains from the following amino acids: asparagine or glutamine, arginine, aspartate or glutamate, lysine, threonine or serine, and tyrosine. While this study does not present an exhaustive analysis of all possible configurations, each structure has been shown to be a minimum on the potential energy surface.

It has been established that MP2 theory gives reasonable results for energies and geometries of hydrogen bonded complexes.[91] However, in order to extend this study to larger complexes, density functional theory (DFT) was chosen over the more time-consuming MP2 calculations. Calculations employing the B3LYP functional have been

shown to yield good agreement with MP2 geometries and structures for hydrogen-bonded complexes, including water clusters.[92] Geometries of neutral N-H··O hydrogen bonded complexes, computed with DFT, have been shown to agree with experimental data.[93] In addition, agreement with experiment for DFT results is well established for strongly bound ionic complexes, involving either negatively or positively charged ions, such as the $OH^-(H_2O)$ and the $H_3O^+(H_2O)$ complexes.[94] Overall, it has been determined that diffuse functions must be included in order to determine accurate binding energies and structures for hydrogen bonded complexes when utilizing DFT.[91,95]

Concanavalin A (Con A) is a widely studied carbohydrate-binding protein (lectin) and has been used frequently as a model to study carbohydrate-protein complexation due to its affinity for oligosaccharides terminating in mannopyranosyl residues.[96] X-ray diffraction data are available for several Con A–carbohydrate complexes, including those with methyl α-D-glucopyranoside,[97] methyl α-D-mannopyranoside,[98] and a methyl 3,6-di-*O*-(α-D-mannopyranosyl)-α-D-mannopyranoside (trimannoside).[99] In all three complexes, water-mediated hydrogen bond networks are formed.[97-100] In the Con A-trimannoside complex, a conserved water interacts with three key amino acids, namely asparagine, arginine, and aspartate. In this study, a single methanol molecule is used to provide a minimal approximation of the carbohydrate ligand.[99,100]

Desolvation entropy will not be accounted for in this study although it is known to play a significant role in carbohydrate binding to Con A.[100-102] The structure of bulk water is known to become more ordered when around methanol relative to that of water, which results in a favorable desolvation entropy in the case of methanol, relative to water.[103] For each cluster the entropic contribution of methanol will be assumed to be

consistently greater in all cases, relative to water, resulting in a more favorable free energy by enhancing the calculated enthalpic differences. Furthermore, variation in enthalpies for the binding of different carbohydrate ligands to Con A correlates directly with the number of hydrogen bonds formed.[100,102] Therefore, by conserving the number of hydrogen bonds in the formation of either water or methanol complexes the enthalpic contribution will reflect the influence of the ligand rather than the number of hydrogen bonds formed.

Methods

All structures were optimized at the B3LYP level of theory[77,78] with the 6-31+G(d) basis set utilizing the Gaussian 98 suite of programs.[75] For all species under investigation, frequency analyses were performed at the same level of theory as the optimizations to ensure that the structures were minima on the potential energy surfaces and to determine the zero point vibration energies (ZPVEs), which were subsequently scaled by an empirical factor of 0.9775.[104] In addition, single point energy calculations were performed with the 6-311++G(3df,3pd) basis set.

All systems may be described by the same series of reactions, Equation 4.1. The binding energy will be referred to as the case in which either methanol or water complexes with a dimer, referred to as **A**, which contains an amino acid fragment and a complexed water, except in the case of the water dimer. In addition, both the dimerization and trimerization energies of these systems were computed as well. Each of these complexation energies was calculated by summing the energy computed from the Boys-Bernardi scheme (E_{BSSE}),[105] the relaxation correction (E_{FRE}),[90,45] and the ZPVE (E_{ZPVE}), Equations 4.2-4.5. Equation 4.1 illustrates the example for the binding energy,

in which the subscript refers to the geometry and the superscript refers to the basis functions. This is the value for the complexation energies that will be referred to in the text unless otherwise noted. Inclusion of the relaxation energy is not standard, but was found to be necessary due to the large structural changes observed between the isolated and complexed geometries.[106] Three body interaction terms were determined utilizing the method of Xantheas *et al.* to assess the role of cooperativity in these complexes, Equation 4.6.[107] The magnitude of the non-additive contributions to the binding energy is extremely relevant for standard non-additive molecular mechanics models because they cannot explicitly account for this energy without introducing polarization.

Binding Energy: $\mathbf{A}_{[\text{AA-H}_2\text{O}]} + \mathbf{B}_{[\text{HOR}]} \rightarrow \mathbf{AB}_{[\text{AA-H}_2\text{O-HOR(R = H, CH}_3)]}$
Dimerization Energy: $\mathbf{A}_{[\text{AA}]} + \mathbf{A}_{[\text{H}_2\text{O}]} \rightarrow \mathbf{A}_{[\text{AA-H}_2\text{O}]}$ [4.1]
Trimerization Energy: $\mathbf{A}_{[\text{AA}]} + \mathbf{A}_{[\text{H}_2\text{O}]} + \mathbf{B}_{[\text{HOR}]} \rightarrow \mathbf{AB}_{[\text{AA-H}_2\text{O-HOR(R = H, CH}_3)]}$

$$\Delta E = E_{NC} + \Delta E_{BSSE} + \Delta E_{FRE} + \Delta E_{ZPVE} \quad [4.2]$$

$$\Delta E_{NC} = E_{AB}^{\alpha\beta}(AB) - E_{A}^{\alpha}(A) - E_{B}^{\beta}(B) \quad [4.3]$$

$$\Delta E_{BSSE} = \left[E_{AB}^{\alpha}(A) - E_{AB}^{\alpha\beta}(A) \right] + \left[E_{AB}^{\beta}(B) - E_{AB}^{\alpha\beta}(B) \right] \quad [4.4]$$

$$\Delta E_{FRE} = \left[E_{AB}^{\alpha}(A) - E_{A}^{\alpha}(A) \right] + \left[E_{AB}^{\beta}(B) - E_{B}^{\alpha}(B) \right] \quad [4.5]$$

$$\begin{aligned} \Delta E_{3body} = {} & E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1A^2B) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^2) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(B) \\ & - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1A^2) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^2) \\ & - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1B) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(A^1) - E_{A^1A^2B}^{\alpha^1\alpha^2\beta}(B) \\ & - E_{A^1A^2B}^{\alpha^1\cup\alpha^2\cup\beta}(A^2B) - E_{A^1A^2B}^{\alpha^1\cup\alpha^2\cup\beta}(A^2) - E_{A^1A^2B}^{\alpha^1\cup\alpha^2\cup\beta}(B) \end{aligned} \quad [4.6]$$

The following nomenclature was adopted throughout the rest of the text: The amino acid water complex is represented by the one-letter amino acid abbreviations. In the case that there is a second, less stable dimer it is indicated by a prime. Inclusion of a superscript C represents a cyclic motif, while a superscript L a linear motif. In the cyclic

complexes the binding moiety is specified with a superscript a, b, or c. For example, the notation to illustrate the linear trimer formed by the binding of HOR (R=CH_3) to the tyrosine-H_2O dimer will be abbreviated as $\mathbf{Y}^{\mathbf{L}}_{\mathbf{R=CH_3}}$. In general, the hydrogen bonding distances are defined as the distance from the oxygen acceptor to the donating hydrogen unless otherwise stated.

In the case of $\mathbf{Y}^{\mathbf{L}}$, the expected linear cluster, in which tyrosine acted as a hydrogen-bond donor, collapsed to the cyclic cluster. Therefore, the binding energy is calculated with respect to the dimer that has the same hydrogen bonding arrangement, $\mathbf{Y'}$. Additionally, five more of the neutral clusters, $\mathbf{N}^{\mathbf{C}}_{\mathbf{R_a}}$, $\mathbf{N}^{\mathbf{C}}_{\mathbf{R_b}}$, $\mathbf{N}^{\mathbf{L}}\mathbf{'}$, $\mathbf{Y}^{\mathbf{C}}_{\mathbf{R_b}}$, and $\mathbf{T}^{\mathbf{C}}_{\mathbf{R_b}}$ were calculated with respect to, $\mathbf{N'}$, $\mathbf{N''}$, $\mathbf{N''}$, $\mathbf{Y'}$, and $\mathbf{T'}$ to maintain consistency between the hydrogen-bonding arrangement in the trimer and dimer. In the cyclic complexes, more than just the lowest energy complex was analyzed to provide an overview of the different hydrogen bond arrangements as well as the small energy differences between the possible complexes.

The Cartesian coordinates for the Con A-trimannoside complex were retrieved from the RCSB Protein Database (1CVN.pdb).[99] Hydrogen atoms were added to the X-ray coordinates using the Builder Module of InsightII. The protein complex was simplified to contain only the truncated amino acids and the trimannoside ligand that participate in the hydrogen bonding interactions at the binding site. Two copies were made of the simplified complex. In the first copy, the trimannoside ligand was replaced by methanol, aligned with the carbon atom (C2) and the hydroxyl group (HO-2) of the central mannosyl residue. Hydroxyl group HO-2 is involved in hydrogen bonding interactions with the protein, via a conserved water molecule. In the second copy, the

trimannoside ligand was replaced by a water molecule, which was formed with hydroxyl group HO-2. Initially, the non-hydrogen atoms of the amino acids as well as the conserved water were frozen during geometry optimization in order to maintain the configuration present in the crystal structure while allowing the hydrogen positions to relax. An additional optimization allowed the whole complex to be fully optimized in order to determine the magnitude of the correction for ZPVE.

Water Clusters

At the B3LYP/6-31+G(d) level the average internuclear O···O distance was 2.77 Å in the water trimer, which compared favorably with theoretical results obtained at the MP2//aug-cc-pVDZ level (2.80 Å),[80] although both values underestimate the experimental value of 2.85 Å.[82] Upon substituting methanol for a single water in $(H_2O)_3$, the hydrogen bond formed between methanol, acting as an acceptor, and the neighboring water decreased in length by 0.02 Å relative to that in the water trimer. In contrast, when methanol acted as a donor, the hydrogen bond distance increased by the same amount, relative to the corresponding hydrogen bond in the water trimer. These results agree with previous studies that showed that the strongest hydrogen bond in the $(H_2O)_2$-CH_3OH trimer occurs when methanol acts as an acceptor, whereas in the weakest methanol acts as a donor.[87]

The binding energy for the water trimer from the water dimer and water monomer was -5.96 kcal·mol^{-1}. Upon substitution of a single water monomer for methanol, three different $(H_2O)_2$-CH_3OH trimers are formed, the a-, b-, and c- complexes shown in Figure 4.1. The average binding energy of methanol to the water dimer was more favorable than water with a value of -6.82 kcal·mol^{-1}, shown in Table 4.1. This difference in energy is in

agreement with previous theoretical results that showed the dissociation energy to be about 1 kcal·mol^{-1} more negative for $(H_2O)_2$-CH_3OH than for $(H_2O)_3$.[87]

Table 4.1 Binding energy and cooperativity terms.[a]

Complex	binding energy		three body interaction energy		hydrogen bonds formed[b]
	R=H	R=CH_3	R=H	R=CH_3	
water					
a-	-5.95	-6.83	-2.94	-3.37	2
b-	-5.97	-6.78	-2.94	-2.89	2
c-	-5.97	-6.86	-2.94	-2.86	2
tyrosine					
$\mathbf{Y}^{C}_{R_a}$	-5.55	-6.38	-2.74	-2.77	2
$\mathbf{Y}^{C}_{R_b}$	-7.22	-8.25	-2.74	-2.62	2
$\mathbf{Y}^{L}$	-3.69	-4.24	-1.55	-1.49	1
asparagine					
$\mathbf{N}^{C}_{R_a}$	-8.80	-9.62	-4.31	-4.19	2
$\mathbf{N}^{C}_{R_b}$	-10.68	-11.59	-4.31	-4.13	2
$\mathbf{N}^{L}$	-1.61	-2.30	0.14	0.13	1
$\mathbf{N}^{L'}$	-3.69	-4.59	-1.26	-1.41	1
threonine					
$\mathbf{T}^{C}_{R_a}$	-5.21	-6.81	-2.84	-2.75	2
$\mathbf{T}^{C}_{R_b}$	-7.41	-7.43	-2.84	-2.77	2
$\mathbf{T}^{L}$	-4.28	-4.92	-1.92	-1.88	1
lysine					
$\mathbf{K}^{B}$	-13.19	-15.06	2.20	2.83	1
$\mathbf{K}^{L}$	-11.11	-12.78	-3.76	-4.45	1
aspartate					
$\mathbf{D}^{C}$	-11.94	-13.42	1.22	1.86	2
arginine					
$\mathbf{R}^{C}_{R_a}$	-7.53	-9.36	0.18	-0.01	2
$\mathbf{R}^{C}_{R_b}$	-7.51	-8.75	0.18	0.73	2
$\mathbf{R}^{L}$	-9.64	-11.00	-3.71	-4.41	1
Con A[c]	-12.83	-12.85	-1.13	-0.75	2
Con A[d]	-5.70	-5.94	-0.84	-0.65	2

[a]All energies are in kcal·mol^{-1}. Positive values are repulsive terms.
[b]Number of hydrogen bonds formed involving the binding moiety.
[c]Partially optimized complex.
[d]Fully optimized complex.

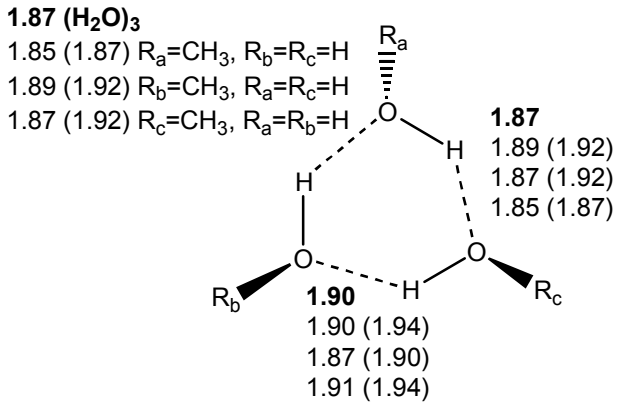


Figure 4.1 Hydrogen bond distances (Å) in the water trimer (in bold), as well as in the three possible $(H_2O)_2$-CH_3OH trimers (B3LYP/6-31+G(d) level). The subscripts a, b, and c refer to the position of the binding species in the cyclic complex. The values in parentheses were previously obtained at the MP2(full)/6-311+G(d,p) level.[87]

The cooperativity terms are favorable for the binding of methanol, or water, to the water dimer. The cooperativity term was -2.94 kcal·mol^{-1} for the water trimer, which is in agreement with previous calculations at the MP2/aug-cc-pVTZ level that yielded -2.45 kcal·mol^{-1}.[89] In contrast, when methanol bound to the dimer the magnitude of the cooperativity term depended on which of the three complexes was formed. The largest enhancement (-3.37 kcal·mol^{-1}) was seen in the a-trimer, whereas the enhancement in the b- and c-trimers was approximately -2.88 kcal·mol^{-1}. This was expected because the methyl group in the a-trimer projected into the plane opposite to both of the water protons that were not participating in the hydrogen bonding interactions, thus avoiding any possible unfavorable steric interactions. Previous calculations of the 3-body interaction

energy for the a-trimer, at the G2(MP2) and B3LYP/6-311+G(3df,2p) levels, yielded -3.0 and -2.9 kcal·mol^{-1}, respectively.[87]

Uncharged Clusters

For each of the three uncharged dimers more than one local minima was determined to be consistent with the hydrogen bonding networks of the larger clusters. The differences in the cluster geometries of both the tyrosine (**Y**) and threonine (**T**) dimers were based on the ability of the hydroxyl groups in these amino acid side chains to act as either donor or acceptor atoms in hydrogen bond formation. The threonine side chain preferred to be an acceptor, by 0.94 kcal·mol^{-1}, while the tyrosine side chain preferred to be a donor, by 2.15 kcal·mol^{-1}, which is expected based on the lower pK_a of the phenolic proton (Table 4.2). In both cases, the hydrogen bond distances were shorter in the more energetically favored structures, although the O-H···O angles were also slightly more distorted from linearity, to values of 172.40° and 170.26° for **Y** and **T**, respectively.

Table 4.2 Energetic differences between the low energy neutral AA-(conserved water) clusters.

Complex	Energy[a]	ΔE[b]
T	-231.579570	0.00
T'	-231.578074	0.94
Y	-423.385781	0.00
Y'	-423.382359	2.15
N	-285.782341	0.00
N'	-285.778791	2.23
N''	-285.774871	4.69

[a] hartrees
[b] kcal·mol^{-1}

Within all four linear trimers, $\mathbf{Y}$, $\mathbf{N^L}$, $\mathbf{N^{L'}}$ and $\mathbf{T^L}$, similar structural trends were observed, and are shown in Figure 4.2. For example, upon substituting methanol for water, the AA-(conserved water) hydrogen bond typically did not change in length significantly, with only a slight decrease in the $CH_3OH\cdots OH_2$ hydrogen bond length relative to that of $HOH\cdots OH_2$, of approximately 0.02 Å, and an average deviation of 1.23° in the $O\text{-}H\cdots O_{meoh/water}$ angle. This modest geometrical difference is consistent with the 0.69 kcal·mol^{-1} favorable increase in binding energy when methanol is the ligand, relative to water. The uncharged cyclic trimers ($\mathbf{Y^C}$, $\mathbf{N^C}$, and $\mathbf{T^C}$) behaved similarly to the $(H_2O)_2$-CH_3OH trimers. For example, when the ligand acted as an acceptor, the binding of methanol resulted in a stronger hydrogen bond, relative to the binding of water. In contrast, when the ligand acted as a donor, the hydrogen bond was stronger for water than for methanol.

Complexation of either methanol or water to the uncharged amino acid water clusters resulted in binding energies that ranged from –11.59 to –1.61 kcal·mol^{-1}. The same energetic trend is not seen in both cyclic and linear clusters. In the tyrosine and threonine cyclic clusters, single atoms act as both a donor and acceptor, this is energetically less favorable than if the donor and acceptor are separate atoms, as in the cyclic complex containing asparagine. Formation of uncharged cyclic clusters is more favorable than for the corresponding neutral linear clusters. Inclusion of the relaxation energy, which is generally larger for the cyclic clusters due to the inherently strained geometry in the cyclic clusters, decreased the separation between the two types. The three clusters, $\mathbf{Y^C_{R_b}}$, $\mathbf{T^C_{R_b}}$ and $\mathbf{N^{L'}}$, were higher in energy than $\mathbf{Y^C_{R_a}}$, $\mathbf{T^C_{R_a}}$, and $\mathbf{N^L}$, respectively (Table 4.3), due to the less favorable hydrogen-bonding network present. As

a result, the dimer utilized in the binding energy calculations is consistent with the hydrogen bond network present in the trimer, but is not the lowest energy structure. In all cases this results in a more favorable binding energy relative to that of the corresponding calculation utilizing the lowest energy dimer.

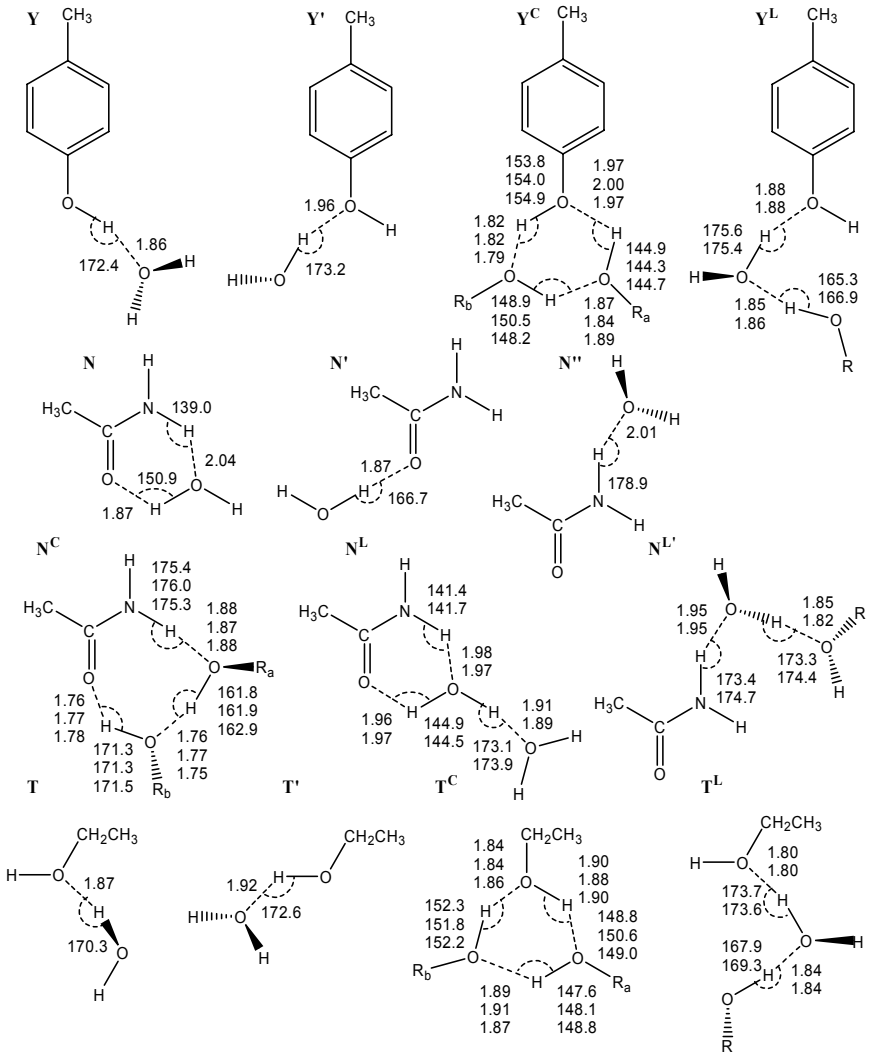


Figure 4.2 Geometries for polar uncharged clusters, where hydrogen bond distances (Å) and angles (°) are defined top to bottom as R_a=R_b=H, R_a=CH_3 and R_b=H, and R_a=H and R_b=CH_3. The **N**/**Q** and **S**/**T** clusters refer to representative fragments from asparagine/glutamine and serine/threonine, respectively, but are referred to as **N** and **T**.

Overall, neutral clusters exhibited favorable three-body interaction energies, except for the case of $\mathbf{N}^{\mathbf{L}}$. The conserved water acts as both a donor to the carbonyl oxygen and an acceptor for the amide group of the asparagine side chain, making $\mathbf{N}^{\mathbf{L}}$ unique. Addition of the second water increases the length of the hydrogen bond to the carbonyl group and slightly decreases the length of the hydrogen bond involving the amide group. Since the three-body term is unfavorable, the weakening of the carbonyl-(conserved water) hydrogen bond dominates the energy.

Table 4.3 Energy comparison between low energy AA-(conserved water)-HOR (R=H,CH_3) clusters.

Complex	Energy[a]	ΔE[b]	Cluster	Energy[a]	ΔE[b]
$\mathbf{Y}^{C}_{R=H}$	-499.864594	0.00	$\mathbf{Y}^{C}_{R_a=CH_3}$	-539.174669	0.00
$\mathbf{Y}^{L}_{R=H}$	-499.857603	4.39	$\mathbf{Y}^{C}_{R_b=CH_3}$	-539.174657	0.01
			$\mathbf{Y}^{L}_{R=CH_3}$	-539.167162	4.71
$\mathbf{N}^{C}_{R=H}$	-362.263666	0.00	$\mathbf{N}^{C}_{R_a=CH_3}$	-401.573377	0.16
$\mathbf{N}^{L}_{R=H}$	-362.252544	6.98	$\mathbf{N}^{C}_{R_b=CH_3}$	-401.573635	0.00
$\mathbf{N}^{L'}_{R=H}$	-362.24968	8.78	$\mathbf{N}^{L}_{R=CH_3}$	-401.562891	6.74
			$\mathbf{N}^{L'}_{R=CH_3}$	-401.560236	8.41
$\mathbf{T}^{C}_{R=H}$	-308.059153	0.00	$\mathbf{T}^{C}_{R_a=CH_3}$	-347.369102	0.00
$\mathbf{T}^{L}_{R=H}$	-308.055461	2.32	$\mathbf{T}^{C}_{R_b=CH_3}$	-347.369029	0.05
			$\mathbf{T}^{L}_{R=CH3}$	-347.365001	2.57
$\mathbf{K}^{B}_{R=H}$	-249.237748	0.00	$\mathbf{K}^{B}_{R=CH_3}$	-288.549756	0.00
$\mathbf{K}^{L}_{R=H}$	-249.235109	1.66	$\mathbf{K}^{L}_{R=CH_3}$	-288.546821	1.84
$\mathbf{D}^{C}_{R=H}$	-381.59955	0.00	$\mathbf{D}^{C}_{R=CH3}$	-420.910217	0.00
$\mathbf{R}^{C}_{R=H}$	-398.145732	1.82	$\mathbf{R}^{C}_{R_a=CH_3}$	-437.457414	1.62
$\mathbf{R}^{L}_{R=H}$	-398.148638	0.00	$\mathbf{R}^{C}_{R_b=CH_3}$	-437.456554	2.16
			$\mathbf{R}^{L}_{R=CH_3}$	-437.459989	0.00

[a] hartrees
[b] kcal·mol^{-1}

Enthalpically, the complexation of the amino acid-conserved water complex is less favorable than the binding of the second water to the aforementioned complex. Formation from the lowest energy dimers leads to binding energies that are less than two times the magnitude of the dimerization energy, Table 4.4, while formation from higher energy dimers leads to binding energies that are more than two times greater in magnitude. Clearly, the additional water helps stabilize the less favorable dimers, while the more optimal hydrogen bond arrangements in the low energy dimers is weakened by the additional water. Overall, formation of the amino acid-H_2O-HOR complex from two entities (amino acid-H_2O complex and HOR) is consistently less favorable than the corresponding formation from three individual monomers due to the disruption of the favorable interactions in the amino acid-conserved water complex, Table 4.5.

Charged Clusters

For the three charged clusters investigated here, only one energy minimum was found for each dimer. The aspartate-H_2O and lysine-H_2O dimers each had one distinct structure, **D** and **K**, respectively. The arginine-H_2O dimer had the possibility of a second structure, in which only one hydrogen bond was present in addition to that present in structure **R** (Figure 4.3). At the level of theory utilized here, this second putative structure collapsed to **R**, which agrees with previous studies of related systems where the minimum structure contained two hydrogen bonds.[108] Linear and cyclic structures were found to be minima for the arginine-$(H_2O)_2$ cluster, while the linear structure for aspartate-$(H_2O)_2$ collapsed to the cyclic $\mathbf{D}^{C}$ cluster during geometry optimization.

Table 4.4 Dimerization energies.

Neutral complex	Energy (kcal·mol^{-1})	Charged complex	Energy (kcal·mol^{-1})
tyrosine		lysine	
Y	-3.26	**K**	-15.69
Y'	-1.63		
asparagine		arginine	
N	-5.24	**R**	-12.40
N'	-4.04		
N"	-2.48		
threonine		aspartate	
T	-2.97	**D**	-13.90
T'	-1.87		

Table 4.5 Trimerization energies.

Neutral complex	Energy (kcal·mol^{-1})		Charged complex	Energy (kcal·mol^{-1})	
	R=H	R=CH_3		R=H	R=CH_3
tyrosine			lysine		
$\mathbf{Y}^{C}_{R_a}$	-7.94	-8.77	$\mathbf{K}^{B}$	-28.74	-30.27
$\mathbf{Y}^{C}_{R_b}$	-7.94	-8.96	$\mathbf{K}^{L}$	-25.68	-26.87
$\mathbf{Y}^{L}$	-4.69	-5.30			
asparagine			arginine		
$\mathbf{N}^{C}_{R_a}$	-11.21	-12.07	$\mathbf{R}^{C}_{R_a}$	-19.57	-21.01
$\mathbf{N}^{C}_{R_b}$	-11.21	-12.15	$\mathbf{R}^{C}_{R_b}$	-19.57	-20.68
$\mathbf{N}^{L}$	-6.79	-7.45	$\mathbf{R}^{L}$	-21.26	-22.28
$\mathbf{N}^{L'}$	-5.74	-6.55			
threonine			aspartate		
$\mathbf{T}^{C}_{R_a}$	-7.87	-8.76	$\mathbf{D}^{C}$	-25.58	-26.84
$\mathbf{T}^{C}_{R_b}$	-7.87	-8.69			
$\mathbf{T}^{L}$	-6.36	-7.05			

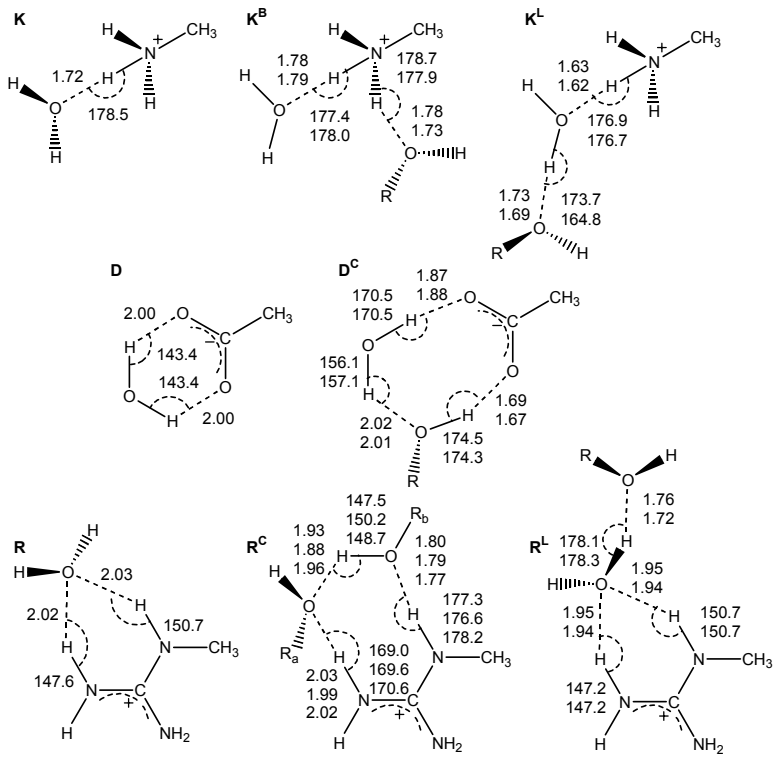


Figure 4.3 Geometry of the charged clusters, where the hydrogen bond distances are defined from top to bottom as R_a=R_b=H, R_a=CH_3 and R_b=H, and R_a=H and R_b=CH_3, and all angles are in degrees. The D/E clusters represent the representative fragments from the aspartate and glutamate clusters, but will only be referred to as D in the text.

The $\mathbf{K^B}$ cluster was unique in that both waters were bound directly to the amino acid side chain without a water-water hydrogen bond. In this branched structure both of the N-H···O angles remained close to linearity (Figure 4.3).[109] In both of the lysine-(H_2O)-(CH_3OH) clusters, as well in the linear arginine-(H_2O)-(CH_3OH) cluster, the hydrogen bond that directly involved methanol was stronger than the corresponding interaction found in the lysine-$(H_2O)_2$ cluster, revealing a more favorable interaction with methanol.

Both aspartate and arginine formed true cyclic structures, which are comparable geometrically to the water trimer. Upon formation of $\mathbf{D^C}$ and $\mathbf{R^C}$, the average hydrogen bond lengths decreased relative to those in the isolated dimers. This was mostly due to

the strained configuration of the dimers, in which water acted as either a double acceptor or a double donor. In both structures, the shortest hydrogen bond corresponded to the one involving an N-H···O or O-H···O angle that remained close to linearity. The hydrogen bond distances in both cyclic structures were similar to those in the water trimer. Small fluctuations in hydrogen bond lengths of less than 0.05 Å were observed upon the replacement of water by methanol. In contrast to the neutral clusters, the hydrogen bond was not always stronger when methanol was acting as an acceptor. For example, in $\mathbf{D}^{\mathbf{C}}$, the strongest hydrogen bond existed when methanol donated to the aspartate.

The binding energies of water to each of the charged AA-H_2O dimers ranged from 1.55 to 7.23 kcal·mol^{-1} more favorable than the binding energy of water to $(H_2O)_2$. According to the binding energies, the strength of the water interactions with the complexes may be ranked from weakest to strongest as: $\mathbf{R}^{\mathbf{C}}_{\mathbf{R}_b} > \mathbf{R}^{\mathbf{C}}_{\mathbf{R}_a} > \mathbf{R}^{\mathbf{L}} > \mathbf{K}^{\mathbf{L}} > \mathbf{D}^{\mathbf{C}} > \mathbf{K}^{\mathbf{B}}$. The formation of the two cyclic arginine clusters relied on the breaking of one of the hydrogen bonds in the dimer, in order to form two new hydrogen bonds in the trimer. This bond breaking may have contributed to the 2.08 kcal·mol^{-1} destabilization of the complex relative to the linear arginine cluster. The binding energy of water to either the lysine or aspartate clusters was at least 1.5 kcal·mol^{-1} more favorable than that to the arginine cluster. This may be due to the more localized charge in the lysine and aspartate clusters, relative to the case of arginine. Further, the more favorable binding of water to $\mathbf{D}^{\mathbf{C}}$, relative to $\mathbf{R}^{\mathbf{L}}$, is consistent with the observation that water prefers to act as a bi-donor rather than a bi-acceptor.[88] Theoretical measurements and experimental calculations have shown that for related structures of $(H_2O)_2OH^-$ and $(H_2O)_2H_3O^+$, the

formation of the positively charged cluster is more favorable, which agrees with our findings that the formation of the cyclic lysine cluster (positive charge) was more favorable than that of the cyclic aspartate cluster (negative charge).[110]

The binding energy of the branched lysine cluster (-13.2 kcal·mol^{-1}) was more favorable than that for the linear cluster (-11.1 kcal·mol^{-1}), which is also in agreement with previous calculations that revealed the branched structure to be 2.5 kcal·mol^{-1} more stable than the linear structure.[109] The conserved water was in approximately the same position in both lysine trimers. Therefore, the majority of the difference in binding energy is related to the interactions involving the bound water. In the branched structure the bound water directly interacted with the charged entity, helping to stabilize the complex, while the bound water in the linear structure was not involved in direct interactions with the ammonium group.

The binding energies of methanol to each of the charged AA-H_2O dimers were from 1.92 to 8.23 kcal·mol^{-1} more favorable than the binding of methanol to the water dimer. The binding energies followed the same trend regardless of whether the ligand was methanol or water, however the differences in binding energies were more substantial in the case of methanol. Overall, the average binding energy of methanol to the charged clusters was 1.58 kcal·mol^{-1} more favorable than that of water, which represents a favorable increase of 0.78 kcal·mol^{-1} from the uncharged clusters. In contrast to the neutral clusters, dimerization of the amino acid and conserved water was energetically more favorable than the complexation of the bound water to the amino acid-conserved water complex, which reveals the additional water actually enthalpically destabilizes the complex. In addition, the large magnitude of the difference in

trimerization and binding energy for both the charged and neutral clusters illustrates that the formation of three body clusters is most favorable from individual moieties as opposed to from smaller clusters.

The magnitude of the cooperative effects varied significantly among the charged clusters. For example, both $\mathbf{K}^{\mathbf{L}}$ and $\mathbf{K}^{\mathbf{B}}$ contain the lysine side chain, but had three-body terms that were opposite in sign. Only in the case of the linear charged clusters was cooperativity consistently favorable. In the cases of $\mathbf{K}^{\mathbf{B}}$, $\mathbf{R}^{\mathbf{C}}$, and $\mathbf{D}^{\mathbf{C}}$, the addition of the ligand to form the cyclic or branched structures involved a direct interaction between the ligand and the amino acid side chains. Although this helped to stabilize the charged side chains, the (conserved water)···AA hydrogen bond was weakened, resulting in unfavorable cooperativity. In contrast, when the ligand bound to form the linear clusters $\mathbf{K}^{\mathbf{L}}$ and $\mathbf{R}^{\mathbf{L}}$, the ligand interacted only with the conserved water and, as a result, the conserved water was polarized and the (conserved-water)···ligand hydrogen bond was strengthened. This configuration resulted in favorable cooperativity. In general, the charged linear clusters exhibited more cooperativity than the charged cyclic clusters, which is in contrast to the results for the uncharged clusters.

Protein – Carbohydrate Interactions

Due to the absence of the hydrogen atoms in the X-ray diffraction data, donor-acceptor hydrogen bond relationships had to be inferred. By assuming that Arg-228 and Asp-14 are donor groups and Asp-16 provides the acceptor group, it was possible to locate the approximate positions of all hydrogen atoms. The geometry of the fully optimized Con A-(conserved water)-CH_3OH complex did not differ significantly from that of the Con A-(conserved water)-H_2O complex, with a RMSD of 0.27 Å for the heavy

atoms, Figure 4.4. The RMSD between the crystal structure and the Con A-(conserved water)-H_2O complex was 0.63 Å, while that with the Con A-(conserved water)-CH_3OH complex was 0.65 Å.

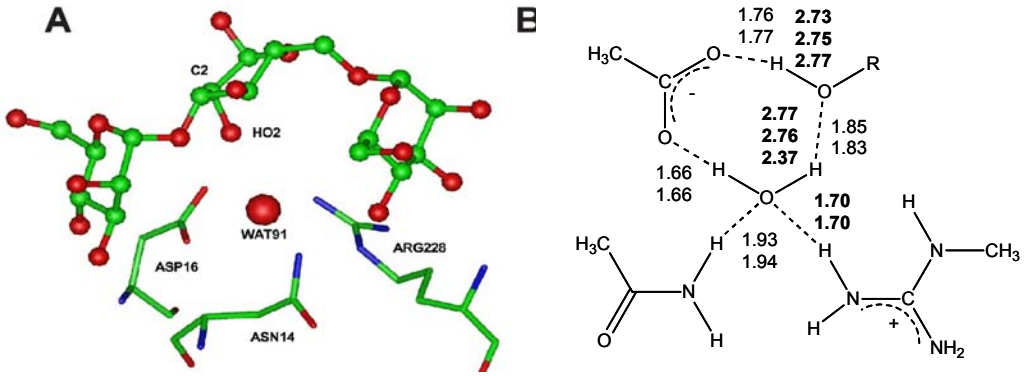


Figure 4.4 The cluster representative of the Con A binding site, with the trimannoside in A and the truncated sugar (R=CH_3) in B. The three truncated amino acids shown are asparagine, aspartate, and arginine. The O··O distances are shown in groups of three (in bold), with R=H, R=CH_3, and the experimental value from top to bottom respectively, while the O··H distances are shown for R=H and R=CH_3 from top to bottom respectively.

The interactions between the aspartate side chain and both the conserved and bound water in the Con A complex is similar to $\mathbf{D}^{\mathbf{C}}$, however, in the Con A complex the conserved water is further coordinated to other side chains. The extra interactions decrease the strength of the hydrogen bond between the aspartate side chain and the ligand, while increasing the strength of the two hydrogen bonds involving the conserved water. The result is a conserved water that is more polarized, yielding a more favorable configuration in the Con A complex. Comparison of the asparagine side chain in the Con A complex to that in $\mathbf{N}^{\mathbf{L}}$ reveals no significant differences. The arginine side chain in the Con A complex is in a less favorable configuration than in $\mathbf{R}^{\mathbf{L}}$ due to the presence of

only one hydrogen bond in the Con A complex relative to two in $\mathbf{R}^{\mathbf{L}}$. Therefore, the presence of additional side chains makes a larger impact on the geometries of the charged residues than on the neutral residues.

Initially, the Con A complex was partially optimized, while holding the heavy atoms of the side chains and the conserved water in their crystallographically determined positions to determine a biochemically relevant binding energy. Analogous to the smaller clusters, one would expect the hydrogen bonds in the Con A system to strengthen upon the binding of methanol, in comparison to the binding of water. However, the hydrogen bond distances for both methanol and water were equivalent. There was only a negligible difference of 0.02 kcal·mol^{-1} in the binding energy of methanol to the Con A cluster, relative to the case of water, for the partially optimized complex corresponding to the lack of difference in the geometry of the two complexes. Allowing the complex to optimize fully yielded similar geometries, but the difference increased to 0.24 kcal·mol^{-1}, which can therefore be attributed mostly to the inclusion of ZPVE. As a result, the magnitude of the binding energy most likely corresponds to that of the partially optimized complex, while the difference corresponds to the fully optimized complex.

Overall, the interplay between neutral and charged amino acids with the conserved water may be relatively insensitive to the ligand that is present (where the number of hydrogen bonds and net charge is equivalent for both structures) but is dependent on the number and type of amino acid side chains. This small energetic difference suggests a situation where the binding equilibrium may be dominated by entropic contributions, which have been shown to be favorable and significant for carbohydrate-lectin binding.[100-102,111]

Conclusions

Binding of HOR to the uncharged linear clusters was either approximately the same as, or not as favorable as, HOR binding to $(H_2O)_2$. In contrast, the binding of HOR to the charged linear clusters was always favorable relative to $(H_2O)_2$. This correlated directly with the consistently more favorable three-body interaction energy in the linear charged clusters. In contrast, the binding of both water and methanol was more favorable for the charged cyclic clusters than for the uncharged cyclic clusters, despite the possible unfavorable three-body term that is mainly due to the charge stabilization over the complex. The binding of either methanol or water to the Con A complex, which contained both neutral and charged amino acid side chains with a total net charge of zero, was found to be similar to that of binding to the water dimer. Overall, the formation of neutral complexes exhibited cooperativity, with the exception of the previously noted asparagine complex, while the exhibition of cooperativity in charged clusters depended on both geometry and presence of charge.

The corrections to the binding energies had a substantial effect, especially in the case of the ZPVE and relaxation corrections. It has been established that the inclusion of ZPVE is necessary to accurately determine the binding energy of hydrogen-bonded water complexes.[112] In general, the exclusion of ZPVE yielded a binding energy for methanol that was slightly more negative than that for the binding of water. For all trimers, the ZPVE was larger in the case where water was the ligand, relative to methanol, increasing these energetic differences. This is clearly illustrated in the difference in binding energies computed from the partially and fully optimized Con A geometries, where ZPVE was only computed for the fully optimized structure. Relaxation was extremely

important in the cases where the AA-H_2O dimers could form two hydrogen bonds. Also, for the AA-H_2O dimers investigated here, if it was possible to allow the formation of two hydrogen bonds, the resulting complex was, not surprisingly, energetically more favorable than the complex containing only one hydrogen bond.

The addition of methanol consistently stabilized all of the complexes relative to the addition of a second water molecule. This is in agreement with experimental and theoretical findings that clusters formed with methanol are energetically more favorable than pure water clusters.[113] Therefore, to the extent that methanol served as a model carbohydrate, our results confirmed that water-carbohydrate interactions were generally more favorable than water-water interactions. However, this energetic difference was quite small, reaffirming that entropic contributions are significant in the binding of carbohydrates to proteins. Moreover, a test set was established for further development of molecular mechanics models.

5. VALIDATION AND TESTING OF THE LONE-PAIR MODEL

Introduction

Carbohydrate structure is notably influenced by localized interactions with water, but the strength of the interactions within the oligosaccharide or lectin/carbohydrate complex are also extremely sensitive to discrete water.[7,22,114-117] In order to correctly gauge the magnitude of the interactions between the solvent and solute, reproduction of the relative energies and structures pertaining to a series of gas-phase hydrogen bonded clusters is the first step.[27,37] Examination of the relative energies for the approach of a water molecule to a compound containing a hydrogen bond acceptor or donor provide a useful test of the electrostatic model and illustrate the compatibility between the chosen water model and the electrostatic model for the solute. Secondly, crystalline-phase MD simulations of pyranoses explicitly demonstrate how well the non-bonded parameters perform and the precision of the successful models will transfer over to the accurate modeling of conformational properties of oligosaccharides in aqueous solutions. Neutron diffraction data is abundant for small monosaccharides, yields defined unit cell lengths and examines only a single conformation for each hexose. Therefore, modeling crystalline unit cells provides a computationally feasible test case as it does not suffer from the conformational sampling issues present in solution and exact hydrogen bond donor and acceptor relationships are known.

The intensities yielded from Nuclear Overhauser Effect (NOE) NMR spectroscopy are approximately inversely proportional to the sixth power of the separation between two corresponding protons. An additional facet of NMR

spectroscopy relates the magnitude of the nuclear spin splitting between two nuclei separated by three bonds to its respective dihedral angle. Together, this data provides distances across the pyranose ring and therefore helps elucidate the 3D conformation of the saccharide. Monitoring the distances and angles that correspond to experimentally determined NMR parameters for α-Glc*p*-OMe, α-Gal*p*-OMe, α-Glc*p*-(1→4)-α-Glc*p*-OMe (methyl α-maltoside) and α-Glc*p*-(1→6)-α-Glc*p*-OMe (methyl α-isomaltoside) over the course of explicitly solvated MD simulations show that, on average, GLYCAM04 and GLYCAM04-LP perform as well or better than earlier versions of GLYCAM and as well as other carbohydrate parameter sets.

Previously, Corzana *et al.* determined how well carbohydrate parameter sets can reproduce experimentally determined properties for two α-glycans in solution as well as to each other when experimental data were not available.[118] A notable division in the properties computed with each force field occurred between the ability to predict the structural organization of solvent around the solute and the ability to correctly predict the dynamic behavior of the solute in water. On average, GLYCAM2000a predicted conformational properties for both α-glycans with the smallest deviation from experiment compared to the other parameter sets investigated, CSFF and HGFB.[118] In contrast, the dynamic properties of the disaccharides modeled with GLYCAM2000a and the TIP3P water model predicted the overall water structure to be less ordered around the solute relative to the other parameter sets as well as relative to bulk water.[36,118,119] In parallel, the orientational motion of the solute in water was too fast relative to experiment as well as the other parameter sets investigated. The dynamic properties of the water model may contribute to the lack of structure exhibited by the combination of the TIP3P water model

and the GLYCAM2000a parameter set since the diffusion constant at 298 K for TIP3P (5.19 x 10^{-5} $cm^2 \cdot s^{-1}$ ± 0.08) is nearly twice the experimental value of 2.30 x 10^{-5} $cm^2 \cdot s^{-1}$.[40,44] Therefore, implementation of the TIP4P (D = 3.31 ± 0.08 x 10^{-5} $cm^2 \cdot s^{-1}$) or TIP5P (D = 2.62 x 10^{-5} $cm^2 \cdot s^{-1}$ ± 0.04) water models as well as the new lone-pair inclusive parameter set will elucidate the origin of this problem.[33,40]

Methods

Quantum Mechanical Computations All quantum mechanical computations were performed with Gaussian 98, version A.11.3.[75] Two different approach trajectories for the *N*-methyl acetamide (NMA)-water complex, referred to as 120N and 120C, were optimized at the B3LYP/6-31+G(d,p) level. During the approach trajectory the $O_{H_2O} \cdots O_{NMA}$ distance was constrained to form the potential energy surface along this coordinate. In addition, in order to obtain a smooth QM approach trajectory the inter-molecular angles and dihedrals were restrained. Two approach trajectories of a water – methanol complex were obtained from Kirschner *et al.* at the HF/6-31G(d) level.[84] In one complex the water molecule acts as the hydrogen bond donor (WdM) and in the second complex methanol behaves as the hydrogen bond donor (MdW). The unrestrained local minima for MdW, WdM, 120C, and 120N were optimized and their relative energies were computed at the B3LYP/6-31G(d,p) level.

Molecular Mechanics Energy minimizations were performed with the SANDER module of AMBER7 with a dielectric constant of unity for 10,000 cycles (9,000 steepest descent followed by 1,000 conjugate gradient).[49] All atoms were included in the calculation of non-bonded interactions and each energy minimization was initiated from the *ab initio* geometry to which it was compared. In each of the approach trajectories, the

intermolecular O···O distance was restrained at the QM value along the potential energy surface. In addition, the methyl group was restrained from rotating away from the *ab initio* value in the NMA-water complex with a harmonic restraint. The only additional restraints were employed as needed to ensure the water molecule remained in the QM orientation relative to the methanol or NMA molecule. No restraints were employed during the energy minimization for any other complex not already explicitly specified.

Crystal Simulations The unit cells for the orthorhombic crystalline structures of α-D-Glc*p* (GLUCSA10[120]), α-D-Glc*p*-OMe (MGLUCP11[121]), α-D-Man*p*-OMe (MEMANP[121]), and β-D-Gal*p*-OMe (MBDGAL02[122]), were each transformed using the $P2_12_12_1$ symmetry operators to generate crystal lattices containing 64 monosaccharides.[120-122] Similarly, the $P2_1$ symmetry operators were utilized to generate a 2x4x2 system for α-D-Glc*p* monohydrate (GLUCMH11[123]) and a 4x4x4 system for α-D-Glc*p*NAc (ACGLUA11[124]). [123,124] The experimental monosaccharide conformation was employed in the initial configuration. Subsequently, no position or symmetry restraints were applied and the box dimensions were allowed to vary over the course of the MD simulations as implemented in AMBER8.[125] For each system governed by the $P2_12_12_1$ symmetry operator, the structure was obtained from neutron diffraction data. The $P2_1$ structures were obtained by experimental crystal density difference experiments, no neutron diffraction data sets were available, in which the atomic position for each hydrogen atom in the monosaccharide was optimized to the difference data and included in the refined structure. Therefore, the data sets obtained by neutron diffraction, as well as those by crystal density difference experiments, have the atomic position for each atom

in the monosaccharide and no approximations were made to generate the initial starting configuration.

For each monosaccharide, ensemble-averaged charges were derived utilizing the same 100 conformers employed previously in the generation of ensemble-averaged charges in the GLYCAM2000a and 04 parameter sets.[13,66] In contrast to GLYCAM2000a, the aliphatic hydrogen atoms are excluded from the charge fitting in GLYCAM04. Exclusion of the aliphatic hydrogen atoms and inclusion of lone-pairs during the charge fitting led to a robust partial charge set, with nearly identical values for the atoms in the same chemical environment within the monosaccharide, Figure 5.1.[69]

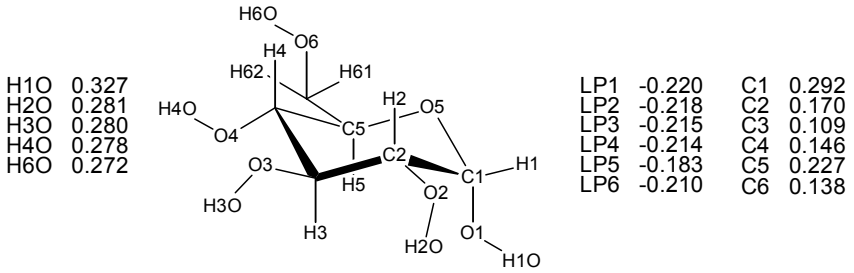


Figure 5.1 Partial charges for the GLYCAM04-LP model of α-D-Glc*p*.

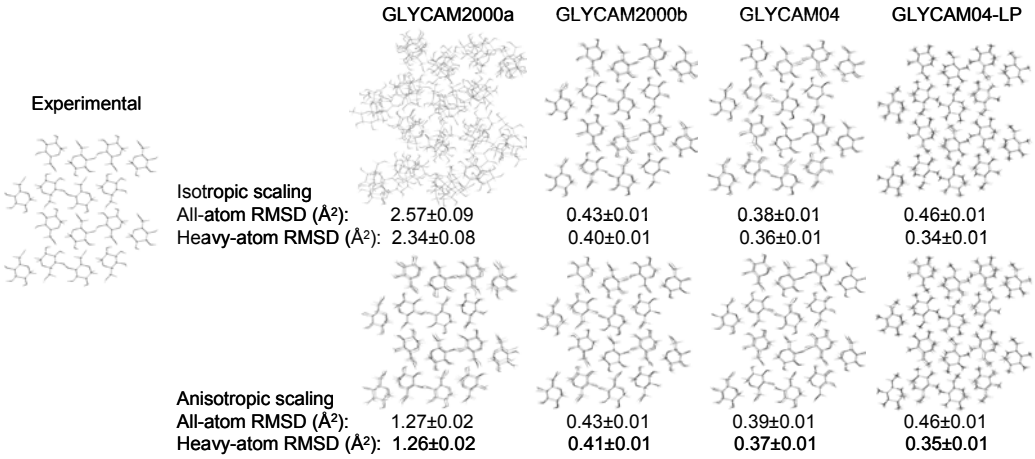


Figure 5.2 Effect of isotropic and anisotropic pressure scaling on the crystal disorder during crystalline-phase MD simulations.

Table 5.1 Comparison of anisotropic and isotropic pressure scaling for MD simulations of α-Glc*p*-OMe[121] (A), β-Gal*p*-OMe[122] (B), and α-Man*p*-OMe[121] (C).

	Exp.	GLYCAM2000a		GLYCAM04		GLYCAM04-LP	
		isotropic	anisotropic	isotropic	anisotropic	isotropic	anisotropic
A							
A (Å)	22.62	23.76±0.02	24.73±0.03	22.69±0.02	22.91±0.03	22.64±0.01	22.70±0.03
B (Å)	29.56	31.05±0.03	32.48±0.03	29.65±0.02	30.02±0.05	29.58±0.02	30.40±0.04
C (Å)	21.12	22.19±0.02	20.00±0.03	21.19±0.01	20.55±0.05	21.14±0.01	20.25±0.03
ΔA (%)		5.05	9.34	0.31	1.30	0.08	0.33
ΔB (%)		5.05	9.86	0.31	1.55	0.07	2.84
ΔC (%)		5.05	-5.32	0.31	-2.68	0.09	-4.10
Mean |Δ| (%)		5.05	8.17	0.31	1.84	0.08	2.42
$\langle E_{ELEC} \rangle_{(c)}$		-6245±25	-6299±21	-8914±26	-8872±25	-4515±21	-4466±21
$\langle E_{VDW} \rangle_{(c)}$		-1165±11	-1204±10	-1191±17	-1227±10	-1137±17	-1186±19
$\langle E_{ELEC} \rangle_{(g)}$		-88±3	-88±3	-114±3	-114±3	-43±2	-43±2
$\langle E_{VDW} \rangle_{(g)}$		-1.8±0.6	-1.3±0.5	-2.1±0.4	-2.1±0.5	-2.3±0.4	-2.4±0.3
$\langle E_{LATTICE} \rangle$	-34	-26	-27	-42	-42	-44	-44
O2···H3O (Å)	1.770	4.08±0.39	4.05±0.31	2.05±0.25	2.13±0.29	1.91±0.18	1.97±0.22
O3···H6O (Å)	1.772	2.56±0.65	2.44±0.36	1.88±0.16	1.91±0.18	1.86±0.14	1.87±0.14
O6···H2O (Å)	1.738	2.33±0.25	2.36±0.23	1.79±0.11	1.79±0.12	1.77±0.11	1.79±0.11
B							
A (Å)	31.12	32.94±0.03	33.95±0.07	31.30±0.02	32.05±0.07	31.21±0.02	31.89±0.17
B (Å)	17.07	18.07±0.02	16.52±0.03	17.17±0.01	16.87±0.04	17.12±0.01	16.69±0.08
C (Å)	26.26	27.80±0.03	28.96±0.06	26.42±0.02	26.09±0.03	26.34±0.02	26.26±0.06
ΔA (%)		5.84	9.10	0.58	2.99	0.29	2.46
ΔB (%)		5.85	-3.20	0.60	-1.15	0.30	-2.25
ΔC (%)		5.86	10.29	0.60	-0.63	0.31	-0.02
Mean |Δ| (%)		5.85	7.53	0.60	1.59	0.30	1.58
$\langle E_{ELEC} \rangle_{(c)}$		-6435±24	-6328±24	-8311±26	-8220±23	-4599±20	-4596±19
$\langle E_{VDW} \rangle_{(c)}$		-1147±12	-1195±12	-1225±14	-1251±13	-1264±10	-1278±12
$\langle E_{ELEC} \rangle_{(g)}$		-88±5	-86±5	-102±3	-100±3	-44.1±1	-43±1
$\langle E_{VDW} \rangle_{(g)}$		-2.0±0.3	-1.9±0.4	-2.2±0.2	-2.2±0.2	-2.3±0.2	-2.30.2
$\langle E_{LATTICE} \rangle$		-38	-29	-45	-46	-45	-46
O2···H6O (Å)	1.860	4.38±1.33	4.03±0.87	2.81±0.72	3.07±0.47	2.03±0.27	2.40±0.44
O6···H4O (Å)	1.739	3.90±1.34	3.06±0.68	1.83±0.13	1.84±0.13	1.85±0.12	1.84±0.12
O4···H2O (Å)	1.773	2.98±0.76	2.39±0.24	1.97±0.19	2.03±0.20	1.89±0.14	1.98±0.17
C							
A (Å)	18.86	21.34±0.01	19.81±0.03	18.99±0.01	19.28±0.02	18.89±0.01	19.16±0.02
B (Å)	37.26	39.54±0.02	38.91±0.04	37.52±0.02	37.22±0.03	37.32±0.02	37.03±0.03
C (Å)	20.11	21.34±0.01	21.78±0.03	20.25±0.01	20.17±0.02	20.14±0.01	20.07±0.02
ΔA (%)		6.11	5.06	0.69	2.25	0.16	1.57
ΔB (%)		6.12	4.43	0.70	-0.11	0.17	-0.63
ΔC (%)		6.12	8.29	0.70	0.27	0.17	-0.20
Mean |Δ| (%)		6.12	5.93	0.70	0.88	0.17	0.80
$\langle E_{ELEC} \rangle_{(c)}$		-5935±19	-5928±20	-7892±29	-7916±30	-4274±18	-4284±17
$\langle E_{VDW} \rangle_{(c)}$		-969±12	-983±12	-1209±14	-1196±14	-1255±12	-1244±13
$\langle E_{ELEC} \rangle_{(g)}$		-76±2	-76±2	-99±3	-99±3	-42±2	-42±2
$\langle E_{VDW} \rangle_{(g)}$		-1.8±0.5	-1.7±0.4	-2.3±0.4	-2.5±0.3	-2.5±0.3	-2.5±0.3
$\langle E_{LATTICE} \rangle$		-30	-30	-41	-41	-42	-42
O3···H6O (Å)	1.917	2.35±0.22	2.32±0.21	1.90±0.16	1.92±0.17	1.96±0.19	2.00±0.21
O4···H3O (Å)	1.810	2.16±0.15	2.19±0.17	1.82±0.12	1.82±0.11	1.86±0.13	1.85±0.12
O5···H4O (Å)	2.052	2.39±0.24	2.41±0.25	2.19±0.26	2.10±0.24	2.15±0.21	2.07±0.20

[a] A negative % deviation indicates a decrease in that dimension, while a positive value indicates expansion.

Unit scaling factors for 1,4 non-bonded interactions (S_{vdw} = S_q = 1) were employed during the MD simulations of carbohydrates, which is consistent with the GLYCAM2000a and 04 parameter sets.[13,70] The Particle Mesh Ewald algorithm was implemented for treatment of long-range interactions as each system was heated from 5 to 300 K (experimental temperature range was 283-303 K) over 50 ps and subsequently maintained at 300 K for 1 ns utilizing the Berendsen temperature coupling scheme.[53,120,122-124] The effects of both isotropic and anisotropic pressure scaling were examined to establish the effect these two pressure models had on the lattice structure, Table 5.1 and Figure 5.2. The deviation from experiment was amplified with the anisotropic model, however the same relative trends in cell distortion were observed with both models. Due to program limitations, anisotropic scaling could not be implemented in the crystals governed by $P2_1$ symmetry. In all cases, data from only the last 500 ps of the 1050 ps MD simulations were utilized to determine the average cell dimensions and hydrogen-bond distances. The deviation in the cell dimensions was found to have established an equilibrium well within that amount of time.[67,120,126]

The lattice energy was computed by taking the difference between the energy contribution from the non-bonded terms (electrostatics and van der Waals) in the crystalline and gas phase, Equation 5.1.

$$E_L = \frac{E_{ELEC,c} + E_{VDW,c}}{N} - \overline{E_{VDW,g}} \qquad [5.1]$$

In order to determine this difference, individual residues were extracted from the crystal and their isolated, unminimized energies were computed. Five evenly spaced snapshots were selected over the course of the last 500 ps of the MD simulation, leading to the use

of 320 structures to obtain the average gas-phase energy for the 64 molecule system. The lattice energy was estimated from the reaction $-\Delta H_c$:

$$(6H_2O_{(g)} + 6CO_{2(g)} \rightarrow C_6H_{12}O_{6(g)} + 6O_{2(g)}) \quad [5.2]$$

and the reaction ΔH_{rxn}:

$$(6H_2O_{(g)} + 6CO_{2(g)} \rightarrow C_6H_{12}O_{6(s)} + 6O_{2(g)}) \quad [5.3]$$

utilizing experimental data, except for the value of ΔH_f for each monosaccharide in the gas phase, since this value cannot be obtained experimentally, Figure 5.3.

$$6CO_{2(g)} + 6H_2O_{(g)} \xrightarrow{\Delta H_{rxn}} C_6H_{12}O_{6(g)} + 6O_{2(g)}$$

$$\uparrow \quad \uparrow \qquad \uparrow \quad \uparrow$$

$$\Delta H=0 \quad \Delta H=0 \qquad \Delta H \neq 0 \quad \Delta H=0$$

$$6CO_{2(g)} + 6H_2O_{(g)} \xrightarrow{-\Delta H_c} C_6H_{12}O_{6(s)} + 6O_{2(g)}$$

$$\Delta H_{rxn} = \Delta H_f(C_6H_{12}O_{6(g)}) + \Delta H_f(6O_{2(g)}) - \Delta H_f(6CO_{2(g)}) - \Delta H_f(6H_2O_{(g)})$$

$$\Delta H_{rxn}(kcal \cdot mol^{-1}) = -260.95(MM3) + 0 - 6(-94.051 \pm 0.031) - 6(-57.7978 \pm 0.0096)$$

$$\Delta H_{rxn} = -650.1428\ kcal \cdot mol^{-1}$$

$$\Delta H_c(C_6H_{12}O_{6(s)}) = -669.17 \pm 0.56\ kcal \cdot mol^{-1}$$

$$E_{LATTICE}(C_6H_{12}O_6) = \Delta H_c(C_6H_{12}O_{6(s)}) - \Delta H_{rxn} = -19.0272\ kcal \cdot mol^{-1}$$

$$\Delta H_f(C_7H_{14}O_{6(g)}) = -254.72\ kcal \cdot mol^{-1}(MM3)$$

$$\Delta H_c(C_7H_{14}O_{6(s)}) = -841.80 \pm 0.20\ kcal \cdot mol^{-1}$$

$$E_{LATTICE}(C_7H_{14}O_6) = -33.5784\ kcal \cdot mol^{-1}$$

Figure 5.3 Algorithm for estimating the lattice energy of α-D-Glc*p* and α-D-Glc*p*-OMe.

Explicitly Solvated MD Simulations Ensemble-averaged charges were calculated with and without partial charges on the aliphatic hydrogen atoms leading to the GLYCAM2000a-LP and GLYCAM04-LP parameter sets. For the MD simulations of the monosaccharides, α-Glc*p*-OMe and α-Gal*p*-OMe, either the GLYCAM2000a or 04 parameters with the TIP3P water model or the GLYCAM2000a-LP or 04-LP parameters with the TIP5P water model were employed. The NPT MD simulations were explicitly solvated with 400-700 water molecules and performed with the SANDER module of

AMBER7.[31,34] First, the system was minimized and then heated from 5 to 300 K over 50 ps and a 1 fs time step was implemented for each LP-containing system, while a 2 fs time step was implemented in the absence of lone-pairs. Initially, the system was equilibrated at 300 K for 2 ns without restraints. Subsequently, umbrella sampling was performed for the O5-C5-C6-O6 ω-angle in α-Glc*p*-OMe and α-Gal*p*-OMe, in which the ω-angle was restrained with a 15 kcal·mol^{-1} harmonic restraint and rotated in 5° increments for the entire 360° rotation. During umbrella sampling (restraints on), a second equilibration period was employed ranging from 200-500 ps, which was followed by the production run. The Weighted Histogram Analysis Method (WHAM[127]) was implemented to determine the free energy change as well as the different population ratios for each ω-angle conformer from the production run.

MD simulations of methyl α-maltoside and methyl α-isomaltoside were performed with three different water models, TIP3P, TIP4P-EW, and TIP5P-EW, in which EW refers to the modification of the original TIP model for use in MD simulations that employ the Ewald summation algorithm is employed for calculation of long-range non-bonded interactions. Methyl α-maltoside was explicitly solvated with 323 water molecules and methyl α-isomaltoside was solvated with 310 water molecules regardless of the water or solute parameters employed. Each molecularly identical system was initiated from the same energy minimized starting structure and was heated from 5 to 300 K over 50 ps. The system was maintained at 300 K for 200 ns during MD simulations of the two disaccharides implementing the GLYCAM04 parameter set and for 100 ns with the GLYCAM04-LP parameter set. The conformational and molecular relaxation properties were analyzed over the full course of each trajectory, while the explicit solute-

solvent interactions were monitored over the first 10 ns of each MD simulation. Previous studies of the solute-solvent interactions have shown that MD simulations of less than 1 ns are long enough in length for convergence of the solute-solvent properties.[128]

The radial distribution functions (RDFs) were determined between individual oxygen atoms in the disaccharide and the oxygen atom in each water molecule of the solvent. Experimentally, the RDFs for mono- and disaccharides have not been determined, with one exception, D-Glc*p*.[129] The classical molecular mechanics force fields investigated here do not model the act of bond breaking and therefore the experimentally observed anomeric mixture of D-Glc*p* cannot be modeled. Therefore, the RDFs obtained via neutron diffraction cannot be quantitatively compared to the results from classical models. Therefore, the experimental studies for D-Glc*p* and water as well as previous modeling studies were qualitatively compared to the results presented here. The area under the peak corresponding to the first solvation shell can be integrated to yield the number of waters in the first hydration shell, η_h. In addition to the number of waters present, the average residency time of these waters was also determined.

The overall rotational motion of the disaccharide is inversely related to T_1 type relaxation, which can be determined via NMR spectroscopy based on the intensity of the relevant signal, *I(t)*, and subsequently converted to the global correlation time (τ_c) through fitting to an exponential decay function, Equation 5.4.

$$I(t) = e^{(-t/\tau_c)} \qquad [5.4]$$

Based on the displacement of an intramolecular vector over the course of time, the magnitude of the vector in one dimension (C(t)) can be monitored utilizing an

autocorrelation function, in which v is the vector chosen and b is the length of this vector, Equation 5.5.

$$C(t) = \frac{< P2(v(0) \bullet v(t)) >}{< (b^3(0) b^3(t)) >}$$
$$P2(v(0) \bullet v(t)) = \frac{1}{2}(3(v(0) \bullet v(t))^2 - 1) \qquad [5.5]$$
$$C(t)\left(\frac{1}{b^6}\right) = I(t)$$

Since τ_c is on the order of 150 ps for the two α-glycans under investigation, in order to minimize the signal to noise ratio the MD simulation should an order of magnitude longer in duration. Each τ_c value reported is the average τ_c value determined over the course of the entire MD simulation for each 100 ps increment.

Analysis of Hydrogen Bonded Gas-Phase Clusters

In order to assess the accuracy of the lone-pair model, the geometry of several neutral hydrogen-bonded clusters, described in detail within Chapter 4, were examined *in vacuo*. Each cluster contains at least one water molecule, which, if no LPs were present, was modeled as TIP3P, or, when the lone-pair model was employed, as TIP5P. The B3LYP/6-31+G(d) determined geometry was used as the starting point for each force field energy minimization.[115] Two root mean squared deviations (RMSDs) were determined from the energy minimized molecular mechanics structure to the quantum structure, one for all of the atoms and another for only the non-hydrogen atoms. Overall, for the 21 clusters the RMSD for the standard non-LP containing model and the new LP model was 0.58 Å (0.30 Å) and 0.49 Å (0.21 Å), respectively, computed for all atoms with the non-hydrogen atom RMSD shown in parenthesis.

At the MP2/aug-cc-pVQZ//MP2-aug-cc-pVTZ level, the interaction energy of the dimer in which water acted as the hydrogen bond donor, WdM, was -5.72 kcal·mol^{-1},

which was more favorable than the -4.95 kcal·mol^{-1} computed for the dimer in which methanol was the hydrogen bond donor, MdW. While, at the HF/6-31G(d) level, which is the same level at which the electrostatics were determined, no difference between the energies of the two different configurations was obtained.[84] The relative energetic order of the two dimers, MdW and WdM, was reversed with the TIP3P/no-LP and TIP5P/LP models, but the energetic difference was decreased from 0.58 kcal·mol^{-1} to 0.32 kcal·mol^{-1} when LPs were added to the MM model, Figure 5.4. Despite the incorrect energetic ranking, the resultant energies of the LP model were much closer to the quantum mechanical energies than if LPs were not included.

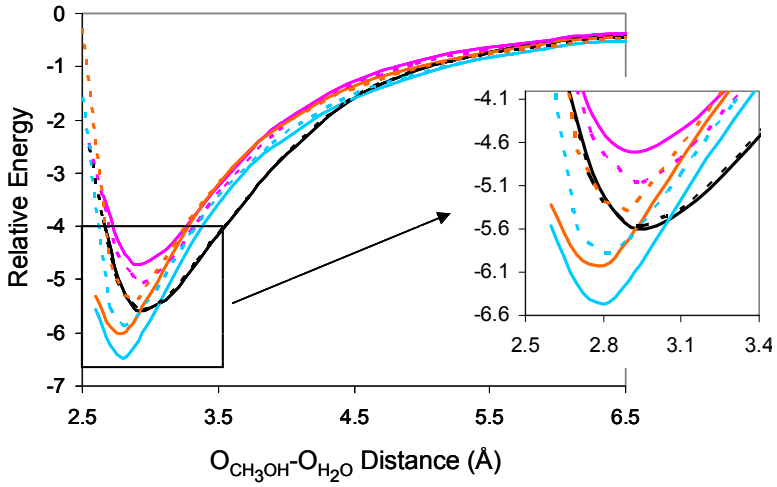


Figure 5.4 Approach of methanol to water, (MdW – solid, WdM - dashed) at the HF/6-31G(d)//HF/6-31G(d) (black) level, B3LYP/6-31++G(2d,2p)//HF/6-31G(d) (magenta) level, TIP3P water model and a methanol model without lone-pairs (blue), and the TIP5P water model with the LP model for methanol (red).[84]

The B3LYP structure of the MdW minimum was reproduced equally well in both models, with a non-hydrogen atom RMSD (with the all-atom RMSD in parenthesis) of .09 Å (0.20 Å) and 0.08 Å (0.15 Å) for the TIP5P/LP model and the TIP3P/no-LP model, respectively. In contrast, when water acted as the donor to methanol, there was a

substantial difference in the all-atom RMSDs, which is clearly illustrated in Figure 5.5. The all-atom RMSDs increased from 0.15 Å to 0.29 Å when the LPs were removed, while the non-hydrogen atom RMSD was again quite similar at 0.05 Å and 0.06 Å. Further work on the inclusion of a van der Waals term on the hydroxyl hydrogen may be necessary to reproduce higher-level quantum results. For example, inclusion of a van der Waals parameter on the hydrogen atoms increased the ability of the TIP3P water model to reproduce bulk water properties when the particle mesh ewald algorithm was employed.[130]

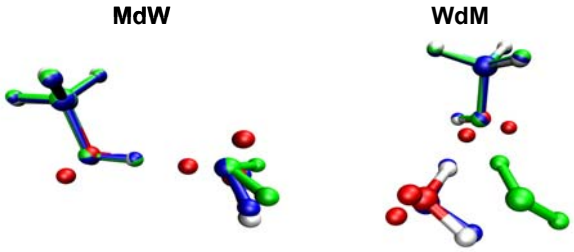


Figure 5.5 Fully optimized local minima for the methanol-water complexes at the B3LYP/6-31+G(d,p) level (colored by atom type), TIP3P and the methanol model without LPs (green), and TIP5P with the LP model of methanol (blue, with LPs in red).

Reproduction of the relative energies associated with the approach of a water molecule along each of the lone-pair axis in NMA would ensure that inclusion of asymmetric lone-pairs degrades neither the geometry nor the energetics of the hydrogen-bonded clusters. The approach along the axis on the same side of the molecule as the methyl group, 120C, was energetically more favorable by 1.3 kcal·mol^{-1} at the B3LYP/6-31+G(d,p) level. Starting from the quantum mechanically determined structures, energy minimizations utilizing the non-LP model with TIP3P reversed the quantum mechanically determined ranking, while inclusion of lone-pairs with the TIP5P water model reproduced the energetic favoring of 120C, Figure 5.6. For the optimized local

minima without restraints, the LP inclusive model was closer to the 120C QM derived structure with an RMSD of 0.25 Å (0.55 Å) compared to 0.45 Å (0.57 Å), for the LP and non-LP model, respectively, Figure 5.7. In contrast, both models yield similar RMSDs to the QM determined structure of 120N of 0.16 Å (0.47 Å) and 0.15 Å (0.47 Å), for the LP and non-LP model, respectively.

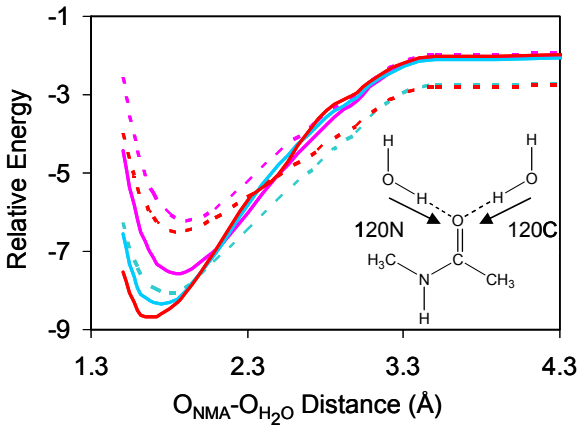


Figure 5.6 Approach of water to NMA, in which water can approach from the nitrogen containing side of the molecule (120N, dashed line) and from the side containing the methyl group (120C, solid line) determined at the B3LYP/6-31+G(d,p) (magenta) level and utilizing the non-LP/TIP3P (blue) and LP/TIP5P (red) models.

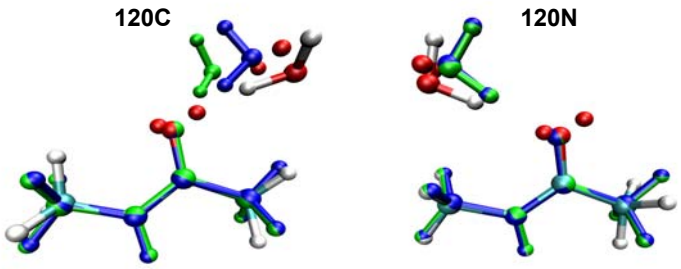


Figure 5.7 Local minima for the NMA-water complexes at the B3LYP/6-31+G(d,p) level (atom colored), TIP3P and the NMA model without LPs (green), and TIP5P with the LP model of NMA (blue, with the LPs shown in red).

MD Simulations of Monosaccharide Crystals

MD simulations of monosaccharide crystals provide a sensitive method for testing non-bonded parameters in condensed phase and have been performed previously for carbohydrates.[67,126] If the non-bonded parameters are too attractive it leads to a more tightly packed crystal, while if the interactions are underestimated, the crystal cell will expand. In addition to the quantitative information related to the structure of the unit cells, the energy associated with the crystal packing is also relevant and was monitored by computing the lattice energies. A combination of theoretical and experimental data were utilized to estimate the lattice energy, and as such this value provides qualitative insight. The addition of lone-pairs is purely a variation in the electrostatic model and monitoring the behavior of cell lengths over the course of MD simulations, for both the non-LP and LP models, provides a sound test for the effects of lone-pair inclusion.

Initially, for each of three methyl glycosides, α-D-Glc*p*-OMe, β-D-Gal*p*-OMe, and α-D-Man*p*-OMe, a 64 and 256 molecular unit system was generated. Two different size systems were employed with the GLYCAM04 parameter set to determine if the smaller system size introduced any artifacts in the distortion of the cell dimensions. The average difference in cell lengths for the three methyl-glycosides in the large (256 molecules) cell relative to the 64 molecule cell was 0.01%. Non-bonded cut-offs of 8, 9, and 10 Å were investigated while employing the particle mesh ewald algorithm in the 256 molecule cells, yielding less than a 0.02 and 0.03% difference for the 9 and 10 Å cut-off, respectively, relative to the 8 Å cut-off for the three methyl-glycosides. A time-step of 0.5 fs was implemented as a standard to determine if longer time-steps would be appropriate. Due to the negligible difference in the results obtained with a time step of

1 fs relative to the 0.5 fs initial time-step, the larger time-step was employed along with an 8 Å cut-off and the smaller lattice size.

The majority of the improvement observed from switching from GLYCAM2000a to GLYCAM04 and 04-LP can be attributed to the modification of the non-bonded parameters for the hydroxyl oxygen atom. The van der Waals' radius in GLYCAM2000a, 1.961 Å, is larger than the OPLS value of 1.7210 Å, which is also implemented in the Cornell *et al.* and GLYCAM04 and 04-LP parameter sets.[13,131] The larger van der Waals' radius, as well as the smaller well depth (0.14 versus 0.21 x 10^{-3} kcal·mol^{-1}), both contributed to the expansion observed in GLYCAM2000a versus 04 and 04-LP. The different partial charge arrangement in 04 versus 04-LP would be expected to have a relatively small effect on the overall change in cell dimensions since each charge set was derived from the same ensemble of conformers, derived from the same quantum mechanical MEP, and implemented identical RESP weightings of 0.01. In order to observe the sensitivity purely due to electrostatics, the OPLS van der Waals parameters for the hydroxyl oxygen atom were implemented in GLYCAM2000a to yield GLYCAM2000b, Table 5.2. Again, GLYCAM04-LP yields the smallest average cell deviation of the GLYCAM parameter sets as well as that reported for the GROMOS and HGFB parameter sets, Table 5.2.[126]

Table 5.2 The effect of different parameter sets and pressure scaling on the ability to reproduce the experimental cell dimensions and lattice energies for α-Glc*p*.

Force Field	T (K)	P[a]	Geometry							Energetics					time
			A (Å)	B (Å)	C (Å)	ΔA (%)	ΔB (%)	ΔC (%)	Mean |Δ| (%)	$\langle E_{ELEC(s)} \rangle$	$\langle E_{VDW(s)} \rangle$	$\langle E_{ELEC(g)} \rangle$	$\langle E_{VDW(g)} \rangle$	$E_{LATTICE}$	(hrs)
Exp.	283-303		20.372	29.701	19.901									-19	
2000a	298	I	22.14±0.03	32.28±0.05	21.63±0.03				8.69	-6477±33	-917±13	-89±5	-0.7±0.8	-26	3.8
		A	21.25±0.04	29.86±0.05	23.73±0.03	4.32	0.53	19.22	8.02	-6604±24	-839±13	-84±2	-0.8±0.5	-32	3.8
2000b	298	I	20.68±0.01	30.16±0.01	20.21±0.01				1.54	-7676±25	-917±13	-85±2	-1.3±0.3	-48	4.1
		A	20.74±0.02	29.70±0.02	20.57±0.02	1.80	-0.01	3.35	1.72	-7706±24	-910±13	-85±2	-1.3±0.3	-48	4.1
04	298	I	20.65±0.01	30.10±0.01	20.17±0.01				1.36	-9637±26	-887±13	-112±3	-1.4±0.3	-51	4.1
		A	21.04±0.02	30.14±0.02	19.65±0.02	3.28	1.49	-1.27	2.01	-9629±26	-901±13	-113±3	-1.4±0.3	-51	4.2
04LP	298	I	20.57±0.01	29.99±0.01	20.10±0.01				0.98	-5305±18	-887±14	-45±2	-1.5±0.3	-50	7.2
		A	20.80±0.02	30.03±0.03	19.76±0.02	2.09	1.11	-0.71	1.30	-5310±19	-897±15	-45±2	-1.6±0.3	-50	7.4
GROMOS[126]	293.8	A				-0.8	-3.0	-0.9	1.57						
HGFB[126]	290.3	A				-3.9	0.8	-0.6	1.77						

[a] Prefers to the pressure scaling, in which I stands for isotropic and A represents anisotropic scaling.

Table 5.3 Structural properties in the isotropic simulations of α-D-Glc*p*···H_2O.

Force Field	Exp.	GLYCAM2000a		
Water model		TIP3P	TIP4P-EW	TIP5P-EW
A (Å)	17.61	18.90±0.03	18.76±0.02	18.73±0.02
B (Å)	20.34	21.83±0.04	21.68±0.02	21.64±0.02
C (Å)	19.42	20.84±0.04	20.69±0.02	20.66±0.02
ΔA,B, and C (%)		7.33	6.57	6.39
Mean |Δ| HB_{dist} (Å)		0.49±0.18	0.36±0.14	4.17±0.73
Mean |Δ| HB_{dist} (%)		23.68	16.97	203.46
Mean |Δ| HB_{angle} (°)		72±34	15±17	29±25
Mean |Δ| HB_{angle} (%)		44.25	9.54	16.50
Force Field	Exp.	GLYCAM2000b		
Water model		TIP3P	TIP4P-EW	TIP5P-EW
A (Å)	17.61	17.77±0.01	17.74±0.01	17.66±0.01
B (Å)	20.34	20.54±0.01	20.49±0.01	20.40±0.01
C (Å)	19.42	19.60±0.01	19.56±0.01	19.47±0.01
ΔA,B, and C (%)		0.96	0.75	0.29
Mean |Δ| HB_{dist} (Å)		0.17±0.09	0.19±0.06	0.20±0.11
Mean |Δ| HB_{dist} (%)		7.78	8.38	9.17
Mean |Δ| HB_{angle} (°)		8±12	6±10	7±12
Mean |Δ| HB_{angle} (%)		4.70	3.78	4.68
Force Field	Exp.	GLYCAM04		
Water model		TIP3P	TIP4P-EW	TIP5P-EW
A (Å)	17.61	17.75±0.01	17.72±0.01	17.62±0.01
B (Å)	20.34	20.50±0.01	20.48±0.01	20.36±0.01
C (Å)	19.42	19.57±0.01	19.55±0.01	19.44±0.01
ΔA,B, and C (%)		0.81	0.67	0.11
Mean |Δ| HB_{dist} (Å)		0.18±0.06	0.17±0.07	0.16±0.08
Mean |Δ| HB_{dist} (%)		8.41	8.00	7.55
Mean |Δ| HB_{angle} (°)		7±11	6±10	6±10
Mean |Δ| HB_{angle} (%)		4.42	3.81	3.56
Force Field	Exp.	GLYCAM04-LP		
Water model		TIP3P	TIP4P-EW	TIP5P-EW
A (Å)	17.61	17.62±0.01	17.67±0.01	17.61±0.01
B (Å)	20.34	20.36±0.01	20.42±0.01	20.35±0.01
C (Å)	19.42	19.44±0.01	19.49±0.01	19.42±0.01
ΔA,B, and C (%)		0.11	0.39	0.02
Mean |Δ| HB_{dist} (Å)		0.19±0.07	0.17±0.09	0.16±0.09
Mean |Δ| HB_{dist} (%)		8.68	8.06	7.53
Mean |Δ| HB_{angle} (°)		7±12	7±11	7±11
Mean |Δ| HB_{angle} (%)		4.60	4.13	4.56

Previous studies have examined the ability of carbohydrate force fields to reproduce solvent-solute properties, revealing that GLYCAM2000a underestimates the hydrogen bonding interaction between the hydroxyl groups in the pyranoside and the TIP3P water model.[118] Simulation of the monohydrate glycoside crystal structure

required the inclusion of a water molecule, which allowed the direct assessment of the compatibility between the carbohydrate parameter set and the chosen solvent model. Here, the TIP series of models were implemented, with the smallest deviation from experiment observed when GLYCAM04-LP was utilized with the TIP5P water model, Table 5.3. Notably, GLYCAM2000b and GLYCAM04 yield the smallest distortion of the cell dimensions when the TIP5P water model is implemented, with TIP3P being the worst, although still in good agreement with the experimental values. Therefore, it is clear that under these conditions, TIP5P performs better than TIP3P and TIP4P, but also that the inclusion of lone-pairs in the carbohydrate force field results in a substantial improvement in the model as well. Similar results were obtained for hydrogen bond distances and angles in all models, but were not a focal point of the study of this crystal since the structure was not obtained by neutron diffraction.

In order to assess the performance of the asymmetric LP model, the α-D-GlcNAc*p* crystal was simulated with the GLYCAM04 and 04-LP parameter sets. The deviation in cell dimensions was nearly halved from 0.89 Å to 0.54 Å with the inclusion of the asymmetric LPs. Initially, the HNCO torsion angle deviated from the experimental geometry to adopt a more planar structure with both parameter sets. The weak hydrogen bond formed between the amide side chains of two neighboring, stacked pyranosides weakened from the experimental value of 2.1 Å to 2.4 Å and 2.8 Å in GLYCAM04 and 04-LP, respectively, as the HNCO angle rotated from 190° to 184° and 183°, for 04 and 04-LP, respectively. Since the goal of the crystal simulations was to assess non-bonded interactions, a small restraint was added to the HNCO angle to maintain the internal geometry of the experimental crystal structure. Inclusion of this

restraint resulted in reproduction of the two weak interactions between the amide groups. The small improvement in the HNCO angle led to a significant improvement in the intermolecular hydrogen bond angle between the two amide groups as well, from an underestimation of over 30°, to an overestimation of less than 6° for both parameter sets, Table 5.4.

Table 5.4 The effect of asymmetric LPs on the hydrogen bonding in α-D-GlcNAc*p*$_{(s)}$.

Force Field	Exp.	Without Restraints		Restraint on HNCO Angle[a]	
		GLYCAM04	GLYCAM04-LP	GLYCAM04	GLYCAM04-LP
A (Å)	46.29	46.70±0.01	46.54±0.01	46.70±0.01	46.53±0.01
B (Å)	19.58	19.75±0.01	1968.±0.01	19.75±0.01	1968.±0.01
C (Å)	38.96	39.30±0.01	38.96±0.01	39.30±0.01	38.16±0.01
ΔA,B, and C (%)		0.89±0.01	0.54±0.01	0.87±0.01	0.52±0.01
HNCO Angle (°)	190	184±10	183±11	188±7	187±7
H1O···O5 (Å)	1.82	1.80±0.12	1.77±0.12	2.39±0.25	2.80±0.37
H3O···O3 (Å)	1.90	2.06±0.25	2.04±0.21	2.02±0.20	1.84±0.19
H6O···O4 (Å)	1.88	1.88±0.13	1.82±0.12	1.88±0.13	1.82±0.12
H4O···O2N (Å)	2.07	2.04±0.21	1.84±0.21	2.06±0.26	2.04±0.21
H2N···O2N (Å)	2.10	2.42±0.25	2.83±0.37	1.80±0.11	1.77±0.12
O1-H1O···O5 (°)	168	163±9	164±10	127±8	113±11
O3-H3O···O3 (°)	150	146±15	148±10	155±9	156±10
O6-H6O···O4 (°)	172	165±8	165±8	165±78	165±8
O4-H4O···O2N (°)	163	155±10	156±10	145±15	148±10
O2N-H2N···O2N (°)	158	126±9	112±11	163±8	164±10

[a] Harmonic restraint of 15 kcal·mol^{-1} was employed.

Experimental enthalpies of combustion, $\Delta H_{combustion}$, are available for α-D-Glc*p* and α-D-Glc*p*-OMe, while ΔH_f was estimated for these two compounds using MM3 as implemented in Sybyl.[132] A smaller difference between the estimated and computed lattice energies was observed for the methyl glycoside relative to α-D-Glc*p*, which may have been the result of the chemically more complex nature of α-D-Glc*p* relative to α-D-Glc*p*–OMe, which could effect both the experimental and MM3 value. Specifically, if the difference between the lattice energy predicted by the GLYCAM04/04-LP parameter sets and the theoretically estimated value was only attributed to the error in the

MM3 computed value for ΔH_f, then the error in the in the MM3 value would be 11% for α-D-Glc*p* and only 4% for the methyl glycoside.

Typically, a purely theoretical or experimental value is preferred over a combination of the two due to subtraction of errors. Utilizing a purely theoretical approach, the lattice energy of α-D-Glc*p* was approximated to be -37 kcal·mol^{-1}.[133] A minicrystal of α-D-Glc*p* was created by surrounding a central α-D-Glc*p* residue with several additional α-D-Glc*p* residues mimicking a crystalline lattice.[133] The MM3 energy was then obtained for the whole lattice and subsequently subtracted from the isolated central molecule (gas phase) and the remaining miniature crystal without the central residue to yield the lattice energy.[133] The value is nearly twice the estimated theoretical value and much closer to the lattice energies predicted with the GLYCAM parameter sets.[133] Therefore, it is not clear which value is more accurate, but only that all models predict lattice energies on the same order of magnitude.

Although the lattice energies are similar for these three GLYCAM parameter sets, the electrostatic energies contributing to these values are very different. In general, GLYCAM04 has the most favorable inter- and intramolecular electrostatics in the solid state at -151 kcal·mol^{-1}, followed by GLYCAM2000b (-120 kcal·mol^{-1}), and then GLYCAM04-LP (-83 kcal·mol^{-1}) per glucose residue and the percentage of the electrostatics in the solid phase that are due to intramolecular interactions versus intra- and intermolecular electrostatics follows the same trend. Both GLYCAM04 and GLYCAM2000b are in the same range at 74% and 71%, respectively, but the 04-LP value has substantially decreased to 54% for α-D-Glc*p*. These differences are cancelled out because the magnitude of the van der Waals contribution is similar for all models

since the same van der Waals parameters were implemented in each of these systems. The greater non-bonded electrostatic energy in the GLYCAM04-LP model relative to the non-LP models is consistent with the greater interaction energy of both methanol – water complexes, MdW by 0.4 kcal·mol^{-1} and WdM by 0.7 kcal·mol^{-1}, for the LP/TIP5P model. Therefore, despite the resultant similarity in the lattice energy, the role a small modification in the electrostatic model can have, despite being derived from the same quantum mechanical MEP, is affirmed, Table 5.5.

Table 5.5 Comparison of the estimated lattice energy and its average energetic contributions (kcal·mol^{-1}) for different force fields to the theoretically predicted value for α-D-Glc*p* and α-D-Glc*p*-OMe.

		α-D-Glc*p*					α-D-Glc*p*-OMe				
	P[a]	$E_{ELEC(s)}$	$E_{VDW(s)}$	$E_{ELEC(g)}$	$E_{VDW(g)}$	lattice energy[b]	$E_{ELEC(s)}$	$E_{VDW(s)}$	$E_{ELEC(g)}$	$E_{VDW(g)}$	lattice energy[b]
Exp.[a]						-19.02					-34
GLYCAM	I	-6477±33	-917±13	-89±5	-0.7±0.8	-26±1.1	-6245±25	-1165±11	-88±3	-1.8±0.6	-26±0.3
2000a	A	-6604±24	-839±13	-84±2	-0.8±0.5	-32±0.6	-6299±21	-1204±10	-88±3	-1.3±0.5	-27±0.4
GLYCAM	I	-7676±25	-917±13	-85±2	-1.3±0.3	-48±0.2					
2000b	A	-7706±24	-910±13	-85±2	-1.3±0.3	-48±0.2					
GLYCAM	I	-9637±26	-887±13	-112±3	-1.4±0.3	-51±0.2	-8914±26	-1191±17	-114±3	-2.1±0.4	-42±0.2
04	A	-9629±26	-901±13	-113±3	-1.4±0.3	-51±0.2	-8872±25	-1227±10	-114±3	-2.1±0.5	-42±0.2
GLYCAM	I	-5305±18	-887±14	-45±2	-1.5±0.3	-50±0.2	-4515±21	-1137±17	-43±2	-2.3±0.4	-44±0.2
04-LP	A	-5310±19	-897±15	-45±2	-1.6±0.3	-50±0.2	-4466±21	-1186±19	-43±2	-2.4±0.3	-44±0.1
MM3[132]						-37.4[133]					

[a] Prefers to the pressure scaling, in which I and A stand for isotropic and anisotropic scaling.
[b] lattice energy = $\Delta_f H$ [$C_6H_{12}O_6$(g)] – 6 ($\Delta_f H$ [H_2O(g)]) – 6 ($\Delta_f H$ [CO_2(g)]) + $\Delta_c H$ [$C_6H_{12}O_6$(s)] Using the MM3 value of -260.95 kcal·mol^{-1} for $\Delta_f H$ [$C_6H_{12}O_6$(g)] and -254.72 kcal·mol^{-1} for $\Delta_f H$ [$C_7H_{14}O_6$(g)] with the experimentally determined values for $\Delta_f H$ [CO_2], $\Delta_f H$ [H_2O], $\Delta_c H$ [$C_6H_{12}O_6$(s)], and $\Delta_c H$ [$C_7H_{14}O_6$(s)], -94.051, -57.7978, -669.1428, and -841.80 kcal·mol^{-1}, respectively.[134]

The partial charges for all three of these force fields were determined from the same quantum mechanical MEP, which enables the origin of the differences in lattice energy to be isolated, Table 5.6. The difference between GLYCAM2000b and GLYCAM04 is directly related to the non-zero charge on the aliphatic hydrogen atoms in

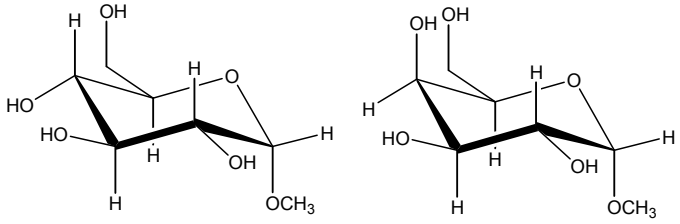
the earlier GLYCAM version, which, in general, led to smaller partial charges on the oxygen atoms in GLYCAM2000b relative to GLYCAM04. The difference between GLYCAM04 and GLYCAM04-LP is only related to the inclusion of LPs. The 1,4 interactions that the oxygen atom is involved in are transferred to each lone-pair so the GLYCAM04 and GLYCAM04-LP systems have the same 1,4 non-bonded interactions. Only the decrease in the magnitude of the partial charge on the oxygen atom in GLYCAM04 relative to the sum of the charges of the lone-pairs attached to this oxygen atom and the difference in distance between a given atom and the oxygen atom versus each lone-pair remains as the possibility for the difference in the electrostatic energy. For example, for O6, GLYCAM04 has a partial charge of -0.682 *e*, while GLYCAM04-LP has a partial charge of -0.210 *e* on each LP, yielding a sum of -0.420 *e*, which is less than two-thirds of the magnitude.

Table 5.6 The internal bond, angle, and dihedral energy as well as the magnitude of the internal 1,4 non-bonded interactions for α-Glc*p* (kcal·mol^{-1}).

GLYCAM VERSION		$E_{INTERNAL}$				1,4 non-bonded		
		$\langle E_{BOND(s)} \rangle$	$\langle E_{ANGLE(s)} \rangle$	$\langle E_{DIHEDRAL(g)} \rangle$	$\langle E_{INT} \rangle$	$\langle E_{1,4VDW(s)} \rangle$	$\langle E_{1,4EELEC(s)} \rangle$	E_{TOT}
2000a	I	491±19	844±21	25±9	1360	1106±17	9825±31	6038±9
	A	506±20	824±20	50±6	1381	1109±17	9730±23	5916±4
2000b	I	362±15	704±18	84±5	1149	555±11	9906±26	4154±1
	A	359±14	700±18	79±5	1138	553±11	9926±23	4138±2
04	I	423±17	933±22	133±5	1490	489±11	14540±31	7137±1
	A	423±16	929±21	137±5	1489	485±10	14536±29	7116±2
04-LP	I	337±15	804±18	127±6	1268	542±11	5531±16	2290±2
	A	337±15	806±20	127±6	1270	539±11	5536±16	2272±2

Conformational Properties of Monosaccharides

The intricate balance of intermolecular hydrogen bonding interactions between the solvent and solute and the intramolecular hydrogen bonds within the monosaccharide is necessary to reproduce the experimentally observable populations for the ω-angle. The ω-angle, which is defined as the rotation around the C5-C6 bond in a pyranoside, is sensitive to the configuration of the hydroxyl group at C4, and is therefore substantially different for glucose and galactose, Figure 5.8 and Table 5.7. The potential of mean force (PMF) calculation ensures full conformational sampling for each of the GLYCAM parameter sets. Since all charge sets were obtained via fitting to the same MEP, the same parameters may be utilized for the non-LP and LP models, although further improvement can be obtained via refitting the parameters with the inclusion of lone-pairs.



I **II**

Figure 5.8. Structure of α-Glc*p*-OMe (**I**) and α-Gal*p*-OMe (**II**).

The equilibration and simulation time was examined for all MD simulations to ensure the analysis was performed at a sufficiently equilibrated point with enough sampling. The availability of population data from 50 ns (GLYCAM2000a) and 100 ns (GLYCAM04) MD simulations of both monosaccharides investigated here will also ensure the correct restraints are implemented during the PMF calculation, as in the limit both of these methods should converge to the same answer. The estimated error in the

experimental calculation of the scalar couplings is approximately 0.2-0.3 Hz, leading to ±10% error in the derived rotamer populations.[135]

Table 5.7 The ω-angle populations for D-Glc*p* and D-Gal*p*.

		D-Glc*p*			D-Gal*p*		
Force Field	Method	*tg*	*gt*	*Gg*	*tg*	*gt*	*gg*
Exp.	[9,10]	7	41	52	32	56	12
	[136]	2	45	53	21	61	18
PHLB [137]	PMF	0	15	85	8	91	19
	MD (10 ns)	-	-	-	8	91	1
CSFF [137]	PMF	1	33	66	21	75	4
	MD (10 ns)	0	31	69	14	80	6
OPLS-AA [30]	MD (10 ns)	4	51	45	46	52	2
	MD (10 ns)	11	77	12	33	64	3
OPLS-AA-SEI[a] [30]	MD (10 ns)	4	29	67	38	53	9
	MD (10 ns)	4	27	69	61	35	4
GLYCAM2000a[b] [13,138]	MD (50 ns)	6	40	54	28	64	8
GLYCAM2000a[b]	PMF	8	44	48	38	57	5
GLYCAM2000a-LP[b,c] [138]	PMF	5	49	46	28	65	7
GLYCAM04[c] [70]	MD (100 ns)	6	45	49	27	62	11
GLYCAM04[c,d]	PMF	6	44	50	32	63	5
GLYCAM04-LP[b,c,d]	PMF	4	47	49	20	71	9

[a] Simulations initiated with different starting conformations for the ω-angle.

[b] Explicitly solvated with TIP5P, while all others utilize TIP3P.[31,34]

[c] The methyl glycoside was utilized.

[d] Average value taken from 9 data sets which differed in equilibration and analysis times for the PMF calculation.

For α-Glc*p*-OMe, the population ratios for the both the 50 and 100 ns MD simulations are quite close to that of the potential of mean force calculations, while those for α-Gal*p*-OMe are farther away, Table 5.7. This is consistent with both the PMF data as well as the MD data, which show that the convergence of the rotamer populations for the angle is slower for α-Gal*p*-OMe than α-Glc*p*-OMe.[13] For the α-Glc*p*-OMe system, all results computed with the GLYCAM parameter sets led to values within experimental error and have no substantial differences between them. Focusing on the PMF results for

the GLYCAM2000a and 2000a-LP parameter sets, the addition of lone-pairs to α-Gal*p*-OMe resulted in closer agreement with NMR data, but both were within experimental error. The results for the 50 ns GLYCAM2000a MD simulation and the PMF computation were substantially different with respect to the *tg* conformation, while the 100 ns GLYCAM04 MD simulation and PMF computation were in better agreement. As a result, the 50 ns MD simulation was most likely not statistically converged.

A decrease in the *tg* population coupled with an increase in both *gt* and *gg* rotamer populations was observed upon addition of LPs during the PMF computations of α-Gal*p*-OMe with the 2000a to 2000a-LP and 04 to 04-LP parameter sets. The largest population difference is in the *tg* rotamer, but the experimental values also have a wide berth. The increase in the *gt* rotamer upon addition of LPs yields a population near the edge of the experimental range, but the increase in the *gg* rotamer upon addition of LPs brings the computed population ratio in closer agreement with the experimental data. Therefore, the conformational properties investigated here are not adversely affected by the addition of lone-pairs, although the magnitude of the improvement can not be quantified. Application of this type of population analysis to larger systems not considered during parameter derivation will fully examine the benefits of including lone-pairs in the carbohydrate parameter set.

Solution Properties of Disaccharides

Solute-Solvent Interactions Previously, the RDFs computed with GLYCAM2000a, were shown to disagree with other carbohydrate force fields that possessed a more structured solvent shell around isomaltoside, Figure 5.9.[118] In contrast, both GLYCAM04 and GLYCAM04-LP revealed similar results to those obtained with

the CHARMM carbohydrate force fields, HGFB and CSFF, which have been known to accurately model solute-solvent interactions. The more structured solvent around the O3 and O3' oxygen atoms of isomaltoside observed with the GLYCAM04 and 04-LP parameters, relative to GLYCAM2000a, was due to the introduction of the OPLS-AA van der Waals parameters for the hydroxyl oxygen atom, Figure 5.10. Significantly different RDFs were observed for different oxygen atoms in the disaccharide, which can be divided into two categories: those possessing either 1)A structured first solvation shell, g(OO) > 1 or 2)A less ordered first hydration shell, g(OO) < 1, such as O5 and O1. These two groups are consistent with the average hydration numbers computed for each oxygen atom, with larger hydration numbers associated with oxygen atoms that have a more structured solvent shell, Table 5.8. Therefore, the overall solvation of isomaltoside was best described as anisotropic in which the local chemical environment of each functional group of the disaccharide substantially affects the local water structure, which is consistent with previous studies on a variety of monosaccharides and generally accepted for amphiphilic molecules.[117,128,139,140]

Table 5.8 Average residency time[a] for each oxygen atom in methyl α-isomaltoside in ps and hydration numbers (η_H) for methyl α-isomaltoside.

	O1'	O2'	O3'	O4'	O5'	O1	O2	O3	O4	O5	O6	η_H 2.8 Å	η_H 3.4 Å
GLYCAM04													
TIP3P	1.8	3.0	2.8	3.5	3.0	2.2	3.2	2.8	3.0	2.4	2.6	14.7	25.0
TIP4P-EW	1.8	3.0	2.8	3.5	3.0	2.2	3.2	2.8	3.0	2.4	2.6	13.9	25.5
TIP5P-EW	2.1	3.9	3.7	4.0	2.6	2.4	4.1	3.7	3.8	2.6	3.7	18.8	29.8
GLYCAM04-LP													
TIP3P	1.9	3.9	3.4	3.6	3.8	2.1	4.2	3.4	3.5	2.6	4.0	15.4	26.7
TIP4P-EW	2.0	5.2	4.3	4.8	5.6	4.0	5.7	4.4	4.3	2.8	5.6	14.3	25.0
TIP5P-EW	2.2	4.3	3.9	4.1	3.0	2.7	4.3	3.9	3.8	2.6	3.8	19.0	29.4
GLYCAM2000a													
TIP3P	19.7	7.6	7.6	7.6	7.7	7.6	7.6	7.6	6.7	6.0	6.3	9.6	22.7

[a] A water molecule is considered to be resident if it is within 3.4 Å of an oxygen atom in the solute.

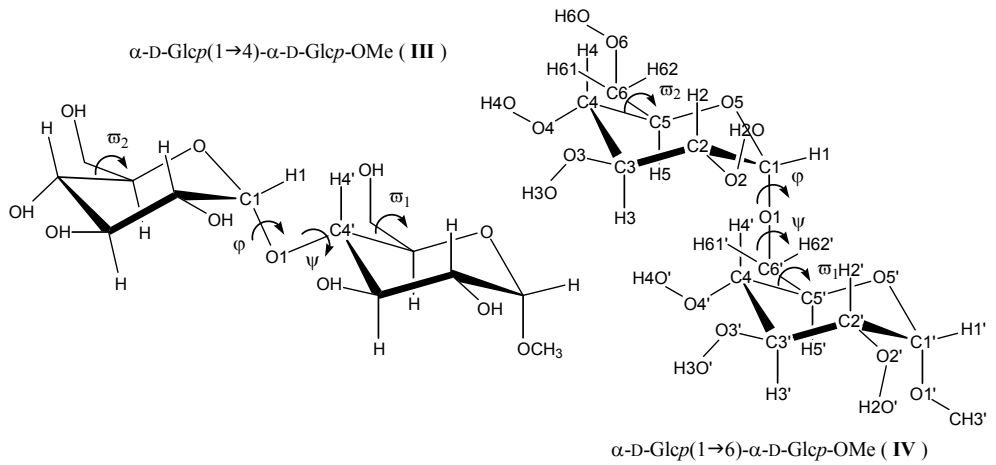


Figure 5.9 Structure and nomenclature for maltoside (**III**) and isomaltoside (**IV**).

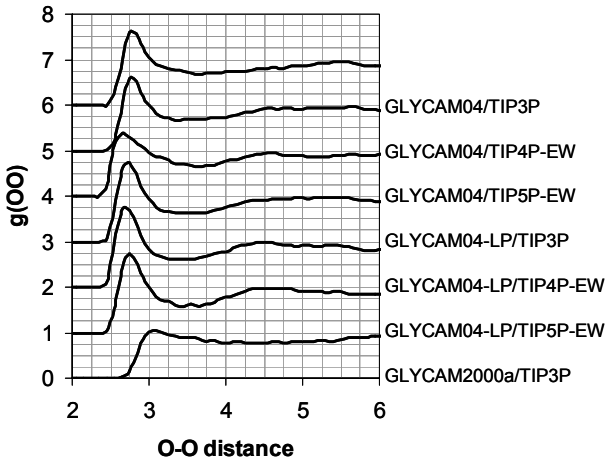


Figure 5.10 RDFs for O3···O_{WAT} (Å) in methyl isomaltoside. Note that all structure factors, g(XX), are per water molecule and here each model is offset by one unit.

The first peak corresponding to the g(OO) component of the 1 molal glucose solution was observed experimentally at ~2.8 Å and was primarily comprised of the bulk water interaction, $g(O_{H_2O}O_{H_2O})$, which is at ~2.8 Å, and a minor percentage of the interactions involving the sugar, $g(O_{H_2O}O_{Glcp})$ and $g(O_{Glcp}O_{Glcp})$.[119,129,141] In general, two model combinations resulted in an underestimation of the first solvation peak for the possible O_{H_2O}-O_{Glcp} interactions at 2.7 Å; the non-LP model with TIP5P-EW and the LP model with TIP4P-EW. Neither of these combinations was proposed as ideal, but

included to illustrate the ideology that not all additive water models can be interchanged with different additive biomolecular parameter sets. In general, the non-LP model and TIP3P or TIP4P-EW yielded first solvation shell peaks with the highest magnitude for the O2' and O2 oxygen atom types, with an average g(OO) of 1.79. A slightly more intense peak with an average g(OO) of 1.84 was observed for the same oxygen atoms with GLYCAM04-LP and TIP3P and TIP5P-EW, but the most intense peak corresponded to the interaction between O6 and O_{H_2O}, with the highest observed g(OO) values of 1.91 and 1.81, respectively with the 04-LP/TIP3P and 04-LP/TIP5P models.

Therefore, the GLYCAM04-LP/TIP5P-EW parameter set yields a more ordered first solvation shell relative to the non-LP GLYCAM models and is the most consistent with the non-GLYCAM parameter sets investigated by Corzana *et al.*[118] The models yielding a solvation peak less than that of bulk water are inconsistent with the linear entropy/enthalpy correlation as described in the *Lemieux effect*, in which the movement of the water molecules away from the hydration layer of the sugar into the bulk solvent is entropically unfavorable, but enthalpically favorable.[117] Therefore, generally speaking, in order to maintain the correct relative enthalpy, the water – (sugar hydroxyl) hydrogen bond should be similar or slightly elongated relative to the water – water interaction. Therefore, the models revealing a maximum in the first solvation shell lower than that of water are considered to inaccurately model these solvation effects. In addition, if each hydroxyl type oxygen atom behaves as a donor and acceptor and each ether type oxygen atom behaves as an acceptor, isomaltoside can form 18 strong hydrogen bonds and possibly a few additional interactions in which the oxygen atom can behave as a double acceptor. In general, with the exception of O6 in the 04/TIP3P and 04/TIP4P-EW

models, the ether oxygen atoms possess less structured solvation shells and lower residency times relative to the hydroxyl oxygen atoms regardless of the model employed. Both the GLYCAM04/TIP5P-EW and GLYCAM04-LP/TIP5P-EW models yield hydration numbers of just over 18 for the first solvation shell, which is consistent with maximizing the number of solute-solvent hydrogen bonds. In conclusion, from this study it has been shown that GLYCAM04 with either of the TIP3P, TIP4P-EW, and TIP5P-EW water models and the GLYCAM04-LP/TIP4P-EW combination of models all yield solute-solvent interactions that are in worse agreement with experiment than the GLYCAM04-LP/TIP3P or TIP5P-EW models.

Conformational Properties The three dimensional structure of a disaccharide is defined by four angles: φ, ψ, ω_1, and ω_2, Figure 5.9. The $^3J_{C,H}$ vicinal coupling constants were obtained for the φ and ψ angles in both maltosides utilizing the Karplus equation of Tvaroska *et al.*, Figure 5.11, and the $^3J_{H,H}$ couplings were determined for the ω_1 and ω_2 angles via the Stenutz *et al.* Karplus-type equation, Figure 1.1.[6,142] Unlike the examination of the solvent-solute interactions, the conformational analysis examined MD simulations of 200 ns in length when the GLYCAM04 model was employed and 100 ns in the case of the 04-LP model.

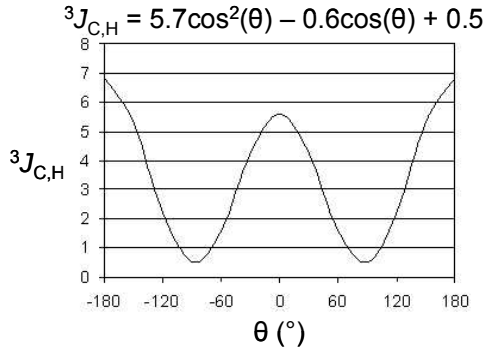


Figure 5.11 Illustration of the Karplus equation derived for the $^3J_{C,H}$ couplings (Hz) for the φ and ψ angles derived by Tvaroska *et al.*[142]

Table 5.9. Average ^3J couplings for **III** (top) and **IV** (bottom), for the GLYCAM04 (200 ns) and 04-LP models (100 ns).

III	$^3J_{C4'\text{-}H1}$	$^3J_{C1\text{-}H4'}$	$^3J_{H5\text{-}H6R}$	$^3J_{H5\text{-}H6S}$	$^3J_{H5'\text{-}H6'R}$	$^3J_{H5'\text{-}H6'S}$	RMSD
			Experiment[118]				
	4.1	4.9	5.0	1.8	5.0	2.2	
			GLYCAM04				
TIP3P	4.1±1.3	4.4±0.9	9.0±2.5	1.9±1.4	2.6±3.6	2.1±1.8	1.9
TIP4P-EW	4.0±1.2	3.7±0.9	9.0±2.5	2.0±1.4	7.6±2.9	2.2±1.8	2.0
TIP5P-EW	3.5±1.2	4.1±0.9	8.9±2.6	2.0±1.6	7.7±3.0	2.0±1.5	2.0
			GLYCAM04-LP				
TIP3P	4.6±1.0	4.6±0.8	5.3±4.6	2.1±1.3	5.8±4.5	2.1±1.6	0.4
TIP4P-EW	4.5±1.0	4.6±0.8	6.2± 2.7	2.2±1.4	5.9±2.5	2.3±1.7	0.7
TIP5P-EW	4.1±1.1	4.5±0.8	7.1± 2.9	2.0±1.3	7.1±2.9	2.0±1.6	1.2
			GLYCAM2000a[118]				
TIP3P	4.0	4.9	5.8	2.0	2.3	1.7	1.2
			CSFF[118]				
TIP3P	3.4	3.9	4.1	2.8	2.1	2.0	1.4
			HGFB[118]				
TIP3P	1.9	2.9	4.1	6.5	3.4	4.1	2.5

IV	$^3J_{C6'\text{-}H1}$	$^3J_{C1\text{-}H6'R}$	$^3J_{C1\text{-}H6'S}$	$^3J_{H5\text{-}H6R}$	$^3J_{H5\text{-}H6S}$	$^3J_{H5'\text{-}H6'R}$	$^3J_{H5'\text{-}H6'S}$	RMSD
				Experiment[118]				
	3.2	3.3	3.8	5.0	1.8	5.0	2.2	
				GLYCAM04				
TIP3P	3.1±1.1	2.4±1.4	1.8±1.2	9.9±1.9	1.7±1.7	7.6±2.9	2.1±1.8	2.3
TIP4P-EW	3.1±1.1	2.4±1.4	1.8±1.2	9.2± 2.6	2.0±2.3	7.6±2.9	2.1±1.8	2.0
TIP5P-EW	3.1±1.1	2.5±1.4	1.8±1.2	9.5± 2.3	1.8±2.1	8.5±2.6	2.0±1.5	2.3
				GLYCAM04-LP				
TIP3P	3.5±1.2	2.4±1.5	2.0±1.3	8.9±2.9	2.3±2.6	6.3±2.7	2.2±1.7	1.7
TIP4P-EW	3.6±1.2	2.1±1.3	2.0±1.2	10.3±1.2	1.9±1.4	6.0±2.7	2.2±1.8	2.2
TIP5P-EW	3.5±1.2	2.7±1.6	1.9±1.3	9.0±2.5	1.9±2.0	6.9±2.8	2.0±1.6	1.8
				GLYCAM2000a[118]				
TIP3P	2.4	2.8	2.0	6.5	1.9	5.7	1.9	1.0
				CSFF[118]				
TIP3P	2.7	3.1	2.0	3.8	2.3	9.3	1.4	1.9
				HGFB[118]				
TIP3P	2.5	2.4	2.1	4.5	7.3	3.5	4.5	2.5

Examination of the φ, ψ space sampled in methyl α-maltoside reveals that only one conformation was present over the entire course of the MD simulation for all models investigated, Figure 5.12. Therefore, this conformation can be considered consistent with the experimentally observed geometry since all models are in reasonable agreement with the experimental $^3J_{C,H}$ couplings for the φ and ψ angles, Table 5.9. Similarly, the $^3J_{C,H}$ coupling constant corresponding to the φ angle in methyl α-isomaltoside was accurately

reproduced since only one conformation was present. In contrast, there were two distinct conformations populated for the ψ angle in methyl α-isomaltoside. The C1-O6'-C6'-C5' angle was dominated by the *anti* conformation, which would then correspond to a *gauche* conformation for the C1-O6'-C6'-H6'R and C1-O6'-C6'-H6'S angles, which yields $^{3}J_{\text{C1-H6'R}}$ and $^{3}J_{\text{C1-H6'S}}$ couplings of approximately 1.5 Hz. Accordingly, since the magnitude of both of these coupling constants are underestimated, all of the models examined here are underestimating the populations of the *gauche*+ and *gauche*- conformations about the C1-O6'-C6'-C5' angle. Previously, the CSFF force field was estimated to yield a larger amount of the *gauche*+ conformation than the models investigated here, which resulted in a more accurate reproduction, 3.1 Hz, of the experimental $^{3}J_{\text{C1-H6'R}}$ coupling constant, 3.3 Hz. Although one of the coupling constants corresponding to the ψ angle was improved, the coupling constant for the other, $^{3}J_{\text{C1-H6'S}}$, has not, and similar values were yielded as with the models presented here, supporting the idea that both conformations were underestimated and not just a single conformation.

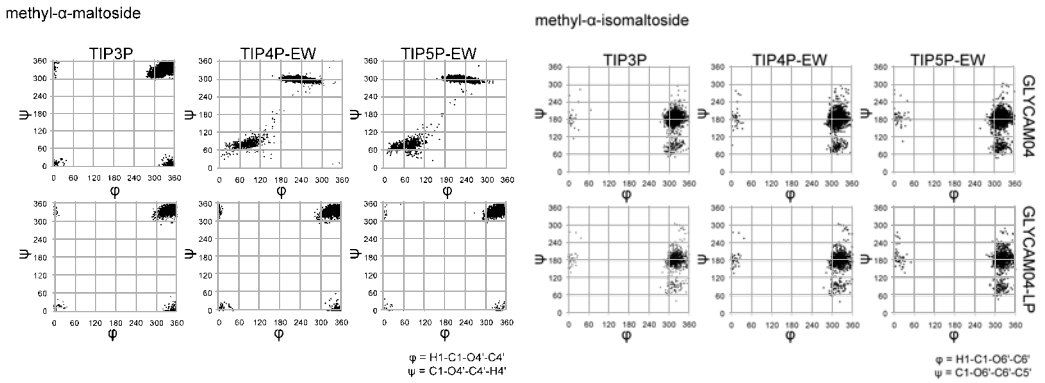


Figure 5.12 Populations of the φ and ψ angles (°) from MD simulations of **III** (left) and **IV** (right) with GLYCAM04 for 200 ns and GLYCAM04-LP for 100 ns.

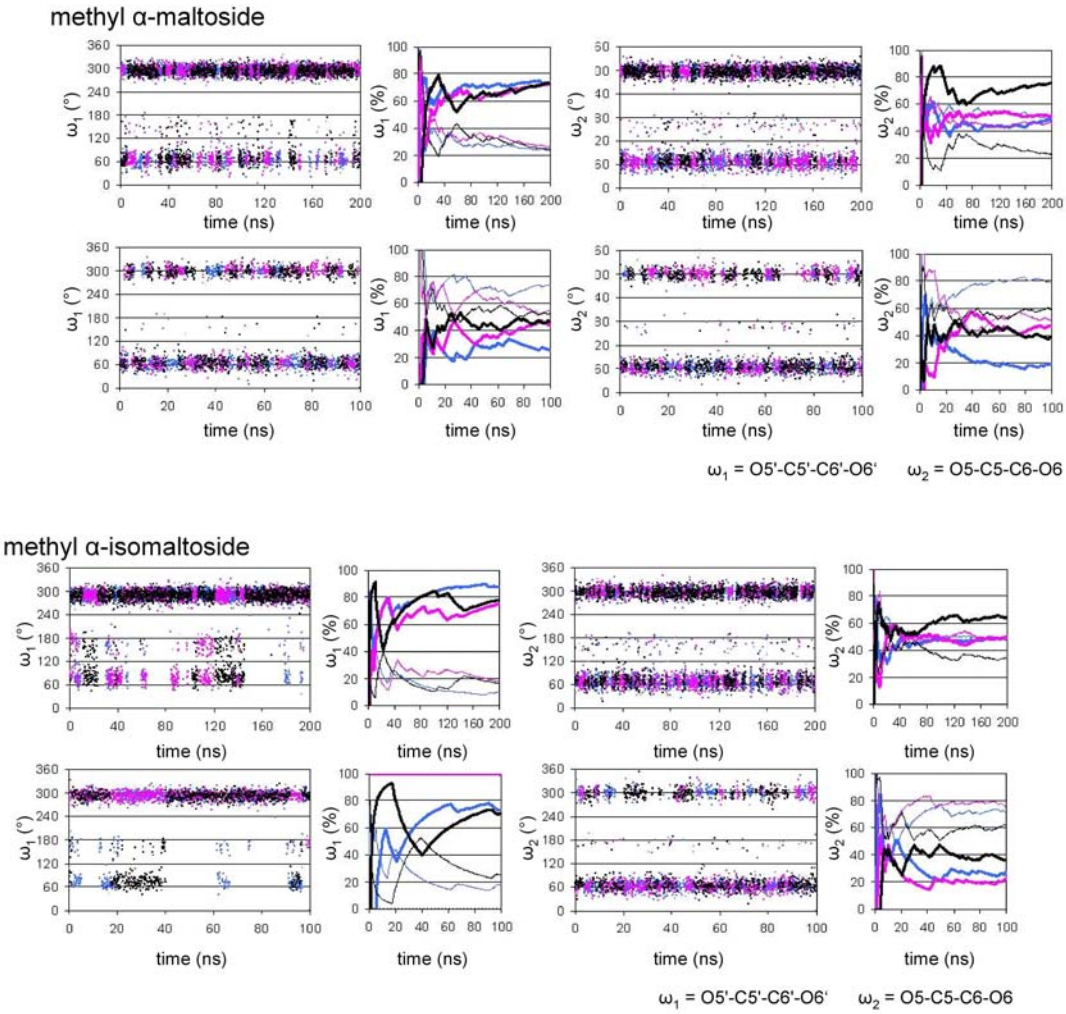


Figure 5.13 Population ratios of the ω_1- and ω_2-angles for **III** (top) and **IV** (bottom) in TIP3P (blue), TIP4P-EW (magenta), and TIP5P-EW (black) with GLYCAM04 or 04-LP. Running average for the *gg* (thin line) and *gt* (thick line) conformer at right.

The rotation about the C5-C6 bond in both maltosides fluctuates between the *gg* and *gt* conformation, with a negligible population for the *tg* conformer for all models, Figure 5.13, similar to that occurring in D-Glc*p*. The *gt* and *gg* conformations correspond to vales of 180° and 60°, respectively, for the angle corresponding to the $^3J_{H5\text{-}H6R}$ coupling and -60° and 60° for the $^3J_{H5\text{-}H6S}$ coupling. Therefore, the $^3J_{H5\text{-}H6S}$ coupling is insensitive to the relative populations of *gg* and *gt* due to the similar magnitudes obtained

for both conformations (Figure 5.11), which is illustrated in the lack of variation over the change in time and conformation ratio present in the MD simulation, Table 5.10.

Table 5.10 Average computed ^3J couplings for **III** (top) and **IV** (bottom) over time.

III	exp.[118]	time	GLYCAM04			time	GLYCAM04-LP		
		(ns)	TIP3P	TIP4P-EW	TIP5P-EW	(ns)	TIP3P	TIP4P-EW	TIP5P-EW
$^3J_{H5\text{-}H6R}$	5.0	40	8.8±2.7	9.1±2.7	9.3±2.9	20	5.7±3.6	6.0±2.7	6.4±2.7
		80	8.8±2.6	8.8±2.6	8.4±2.8	40	5.6±4.6	6.2±2.7	6.4±2.8
		120	8.9±2.6	8.9±2.6	8.5±2.8	60	5.8±4.6	6.4±2.8	6.6±2.8
		160	8.9±2.6	8.9±2.6	8.7±2.7	80	5.2±4.6	6.4±2.8	7.0±2.9
		200	9.0±2.5	9.0±2.5	8.9±2.6	100	5.3±4.6	6.2±2.7	7.1±2.9
$^3J_{H5\text{-}H6S}$	1.8	40	1.9±1.4	1.9±1.5	2.1±1.5	20	2.1±1.4	2.2±1.4	2.0±1.3
		80	2.0±1.4	2.0±1.5	2.0±1.6	40	2.1±1.4	2.2±1.4	2.0±1.3
		120	2.0±1.4	2.0±1.5	2.0±1.7	60	2.1±1.4	2.2±1.4	2.0±1.3
		160	1.9±1.4	2.0±1.4	2.0±1.6	80	2.1±1.4	2.2±1.4	2.0±1.3
		200	1.9±1.4	2.0±1.4	2.0±1.6	100	2.1±1.3	2.2±1.4	2.0±1.3
$^3J_{H5'\text{-}H6'R}$	5.0	40	6.0±3.8	7.0±3.0	7.8±2.9	20	6.0±4.5	6.6±2.5	6.6±2.8
		80	3.6±4.1	7.3±2.9	7.7±2.9	40	5.7±4.5	6.0±2.5	7.7±2.8
		120	3.2±4.0	7.4±2.9	7.8±2.9	60	5.3±4.5	5.8±2.5	7.0±2.9
		160	2.8±3.8	7.3±2.9	7.6±3.0	80	5.9±4.5	5.8±2.5	7.0±2.9
		200	2.6±3.6	7.6±2.9	7.7±3.0	100	5.8±4.5	5.9±2.5	7.1±2.9
$^3J_{H5'\text{-}H6'S}$	2.2	40	2.2±1.9	2.2±1.7	1.8±1.5	20	2.1±1.6	2.3±1.7	2.0±1.6
		80	2.1±1.9	2.2±1.8	1.9±1.6	40	2.1±1.6	2.4±1.6	2.0±1.4
		120	2.1±1.8	2.2±1.8	2.0±1.6	60	2.0±1.5	2.3±1.7	2.1±1.6
		160	2.1±1.9	2.2±1.9	2.0±1.5	80	2.0±1.6	2.3±1.7	2.0±1.6
		200	2.1±1.8	2.2±1.8	2.0±1.5	100	2.1±1.6	2.3±1.7	2.0±1.6

IV	exp.[118]	time	GLYCAM04			time	GLYCAM04-LP		
		(ns)	TIP3P	TIP4P-EW	TIP5P-EW	(ns)	TIP3P	TIP4P-EW	TIP5P-EW
$^3J_{H5\text{-}H6R}$	5.0	40	9.1±2.6	8.6±2.9	8.6±2.9	20	6.9±3.4	10.4±0.6	9.4±2.4
		80	9.5±2.2	9.1±2.6	9.5±2.3	40	8.4±3.3	10.4±0.6	7.6±2.9
		120	9.9±1.9	8.7±2.9	9.7±2.1	60	9.1±2.8	10.4±0.6	8.5±2.8
		160	9.9±1.8	8.9±2.7	9.2±2.5	80	9.0±2.9	10.4±0.6	9.0±2.5
		200	9.9±1.9	9.2±2.6	9.5±2.3	100	8.9±2.9	10.3±1.2	9.0±2.5
$^3J_{H5\text{-}H6S}$	1.8	40	1.9±2.2	2.2±2.6	2.2±2.7	20	3.1±3.6	1.7±1.0	1.9±2.0
		80	1.8±1.9	1.9±2.2	1.9±2.1	40	2.6±2.9	1.7±1.0	2.1±2.6
		120	1.7±1.7	2.2±2.6	1.8±1.9	60	2.3±2.5	1.7±1.0	1.9±2.2
		160	1.7±1.6	2.1±2.5	1.9±2.2	80	2.4±2.6	1.7±1.0	1.9±2.0
		200	1.7±1.7	2.0±2.3	1.8±2.1	100	2.3±2.6	1.9±1.4	1.9±2.0
$^3J_{H5'\text{-}H6'R}$	5.0	40	7.6±2.9	7.8±2.9	8.1±2.7	20	7.1±2.8	6.1±2.7	6.4±2.7
		80	7.4±3.0	7.7±2.9	8.2±2.7	40	6.1±2.7	5.6±2.4	7.0±2.8
		120	7.5±3.0	7.4±2.9	8.5±2.6	60	6.3±2.8	5.9±2.6	7.1±2.8
		160	7.5±2.9	7.6±2.9	8.5±2.6	80	6.3±2.7	5.8±2.6	7.0±2.8
		200	7.6±2.9	7.6±2.9	8.5±2.6	100	6.3±2.7	6.0±2.7	6.9±2.8
$^3J_{H5'\text{-}H6'S}$	2.2	40	2.1±1.8	2.2±1.8	2.0±1.6	20	2.2±1.5	2.3±1.9	2.1±1.9
		80	2.2±1.9	2.1±1.8	2.0±1.6	40	2.2±1.7	2.2±1.8	2.1±1.7
		120	2.1±1.9	2.1±1.8	2.0±1.5	60	2.2±1.7	2.2±1.7	2.1±1.7
		160	2.1±1.9	2.1±1.8	2.0±1.5	80	2.2±1.7	2.2±1.7	2.1±1.6
		200	2.1±1.8	2.1±1.8	2.0±1.5	100	2.2±1.7	2.2±1.8	2.0±1.6

In contrast, a large magnitude of approximately 10 Hz is observed for the $^{3}J_{H5\text{-}H6R}$ coupling constant when the pyranoside populates the *gg* conformation, and a smaller value, < 2Hz, when the *gt* conformation is populated, which correlates to the fluctuation observed over time and conformation ratio. In addition, previous studies examined the same two maltosides, but only examined the conformations of the disaccharides over a 10 ns timeframe, while the resultant populations after 10 ns versus the 100 and 200 ns MD simulations presented here yield contradicting results, illustrated in Figure 5.14.

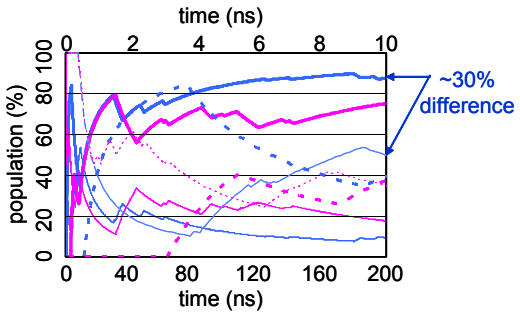


Figure 5.14 Difference between the population of the ω-angle conformers after 10 ns (thin lines) and 200 ns (thick lines). Color scheme same as Figure 5.13.

The overestimation of the $^{3}J_{H5\text{-}H6R}$ coupling constant for ω_1 can be attributed to the overestimation of the g*t* conformation in both maltosides, which was also true for ω_2 as well, but not to as great of an extent. For methyl α-maltoside, both ω_1 and ω_2 were more accurately modeled with the GLYCAM04-LP parameters relative to the 04 parameter set. The agreement for ω_2 worsens in methyl α-isomaltoside for all models relative to methyl α-maltoside, but ω_2 was more accurately modeled with the lone-pair model for the solute. Notably, for the ω_2 angle in isomaltoside modeled with the GLYCAM04-LP and TIP4P-EW parameters populated only the *gg* conformation, leading to the largest overestimation in the coupling constant for these two maltosides.

Therefore, this may be an illustration of the incompatibility of the TIP4P-EW water model and GLYCAM04-LP, analogous to the incompatibility illustrated in the underestimation of the location of the first solvation shell of isomaltoside for this same combination of models.

Relaxation Properties Corzana *et al.* reported that the GLYCAM2000a parameters underestimate interactions between saccharides and water, which was confirmed here in results obtained for the RDFs and η_H of **III** and **IV**, Table 5.11.[118] A clear improvement in accuracy was observed in the 04 and 04-LP parameters, in which, in addition to modified partial charges, the van der Waals parameters for the hydroxyl oxygen atom were significantly altered. In addition to the organization of the water molecules around carbohydrates, the overall motion, or tumbling, of the carbohydrate is also important. The overall tumbling is related to the strength of solute-solvent interactions, but is also dependent on other properties. Therefore, an additional measure of the accuracy of solute-solvent interactions is given by the time required for a molecule to randomize its orientation, referred to as the correlation time, τ_C.

For both of the disaccharides investigated here a molecular correlation time of approximately 150 ps has been reported, from fitting to an NOE buildup curve, which is obtained from measuring NOE intensities at different mixing times. To determine τ_C from MD simulations, the autocorrelation functions of several different non-exchangeable proton-proton vectors were examined giving rise to an average overall relaxation time for each combination of parameter sets, Figure 5.15. The exponential decay of this function deviates from a single, pure exponential form during the initial, short time period of less than 10 ps, Figure 5.16. Therefore, two line fitting schemes

were implemented, in which the requirement that the function pass through the origin was varied to take into account the rapid non-exponential decay observed at short-correlation times. In addition, the background noise increases as the function decays to zero and therefore long time lengths, over 100 ps, will not be included in the analysis, as illustrated specifically for the TIP3P water model in Figure 5.16-B.

Table 5.11 Correlation times (ps) for each model combination for **III** and **IV**.

	GLYCAM04-LP						GLYCAM04					
	B=0	B≠0	B=0	B≠0	B=0	B≠0	B=0	B≠0	B=0	B≠0	B=0	B≠0
vectors	TIP5P-EW		TIP4P-EW		TIP3P		TIP5P-EW		TIP4P-EW		TIP3P	
	methyl α-maltoside											
a	90	100	117	132	59	64	98	110	81	89	40	42
b	82	93	107	124	53	58	86	99	73	82	35	36
c	70	85	93	112	50	58	77	92	61	73	33	38
d	99	106	132	142	65	68	103	111	86	91	41	42
e	87	101	114	136	61	68	94	113	79	91	40	43
f	56	84	67	104	46	63	57	89	54	77	33	41
<average>	81	95	105	125	56	63	86	102	72	84	37	40
	methyl α-isomaltoside											
a	89	101	113	128	57	62	100	116	83	93	40	41
b	71	84	90	107	46	52	76	91	66	77	35	38
c	72	92	91	117	51	61	80	104	70	90	38	43
d	78	93	101	117	51	58	78	93	69	76	37	41
e	73	89	106	125	50	57	76	96	62	75	38	42
f	72	92	91	117	49	56	78	94	68	79	35	38
<average>	76	92	99	119	51	58	81	99	70	82	37	41

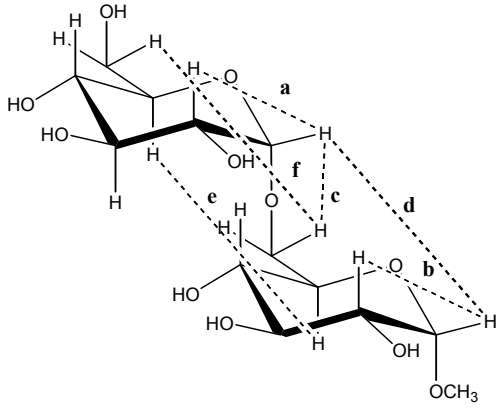


Figure 5.15 Description of vectors monitored for methyl α-isomaltoside (shown) and methyl α-maltoside, in which vector **c** corresponds to H1-H4' in methyl α-maltoside as opposed to H1-H6' as shown here.

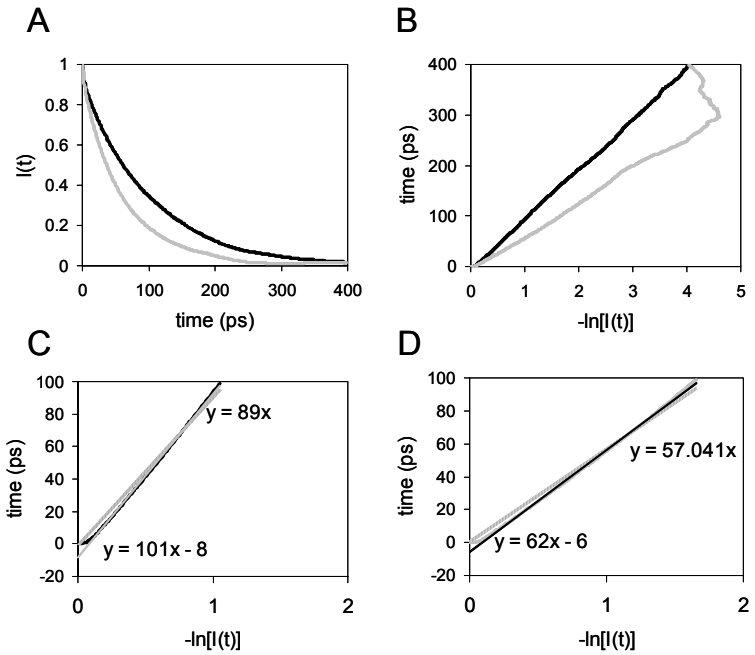


Figure 5.16 A) The autocorrelation of vector **a** in isomaltoside modeled with GLYCAM04-LP and solvated with TIP3P (gray) or TIP5P-EW (black). B) Linear form of Equation 5.4, in which τ_c can be determined from the slope of the line, $\tau_c = -t / \ln(I(t)) + B$. Illustation of the two methods, B = 0 (dotted) and B ≠ 0 (solid), yield for TIP5P-EW (C) and TIP3P (D).

Implementation of the non-LP model for the disaccharides leads to a clear improvement in τ_c values as the water models are varied from TIP3P, TIP4P-EW, to TIP5P-EW. A substantial increase in τ_c, on the order of 200%, was observed for TIP3P relative to the TIP4P-EW water model, Table 5.11. A smaller increase, averaging 15 ps, was observed on replacing TIP4P-EW with the TIP5P-EW water model. Implementation of the GLYCAM04-LP parameter set reveals the same relative trend between TIP3P and TIP4P-EW, with a small increase in correlation time over the respective GLYCAM04 MD simulations. Notably, introduction of the TIP5P-EW water model with the GLYCAM04-LP parameter set does not lead to a longer τ_c over TIP4P-EW or the respective GLYCAM04/TIP5P-EW combination of parameters. Therefore, introduction of tetrahedrally placed LPs on the water model results in a different interaction between

the solute and solvent, relative to the water models without tetrahedrally placed LPs, *i.e.* TIP3P and TIP4P-EW. Overall, of all the model combinations investigated here, the GLYCAM04-LP parameter set with TIP4P-EW water model yielded correlation times that were the closest to experiment with the models containing TIP5P-EW the next closest. Notably, the experimental values were obtained in D_2O, while the MD simulation was performed in H_2O. The correlation time is also dependent on the viscosity of the solvent in which it is determined as shown in Equation 5.6.[143] Since the viscosity, η, at 300 K of H_2O is 0.00089 Pa·s and that of D_2O is 0.0011 Pa·s, the models presented here become closer to the theoretically predicted value for τ_c of approximately 121 ps.

$$\tau_c = \frac{4\pi\eta r_H^3}{3k_B T} \qquad [5.6]$$

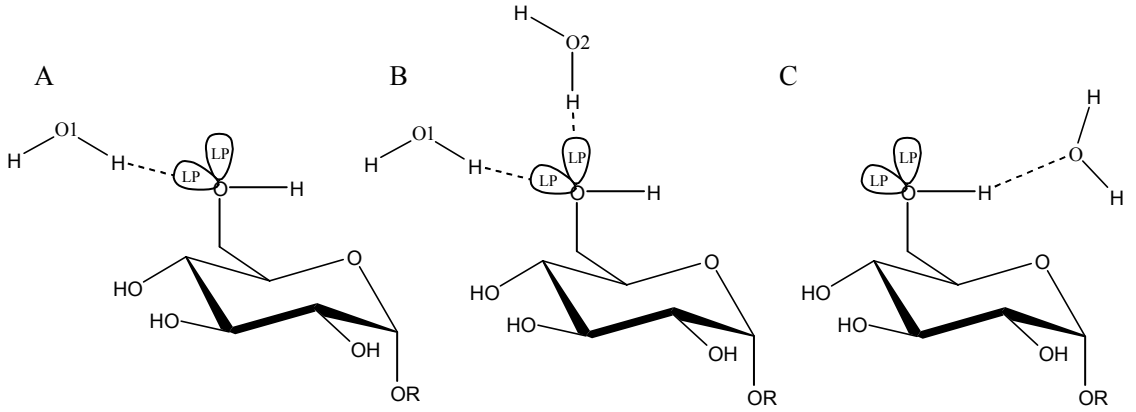


Figure 5.17 Illustration of the possible hydrogen bond relationships present for each carbohydrate hydroxyl group, in which it can behave as and acceptor to one water (A), an acceptor to two water molecules, referred to a three center hydrogen bond (B), and as a donor to one water molecule (C), R = Glc-OMe.

In order to determine the origin of the difference in the interaction of TIP5P-EW with the solute versus TIP3P and TIP4P-EW, the nature of the solute-solvent interaction was monitored in terms of intermolecular hydrogen bonding. Specifically, the three

possible unique interactions between a single hydroxyl group in the solute with the surrounding water molecules were monitored, Figure 5.17. During MD simulations that implemented the GLYCAM04 parameter set and either TIP3P or TIP4P-EW, the hydroxyl groups in the disaccharide behaved as both hydrogen bond acceptors and donors with the same relative occupancy. There was a slightly higher percentage of hydrogen bond acceptor behavior in the solute with the TIP4P-EW model relative to TIP3P, while the percentage of donor behavior decreased relative to TIP3P. Notably, the donor behavior of the hydroxyl groups with TIP5P-EW averaged over 95%, which is over 30% higher than with the other water models, while the acceptor behavior decreased over 10%. For all water models, the occurrence of three center hydrogen bonds in which a carbohydrate hydroxyl group acted as the acceptor was under 10%, Table 5.12.

Previously, theoreticians observed the tetrahedral hydration character of the hydroxyl groups in carbohydrates utilizing the HGFB carbohydrate force field in the CHARMM program with the SPC water model.[21,144] Deviation from a tetrahedral-type hydrogen bond arrangement was found to be related to the proximity of neighboring hydroxyl groups. Therefore, if the lack of tetrahedral hydration is due to the proximity of neighboring groups, there will generally be a lower percentage of three center hydrogen bonds present since this would be the energetically most favored configuration. In addition, presence of only a small amount of double-acceptor character would also be consistent with experimental results probing the donor-acceptor character found in bulk water.

Table 5.12 Hydrogen bond relationship between the hydroxyl groups in methyl α-isomaltoside and the surrounding water molecules.[a]

		O2'	O3'	O4'	O5'	O6'	O2	O3	O4	O5	O6
GLYCAM04											
TIP3P											
	% HB_{donor}	65	62	64	-	-	66	60	63	-	54
	% 1 $HB_{acceptor}$	59	53	56	43	2	58	52	56	29	53
	% 2 $HB_{acceptor}$	6	3	2	0	0	5	3	3	0	6
	% HB	87	82	84	43	2	87	81	84	29	81
TIP4P-EW											
	% HB_{donor}	59	54	58	-	-	60	56	59	-	53
	% 1 $HB_{acceptor}$	67	63	68	43	3	68	64	69	38	65
	% 2 $HB_{acceptor}$	7	3	2	0	0	6	3	3	0	6
	% HB	89	83	87	43	3	89	84	88	39	86
TIP5P-EW											
	% HB_{donor}	97	95	96	-	-	96	94	96	-	94
	% 1 $HB_{acceptor}$	41	35	39	23	4	41	34	38	.8	41
	% 2 $HB_{acceptor}$	4	2	2	0	0	4	2	1	0	4
	% HB	98	97	97	23	4	98	96	97	18	96
GLYCAM04-LP											
TIP3P											
	% HB_{donor}	48	47	50	-	-	52	46	53	-	40
	% 1 $HB_{acceptor}$	58	62	64	59	1	58	63	65	38	49
	% 2 $HB_{acceptor}$	30	23	16	2	0	30	22	15	1	43
	% HB	93	92	90	60	3	93	92	90	39	95
TIP4P-EW											
	% HB_{donor}	46	42	46	-	-	53	43	46	-	36
	% 1 $HB_{acceptor}$	55	65	71	75	0	56	64	70	39	39
	% 2 $HB_{acceptor}$	40	27	18	6	0	39	27	19	1	58
	% HB	97	95	94	81	0	97	95	94	40	98
TIP5P-EW											
	% HB_{donor}	82	78	80	-	-	81	78	81	-	72
	% 1 $HB_{acceptor}$	49	47	45	35	1	49	46	44	19	49
	% 2 $HB_{acceptor}$	9	7	4	1	0	9	8	4	0	11
	% HB	91	89	90	36	1	91	89	90	19	88

[a] A hydrogen bond was considered occupied when the O···H distance was less than 2 Å.

Experimentally, in 2004, at room temperature 80 ± 20 % of bulk water was shown to engage in only a single strong donor interaction, as opposed to double donor or the absence of donor interactions.[145,146] Although these results conflict the tetrahedral structure of bulk water as found in ice, it has been shown that two types of hydroxyl groups exist in bulk water, one that was involved in a strong hydrogen bond and one that was involved in a weak hydrogen bond.[147] Overall, each water was engaged in two strong hydrogen bonds, one in which the water was a donor and one in which it was an

acceptor, forming hydrogen bond networks through a combination of these two strong hydrogen bonds and weaker hydrogen bonds.[145] Therefore, the difference between the theoretically predicted structure of bulk water versus the most recent experimental findings needs to be assessed as well. For example, is the tetrahedral character of the saccharide solvation with the SPC model due to solute – solvent interactions or is it an artifact in which the SPC water model reversed the 80:20 ratio of two strong hydrogen bonds versus the tetrahedral network, respectively, in bulk water?[21,145]

Since both LP-containing water models induce a substantial increase in correlation time, yielding values closer to experiment, the difference in hydrogen bonding behavior cannot be directly influencing this effect since the behavior is different in both models. The increase in accuracy of τ_c observed with LP-containing water models relative to TIP3P may instead be governed by the inherent accuracy of the bulk properties of the water model, Table 1.1. The aforementioned observation of the overestimation of the diffusion rates obtained with the TIP series of water models for bulk water is coupled with the underestimation of τ_c for bulk water, which both would lead to an overestimation of τ_c for the saccharide. Specifically, despite the large improvement in the accuracy calculated for the diffusion constant with the TIP5P water model, τ_c of bulk water is still underestimated by 40%.

Clearly, from these studies, regardless of the parameter set employed for the solute, the solute – solvent properties modeled with TIP3P are inaccurate. Despite only a small change in the donor/acceptor character of the solute – solvent interactions, the TIP4P-EW model yields a substantial difference and increase in accuracy of the orientational motion of the solute. The increase in accuracy of the computed τ_c value is

most likely directly related to the inherent properties of the TIP4P-EW water model, which correctly predicts the diffusion constant of bulk water. The lack of out of plane character for the TIP4P-EW water model is the reason for the high resemblance to the three – point TIP3P water model, since introduction of out of plane character as lone-pairs in the TIP5P water model nearly doubles the donor behavior of the sugar hydroxyls.

Experimentally, statistical studies have shown that in principal, the lowest energy configuration of a carbohydrate in the crystalline-phase corresponds to the configuration with the maximum number of hydrogen bonds, but due to additional complications that is not always true;[148]

> "As a result, there is no way of predicting, from a structural survey alone, which of the four general patterns described will be adopted by a particular carbohydrate molecule."[148]

In accord with the experimental results from bulk water, assuming that the hydroxyl groups behave similarly to that of water, if a three-center hydrogen bond was present the strong hydrogen bond in which the carbohydrate is the donor should remain occupied during this three-center interaction. Since the only two models that have accurately modeled the solute-solvent interactions investigated here are the 04-LP/TIP3P and 04-LP/TIP5P-EW models, based on this experiment the 04-LP/TIP5P-EW water model is more consistent with the experimental data, Table 5.13.

Table 5.13 Duality of the role of the O6 atom in methyl α-isomaltoside played in hydrogen bonding with water. Percent population of each type of hydrogen bond O6 formed with water is shown.

	donated 1 & accepted 1 hydrogen bond			donated 1 & accepted 2 hydrogen bond		
water model	TIP3P	TIP4P-EW	TIP5P-EW	TIP3P	TIP4P-EW	TIP5P-EW
GLYCAM04	37	36	38	1	3	4
GLYCAM04-LP	18	13	35	18	22	9

Therefore, since the hydrogen bonding behavior was so different with these three water models it is not clear which combination of models corresponds to the most accurate behavior due to lack of experimental data to compare with. Based on the accuracy of the GLYCAM04-LP/TIP5P combination in the simulation of the α-Glc monohydrate crystal lattice over the other models, the hydrogen bonding in this system can be projected to be the most accurate. Overall, the inability of the TIP3P water model to accurately model this property was explicitly illustrated relative to the other two models. Therefore, since the HGFB and CSFF CHARMM parameter sets yielded correlation times on the order of the experimental value with TIP3P, utilization of the TIP4P-EW or TIP5P-EW models would be projected to lead to an overestimation of τ_c.[118]

Conclusion

Inclusion of the new lone-pair model with TIP5P consistently increased the accuracy for MM energy minimizations over the respective non-LP model with TIP3P. In addition, the crystalline-phase MD simulations clearly illustrated the improved reproduction of electrostatic interactions when lone-pairs are included. Typically, inclusion of LPs increased the accuracy of the model, but in the cases where improvement was not explicitly observed, the inclusion of lone-pairs did not deteriorate the model.

Our studies reveal that the water model has a substantial impact on the dynamic properties of the solute in explicit solvent. Future studies will be needed to clearly address the discrepancies noted in describing the overall tumbling of saccharides in the aqueous phase if the correlation time computed with the HGFB force field and the TIP3P water model is doubled due to the strength of the hydrogen bonding interactions between

the solute and solvent and if this model deteriorates when progressing to a more accurate water model such as TIP4P or TIP5P. The expected deterioration of the HGFB model with other water models than the TIP3P water model emphasizes the importance of the inability to freely interchange solute and solvent models, since the HGFB parameter sete is based on the CHARMm methodology which includes the explicit inclusion of a water model, in this case TIP3P. Due to the profound effect the water model has on the molecular rotation dynamics, additional studies are needed to verify the correct hydrogen bonding interactions are occuring in MD simulations of bulk water. The focal point of these studies should utilize QM/MD methodologies to determine the proper donor and acceptor interactions that occur in the aqueous phase and their agreement with experimental data.

6. CARBOHYDRATE-LECTIN COMPLEXES MEDIATED BY BOUND WATER

Introduction

X-ray crystallography has been used to elucidate the location of buried water molecules in a variety of biomolecular crystals.[98,99,149-151] For example, the crystal structure of ubiquitin was shown to have four buried water molecules that are conserved throughout a variety of different crystalline forms.[151] The buried water molecules are located in the interior of the protein and separated from the bulk solvent. Despite the presence of the buried water molecules in crystal structures, confirmation of their presence under biologically relevant conditions was of interest. Therefore, solution state NMR spectroscopy was applied to investigate the dynamic motion of the buried water molecules as well as the water molecules located in the hydration shell around ubiquitin.[152] The sign of the peaks obtained from NMR experiments based on the Nuclear Overhauser Effect (NOE) were found to be dependent on the residency time for these water molecules. Overall, hydration waters corresponded to negative NOE cross-peaks, correlating to a residency time on the order of ~100-300 ps. In contrast, the bound waters in ubiquitin each corresponded to a positive water-protein cross peak, yielding a residency time longer than 1 ns, confirming the different behavior of buried vs. hydration water in biomolecular systems.[152]

Generally, the interaction between carbohydrates and proteins is a weak-binding event.[153] In specific cases, it has been noted that the interaction between these two biomolecular species is mediated by one or more water molecules, which are referred to as "bound water".[98-100,116,149,150,154] Since the waters of hydration and buried water

molecules exhibit different properties from the bulk solvent and each other, it is expected that the bound water molecules also exhibit unique properties.[152] The role of bound water and how it arrives at its role of a mediator of hydrogen bonding interactions between the protein and the ligand is not well understood, but the importance in binding has been confirmed in the aqueous phase for a lectin – saccharide complex.[100] Three distinct possibilities exist. Is it necessary for the bound water to be interacting with the protein for binding? Is the water complexed with the ligand and brought to the binding site with the ligand? Or is it a more complex combination of events?[117,140,155]

The behavior of the bound water once the protein – ligand complex has been formed is the subject of the present study. Specifically, the ability of the bound water molecule to interchange with the solvent and the frequency with which it occurs as well as how the overall interactions in the binding site change with the movement of the bound water will be addressed. NMR spectroscopy can provide a general picture of the behavior of water that slowly exchanges with the bulk solvent, but is unable to distinguish between individual interactions at the atomic level for water molecules. In order to understand the dynamic motion of the bound water, MD simulations have been implemented that focus on the binding site. The investigation of two lectin – saccharide complexes that are each mediated by a bound water molecule revealed the behavior of the bound water molecule to be entirely dependent on the parameter sets employed. Due to the disagreement observed between the results from MD simulations with different parameter sets a final assessment of the role the bound water plays can not be made.

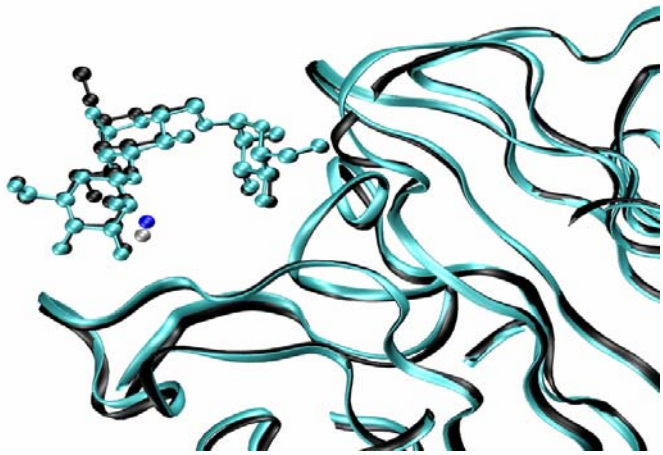


Figure 6.1 Alignment of the backbone atoms in the crystal structure of 1CVN (cyan) and TBD (black). The oxygen of atom of the single water molecule resolved in the binding site is shown for both the natural and synthetic ligand, in blue and gray, respectively.

The mannose-specific legume lectin Concanavalin A (Con A) was originally crystallized in 1919 (although the structure was not determined until later) and was the first lectin to be crystallized with a carbohydrate (monosaccharide) in 1989.[98,156,157] Con A recognizes the core trimannoside region of *N*-linked glycans. The first oligosaccharide – Con A complex (1CVN) was determined in 1996, in which the ligand included the core trimannoside region of *N*-linked glycans, Man*p*-α-(1→3)[Man*p*-α-(1→6)]-Man*p*, which has a binding affinity, K_a, of 3.74 x 10^{-5} M^{-1} at 300 K.[99,116,153]

All three mannose residues have direct interactions with the protein in 1CVN. The reducing sugar forms a hydrogen bond between O4 and Tyr12(OH), but is also part of a water-mediated interaction between O2 and Asn14, Asp16, and Arg228.[99] This bound water is conserved in the crystal structures of other Con A complexes as well as the free protein as a hydration water.[157,158] Therefore, it was proposed that modification of the hydroxyl group at the 2-position of the reducing sugar to an ethylhydroxyl group

may disrupt or alter the water – mediated interactions observed with the natural ligand corresponding to a loss in binding affinity, K_a of 0.49 x 10^{-5} M^{-1} at 298.15 K.[116] The crystal structure of the Con A – α-(1→3)Man*p*-[α-(1→6)Man*p*]-2-deoxy-2-ethylhydroxy-Man*p*-OMe complex revealed that the bound water was also conserved in this complex, Figure 6.1. This complex will be referred to as TBD (to be deposited).[159]

Superposition of the protein backbone of the two Con A complexes reveals a similar structural conformation for the protein and ligand, with a 0.7 Å root mean square deviation (RMSD) for the backbone atoms in the protein, a 1.5 Å shift for the bound water molecule, and a 1.4 Å RMSD for the non-hydrogen atoms in the trimannoside, excluding the substituent at the C2 position of the reducing sugar. Examination of each mannose residue in the trimannoside yielded an RMSD of 2.2 Å for the reducing residue (241) and 1.1 Å for the 1→3 linked mannose (243) and 0.5 Å for the 1→6 linked mannose (242). The RMSD in residue 243 results from the difference in the rotation around the ω-angle, while the major difference between the reducing mannosyl residues (241) is the ring puckering due to the synthetic modification made at the C2 position, Figure 6.2.

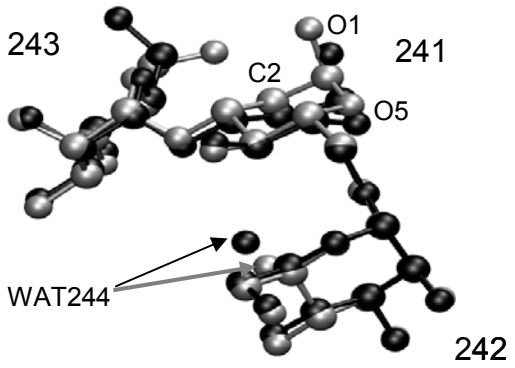


Figure 6.2 Overlay of the trimannoside from 1CVN (black) and TBD (gray) utilizing the non-hydrogen atoms of the sugar, excluding the substituents at C2. The bound water (244) is shown, but not included in the fit.

The hydrogen bond between 241(O4) and Tyr(OH) is retained at 2.6 Å in the crystal structure containing the modified ligand (TBD), which is perceived to be crucial for binding. The short distance between 241(O2) and the bound water, residue 244, of 2.4 Å in 1CVN is extended to 3.0 Å in TBD. In contrast, the remaining interactions involving O2, HOH244(O), Asn14(O), Asp16(O), and Arg228(NH) all shorten in TBD, relative to 1CVN, Figure 6.3. Most substantially, in the TBD structure a hydrogen bond is present between Arg228(NH) and the oxygen atom at the C2 position, which was not present in 1CVN. In general, the same interactions between the protein and carbohydrate are present in both complexes, Figure 6.4.

Through the analysis of data from MD simulations, the behavior exhibited by the bound water molecules in the Con A complexes has been compared and analyzed. Understanding how bound water behaves in the binding site would help provide a rationale for the treatment of bound water during the ligand design process.

Methods

The water molecules resolved by experiment and designated as those which mediate interactions between the protein and the carbohydrate were included as part of the initial starting configuration (Figure 6.3 and 6.5), along with the protein and carbohydrate. A central point in the binding site was chosen for each complex and all residues within 12 Å were allowed to move, referred to as the belly region, while the rest of the protein was frozen. Explicit water was included around the same central point extending out 30 Å. Standard simulated annealing protocol was followed with equilibration of the solvent, not including the X-ray resolved waters. Specifically, the bound water determined experimentally is referred to as residue 244 for the complexes

formed with Con A. For comparison, the experimental label is WAT85 for the Con A – trimannoside complex (1CVN). [99,150]

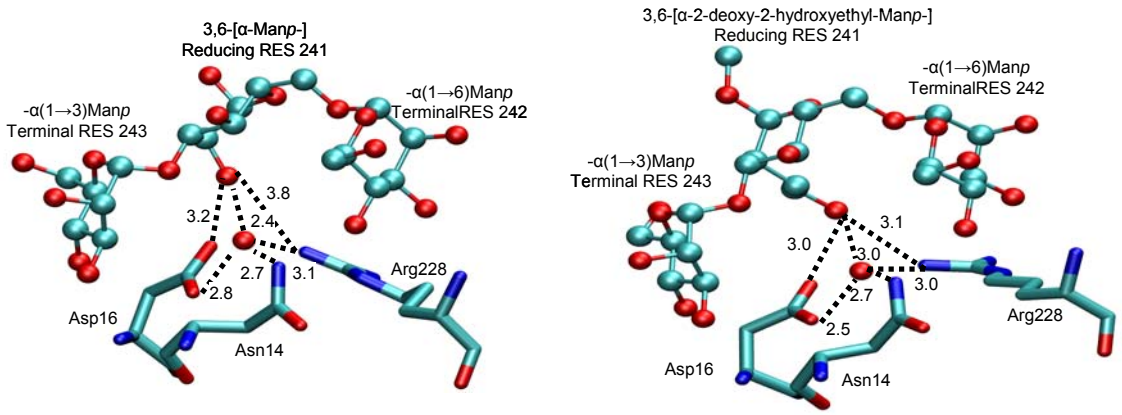


Figure 6.3 Interactions (Å) involving the bound water in the crystallographic binding site of Con A complexed with an α-trimannoside (1CVN, left) and with the synthetically modified α-trimannoside (TBD, right).

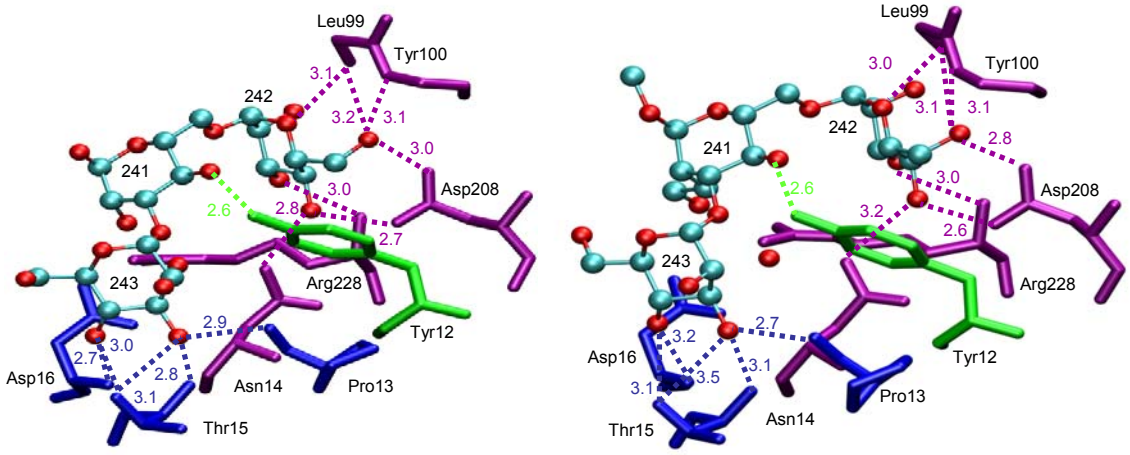
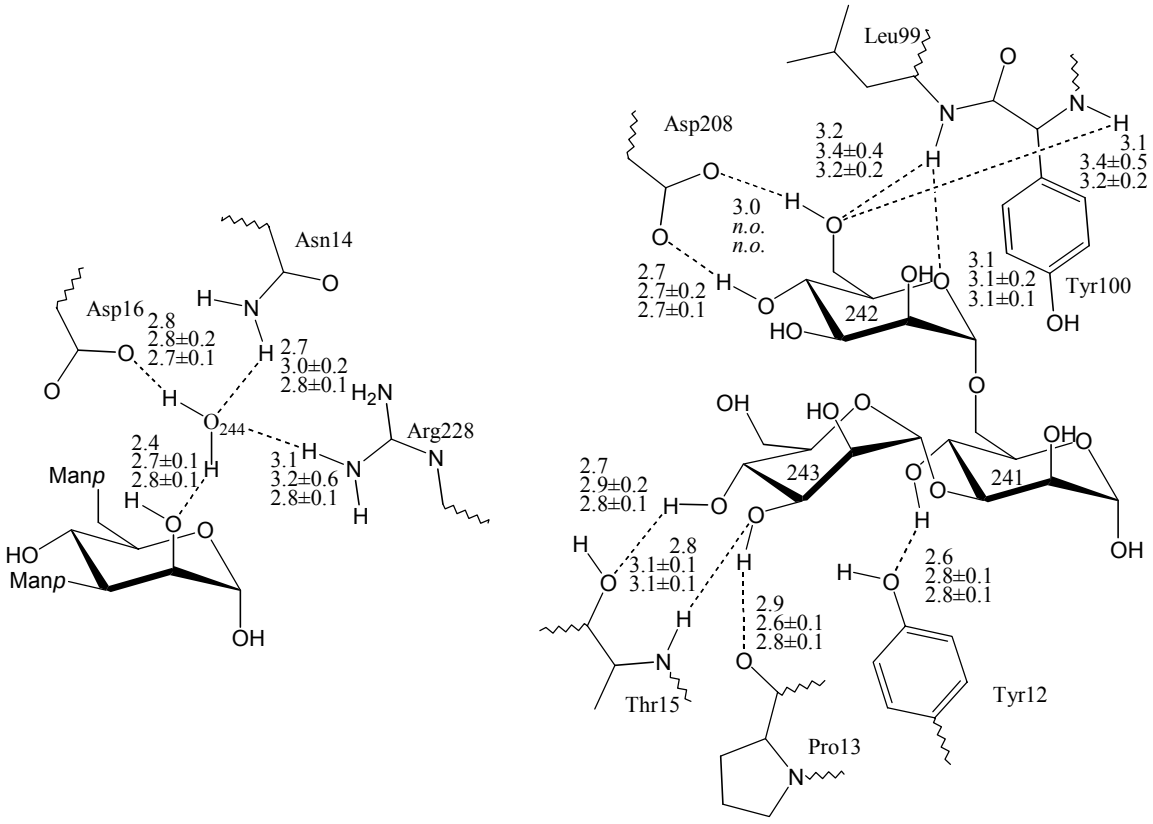


Figure 6.4 Binding site of Con A complexed with the α-trimannoside (1CVN, left) and the synthetically modified α-trimannoside (TBD, right). Hydrogen bond distances (Å) between the protein and mannosyl residues **243**, **241**, and **242**.

All solvent water, hydrogen atoms in the bound water and hydroxyl groups in the ligand, and all residues in the belly region, were minimized for 10000 steps of steepest descent followed by 10000 steps of conjugate gradient or until the energy gradient was less than 0.001 kcal·mol^{-1}. An additional minimization allowing the solvent water, bound water, ligand, and all residues in the belly region to relax was again performed for 10000 steps of steepest descent followed by 10000 steps of conjugate gradient until the energy gradient was less than 0.001 kcal·mol^{-1}. Over 100 ps the system was heated to 300 K and maintained at this temperature for at least 5 ns. A time step of 1 fs was implemented along with an 10 Å cutoff. The 1,4 electrostatic and 1,4 van der Waals interactions were scaled by the standard scaling factor implemented in the non-polarizable and non-LP containing PARM99 force field, which was utilized for the protein in all cases, corresponding to values of 1/1.2 and 1/2.0, respectively.[160]

For each ligand-protein complex two MD simulations were performed, initiated from the same starting structure and following the identical protocol. First, the GLYCAM04 parameter set and TIP3P water model were implemented for the carbohydrate and water model, respectively.[31,70] Secondly, the GLYCAM04-LP parameter set and the TIP5P water model were implemented for the carbohydrate and water model, respectively.[34,138] The MD simulation will be identified by referring to the parameter set employed for the carbohydrate, *i.e.* GLYCAM04 (04). Both the CARNAL and PTRAJ programs were implemented for data analysis.[125] Data analysis was performed from the first point of heating after the simulated annealing, since hydrogen bond formation and breaking may have occurred during these initial stages.

I. Hydrogen bond distances in 1CVN (top) and averaged over the simulation with GLYCAM04 (middle) and 04-LP (bottom).



II. Overlay of the final snapshot from 04 (pink), 04-LP (blue, and 1CVN (black/gray).

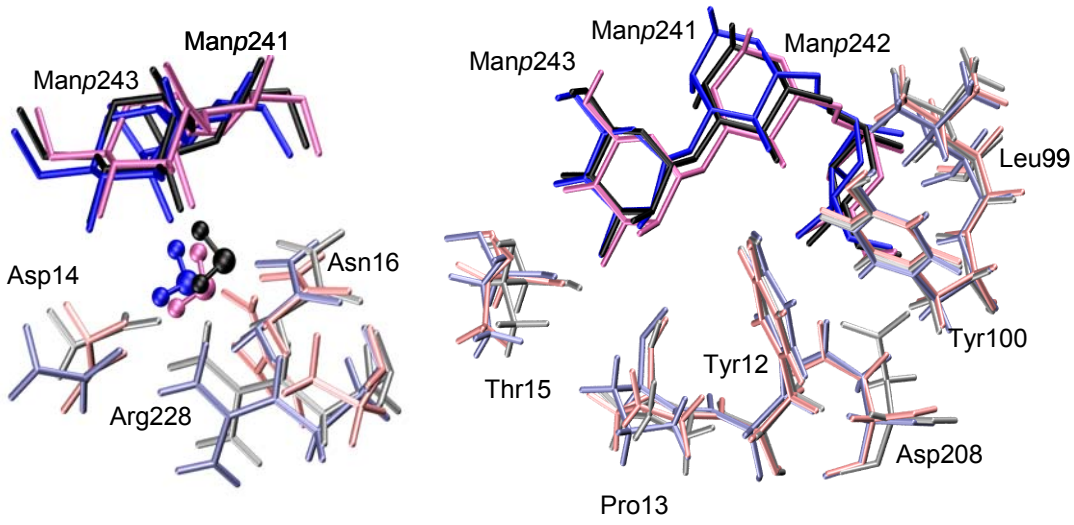


Figure 6.5 Hydrogen bonding (I) and motion (II) observed in the binding site of 1CVN over the course of MD simulations.

Results

The bound water remained in the binding site throughout all MD simulations of the Con A complexes, except the GLYCAM04-LP MD simulation of TBD. Despite the ability of both parameter sets to reproduce the stability of the bound water in 1CVN, there are substantial differences in the configuration of the residues in the binding site of 1CVN. The side chains of Asp208 and Thr15 rotated from their initial configuration in the crystal structure, resulting in the breaking of the 3.0 Å Asp208(Oδ2)···Man*p*242(O6) hydrogen bond and a weakening of the Thr15(N)···Man*p*243(O3) hydrogen bond from 2.8 to 3.1 Å, Figure 6.5. An additional difference was observed in the 04-LP simulation, in which the side chain of Asp16 has rotated, breaking the weak interaction of 3.2 Å with mannose 241 in the crystal structure. This additional deviation from the crystal structure is most likely due to the imbalance of interactions initiated from the inclusion of lone-pairs on the ligand and solvent, but not the protein.

The conclusion that breaking of the weak interaction between Asp16 and mannose 241 is due to an imbalance caused by the introduction of lone-pairs to only a portion of the system was determined due to the occurrence of the same interaction during the GLYCAM04-LP simulation of TBD. Similar to the case of 1CVN/04-LP, the side chain of Asp16 rotated, breaking the weak interaction with mannose 241, and eventually resulted in the breaking of the hydrogen bond between the bound water (244) and residue 241 (O2). At this point, Asp16 formed an additional hydrogen bond with the bound water, allowing both oxygen atoms in the carboxylate group to accept a hydrogen bond from the bound water. This configuration was unstable and resulted in an increase in the motion of the residues in the binding site, exemplified by the increase in RMSD for

the trimannoside relative to 1CVN/04-LP (Table 6.1), in which the hydrogen bond between mannose 241 and the bound water was briefly reformed before the bound water was released into the bulk solvent. Once the water is released into the bulk solvent the ethylhydroxyl group rotated into the binding site and remained there throughout the rest of the simulation, Figure 6.6.

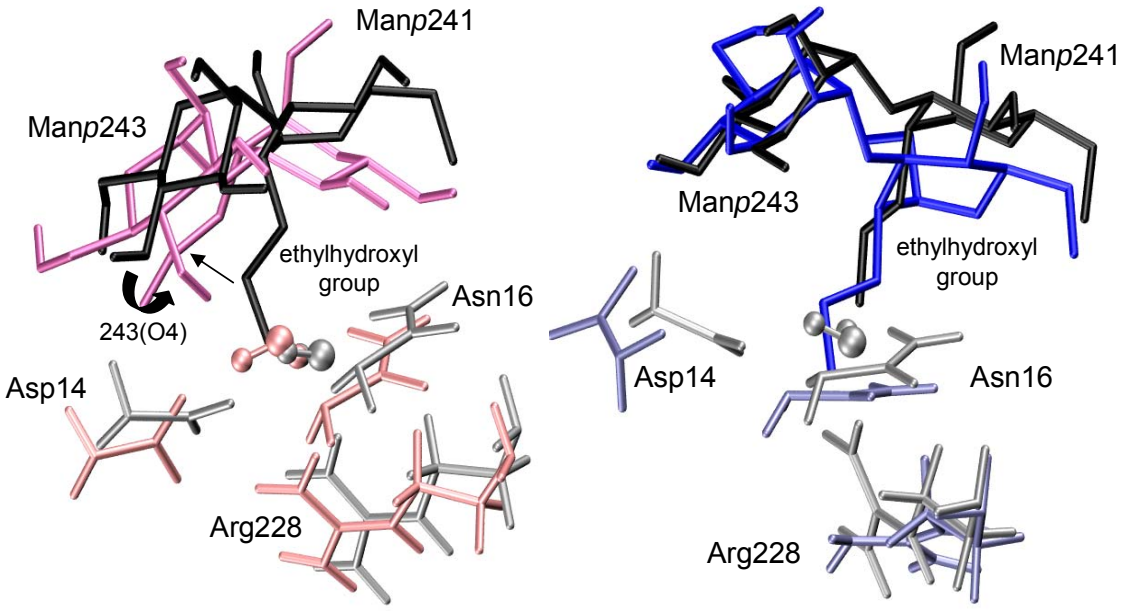


Figure 6.6 Overlay of the final snapshot from 04 (pink), 04-LP (blue) and TBD (black/gray). The bound water molecule was not present in the final snapshot of 04-LP.

Since the bound water is present in the crystal structure TBD, this simulation incorrectly predicts the configuration of the binding site. Although different interactions formed with different parameter sets for 1CVN and TBD complexes, the overall conformation of the ligand did not change. The linkage angles are maintained in the crystal conformation throughout the entire simulation, regardless of carbohydrate and protein parameter set and solvent model, Figure 6.7. The (1→3) linked mannose (243) in TBD exhibits a wider range of motion than in 1CVN, which remained centered tightly at

the value determined in the crystal structure. The larger degree of fluctuation was due to the interaction of the bound water molecule (244) with hydroxyl groups of the (1→3) linked mannose (243) in the TBD and not the 1CVN MD simulations with GLYCAM04. Specifically, mannose residue 243 (O4) accepted a hydrogen bond from the bound water 244, which resulted in the dynamic switching of the conformation of the ethylhydroxyl group from the initial configuration in which it played the role of hydrogen bond acceptor, to hydrogen bond donor to the bound water, Figure 6.8.

Table 6.1 RMSDs for all residues and the bound water in the binding site of the Con A complexes from their respective crystal structures, 1CVN and TBD.

Residue	1CVN		TBD	
Number	04	04-LP	04	04-LP
12	0.4±0.1	0.4±0.1	0.9±0.2	0.8±0.1
13	0.5±0.1	0.6±0.1	1.3±0.4	1.7±0.2
14	0.4±0.1	0.7±0.1	1.1±0.1	1.1±0.1
15	0.4±0.1	0.8±0.2	1.3±0.2	1.3±0.3
16	0.8±0.1	1.1±0.2	1.3±0.3	2.0±0.5
99	0.4±0.1	0.4±0.1	0.9±0.2	0.8±0.2
100	0.3±0.1	0.4±0.1	0.7±0.2	1.0±0.3
208	1.1±0.1	1.0±0.1	1.2±0.1	1.1±0.1
228	1.2±0.3	0.5±0.1	0.9±0.2	0.6±0.2
12…228	0.7±0.1	0.7±0.1	1.1±0.1	1.2±0.1
trimannoside	0.4±0.1	0.5±0.1	1.7±0.3	1.2±0.3
3,6[α-Man*p*-]	0.3±0.1	0.6±0.1	1.9±0.5	1.4±0.4
α(1→6)-Man*p*(242)	0.5±0.1	0.4±0.1	1.1±0.2	0.7±0.2
α(1→3)-Man*p*(243)	0.4±0.1	0.4±0.1	1.8±0.4	1.2±0.5
bound water(244)	1.1±0.2	0.9±0.1	1.3±0.4	18±12

Since the behavior of the bound water with the LP and non-LP model was substantially different in the TBD complex for the two parameter sets, the two models must exhibit different behavior of the substituent at C2 since the overall shape of the ligand does not change. Notably, the ethylhydroxyl group reorients in both simulations, resulting in a different hydrogen bond network both in each simulation and the crystal

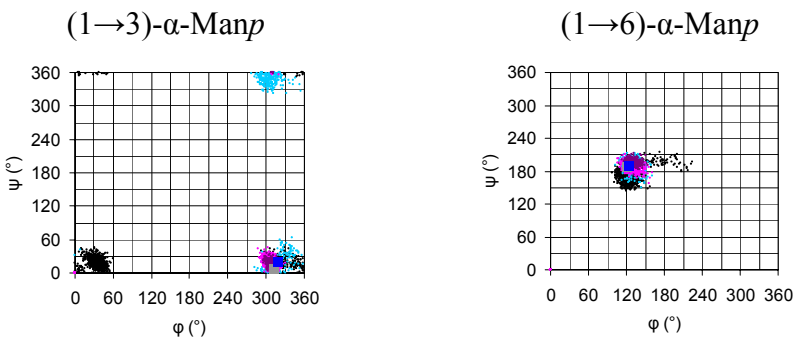


Figure 6.8 φ, ψ plots for the trimannoside obtained with 04 (1CVN-magenta, TBD-black) and 04-LP (1CVN-purple, TBD-light blue) with the experimental value in gray for 1CVN and blue for TBD.

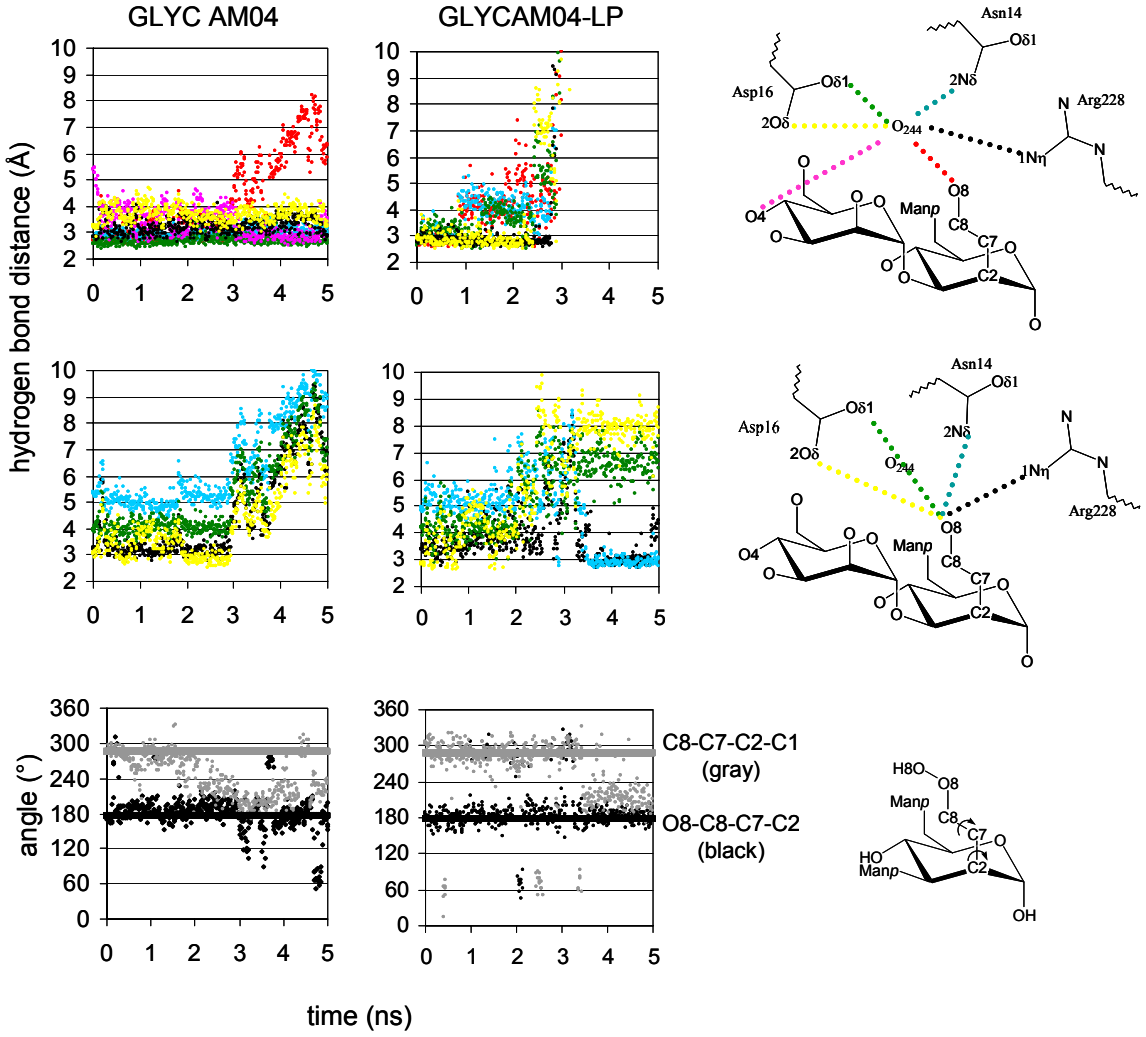


Figure 6.9 Hydrogen bond distances for TBD and the corresponding change in the orientation of the ethylhydroxyl group in mannose 241 for 04 (left) and 04-LP (right),

structure, Figure 6.9. In conclusion, both simulations of TBD resulted in the loss of the initial water-mediated carbohydrate – protein interaction. In the case of GLYCAM04-LP, a new mediated interaction was not formed, but in the GLYCAM04 simulation a new water-mediated interaction formed between residue 243(O4) and the bound water 244.

Conclusions

Due to the large configurational and conformational changes in all systems investigated the role of the bound water cannot be affirmed. Of more attention, the failing of all models illustrates the weaknesses the lone-pair model was explicitly developed to address. Clearly, in order to investigate the heterogenous carbohydrate – protein systems the new lone-pair model must be extended to the protein parameter set. The initial steps in this direction are presented in Chapter 7.

7. PRELIMINARY EXTENSION OF THE GLYCAM LONE-PAIR MODEL TO DEVELOP A MODIFIED AMBER PROTEIN FORCE FIELD

Introduction

Molecular modeling of small peptides has provided enlightenment into folding processes as well as conformational structure predictions, Table 7.1. The dominant problem in the molecular simulation of peptide folding on the atomic scale is the computational time required to reach the folding timescale. A protein of average size, 30 kilodalton (kD), modeled with explicit solvent is typically simulated on the nanosecond (ns) timescale, while the shortest folding event is on the order of microseconds (μs).[161] Two general strategies have been successfully employed to overcome the limited computational sampling time, umbrella sampling and multiple, parallel high temperature MD simulations.[161-163] The synthetic twenty residue peptide referred to as "Trp-cage" has proven to be one of the most ideal test systems for computational methods due to its short folding time, τ_f, of 6 μs into a β-hairpin structure.[164] MD simulations have been able to predict the folded structure from a fully extended conformation as well as the approximate folding time. [163,165,166]

Despite the successes of the non-polarizable protein force fields presented in Table 7.1, these models have been shown to have general, systematic weaknesses. For example, the AMBER PARM94/99 parameter sets have been shown to overestimate the stability of α-helical conformations, while OPLS-AA leads to an overestimation in the stability of β-sheets, and CHARMM (version 27) does not sample the intermediate regions properly, Figure 7.2.[23,167,168] All three of these weaknesses were illustrated through monitoring the φ (C-N-Cα-C) and ψ (N-Cα-C-N) angles (Figure 7.1) adopted in

the alanine dipeptide over the course of an explicitly solvated 6 ns MD simulation.[168] The MM force fields were assessed relative to the results obtained from a QM/MM simulation.[168] On average, the maximum populated ψ value lies between 60° and 180° for all of the tested force fields, including the QM/MM model, except AMBER PARM98.[168] The alanine dipeptide modeled with PARM98 infrequently populates ψ values greater than zero and the maximum populated region had a ψ value ranging from 0° to -60°. Therefore, the overstabilization of α-helices observed in PARM94 (both PARM98 and 94 parameter sets have the same peptide backbone torsion parameters) can be correlated with this overpopulation of a ψ value, which is representative of α-helices. The other MM force fields as well as the QM/MM model predicted conformer distributions for the alanine dipeptide that are consistent with the conformation of alanine residues in crystal structures of proteins.[169,168]

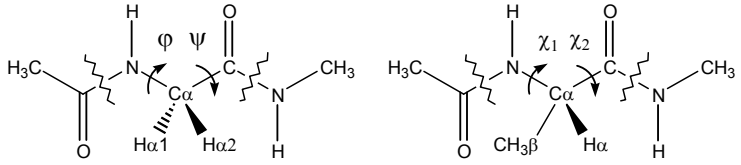


Figure 7.1 The φ, ψ, χ_1 and $\chi_2$torsions in the glycine (left) and alanine dipeptides (right).

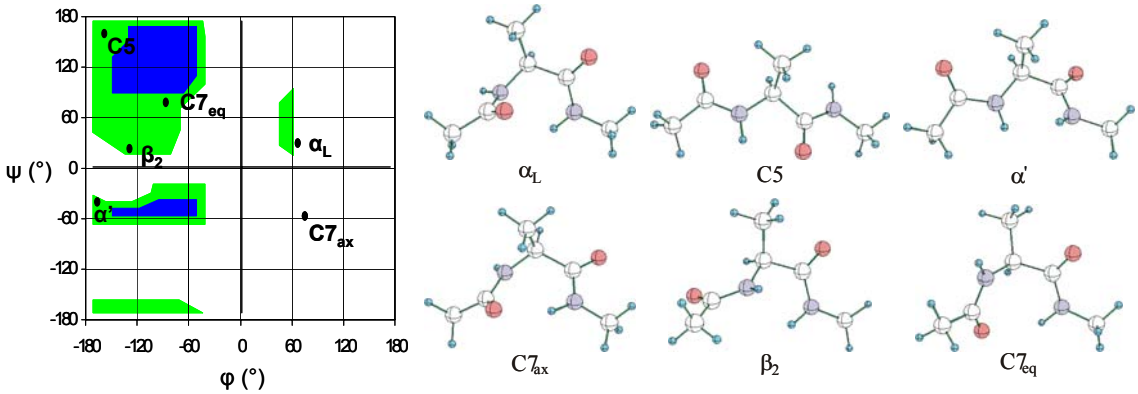


Figure 7.2 Accessible φ,ψ regions (commonly (blue) and generally allowed (green)) observed in proteins. Six low energy conformers of the alanine dipeptide (black, right).

Table 7.1 Conformations of a trpzip (A), β-heptapeptide[a] (B) and trpcage (C) peptides in water by experiment and theory.

A. Trpzip type peptide from the gb1 polypeptide, residues 41-56 (GEWTYDDATKTFTVTE)

Parameter Set/Solvent Model	Temperature	Observations
AMBER-PARM94 (backbone torsions reoptimized)/explicit	270-525	~30%, ~40% β-hairpin at 282 and 300 K, $\tau_f \approx$ 1-7 μs Folding mechanism involved formation of hydrophobic core with minimal hydrogen-bonding.[170]
OPLS/implicit	300	Folding mechanism involved formation of hydrophobic core with minimal backbone hydrogen bonding, possible semi-helical intermediate, many various possibilities for β-hairpin structure.[171]
OPLS-AA/explicit	360/380	No α-helical content, β-hairpin population decreases from 71% at 270 K to 66% at 310 K.[172]
CHARMM19/implicit OPLSAA/implicit	270-690	~75%, ~68%, ~55% β-hairpin at 300, 335, and 395 K, respectively, semi-helical conformers present β-hairpin structure not lowest ΔG, no α-helical content.[173]
Experimental Method	**Temperature**	**Observations**
CD/NMR	278	~42% β-hairpin[174]
NMR	278-308	β-hairpin[175]
Temp. jump spectroscopy	273-328	β-hairpin content decreases from ~80%, to ~40%, and ~0% at 273, 310 and over 360 K, $\tau_f \approx 6$ μs[176]
Temp. jump spectroscopy	273-363	β-hairpin stabilized by 3 hydrophobic interactions, several conformers containing hydrophobic core exist near global minima with varied degrees of hydrogen bonding.[177]
NMR	288	~40% partial hairpin structure[178]

B. β-heptapeptide (H-β-HVal-β-HAla-β-Leu-(*S,S*)-β-Ala(αMe)-β-HVal-β-HAla-β-HLeu-OH)

Parameter Set/Solvent Model	Temperature	Observations
GROMOS96/explicit	298/350/400	3_1-helical fold stable at 298 K and unfolds after 800 ps (400 K) without returning to native structure.[179]
GROMOS43A1/explicit	298/340-360	populated native 3_1-helix above τ_m.[180] (τ_m refers to the melting temperature)
GROMOS43A1/explicit	298/340-360	~97% 3_1-helical fold (298 K)[181]
OPLS-AA[a]/explicit	300-400	$360 < \tau_m < 380$ K[182]
Experimental Method	**Temperature**	**Observations**
CD/NMR (CD_3OD)	293-393	3_{14} helix is stable (at least up to 353 K)[183]

C. Trpcage (NLYIQWLKDGGPSSGRPPPS)

Parameter Set/Solvent Model	Temperature	Observations
AMBER-PARM99mod/ implicit	300/325/400	0.97 Å C_α RMSD to TC5b NMR[165]
FF not specified/ implicit	300	1.5 μs $< \tau_f <$ 8.7 μs[166]
AMBER-PARM94/ implicit	250-630	$\tau_m \approx$ 400 K, free energy global minima is native structure[163]
Experimental Method	**Temperature**	**Observations**
NMR/CD (TC5b PDB id)	273-373	$\tau_m = 315$ K, Contains an α- and 3_{10}-helix[164]
Temp. jump spectroscopy	296	$\tau_f \approx 4.1$ μs[184]

[a] Torsion parameters reoptimized for β-peptides

Since the GLYCAM parameter sets for carbohydrates were developed for utilization within the AMBER force field and in conjunction with the AMBER protein parameter sets, improving the AMBER parameter set for proteins was the focal point of this study. In order to determine how to improve the AMBER parameter sets, the previous AMBER development strategy employed will be summarized. The Cornell *et al.* series of AMBER parameter sets includes PARM94, PARM96, PARM98, and PARM99.[25,185,65] In each of PARM94, 96, 98 and 99 generalized torsion terms for a test set of organic compounds were derived and explicit torsion terms specific to the peptide backbone were added afterwards by fitting to QM data for small polypeptides. Initially, PARM94 was fit to key points on the potential energy surface in φ,ψ space for the alanine dipeptide ($C7_{ax}$, $C7_{eq}$, C5, α_R) and for the glycine dipeptide (C7, C5, α_R) determined at the MP2/TZP//HF/6-31G(d) level.[186] As computations became more efficient, the availability of a greater amount of *ab initio* data led to the inclusion of more points on the surface of the alanine dipeptide for PARM96 as well as the capped alanine tetramer (ala_4) in PARM99, Figure 7.3.[187]

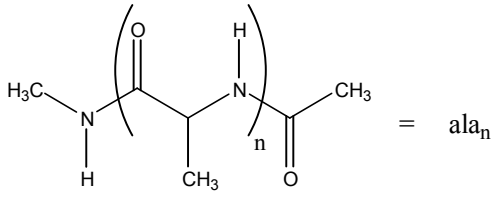


Figure 7.3 Structure of the ala_4, ala_6, and ala_8 polymers.

In general, it was observed that the AMBER 94 parameter set favors α-helical conformations, and small modifications have been made to ameliorate this problem, specifically via modification of the φ- and ψ- torsion parameters.[165,167,188,189] Notably,

generation of a modified parameter set, which was identical to PARM94 with the sole exception in that the φ and ψ torsion terms were removed and set to zero, resulted in improved agreement between computed and experimental structural and thermodynamic data for the Fs peptide (a 21 residue alanine and arginine containing polypeptide).[188] More detailed modifications have been made for increasing the compatibility of the AMBER parameter sets with the GB model, in which partial charges were refit to reproduce the B3LYP/cc-pVTZ//HF/6-31G(d) MEP and φ,ψ torsion terms were readjusted to reproduce relative energies for dipeptides at the MP2/cc-pVTZ//HF-6-31G(d) level with a dielectric constant of 4.[190] Typically, the dielectric constant is set to one during gas-phase quantum mechanical computations for parameter set development and in explicitly solvated MD simulations. As more *ab initio* data became available, the inability of additive force fields to simultaneously reproduce all points on the potential energy surface in φ,ψ space of the alanine dipeptide came to light. As a result, an empirical correlation function was utilized to reproduce the entire *ab initio* alanine, glycine, and proline dipeptide φ,ψ potential energy surface (LMP2/cc-pVX//MP2/6-31G(d), in which X was DZ, TZ, and QZ(-g)), to improve the CHARMm parameter set for proteins.[191]

Despite the ability to reproduce the relative energetics of the φ,ψ potential energy surface of the alanine dipeptide, the appropriateness of using the alanine dipeptide as a model compound for protein force field development is also in question. The values for the φ and ψ torsions that define the secondary structures observed in proteins are not always minima on the potential energy surface of the alanine dipeptide. The secondary structures observed in proteins stabilize the tertiary structure and require polymers larger

than the alanine dipeptide, for example, 4 or more monomers are required to form the α-helix. The hydrogen bond cooperativity exhibited in these structures enhances their relative stability, which is a feature that is absent in dipeptides. Additional problems, such as the observed difference in bond lengths and angles for the alanine dipeptide in the gas phase versus the condensed phase, have led parameter developers away from solely utilizing the alanine dipeptide as the model compound for protein parameter set development.[192]

Removal of the weaknesses in the AMBER parameter sets would lead to more reliable structural predictions. This would significantly impact the ability of the protein parameter set to correctly predict the conformational distributions of flexible peptides and glycopeptides.[193] Due to the noted increase in accuracy of the GLYCAM04-LP parameters for carbohydrates, addition of LPs, in the same fashion, to the protein parameter set was also undertaken. Here we show that additional optimization of the φ, ψ-torsion coefficients within the lone-pair formalism in GLYCAM04-LP, as well as utilization of larger model compounds in the development test set, leads to an improved protein parameter set. Although the charges for the lone-pair containing PARM99EP and the GLYCAM-LP parameter sets were each MEP-derived, the lone-pair placement in PARM99EP was not. In addition, the PARM99EP torsion parameters were not reoptimized after the addition of lone-pairs, rather the PARM99 parameter set was employed.

Many of the successes illustrated in Table 7.1 utilized implicit solvent methods. In general, these methods are computationally fast, but suffer from several weaknesses, such as the disagreement with NMR data in areas of solvent-exposed intramolecular

hydrogen bonds and the overstabilization of salt-bridges and folded states in proteins.[163,173,194] Therefore, explicit solvent will be implemented throughout the development and testing stages of a GLYCAM04-LP model for proteins.

Methods

Quantum Mechanics The Gaussian 98 suite of programs, version A.9.11, was utilized for all quantum mechanical calculations.[75] The 2D HF/6-31G(d) potential energy surface of the alanine and glycine dipeptides was constructed by scanning the φ,ψ angles in 15° increments, Figure 7.1.[76] The initial guess implemented in OPTOR, a torsion parameter optimization program, for the C-N-Cα-C (φ) and N-Cα-CT-N (ψ) angles was determined by fitting to energy slices through the surface of each dipeptide.[195] Full 360° rotational curves for the C-N-Cα-Cβ (χ_1) or N-C-Cα-Cβ (χ_2) torsion angles were determined for *N*-ethyl acetamide (NEA), *N*-isopropyl acetamide (NIA), *N*-*t*-butyl acetamide (NTA), *N*-methyl ethanamide (NME), *N*-methyl isopropylamide (NMI), and *N*-methyl *t*-butylamide (NMT). The O-C-N-H torsion profile was obtained quantum mechanically for *N*-methyl acetamide, but the parameters currently employed in the AMBER parameter sets accurately reproduced the quantum mechanical energy curve, so no adjustments to these parameters were required.

The local minima for the alanine dipeptide were optimized at the HF/6-31G(d) level, starting from the geometries reported in the literature. Six local minima will be utilized in the parameter development. There are two hydrogen-bonded structures, that of $C7_{ax}$ and $C7_{eq}$, in which the 7 refers to the 7-membered circle formed by the internal hydrogen bond and the axial and equatorial notation refers to the orientation of the methyl group at Cα. There are two α-helical conformers, in which α_L refers to the left-

handed helix and α' refers to a helical structure, but is near the edge of the generously allowed φ, ψ region in proteins, Figure 7.2. C5 has torsion angles representative of a fully extended peptide, in the larger alanine polymers the C5 nomenclature is dropped and the conformation is referred to as extended. The β_2 structure is representative of the Type II β-turn found in proteins and the B_{strand} conformation is only stable with larger alanine polymers.

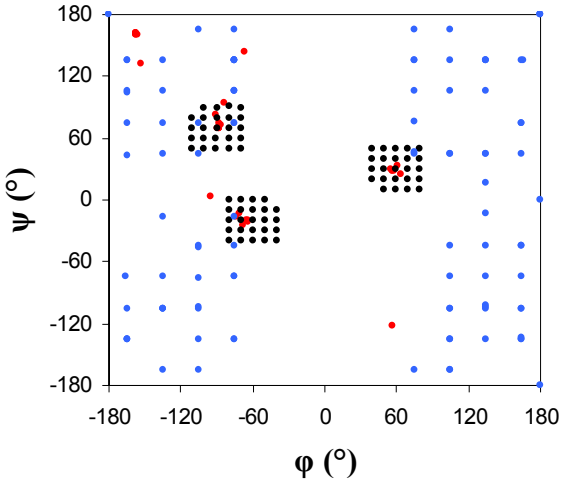
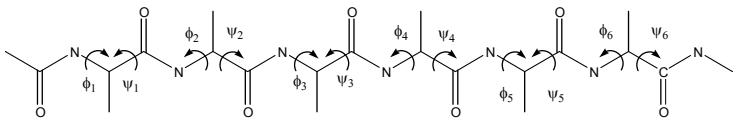


Figure 7.4 The backbone φ and ψ values (°) corresponding to the five lowest energy minima for ala_6 (red), the additional 12 ala_6 conformers (black), and the additional 17 gly_6 conformers (blue). All six φ, ψ pairs are shown for each hexapeptide as illustrated to the right.



The local minima for ala_6 and gly_6 peptides were optimized at the same level as the alanine dipeptide (HF/6-31G(d)). The starting structure for each additional gly_6 conformer was obtained by placing a 30° incremented grid over the accessible φ,ψ space for glycine residues within protein structures, which is quite large due to the absence of a side chain on the glycine residue. The starting points for the ala_6 conformers were

obtained utilizing a cross-like sampling of the subspace centered on minima as previously performed by Kaminski *et al*.[196] The α_L α_R, and β-strand conformers of ala_6 have regular φ,ψ values throughout the set of 6 backbone torsion angles, while β_2 samples a wide variety of backbone angles within one conformer, which is shown in Figure 7.4. Therefore, a 10° spaced grid was centered at the φ,ψ minima of α_L (referred to as conformers I-IV), α_R (referred to as conformers V-VIII), and β-strand conformers (referred to as conformers IX-XII). The HF/6-31G(d) optimized geometries for five minima (α_L, α_R, β_2, β-strand, andextended) of the ala_4, ala_6, and ala_8 peptides were kindly provided by Dr. Jane D. Dyekjaer.[197]

Molecular Mechanics The SANDER module in AMBER7 was utilized to energy minimize each structure initiated from the quantum mechanical optimized geometry. All atoms were included in the calculation of the electrostatic energy, a dielectric constant of unity, and 20,000 steps (9,000 steps of steepest descent followed by 11,000 steps of conjugate gradient) were utilized to energy minimize each peptide. The backbone conformations were maintained by a harmonic restraint that was applied to the φ and ψ angles for all structures. The torsion terms were optimized via a least squares fit to the energy of each peptide using the OPTOR program.[195] After the torsion optimization, each structure was reminimized with the new torsion terms for verification purposes. Partial charges utilized in all AMBER non-polarizable parameter sets are equivalent and were also used to develop the non-LP model here. Addition of asymmetric lone-pairs required derivation of appropriate partial charges for the alanine and glycine residues. A single conformation of each capped peptide was employed during the derivation, in which φ and ψ were restrained at 180° while the rest of the

internal coordinates were optimized at the HF/6-31G(d) level. Partial charges were computed from the HF/6-31G(d) MEP as previously mentioned, with partial charges on the aliphatic hydrogen atoms included.

Table 7.X Partial charges implemented during the derivation of backbone torsion parameters.

Gly	PARM94/96/99	LP	Ala	PARM94/96/99	LP
N	-0.4157	-0.5189	N	-0.4157	-0.5463
H	0.2719	0.2979	H	0.2719	0.3056
Cα	-0.0252	0.0593	Cα	0.0337	0.1004
Hα2	0.0698	0.0619	Hα	0.0823	0.0340
Hα3	0.0698	0.0619	Cβ	-0.1825	-0.0083
C	0.5973	0.4361	Hβ1	0.0603	0.0120
O	-0.5679	0.0000	Hβ2	0.0603	0.0120
LP(LP···O of 0.7 Å)		-0.1594	Hβ3	0.0603	0.0120
LP(LP···O of 0.3 Å)		-0.2722	C	0.5973	0.4920
			O	-0.5679	0.0000
			LP(LP···O of 0.7 Å)		-0.1520
			LP(LP···O of 0.3 Å)		-0.3037

In order to utilize the OPTOR program, initial torsion parameters for the angles that will be optimized must be given. In general, automated fitting algorithms suffer from the weakness of being dependent on this initial value. Here, initial parameters, referred to as G_i or G_iLP, for the φ (C-N-Cα-C) and ψ (N-Cα-C-N) angles were derived to reproduce four slices through the alanine dipeptide and glycine dipeptide surfaces, two that varied φ, one while ψ=60° and one with ψ=180°, and two that varied ψ, one with φ=60° and one with φ=180°. In order to obtain the most optimal fit, different parameters were implemented in the model without LPs versus the model with LPs. Parameters derived for the ψ torsion angle include a negative V_2 term for both models, while the V_1 term was four times as large (1.7 kcal/mol °) for the lone-pair model relative the model without lone-pairs (0.4 kcal/mol °), Table 7.2. In contrast, simply removing the parameters for the φ torsion resulted in a better fit to the dipeptide slices, which is consistent with previous studies that showed removal of the PARM94 backbone torsion

parameters resulted in better agreement with experiment for an oligopeptide.[188] The glycine dipeptide surface and the various gly_6 polymers were included in the data set for the optimization of the φ and ψ angles. Since only glycine polymers were utilized during optimization of the φ and ψ angles, the optimized parameters are denoted with a "G". For consistency, the initial parameters employed in the OPTOR program for the φ and ψ angles will be referred to as G_i.

The full, 360° rotational profiles for C-N-CT-CT (χ_1) were optimized in NEA, NIA, and NTA and the N-C-CT-CT (χ_2) torsion profiles were optimized in NME, NMI, and NMT to obtain the initial guess for χ_1 and χ_2, Figure 7.5, referred to as A_i or A_iLP. Notably, introduction of the asymmetric lone pair model in the PARM94 force field resulted in a substantial improvement in the reproduction of the energy curve for NEA and NIA, even prior to deriving new torsion parameters, while the more symmetric molecules resulted in similar potential energy curves for both models. Similar to the optimization of the φ and ψ angles, optimization of one angle, χ_2, with the lone-pair and non-LP models resulted in the same Fourier coefficients; while different torsion parameters were optimal for the second angle, χ_1, in each model. Only a V_2 term was necessary when the asymmetric lone-pairs are present, while V_1 and V_3 terms were required to reproduce the torsion profile without lone-pairs, Table 7.2. Similar to the nomenclature for parameters regarding the φ and ψ angles, the optimized parameters for the χ_1 and χ_2 systems only use alanine polymers and will therefore be referred to by "A".

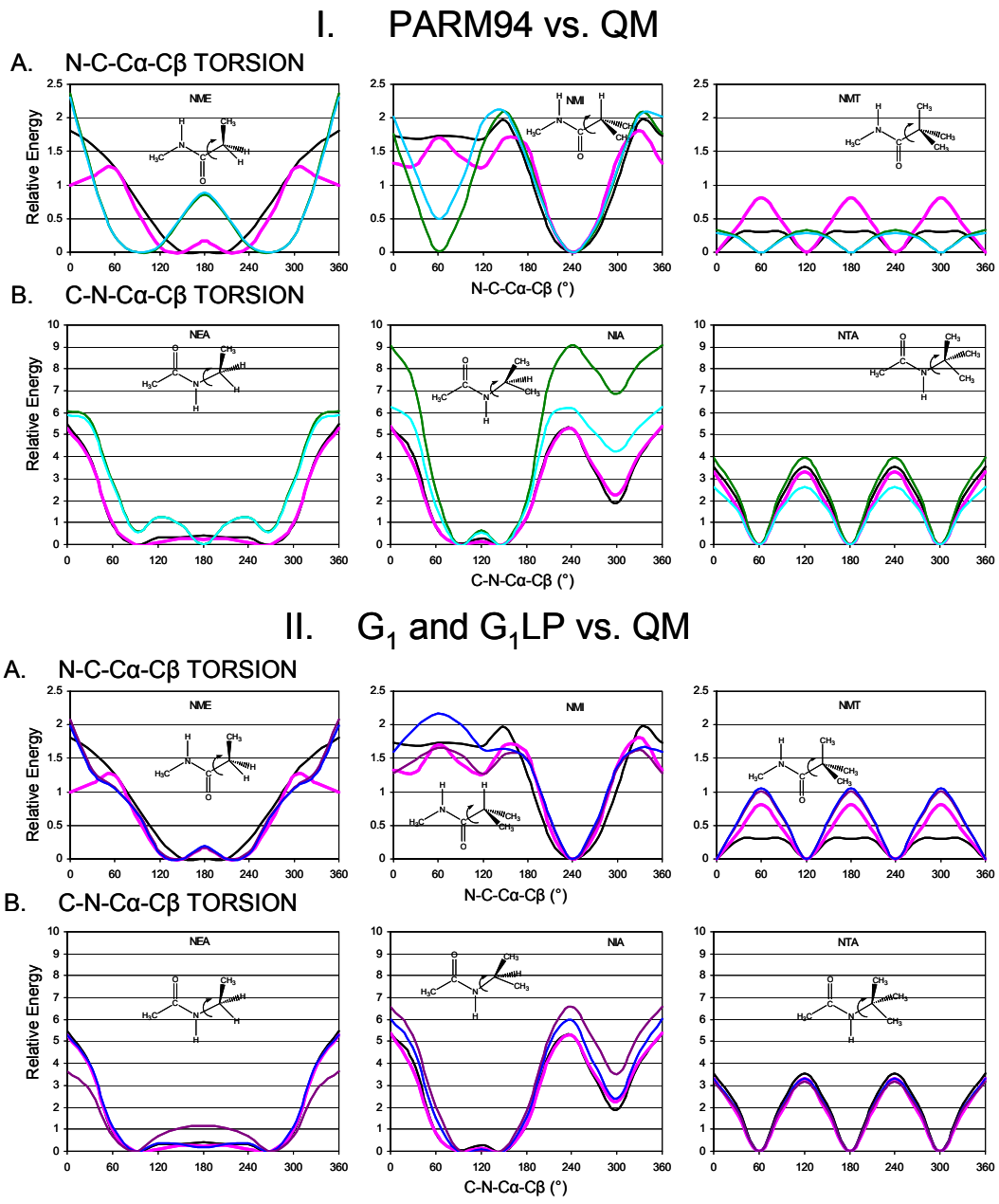


Figure 7.5 Torsion profiles produced with and without LPs for χ_2 (A) and χ_1 (B) the PARM94 (green) and 94 with new LP model (cyan) (I) and the G_i (purple) and G_iLP (blue) parameter sets (II). QM results obtained at the HF/6-31G(d)//HF/6-31G(d) level (black) and the B3LYP/6-31++G(2d,p)//HF/6-31G(d) level (magenta) are shown.

Table 7.2 Parameters utilized in the AMBER parameter sets compared to the newly optimized values presented here.[a]

	φ (C-N-Cα-C)				ψ (N-Cα-C-N)			
	V_1	V_2	V_3	V_4	V_1	V_2	V_3	V_4
PARM94	-	-0.20[b]	-	-	-0.75	-1.35	-	-0.40
PARM96	0.85	-0.30	-	-	0.85	-0.30	-	-
PARM99[b]	0.80	-0.85	-	-	-1.70	-2.00	-	-
G_i	-	-	-	-	0.40	-1.70	-	-
G_iLP	-	-	-	-	1.70	-1.20	-	-
G_1	1.59	0.97	0.34	0.15	0.60	-1.54	-0.22	0.30
G_2	2.17	1.07	-	0.03	0.62	-1.74	0.04	0.44
G_1LP	1.68	1.42	0.75	0.42	0.53	-1.26	-0.51	0.40
G_2LP	2.74	1.90	0.51	0.36	1.51	-1.44	-	0.52
	χ_1 (C-N-CT-CT)				χ_2 (N-C-CT-CT)			
	V_1	V_2	V_3	V_4	V_1	V_2	V_3	V_4
PARM94	0.53	-	-0.15	-0.50	-	0.07	-	0.10
PARM96	0.53	-	-0.15	-0.50	-	0.07	-	0.10
PARM99[b]	0.53	-	-0.15	-0.50	-	0.07	-	0.10
A_i	-1.22	-	-0.22		0.41	-0.51	-0.22	0.08
A_iLP	-	0.20	-	-	0.41	-0.51	-0.22	0.08
G_1A_1	-0.16	0.23	0.14	1.12	1.12	0.47	-0.03	0.30
G_1A_2	0.41	0.43	0.05	0.88	-0.44	0.10	0.72	0.68
G_2A_1	0.02	0.22	0.45	0.68	1.12	0.09	-0.21	0.39
G_2A_2	0.43	0.22	0.45	0.39	-0.43	-0.51	0.24	0.82
G_1A_1LP	0.92	0.49	-0.14	1.42	0.67	0.71	0.02	0.27
G_1A_2LP	2.05	0.58	-0.76	0.60	-2.00	0.38	2.19	2.00
G_2A_1LP	-0.22	-0.44	0.03	0.55	1.33	0.26	-0.58	-
G_2A_2LP	0.74	-1.02	-0.22	-0.63	-0.79	-0.68	1.10	2.37

[a] The "-" indicates a Fourier coefficient of zero.
[b] The PARM99 Fourier coefficients are also implemented in 02 and 02EP, in which the difference becomes the inclusion of atomic polarizabilities in 02 and 02EP as well as the inclusion of lone-pairs in 02EP.

Crystal Simulations The unit cell of the alanine dipeptide zwitterion contains 8 molecules, as determined by experimental crystal density difference diffraction.[198] The 8 molecule unit cell was transformed to a 128 molecule cell utilizing the *I*4 symmetry operators (2x2x4). The experimental conformation was utilized in the initial configuration, which contained all hydrogen atoms except one methyl hydrogen atom. The missing methyl hydrogen atom was added based on the standard tetrahedral configuration of methyl groups. Subsequently, no position or symmetry restraints were

applied to the box dimensions and they were allowed to vary over the course of the MD simulations as implemented in the SANDER module of AMBER8.

As noted previously, exclusion of the aliphatic hydrogen atoms during the charge fitting led to a robust partial charge set for monosaccharides. It is not clear as to whether the Hα atom in amino acids will be required to hold a partial charge. Therefore, two separate implementations of the GLYCAM04-LP asymmetric lone-pair model were made for the monomers in this crystal structure. LP(H) included Hα in the partial charge fitting, while LP(noH) did not, (q(Hα) = 0). Partial charges were obtained utilizing only the crystal structure, as opposed to an ensemble of structures. The hydrogen atoms in the amino group were constrained to have identical charges as were the oxygen atoms in the carboxylate group. For the LP(H) parameter set the Hα atom was predicted to have a partial atomic charge of 0.012 *e* and 0.0615 *e*, in the C-terminal and N-terminal alanine residues, respectively.

Unit scaling factors for 1,4 non-bonded interactions were employed during the MD simulations of carbohydrates, which is consistent with the GLYCAM2000 and GLYCAM04 force fields.[13,70] For comparison purposes, the AMBER recommended values for the scaling factors for the 1,4 van der Waals (1/2.0) and electrostatic (1/1.2) interactions were also employed. The Particle Mesh Ewald algorithm was implemented for treatment of long-range interactions. Each system was heated from 5 to 300 K (experimental temperature range was 283-303 K) over 50 ps and subsequently maintained at 300 K for 1 ns utilizing the Berendsen temperature coupling scheme.[53,198] Data from only the last 500 ps were utilized to determine the average cell dimensions and hydrogen-bond distances.

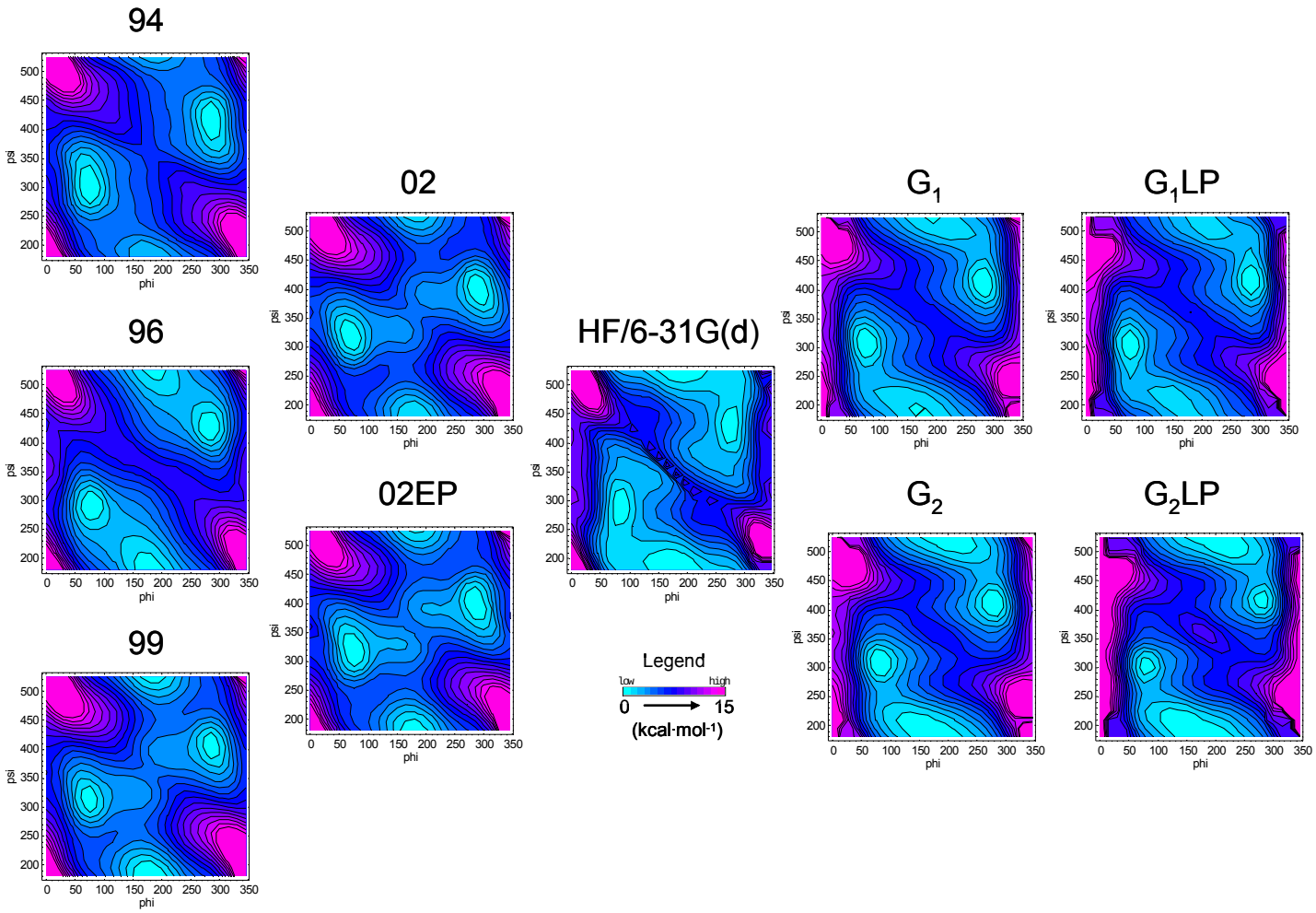
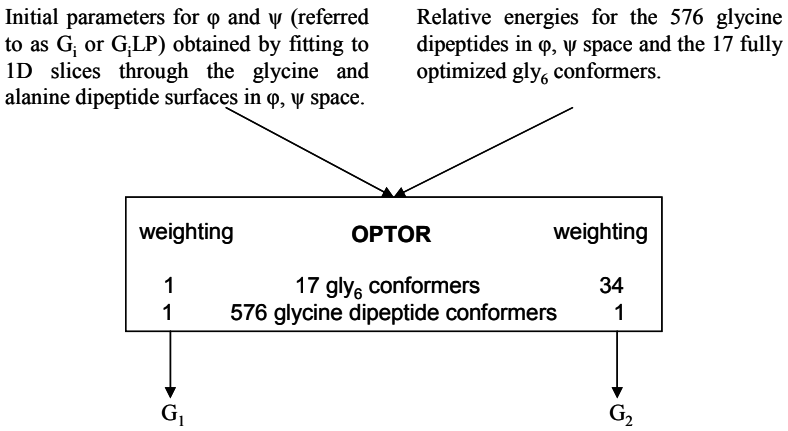


Figure 7.6 Potential energy surfaces of the glycine dipeptide obtained with various parameter sets and quantum mechanically. Each line represents 1 kcal·mol^{-1} and a spread of 15 kcal·mol^{-1} is shown



Flowchart 7.1 Protocol for the modification of the PARM94 parameter set to reproduce the relevant energetics for the glycine dipeptides and gly_6 peptides via modification of the backbone torsion parameters leading to the G_1 and G_2 parameter sets.

Results: Parameters for the φ and ψ Torsions of the Peptide Backbone

Up to this stage, only the alanine dipeptide and ala_4 conformers were utilized in the derivation of backbone torsion parameters for the AMBER parameter sets. Here, the glycine dipeptide and gly_6 peptides were utilized to determine parameters relevant for the φ and ψ torsions, while the alanine dipeptide, ala_4, ala_6, and ala_8 peptides were utilized to determine the torsion parameters for χ_1 and χ_2. The glycine surface has a single plane of symmetry yielding two unique local minima, with one representing the fully extended peptide, Figure 7.6. The relative energies for sixteen gly_6 conformers are reported relative to that of the fully extended polypeptide. Hexapeptides were chosen due to their ability to form structurally similar conformations to larger polypeptides and proteins; a feature that a dipeptide cannot fully achieve. Five out of the sixteen structures formed at least one intramolecular hydrogen bond, Figure 7.7.

Two different weighting schemes were implemented to distinguish their effect on the resultant torsion parameters. Scheme G_1 has more emphasis on the glycine dipeptide, while the emphasis on gly_6 was increased in scheme G_2, Flowchart 7.1. The same

relative trends were observed regardless of the inclusion of LPs for the values of the Fourier coefficients. As the weighting on the gly_6 peptides was increased, the value of V_1 (1.59→2.17) and V_2 (0.97→1.07) increased, while V_3 (0.34→0) and V_4 (0.16→0.03) decreased for the φ angle, and, similarly, the magnitude of V_1 (0.60→0.62), V_2(-1.54→-1.74), and V_4 (0.30→0.44) increased, while V_3 (-0.22→0.04) decreased for the ψ angle. Whereas the V_1 term for the ψ angle increased only negligibly without LPs from G_1 to G_2, the V_1 term tripled from G_1LP to G_2LP making the extended structure of glycine increasingly favorable. This can be clearly seen in the energetically more unfavorable values for structures II-XVII relative to the extended structure computed with the G_2LP parameter set relative to the G_1LP parameter set. In addition, the heavier weighting on the hexapeptides in schemes G_2 and G_2LP results in lower RMSDs from the HF/6-31G(d) energies, 1.7 Å and 2.3 Å, respectively (Table 7.3). The G_2 parameters yield a slightly smaller deviation in the relative energies for the gly_6 conformers than those obtained with G_2LP parameters, which may be due to the symmetry of the glycine residue and the asymmetry of the amide oxygen lone pair model, but both G_2 and G_2LP models are more accurate than the AMBER parameter sets, Table 7.3.

All four newly developed parameter sets yield surfaces for the glycine dipeptide that are in qualitative agreement with the quantum derived surface, even when more emphasis was placed on reproduction of the hexapeptide relative energies. In contrast, only the PARM94 and 96 AMBER parameter sets yields qualitatively correct surfaces, with all other models having a low energy surface expanding towards the center of the 2D plot at φ = 180° and ψ = 0°.

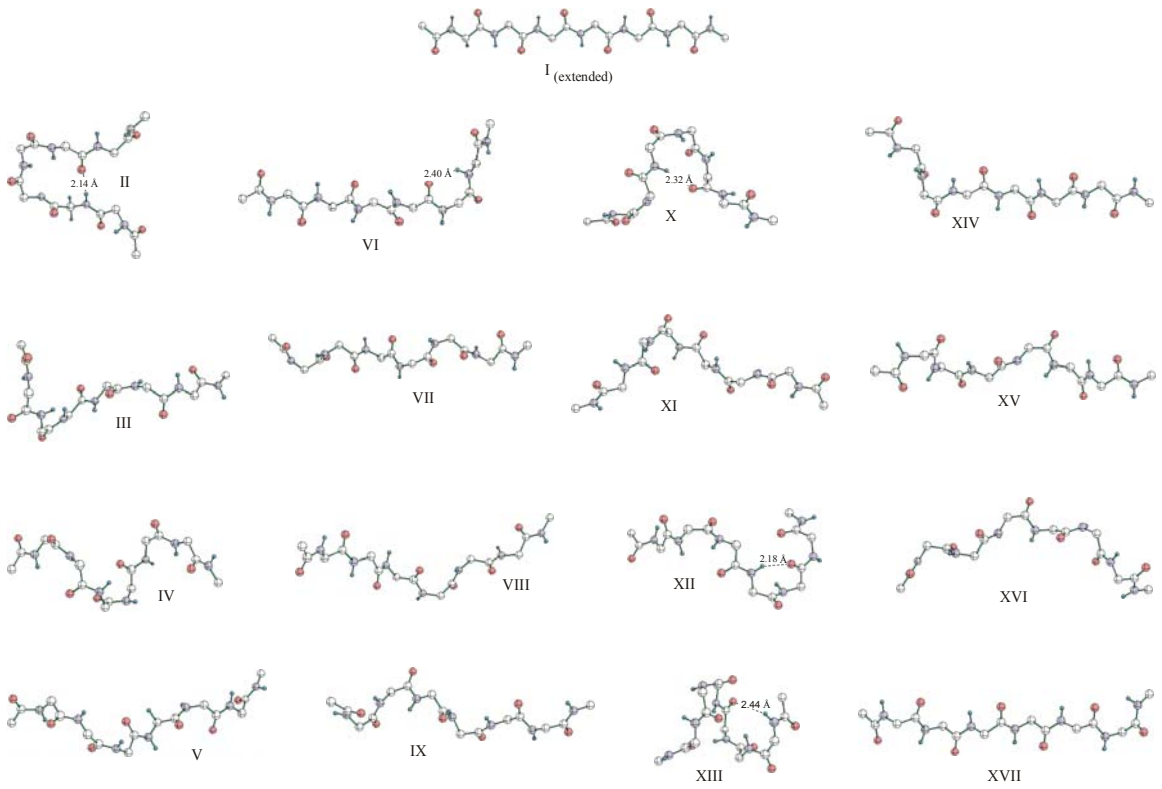


Figure 7.7 HF/6-31G(d) optimized conformations for the 17 gly_6 minima on the φ, ψ grid, with intramolecular hydrogen bonds shown in structures II, VI, X, XII, and XIII.

Table 7.3 Energies for the gly_6 peptides relative to structure I (extended), with the five most accurate models shown in bold.[a]

						AMBER-PARM				
	QM[b]	**G_1**	**G_2**	G_1LP	**G_2LP**	**94**	**96**	99	99[c]	02EP[c]
II	24.5	**22.4**	**24.0**	14.3	**22.6**	**30.4**	**32.6**	36.4	-1.1	-0.8
III	27.9	**23.8**	**24.5**	16.2	**23.0**	**29.1**	**29.3**	33.1	-0.3	-0.1
IV	34.8	**33.7**	**36.1**	25.3	**33.7**	**30.3**	**30.6**	41.3	-0.9	0.2
V	25.4	**25.1**	**26.1**	20.0	**24.0**	**25.5**	**22.7**	41.6	0.1	0.4
VI	15.3	**13.1**	**15.8**	7.4	**16.1**	**13.6**	**13.9**	24.7	3.5	4.2
VII	15.3	**7.6**	**12.7**	3.4	**12.2**	**13.5**	**14.0**	27.4	4.9	5.2
VIII	11.3	**7.4**	**9.3**	4.8	**9.0**	**13.2**	**12.2**	21.3	2.2	2.4
IX	19.0	**17.4**	**18.1**	11.1	**18.9**	**16.8**	**13.5**	23.7	3.5	4.0
X	31.6	**29.5**	**31.7**	21.3	**29.1**	**25.0**	**20.6**	36.4	-0.7	-0.5
XI	31.7	**26.8**	**30.0**	19.0	**27.8**	**26.1**	**22.5**	30.4	-0.6	0.5
XII	24.9	**24.2**	**26.6**	18.6	**24.7**	**25.1**	**25.3**	38.0	0.3	0.5
XIII	27.0	**20.5**	**23.4**	13.5	**23.4**	**22.1**	**25.9**	35.7	-0.2	3.0
XIV	17.8	**16.7**	**18.0**	10.6	**17.5**	**19.8**	**18.0**	25.4	2.0	2.4
XV	16.0	**13.9**	**14.3**	7.9	**14.9**	**9.2**	**11.3**	19.9	3.1	3.9
XVI	13.8	**11.6**	**13.2**	7.6	**13.0**	**12.2**	**17.3**	22.7	3.3	4.4
XVII	14.5	**11.3**	**13.6**	6.7	**13.1**	**17.3**	**15.4**	27.7	3.2	4.0
RMSD		**3.5**	**1.7**	9.3	**2.3**	**3.8**	**4.8**	9.5	22.4	21.7

[a] $kcal \cdot mol^{-1}$
[b] HF/6-31G(d)
[c] Polarization turned on during the energy minimization

Results: Parameters for the χ_1 and χ_2 Torsions

Since the larger alanyl peptides were optimized at the HF/6-31G(d) level, this methodology was employed for consistency in all structures utilized in the parameterization of χ_1 and χ_2. Because reproduction of the ala_2 minima has been a focal point of protein force field development studies, much higher levels of theory have been applied to this system, Table 7.4. Notably, the relative energies of the $C7_{ax}$ and β_2 structures reverse as higher levels of theory are employed; however, the energy difference is less than 1 $kcal \cdot mol^{-1}$. In addition, here larger polymers were utilized in the parameter derivation, which decreased the effect of this error by placing a smaller emphasis on the importance of the alanine dipeptide results.

Table 7.4 Quantum mechanical comparison of low energy minima of the alanine dipeptide relative to $C7_{eq}$.[a]

Energy	Geometry	C5	$C7_{ax}$	B_2	α_L	α'
HF/6-31G(d,p)	HF/6-31G(d,p)	0.95	2.75	2.67	4.31	5.51
HF/cc-pVTZ	HF/6-31G(d,p)	0.84	2.53	2.63	4.28	5.45
MP2/ cc-pVTZ(-f)	HF/6-31G(d,p)	1.33	3.14	2.64	4.40	5.84
MP2/TZP	HF/6-31G(d,p)	1.47	3.25	2.05	4.42	NR
B3LYP/6-31G(d)[c]	B3LYP/6-31G(d)	1.43	2.61	3.18	5.82	6.85
BLYP/ TZVP+[b]	B3LYP/TZVP+	1.16	2.46	2.78	5.81	6.67
MP2/ aug-cc-pvDZ[b]	B3LYP/TZVP+	1.77	2.36	3.36	5.02	6.88
LMP2/ cc-pVTZ(-f)	MP2/6-31G(d)	1.11	2.48	2.78	4.36	5.49
MP2/6-31G(d)[c]	MP2/6-31G(d)	1.76	2.58	3.37	4.6-	6.34
MP2/ cc-pVTZ(-f)	MP2/6-31G(d)	1.61	2.41	3.27	4.38	5.94
MP2/ TZP	MP2/6-31G(d)	1.86	2.13	3.42	4.4-	5.67
MP4/ cc-pVTZ(-f)	MP2/6-31G(d)	1.39	2.48	3.05	4.23	5.62
MP4-BSSE/cc-pVTZ(-f)	MP2/6-31G(d)	0.89	2.55	2.56	4.21	5.17
MP2/aug-cc-pVDZ[b]	MP2/aug-cc-pVDZ	1.91	2.28	3.11	4.50	6.51
MP2/aug-cc-pVTZ[b]	MP2/aug-cc-pVDZ	1.87	1.96	3.00	4.59	6.59
MP2/aug-cc-pVQZ[b]	MP2/aug-cc-pVDZ	1.59	2.37	3.21	4.95	6.72
MP2/CBS-limit[b]	MP2/aug-cc-pVDZ	1.39	2.66	3.35	5.19	6.80

[a] Unless otherwise noted, reference is Beachy *et al.*, 1997 and all values are in kcal·mol^{-1}.[187]
[b] From Vargas *et al.,* 2002.[199]
[c] From Jalkanen *et al.*, 1996.[200]

As larger alanyl peptides containing common secondary structural elements observed in proteins are introduced, the role played by hydrogen bonding is amplified due the larger number of hydrogen bonds present, Figure 7.8. Since hydrogen bonding is not a simple additive effect in peptides due to cooperativity and since hydrogen bonding is integral for the stabilization of protein secondary structure, the inclusion of models containing multiple hydrogen bonds in parameter derivation is desirable.[115] Within the six local minima on the ala_2 potential energy surface, weak hydrogen bonding interactions are observed in $C7_{eq}$ and $C7_{ax}$, in which weak hydrogen bonds are defined here as those having heavy-atom separations between 3 and 4 Å and the X-H···O hydrogen bond angle deviating far from 180°. As the dipeptide size increases to ala_4, helical and β-strand structures each form with three hydrogen bonds, while the β_2 conformer possesses two strong and two weak hydrogen bonds. Increasing polymer size

to ala_6 leads to two additional hydrogen bonds in the helical structures and three in the β-strand structure. In ala_6 the β_2 structure has an additional strong hydrogen bond compared to ala_4, but no further weak hydrogen bonds. Finally, ala_8 has 7 hydrogen bonds in both helical structures and 8 in the β-strand, Figure 7.8. The β_2 structure has four weak and four strong hydrogen bonds. Therefore, the hydrogen bonding contribution in each conformation does not increase systematically as polymer size grows from N=4 to N=8.

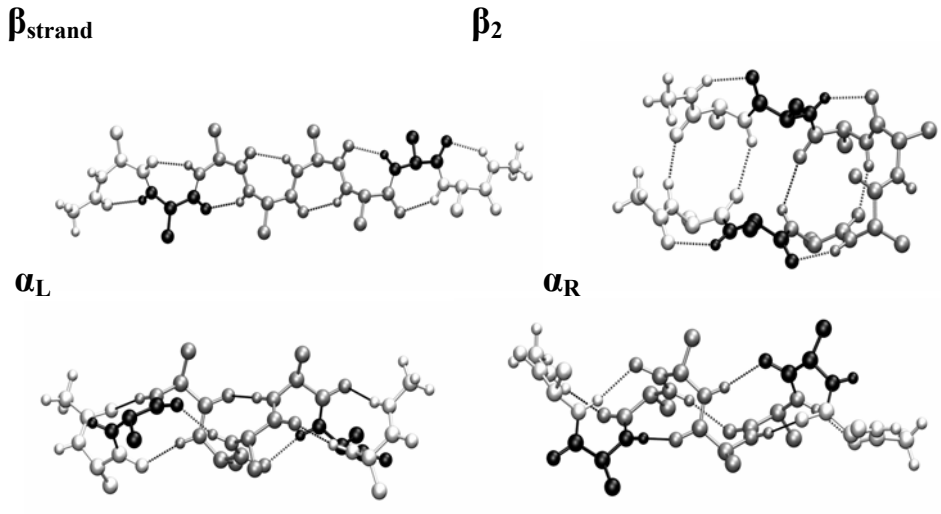
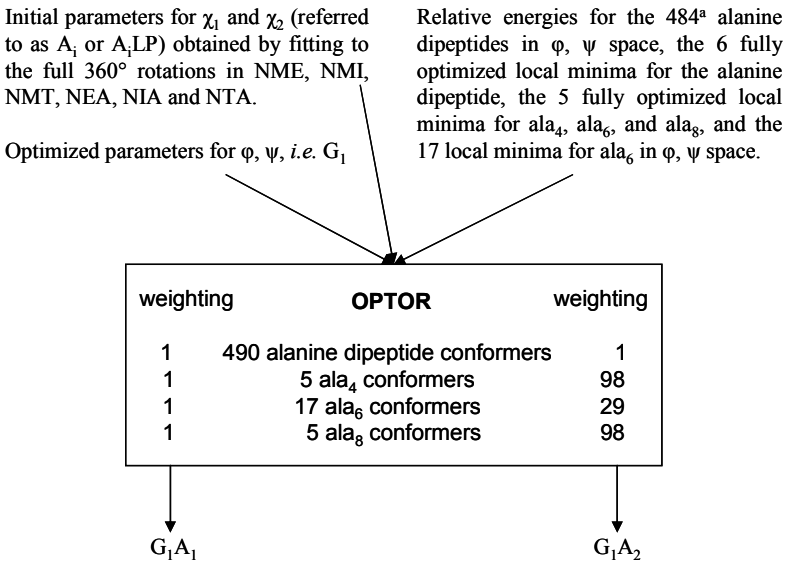


Figure 7.8 Overlay of four conformations of the ala_4 (silver), ala_6 (black), and ala_8 (white) peptides. Hydrogen bonds are illustrated with a dashed black line (HN···O).

The two weighting schemes for the χ_1 and χ_2 torsion angles, Flowchart 7.2, coupled with the G_1 and G_2 parameter sets as well as the inclusion or omission of LPs, leads to a more diverse group of Fourier coefficients, in which sign changes (also referred to as 180° phase shifts) are observed. Notably, there is a much larger difference observed between the A_1 and A_2 parameter sets versus those between the G_1 and G_2 sets, which most likely originates from the increased number of alanine polymers larger than the dipeptide relative to the glycine compounds.



Flowchart 7.2 Protocol for modification of the of the χ_1 and χ_2 torsion parameters in the G_1 parameter set to reproduce the relative energies for a series of alanyl peptides, yielding the G_1A_1 and G_1A_2 parameter sets. The same protocol applied to G_2LP led to G_2A_1LP and G_2A_2LP, *etc*. [a]Only the conformers within 11 kcal·mol^{-1} were included for the alanine dipeptide.

The relative energies of the alanine dipeptide computed utilizing the PARM94 and PARM99 force fields reported by Wang *et al.* were determined through energy minimizations in AMBER5 with no restraints on the torsion angles for $C7_{eq}$, $C7_{ax}$, and C5, while torsion restraints were required for the β_2, α_L, and α'.[160] In contrast, the AMBER7 program was implemented here and torsion restraints were required to maintain the quantum mechanical geometry for all structures. Previously, Wang *et al.* implemented a square bottom potential that allowed 5° of movement in the torsion angle, whereas the harmonic restraint employed here applies a penalty if the angle deviates more than 1° from the QM value.[160] Additionally, the starting points for the MM optimizations reported by Wang *et al.* were the HF/6-31G(d,p) geometry-optimized structures, as opposed to the HF/6-31G(d) geometries used here.[160] The 1,4 scaling factors for electrostatics and van der Waals interactions were 1/1.2 and 1/2.0,

respectively, and all electrostatics were included in the energy calculation in both the study by Wang *et al.* and here. The use of slightly different starting geometries and restraint protocols led to the minor deviation observed between the previous and present results obtained for each of the parameter sets (PARM94, 99, and 02), Table 7.5. If the parameters employed by Wang *et al.* were able to fully reproduce the QM data the deviation between the present values and those of Wang *et al.* would not exist, which is clear since the minimized geometries obtained by Wang *et al.* deviated from the quantum mechanically obtained structures. Using AMBER's polarizable force field, 02, which is presumed to be the most accurate and therefore the closest to reproducing the QM data, the deviation between the results from the two sets of data substantially decreases.

Table 7.5 Relative energies of alanine dipeptide minima relative to the $C7_{eq}$ conformation.[a]

	C5	$C7_{ax}$	β_2	α_L	α'	error[d]
QM[b]	0.41	2.82	2.58	4.72	5.74	
Parameters Presented Here						
G_1A_1	1.12	-0.55	5.69	7.88	8.92	2.70
G_1A_2	0.13	-1.05	5.58	8.50	8.63	2.76
G_2A_1	0.71	0.23	4.86	8.12	8.68	2.30
G_2A_2	0.26	0.11	4.83	8.70	8.82	2.43
G_1A_{1LP}	1.40	-0.60	5.47	7.21	8.59	2.53
G_1A_{2LP}	-0.01	-0.73	4.67	9.27	9.18	2.81
G_2A_{1LP}	0.31	0.52	5.60	7.73	9.04	2.35
G_2A_{2LP}	1.24	2.56	5.04	10.03	11.88	3.00
AMBER-PARM						
94	1.67	1.17	4.81	9.21	8.76	2.53
94[c]	1.5	1.48	3.64	4.5	6.71	
96	0.36	1.91	6.69	11.50	9.11	3.04
99	-0.53	1.27	2.09	7.68	6.02	1.24
99[c]	0.9	2.23	2.58	5.14	5.32	
02	0.21	0.67	2.02	4.90	6.26	0.72
02[c]	1.25	2.7	2.55	5.4	4.87	
02EP	0.24	0.93	2.03	4.89	6.23	0.65

[a] All energies reported in kcal·mol^{-1}
[b] HF/6-31G(d) energies.
[c] Energies reported by Wang *et al.* 2000.[160]
[d] Average unsigned error.

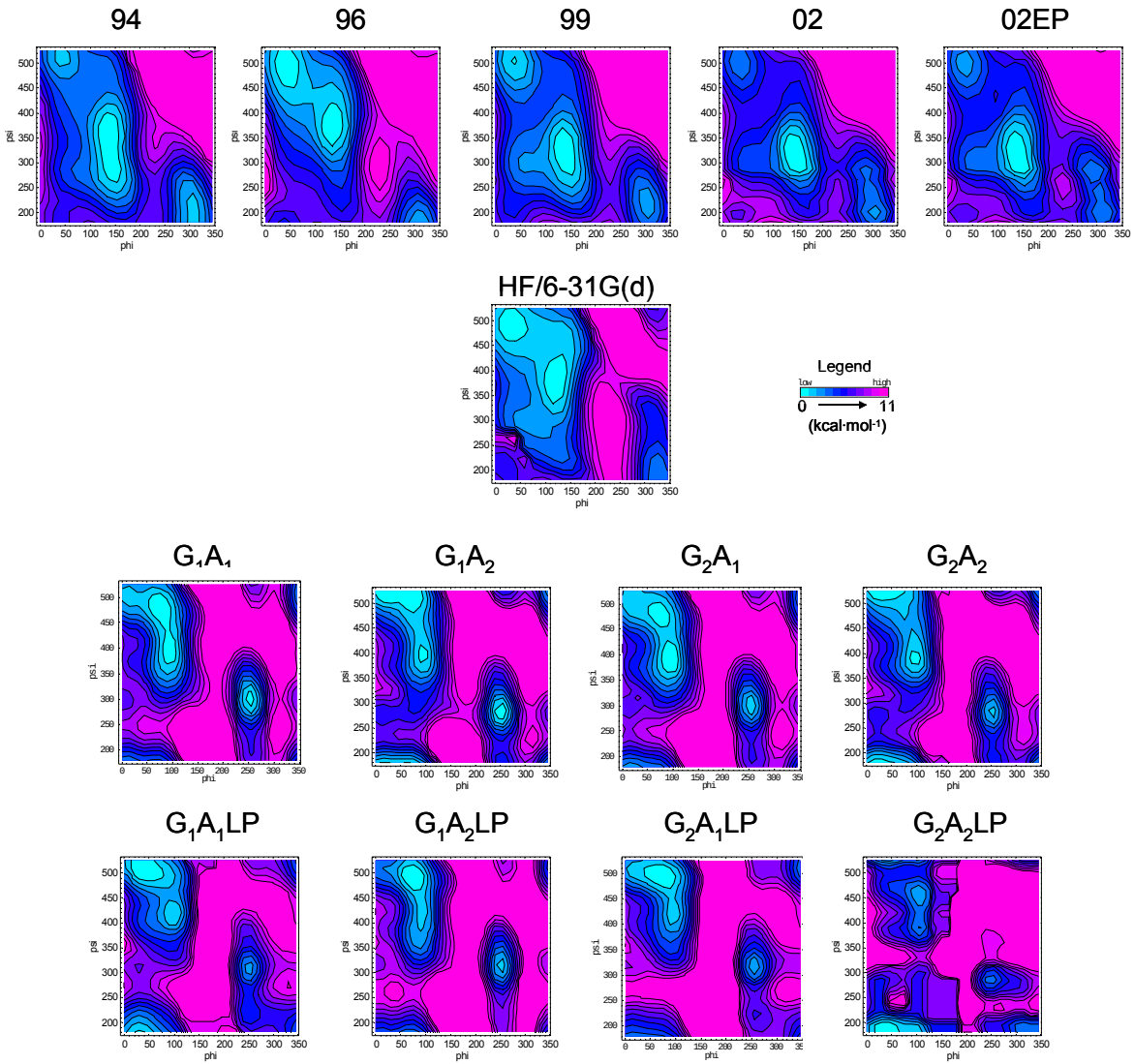


Figure 7.9 Energy surfaces of the alanine dipeptide spanning 11 kcal·mol^{-1} from the global minima (each line is at 1 kcal·mol^{-1}).

Table 7.6 Relative energies and errors for the alanyl peptides computed with various parameter sets and compared to the HF/6-31G(d) determined values.[a]

										AMBER PARM				
	QM[b]	G_1A_1	G_1A_2	G_2A_1	G_2A_2	G_1A_1LP	G_1A_2LP	G_2A_1LP	G_2A_2LP	94	96	99	02[c]	02EP[c]
ala_4														
α_L	10.6	1.0	11.2	0.9	0.9	0.9	0.8	0.9	0.8	12.2	20.3	6.1	0.5	0.1
α_R	4.6	0.3	20.2	0.2	0.2	0.3	0.3	0.3	0.2	11.5	18.4	10.2	1.7	1.2
β_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
β-strand	5.2	-0.2	5.5	-0.2	-0.2	-0.3	-0.2	-0.2	-0.2	9.4	7.6	7.5	0.2	-0.1
Extended	5.7	-0.2	4.6	-0.2	-0.2	-0.3	-0.2	-0.3	-0.3	13.5	6.4	4.5	-0.4	-0.1
error[d]		6.3	4.4	6.3	6.3	6.3	6.3	6.4	6.4	5.1	6.7	3.4	6.0	6.3
ala_6														
α_L	9.3	10.5	13.7	11.1	1.4	9.5	14.6	14.8	15.5	10.0	26.3	3.5	1.4	0.7
α_R	1.0	1.5	4.1	1.9	0.3	-0.7	3.6	4.7	6.1	-11.1	8.3	-13.4	2.3	0.9
β_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
β-strand	5.2	10.8	11.0	10.1	-0.1	10.8	9.3	12.1	10.5	12.9	14.8	12.8	0.1	-0.1
Extended	6.1	16.7	9.7	13.3	-0.1	22.0	4.7	15.3	8.6	19.3	13.0	8.7	-0.7	-2.2
error[d]		4.4	4.2	3.7	5.0	5.8	3.4	6.3	4.8	8.4	10.2	7.6	5.3	5.6
ala_6 (grid)														
I	14.0	1.3	1.2	12.7	14.0	13.9	15.8	14.0	15.8	25.4	22.7	20.9	-1.0	-0.4
II	12.2	1.4	1.2	10.8	11.4	12.4	12.4	13.5	12.9	19.4	16.5	15.2	-0.8	0.6
III	22.4	1.3	1.1	16.5	19.1	16.2	22.4	14.2	20.1	38.1	34.0	33.9	-1.3	-0.7
IV	19.2	1.1	0.9	19.6	22.2	19.2	25.9	17.0	24.3	39.7	34.5	36.2	-1.5	-0.2
V	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VI	14.3	0.4	0.4	9.4	9.9	8.1	9.1	10.4	12.1	9.7	9.6	9.3	-0.5	0.4
VII	6.7	0.3	0.3	2.8	1.0	3.6	-3.1	4.9	-3.6	3.2	2.9	3.2	-0.2	1.0
VIII	10.6	0.4	0.4	6.6	4.1	7.2	-2.4	9.3	2.6	4.0	1.0	3.8	-0.4	0.0
IX	4.4	-0.1	-0.1	2.8	1.9	6.7	2.0	2.8	-2.5	20.1	4.6	20.8	-2.6	-1.5
X	8.1	-0.3	-0.3	9.8	9.6	12.4	7.1	9.6	6.7	29.8	11.9	30.7	-1.8	-1.0
XI	6.8	-0.2	-0.2	8.1	7.7	11.4	6.7	7.8	5.4	26.7	9.6	27.2	-2.1	-0.9
XII	5.7	-0.2	-0.2	5.7	5.3	9.2	4.4	6.0	2.9	25.3	9.1	25.4	-2.2	-1.4
error[d]		10.8	10.9	2.4	2.6	3.1	3.8	2.1	3.9	13.3	6.2	12.1	12.6	11.7
ala_8														
α_L	10.8	11.1	15.8	11.7	15.6	9.8	18.4	13.9	18.3	10.8	28.9	-0.6	1.9	2.0
α_R	0.2	1.1	4.8	1.4	4.4	-3.0	4.7	0.7	6.6	-16.4	5.7	-21.8	3.4	3.3
β_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
β-strand	8.9	15.5	16.4	14.4	16.5	14.1	14.4	13.0	14.9	19.6	18.8	16.9	-0.3	0.7
Extended	10.3	23.6	14.9	19.0	14.5	28.9	7.9	17.2	11.7	28.3	16.6	11.8	-1.3	-2.0
error[d]		5.3	5.4	4.1	5.2	7.0	5.0	3.7	5.4	11.3	10.0	10.7	8.2	8.1
error[d] ($ala_{n;n=2-8}$)		7.0	6.8	3.4	3.8	4.4	4.1	3.6	4.4	9.2	6.8	8.1	7.9	7.5
error[d] ($ala_{n;n=4-8}$)		8.0	7.6	3.6	4.2	4.8	4.4	3.9	4.7	10.7	7.6	9.6	9.4	9.0
error[d] ($ala_{n;n=6,8}$)		8.3	8.3	3.0	3.7	4.5	4.0	3.3	4.4	11.9	7.8	10.9	10.1	9.6

[a] All energies reported in kcal·mol^{-1} and energies highlighted in gray are the best lone-pair and non-LP parameter sets presented here.
[b] HF/6-31G(d) energies.
[c] Energy minimization was done with atom-type polarizabilities.
[d] Average unsigned error.

The AMBER parameter sets are in much better agreement with the quantum mechanical potential energy surface of the alanine dipeptide than are the parameter sets derived here, Figure 7.9. The error from the QM values to the molecular mechanics values computed with parameter sets derived here for the local minima of the alanine dipeptide are approximately three times larger (averaging 2.61 kcal·mol^{-1}) than the error from the 02EP parameters (0.65 kcal·mol^{-1}). The removal of polarization in the AMBER calculation, referred to as PARM99, yields an error of 1.24 kcal·mol^{-1}, which decreases the average error in the parameter sets derived here (type G_xA_x) from three-fold to two-fold worse than the AMBER parameters. Despite the differences in the ability to reproduce the relative energies of the peptide, the energetic ordering for the six conformers was reproduced with the parameter sets presented here with the exception of the occasional reversal of C5 and $C7_{ax}$.

As seen with the potential energy surface of the glycine dipeptide, the qualitative potential energy surface computed with the AMBER parameter sets, with the exception of PARM96, yields an incorrect connectivity between the two lowest energy minima, C5 and $C7_{eq}$, Figure 7.9. In contrast, this intermediate region was reproduced quite well with the parameter sets presented here. Only the energy values within 11 kcal·mol^{-1} of the global minimum are plotted for comparison. Introduction of larger alanine oligomers, which contain φ,ψ torsion angles representative of the secondary structure observed in proteins, results in an increase in the magnitude of alanine dipeptides that are more than 11 kcal·mol^{-1} from the global minima, Figure 7.2. The PARM96 parameter set yields the most accurate alanine dipeptide surface, but has been found to perform the worse in MD simulations of larger peptide systems.[23] Therefore, this highlights the importance of

accurately modeling the low-energy conformations of the alanine dipeptide, while the high-energy regions may not be as important and energetically quite different for larger alanyl polymers.

Application of the newly derived parameters to the conformeres of ala_4 resulted in error values, relative to the HF/6-31G(d)//HF/6-31G(d) energies, that are approximately equal to the relative quantum mechanically computed energy differences. In addition, the relative ranking of the ala_4 conformers is non-existent since they all have nearly the same energy. The AMBER parameters yield errors on the same order of magnitude, but the relative energies between the ala_4 conformers are more substantial. Only in the case of PARM99 does the error decrease below 4 $kcal \cdot mol^{-1}$, but the relative ranking of ala_4 conformers is still incorrect. Notably, inclusion of either polarizability alone, or polarizability with lone-pairs, with the PARM99 parameters (02 or 02EP) greatly reduces the accuracy of these parameters, yielding errors on the order of magnitude as those presented here (6.0 and 6.3 $kcal \cdot mol^{-1}$, respectively). In general, application to larger oligomers yields an increase in accuracy for the parameters developed here relative to the results obtained for the alanine dipeptide and ala_4, while the accuracy of the AMBER parameter sets decreases, Table 7.6.

The most notable difference between the relative energies computed using the optimized parameters developed here and the AMBER parameter sets is the overestimation of the stability of α_R in ala_6 and ala_8 in the 94 and 99 parameter sets, Table 7.6. PARM96 has a 7 $kcal \cdot mol^{-1}$ destabilization of the α_R conformation of ala_4 in comparison to PARM94 and 99, in which the energy of α_R is relative to the β_2 conformer with both parameter sets. The relative destabilization increases in magnitude over 3-fold

from ala_4 to ala_6 and ala_8, which is in agreement with the observed overstabilization of α-helical conformations in systems employing the PARM94 and 99 parameter sets.[165,167,188,189]

In addition to developing an optimized Cornell *et al.* type protein parameter set, the ability of the PARM02EP lone-pair model has been assessed. Since the backbone parameters in PARM02EP were not explicitly developed for lone-pairs (LPs, extra points in (EP) in AMBER nomenclature) the difference between the PARM99 and PARM02EP parameter sets is only the inclusion of lone-pairs and atomic polarizabilities, while the only difference between PARM02 and PARM02EP is the inclusion of lone-pairs. Minimal improvement was observed during the addition of polarizability, in which the error in reproduction of the QM energies for the low energy conformers of ala_2, ala_4, ala_6, and ala_8 decreased from 8.1 to 7.9 kcal·mol^{-1} for the 99 and 02 parameters, respectively. Inclusion of lone-pairs, PARM02EP, further decreases the error to 7.5 kcal·mol^{-1}, but the error is still twice the value of the G_2A_1 and $G2A_1LP$ parameters. Therefore, as currently implemented, without specific reoptimization of the torsion parameters, inclusion of LPs at 0.3 Å around all oxygen atoms does not lead to a substantial improvement over the non-LP model.

The optimal non-LP and LP-inclusive parameter sets developed here are G_2A_1 and G_2A_1LP, respectively. Inclusion of ala_8 was necessary to show that reproduction of ala_6 polymers and not ala_4 and ala_2 was not an anomaly, but required to correctly model larger proteins. The relative ranking between the α_R, β_2 and β-strand structures is most accurately reproduced with the G_2A_1 and G_2A_1LP parameters of all the parameter sets investigated here. Therefore, based on the anticipated behavior of these parameters in

larger polymers, the right energetic balance appears to have been obtained between α-helical and β-strand secondary structures. In conclusion, non-additive effects, such as hydrogen bond cooperativity, that are exhibited in larger polymers, but cannot be accounted for utilizing only the alanine dipeptide and ala_4 peptide as model compounds for non-polarizable protein parameter set development, need to be empirically taken into account during the parameterization process. In the parameter sets developed here, larger oligomers were utilized, whereas in CHARMM a correlation function was introduced in an attempt to account for these effects.[191]

Crystalline-Phase MD Simulation of the Alanine Dipeptide Zwitterion

To assess the effect inclusion of lone-pair has on the distortion of the crystal lattice of peptides, a crystal-phase MD simulation was performed on a small peptide. Since the outcome of the simulations depends almost exclusively on the non-bonded parameters, each of the newly optimized parameter sets was not examined. Therefore, the PARM99 parameters were implemented with the asymmetric lone pair model, which isolates the effect of the lone-pairs from any differences in torsion parameters. Two charge sets were developed utilizing the charge-fitting methodology employed in GLYCAM, with the exception that ensemble averaged partial charges were not employed. To determine the necessity of including a charge on H_α, two charge sets, which both included asymmetric-LPs on the amide carbonyl oxygen atom, were utilized, in which the nomenclature utilized to refer to the charge set when H_α was included in the charge fitting is LP(H), while LP(noH) refers to the exclusion of H_α from the fit.

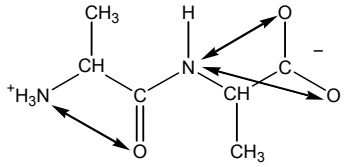


Figure 7.10 Structure of the Ala-Ala (ala_2) zwitterion, with the 1,4 interactions scaled in CHARMM illuminated by the double-headed arrows.

Table 7.7 Deviation in hydrogen bonding and cell dimensions in the ala_2 zwitterion crystal.

				average cell deviations (Å)				average percent cell deviations				ΔHB_{NH-O}	
	A	B	C	ΔA	ΔB	ΔC	<ΔA,B,C>	ΔA	ΔB	Δ C	<ΔA,B,C>	(Å)	(%)
isotropic pressure scaling[a]													
Exp.	35.97	35.97	20.62	-	-	-	-	-	-	-	-		
PARM94	35.47	35.47	20.33	-0.50	-0.50	-0.28	0.43	-	-	-	1.38	0.06	2.93
PARM96	35.50	35.50	20.34	-0.47	-0.47	-0.27	0.41	-	-	-	1.31	0.06	2.92
PARM98	35.47	35.47	20.33	-0.50	-0.50	-0.28	0.43	-	-	-	1.38	0.06	2.93
PARM99	35.52	35.52	20.36	-0.45	-0.45	-0.26	0.39	-	-	-	1.26	0.06	3.12
PARM02	35.44	35.44	20.31	-0.53	-0.53	-0.30	0.45	-	-	-	1.47	0.08	4.19
PARM02EP	35.59	35.59	20.40	-0.38	-0.38	-0.22	0.33	-	-	-	1.07	0.01	0.28
LP(H)	36.22	36.22	20.76	0.25	0.25	0.15	0.22	-	-	-	0.71	0.06	3.00
LP(noH)	36.22	36.22	20.77	0.25	0.25	0.15	0.22	-	-	-	0.72	0.06	2.90
anisotropic pressure scaling													
Exp.	35.97	35.97	20.62	-	-	-	-	-	-	-	-	-	-
PARM94	35.63±0.03	35.63±0.02	20.08±0.02	-0.34	-0.34	-0.54	0.41	0.95	0.94	2.60	1.50	0.06	2.94
PARM96	35.64±0.03	35.64±0.03	20.12±0.02	-0.33	-0.33	-0.49	0.38	0.91	0.92	2.38	1.40	0.06	3.01
PARM98	35.63±0.03	35.63±0.02	20.08±0.02	-0.34	-0.34	-0.54	0.41	0.95	0.94	2.60	1.50	0.06	2.94
PARM99	35.64±0.03	35.64±0.02	20.17±0.02	-0.33	-0.33	-0.45	0.38	0.92	0.93	2.16	1.34	0.06	3.16
PARM02	36.17±0.03	36.16±0.04	19.24±0.03	0.20	0.19	-1.38	0.59	0.56	0.54	6.69	2.59	0.08	4.29
PARM02EP	37.79±0.05	35.10±0.04	19.54±0.03	1.82	-0.87	-1.07	1.25	5.07	2.41	5.20	4.23	<0.01	0.19
LP(H)	35.90±0.02	35.90±0.02	21.20±0.01	-0.07	-0.07	0.58	0.24	0.20	0.19	2.83	1.07	0.06	3.09
LP(noH)	35.89±0.02	35.89±0.02	21.23±0.01	-0.08	-0.08	0.61	0.26	0.21	0.22	2.96	1.13	0.06	3.02

[a] All simulations with isotropic pressure scaling resulted in standard deviations of 0.01 Å in cell dimensions.

The *I*4 space group inherent to the crystal structure of the ala_2 zwitterion, allows MD simulation with both isotropic and anisotropic pressure scaling. The deviation of the crystal cell dimensions was similar with all models if isotropic scaling was employed, Table 7.7. Notably, the LP(H) and LP(noH) models yielded the lowest deviation, while the 02EP model was the next best model, Table 6.7. Inclusion of polarization into the PARM99 parameter set, PARM02, actually resulted in a worse reproduction of the crystal unit cell dimensions. Unlike in the α-Glc*p*NAc crystal MD simulation, no significant deviation from planarity in the HNCO angle of ala_2 was present. The peptide torsion angle is 4° closer to planarity at 173°, relative to a value of 169° for the same torsion angle in α-Glc*p*NAc. All hydrogen bonds are formed between either the amide or amino hydrogen atoms and the oxygen atoms in the carboxylate group. Due to rotation of the amino group during the MD simulation, the hydrogen bonds formed with the amino hydrogen atoms were not included in the hydrogen bond statistics. All models implemented here reproduce the interresidue hydrogen bond distances and angles to within 5% of the experimental values, with PARM02EP yielding the closest agreement, with ≤0.01 Å average deviation in the hydrogen bond distances. Anisotropic pressure scaling on ala_2 peptide had only a subtle, worsening effect on the reproduction of the experimental cell dimensions and hydrogen bond distances, which is in contrast to the large effect anisotropic scaling had on the crystal simulations of monosaccharides.

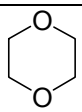
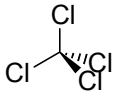
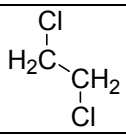
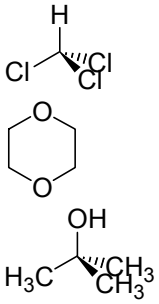
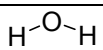
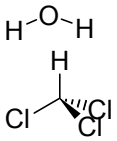
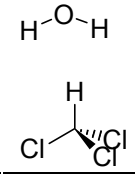
Notably, the LP models resulted in expansion of the cell dimensions for the isotropic simulations, while this expansion was observed only in the cell dimension C if anisotropic pressure scaling was implemented. All of the AMBER parameters resulted in a decrease of cell size for the isotropic simulations and expansion in the anisotropic

simulations was observed only in the A and B dimensions with PARM02 and only the A dimension of PARM02EP. Inclusion of the AMBER 1,4 scaling terms of 1/2.0 and 1/1.2 for van der Waals and electrostatic interactions led to a small contraction of the unit cell relative to the results obtained with the scaling factors of unity, increasing the accuracy of the LP(H) and LP(noH) models, while the AMBER models became more inaccurate since they have already resulted in an over – contraction of the cell. The necessity for the inclusion of 1,4 scaling for peptides was observed in development of the CHARMM force field, in which an atom-pair specific scaling is implemented rather than a general 1,4 scaling. Specifically, during the development of the CHARMM force fields, the interaction between the peptide oxygen and nitrogen was found to be too repulsive and a non-bonded scaling factor was added for this atom-pair, Figure 7.10.

Conclusions

In conclusion, the G_2A_1 and G_2A_1LP protein parameter sets are identified as the most promising parameter sets to continue development and testing. However, to determine the full accuracy of these two models, explicitly solvated simulations of small peptides and proteins need to be performed. A variety of information has been obtained experimentally that is relevant to the structure of the alanine dipeptide in solution. Specifically, that the conformation is solvent dependent and it is does not adopt an entirely disordered structure, Table 7.8. The effect of solvent on the conformations in alanine dipeptide may be analogous to the effect observed for the ω-angle population in monosaccharides.[7,201]

Table 7.8 Experimental assignments for the conformational distribution of ala_2 in a variety of solvents from oldest to most recent reports.

Experimental methods	Year	Solvent	Dielectric	Observations related to the conformational distribution for ala_2
CD[202]	1971		78.5	High population of polyproline II (PPII) conformation
CD[202]	1971		2.2	PPII conformation is unlikely
IR[203]	1972		2.24	70% C7 30% C5
ORD/CD/IR[204]	1974		16.7	> 1 conformer present
CD/NMR[205]	1980		4.8 2.2 10.9 36.6 33 78.5	C7 turns to α_R and PPII as we the solvent becomes more polar, while C5 remains constant. Qualitative estimate in polar solvents, while ±15% in nonpolar solvents
NMR[206]	2003		78.5	PPII dominant
NMR[207]	2004		78.5 4.8	PPII and α_R equal and quickly converting in solution. Only PPII like conformation in crystalline form and $CHCl_3$
Raman spectroscopy[208]	2004		78.5 4.8	PPII, α_R, and $C7_{eq}$ present with the amount of $C7_{eq}$ increasing from water to chloroform.

Two additional avenues for the validation of the G_2A_1 and G_2A_1LP parameters are thus projected. One possibility lies in the monitoring of the percentage of backbone torsion angles that are not in the generally allowed regions for proteins, the Cα RMSD relative to the crystal structure, unfavorable side chain rotamers, and bad contacts a fully

folded protein structure adopts over the course of a room temperature MD simulation. Secondly, MD simulations in which the fully extended peptide folds to the correct structure, such as for those peptides examined in Table 7.1 will also assess the accuracy of the parameters presented here.

Overall, this assessment of the AMBER parameter sets has shown that although each set can reproduce the relative energies for the alanine dipeptide, the agreement quickly dissipates as polymer size is increased. In addition, the overstabilization of α-helical conformations observed in MD simulations of proteins with PARM94/98 is observed in peptides as small as ala_6. The present polarizability and lone-pair model as implemented in AMBER8 does not result in a substantial improvement in the systems investigated here. The G_2A_1 and G_2A_1LP parameter sets developed here offer a significant improvement over the present AMBER parameter sets for the model systems investigated here, and the inclusion of lone-pairs actually results in a more accurate model over its non-LP counterpart, which cannot be said of the PARM02EP parameter set.

8. CONCLUSIONS AND FUTURE DIRECTIONS

In recent years, emphasis has shifted from reproduction of gas-phase properties to solution-phase properties for validation of carbohydrate force fields. The current generation of carbohydrate parameter sets has been shown to reproduce solution-phase properties, such as homo- and heteronuclear ^{3}J-couplings, ω-angle rotamer populations, and rotational correlation times. With the increased amount of experimental data for larger glycans, and developments in force field methodology, a systematic analysis of the ability of each parameter set to reproduce condensed-phase properties of oligosaccharides can be envisioned.

Utilizing the quantum mechanical MEP to determine the distance of the lone-pairs from the respective oxygen atoms has led to a description of the molecular electrostatics that is consistent with the currently available TIP5P water model for the hydroxyl and ether type oxygen atoms. The aforementioned sp^3-type oxygen atoms each have a LP-O distance of 0.7 Å and a charge of zero on the oxygen atom. A shorter LP-O distance of 0.3 Å for ketones and an asymmetric LP arrangement for amides, with LP1-O and LP2-O distance of 0.7 and 0.3 Å, respectively, was determined to give the best fit between the quantum mechanical and classical MEP. For each fitting, constraining the oxygen, as well as the aliphatic hydrogen atoms, to a zero charge led to a robust partial charge set with very similar charges for the various atom types in similar environments, *i.e.* each lone-pair in a secondary alcohol group of α-D-Glc*p* has a partial charge of -0.2 *e*. Inclusion of the new lone-pair model with TIP5P consistently increased the accuracy for MM energy minimized geometries of complexes over those of the non-LP containing

model with TIP3P. In addition, the crystal MD simulations clearly illustrate the improved reproduction of electrostatic interactions when lone-pairs are included.

A complete carbohydrate parameter set must also include the possibility to model ionic sugars, such as sialic acid and uronic acids. These systems present unique developmental challenges and have not been included in comparison of parameter sets. In addition, the majority of protein parameter sets mentioned here are parameterized to be employed in conjunction with three point, planar water models, SPC/E or TIP3P, while a number of more accurate water models exist. Finally, as computer performance continues to increase, polarizable biomolecular MD simulations become a possibility. Therefore, the establishment of a protocol that substantially improves the accuracy of additive models can easily be employed in the development of a polarizable protein, as well as carbohydrate, parameter set.

Overall, the accuracy of carbohydrate parameter sets has improved dramatically in the past decade, and the methodologies have been applied to a variety of biologically relevant problems from structural determinations to the calculation of carbohydrate – protein interaction energies. By requiring the parameter set to reproduce ensemble properties that are inherently dynamic in their origin, carbohydrate parameter sets are being held to more stringent criteria than their protein counterparts. This rigor has the benefit that MD simulations of oligosaccharides can be used as a method to predict 3D structure, and can provide a much needed basis for interpreting and augmenting limited experimental data.

9. WORKS CITED

(1) H. Auchincloss and D. H. Sachs Xenogeneic transplantation *Annu. Rev. Immunol.* **1998**, 16, 433-470; R. A. Dwek Glycobiology: Toward understanding the function of sugars *Chem. Rev.* **1996**, 96(2), 683-720; S. E. Dumitriu Polysaccharides: Structural Diversity and Functional Versatility; Marcel Dekker: New York, 1998; U. Galili, M. R. Clark*, et al.* Evolutionary Relationship between the Natural Anti-Gal Antibody and the Gal-Alpha-1- 3gal Epitope in Primates *Proc. Natl. Acad. Sci. USA* **1987**, 84(5), 1369-1373; A. H. Good, D. K. C. Cooper*, et al.* Identification of Carbohydrate Structures That Bind Human Antiporcine Antibodies - Implications for Discordant Xenografting in Humans *Transplant. Proc.* **1992**, 24(2), 559-562; J. L. Platt and F. H. Bach Discordant Xenografting - Challenges and Controversies *Curr. Opin. Struct. Immun.* **1991**, 3(5), 735-739; M. S. Sandrin, H. A. Vaughan*, et al.* Anti-Pig Igm Antibodies in Human Serum React Predominantly with Gal(Alpha-1-3)Gal Epitopes *Proc. Natl. Acad. Sci. USA* **1993**, 90(23), 11391-11395.

(2) M. R. Wormald, A. J. Petrescu*, et al.* Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, x-ray crystallography, and molecular modelling *Chem. Rev.* **2002**, 102, 371-386; A. Imberty and S. Perez Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations *Chem. Rev.* **2000**, 100, 4567-4588; R. J. Woods Computational carbohydrate chemistry: what theoretical methods can tell us *Glycoconjugate Journal* **1998**, 15, 209-216.

(3) M. Karplus Contact Electron-Spin Coupling of Nuclear Magnetic Moments *J. Chem. Phys.* **1959**, 30(1), 11-15; M. Karplus Vicinal Proton Coupling in Nuclear Magnetic Resonance *J. Am. Chem. Soc.* **1963**, 85(18), 2870-&.

(4) J. Gonzalez-Outeirino, K. N. Kirschner*, et al.* Reconciling solvent effects on rotamer populations in carbohydrates: a joint MD and NMR analysis *Canadian Journal of Chemistry* **2005,** *in press*.

(5) H. Ohrui, Y. Nishida*, et al.* Preferred Conformation About the C5-C6 Bond of N-Acetylneuraminyl(2-6)-D-Galactopyranosides and N-Acetylneuraminyl(2-6)-D-Glucopyranosides in Solution *J Org Chem* **1991**, 56(5), 1726-1731; E. Q. Morales, J. I. Padron*, et al.* Cd and H-1-Nmr Study of the Rotational Population Dependence of the Hydroxymethyl Group in Beta-Glucopyranosides on the Aglycon and Its Absolute-Configuration *J Org Chem* **1995**, 60(8), 2537-2548; I. Tvaroska, F. R. Taravel*, et al.* Quantum mechanical and NMR spectroscopy studies on the conformations of the hydroxymethyl and methoxymethyl groups in aldohexosides *Carbohydr. Res.* **2002**, 337(4), 353-367; A. Roen, J. I. Padron*, et al.* Hydroxymethyl rotamer populations in disaccharides *J Org Chem* **2003**, 68(12), 4615-4630; C. Thibaudeau, R. Stenutz*, et al.* Correlated C-C and C-O bond conformations in saccharide hydroxymethyl groups: Parametrization and application of redundant H-1-H-1, C-13-H-1, and C-13-C-13 NMR J-couplings *J. Am. Chem. Soc.* **2004**, 126(48), 15668-15685.

(6) R. Stenutz, I. Carmichael*, et al.* Hydroxymethyl group conformation in saccharides: Structural dependencies of (2)J(HH), (3)J(HH), and (1)J(CH) spin-spin coupling constants *J Org Chem* **2002**, 67(3), 949-958.

(7) G. D. Rockwell and T. B. Grindley Effect of solvation on the rotation of hydroxymethyl groups in carbohydrates *J. Am. Chem. Soc.* **1998**, 120(42), 10953-10963.

(8) L. Poppe and H. Vanhalbeek Selective, Inverse-Detected Measurements of Long-Range C-13, H-1 Coupling-Constants - Application to a Disaccharide *J. Magn. Reson.* **1991**, 93(1), 214-217; M. Hricovini, I. Tvaroska*, et al.* Solution Behavior of Methyl Beta-Xylobioside - Conformational Flexibility Revealed by Nmr Measurements and Theoretical Calculations *Carbohydr. Res.* **1990**, 198(2), 193-203.

(9) K. Bock, J. B. F. Guzman*, et al.* A 1H and 13C n.m.r. spectroscopic analysis of six pseudohexoses *Carbohydr. Res.* **1988**, 174, 354-359.

(10) K. Bock and J. O. Duus A Conformational Study of Hydroxymethyl Groups in Carbohydrates Investigated by H-1-Nmr Spectroscopy *J. Carbohydr. Chem.* **1994**, 13(4), 513-543.

(11) M. Hoffmann and J. Rychlewski Effects of substituting a OH group by a F atom in D-glucose. Ab initio and DFT analysis *J. Am. Chem. Soc.* **2001**, 123(10), 2308-2316.

(12) B. M. Pinto and R. Y. N. Leung In The anomeric effect and associated stereoelectronic effects; G. R. J. Thatcher, Ed.; American Chemical Society: Washington, DC, 1993.

(13) K. N. Kirschner and R. J. Woods Solvent interactions determine carbohydrate conformation *Proc. Natl. Acad. Sci. USA* **2001**, 98(19), 10541-10545.

(14) F. Franks, P. J. Lillford*, et al.* Isomeric Equilibria of Monosaccharides in Solution - Influence of Solvent and Temperature *Journal of the Chemical Society-Faraday Transactions I* **1989**, 85, 2417-2426.

(15) C. Moller and M. S. Plesset Note on an approximation treatment for many-electron systems *Phys. Rev.* **1934**, 46(7), 618-622.

(16) Ditchfie.R, W. J. Hehre*, et al.* Self-Consistent Molecular-Orbital Methods .9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules *J. Chem. Phys.* **1971**, 54(2), 724-&.

(17) J. H. Lii, K. H. Chen*, et al.* Alcohols, ethers, carbohydrates, and related compounds. II. The anomeric effect *J. Comput. Chem.* **2003**, 24(12), 1473-1489.

(18) J. Dabrowski, T. Kozar*, et al.* Conformational Mobility of Oligosaccharides - Experimental-Evidence for the Existence of an Anti Conformer of the Ga1-Beta-1-3glc-Beta-1-Ome Disaccharide *J. Am. Chem. Soc.* **1995**, 117(20), 5534-5539.

(19) S. J. Angyal and V. A. Pickles Equilibria between Pyranoses and Furanoses .2. Aldoses *Australian Journal of Chemistry* **1972**, 25(8), 1695-&; S. J. Angyal Composition and Conformation of Sugars in Solution *Angewandte Chemie-International Edition* **1969**, 8(3), 157-&.

(20) K. J. Naidoo and J. W. Brady Calculation of the Ramachandran potential of mean force for a disaccharide in aqueous solution *J. Am. Chem. Soc.* **1999**, 121(10), 2244-2252.

(21) B. Leroux, H. Bizot*, et al.* Water structuring around complex solutes: Theoretical modeling of alpha-D-glucopyranose *Chem. Phys.* **1997**, 216(3), 349-363.

(22) K. J. Naidoo and J. Y. J. Chen The role of water in the design of glycosidic linkage flexibility *Mol. Phys.* **2003**, 101(17), 2687-2694.

(23) J. W. Ponder and D. A. Case In Protein Simulations, 2003, p 27-85.

(24) W. L. Jorgensen, D. S. Maxwell*, et al.* Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids *J. Am. Chem. Soc.* **1996**, 118(45), 11225-11236.

(25) W. D. Cornell, P. Cieplak*, et al.* A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules *J. Am. Chem. Soc.* **1995**, 117(19), 5179-5197.

(26) D. A. Pearlman, D. A. Case*, et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, 91, 1-41.

(27) A. D. Mackerell, D. Bashford*, et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins *J. Phys. Chem. B* **1998**, 102(18), 3586-3616.

(28) B. R. Brooks, R. E. Bruccoleri*, et al.* Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations *J. Comput. Chem.* **1983**, 4(2), 187-217.

(29) R. D. Lins and P. H. Hunenberger A new GROMOS force field for hexopyranose-based carbohydrates *J. Comput. Chem.* **2005**, 26(13), 1400-1412.

(30) D. Kony, W. Damm*, et al.* An improved OPLS-AA force field for carbohydrates *J. Comput. Chem.* **2002**, 23(15), 1416-1429.

(31) W. L. Jorgensen, J. Chandrasekhar*, et al.* Comparison of simple potential functions for simulating water *J. Chem. Phys.* **1983**, 79(2), 926-935.

(32) H. J. C. Berendsen, J. R. Grifera*, et al.* The Missing Term in Effective Pair Potentials *J. Phys. Chem.* **1987**, 91, 6269-6271.

(33) W. L. Jorgensen and J. D. Madura Temperature and size dependence for monte carlo simulations of TIP4P water *Mol. Phys.* **1985**, 56, 1381.

(34) M. W. Mahoney and W. L. Jorgensen A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions *J. Chem. Phys.* **2000**, 112(20), 8910-8922.

(35) J. J. C. Berendsen, J. P. M. Postma*, et al.* Intermolecular Forces; Reidel: Dordrecht, 1981.

(36) A. K. Soper and M. G. Phillips A new determination of the structure of water at 25C *Chem. Phys.* **1986**, 107, 47-60.

(37) A. D. Mackerell, J. Wiorkiewiczkuczera*, et al.* An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids *J. Am. Chem. Soc.* **1995**, 117(48), 11946-11975.

(38) W. L. Jorgensen and J. Tiradorives The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin *J. Am. Chem. Soc.* **1988**, 110(6), 1657-1666.

(39) H. W. Horn, W. C. Swope*, et al.* Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew *J. Chem. Phys.* **2004**, 120(20), 9665-9678.

(40) M. W. Mahoney and W. L. Jorgensen Diffusion constant of the TIP5P model of liquid water *J. Chem. Phys.* **2001**, 114(1), 363-366.

(41) S. W. Rick A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums *J. Chem. Phys.* **2004**, 120(13), 6085-6093.

(42) Verhoeve.J and A. Dymanus Magnetic Properties and Molecular Quadrupole Tensor of Water Molecule by Beam-Maser Zeeman Spectroscopy *J. Chem. Phys.* **1970**, 52(6), 3222-&; S. A. Clough, Y. Beers*, et al.* Dipole moment of water from stark measurements of H2O, HDO, and D2O *J. Chem. Phys.* **1973**, 59(5), 2254-2259.

(43) R. S. Fellers, C. Leforestier*, et al. Science* **1999**, 284, 1999.

(44) J. Jonas, T. Defried*, et al.* Molecular Motions in Compressed Liquid Water *J. Chem. Phys.* **1976**, 65(2), 582-588.

(45) P. H. Poole, F. Sciortino*, et al.* Spinodal of Liquid Water *Physical Review E* **1993**, 48(5), 3799-3817; P. H. Poole, F. Sciortino*, et al.* Phase-Behavior of Metastable Water *Nature* **1992**, 360(6402), 324-328.

(46) F. Sciortino and S. Sastry Sound-Propagation in Liquid Water - the Puzzle Continues *J. Chem. Phys.* **1994**, 100(5), 3881-3893.

(47) R. J. Speedy Thermodynamic Properties of Supercooled Water at 1-Atm *J. Phys. Chem.* **1987**, 91(12), 3354-3358; C. A. Angell, M. Oguni*, et al.* Heat-Capacity of Water at Extremes of Supercooling and Superheating *J. Phys. Chem.* **1982**, 86(6), 998-1002; H. Kanno and C. A. Angell Volumetric and Derived Thermal-Characteristics of Liquid D2o at Low-Temperatures and High-Pressures *J. Chem. Phys.* **1980**, 73(4), 1940-1947; C. A. Angell and H. Kanno Density Maxima in High-Pressure Supercooled Water and Liquid Silicon Dioxide *Science* **1976**, 193(4258), 1121-1122; G. S. Kell Density, Thermal Expansivity, and Compressibility of Liquid Water from 0 Degrees to 150 Degreesc - Correlations and Tables for Atmospheric-Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale *Journal of Chemical and Engineering Data* **1975**, 20(1), 97-105.

(48) A. R. Leach Molecular Modelling principles and applications; Addison Wesley Longman Limited: Essex, 1996.

(49) D. A. Case, D. A. Pearlman*, et al.*; University of California: San Francisco, 2002.

(50) T. A. Darden, D. York*, et al.* PME: An N*log(N) method for Ewald sums in large systems *J. Chem. Phys.* **1993**, 98(12), 10089-10092.

(51) L. Verlet Computer 'experiments' on classical fluids. I. Thermodynamical properties of lennard-jones molecules. *Physical Review* **1967**, 159, 98-103.

(52) R. W. Hockney The potential calculation and some applications. *Methods in Computational Physics* **1970**, 9, 136-211.

(53) H. J. C. Berendsen, J. P. M. Postma*, et al.* Molecular-Dynamics with Coupling to an External Bath *J. Chem. Phys.* **1984**, 81(8), 3684-3690.

(54) R. S. Mulliken Electronic population analysis on LCAO-MO moleuclar wave functions. II. Overlap populations, bond orders, and covalent bond energies *J. Chem. Phys.* **1955**, 23(10), 1841-1846.

(55) S. R. Cox and D. E. Williams Representation of the molecular electrostatic potential by a net atomic charge model *J. Comput. Chem.* **1981**, 2(3), 304-323.

(56) P. H. Smit, J. L. Derissen*, et al. Mol. Phys.* **1979**, 37, 521; E. Scrocco and J. Tomasi *Advances in Quantum Chemistry* **1978**, 11, 115; R. S. Mulliken Electronic population analysis on LCAO-MO molecular wave functions.* I *J. Chem. Phys.* **1955**, 23(10), 1833-1840; R. J. Woods, M. Khalil*, et al.* Derivation of net atomic

charges from molecular electrostatic potentials *J. Comput. Chem.* **1990**, 11(3), 297-310.

(57) B. H. Besler, K. M. Merz*, et al.* Atomic charges derived from semiempirical methods *J. Comput. Chem.* **1990**, 11(4), 431-439.

(58) C. M. Breneman and K. B. Wiberg Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis *J. Comput. Chem.* **1990**, 11(3), 361-373.

(59) L. E. Chirlian and M. M. Francl Atomic charges derived from electrostatic potentials: A detailed study *J. Comput. Chem.* **1987**, 8(6), 894-905.

(60) U. C. Singh and P. A. Kollman An approach to computing electrostatic charges for molecules *J. Comput. Chem.* **1984**, 5(2), 129-145.

(61) K. Watanabe and M. L. Klein Effective pair potentials and the properties of water *Chem. Phys.* **1989**, 131, 157-167.

(62) R. J. Woods, R. A. Dwek*, et al.* Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development *J. Phys. Chem.* **1995**, 99(11), 3832-3846.

(63) C. I. Bayly, P. Cieplak*, et al.* A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model *J. Phys. Chem.* **1993**, 97(40), 10269-10280.

(64) W. D. Cornell, P. Cieplak*, et al.* Application of Resp Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation *J. Am. Chem. Soc.* **1993**, 115(21), 9620-9631.

(65) J. M. Wang, P. Cieplak*, et al.* How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, 21(12), 1049-1074.

(66) M. Basma, S. Sundara*, et al.* Solvated ensemble averaging in the calculation of partial atomic charges *J. Comput. Chem.* **2001**, 22(11), 1125-1137.

(67) R. J. Woods and R. J. Chapelle RESP-partial charges for condensed phase MD simulations *J. Mol. Struct.* **2000**, 527, 149-156.

(68) C. Carey, L. E. Chirlian*, et al.* Beyond CHELP: Improved potential derived charges for sugars *Glycoconjugate Journal* **1997**, 14(4), 501-505.

(69) M. M. Francl, C. Carey*, et al.* Charges fit to electrostatic potentials .2. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996**, 17(3), 367-383.

(70) K. N. Kirschner and R. J. Woods An improved force field for carbohydrate modeling **2005**.

(71) R. W. Dixon and P. A. Kollman Advancing beyond the atom-centered model in additive and nonadditive molecular mechanics *J. Comput. Chem.* **1997**, 18(13), 1632-1646.

(72) S. J. Weiner, P. A. Kollman*, et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins *J. Am. Chem. Soc.* **1984**, 106, 765-784.

(73) N. O. J. Malcolm and P. L. A. Popelier The full topology of the Laplacian of the electron density: scrutinising a physical basis for the VSEPR model *Faraday Discussions* **2003**, 124, 353-363; R. F. W. Bader and C. F. Matta Atomic charges are measurable quantum expectation values: A rebuttal of criticisms of QTAIM charges *J. Phys. Chem. A* **2004**, 108(40), 8385-8394; P. L. A. Popelier A robust algoritm to locate automatically all types of critical points in the charge density and its Laplacian *Chem. Phys. Lett.* **1994**, 228, 160-164.

(74) G. A. Kaminski, H. A. Stern*, et al.* Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry *J. Phys. Chem. A* **2004**, 108(4), 621-627.

(75) M. J. Frisch, G. W. Trucks*, et al.*; Gaussian, Inc.: Pittsburgh PA, 1998.

(76) C. C. J. Roothan New developments in molecular orbital theory *Rev. Mod. Phys.* **1951**, 23(2), 69-89.

(77) A. D. Becke Density-functional thermochemistry. III. The role of exact exchange *J. Chem. Phys.* **1993**, 98(7), 5648-5652.

(78) C. Lee, W. Yang*, et al.* Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density *Phys. Rev. B* **1988**, 37, 785-789.

(79) D. E. Woon and T. H. Dunning Gaussian-Basis Sets for Use in Correlated Molecular Calculations .3. The Atoms Aluminum through Argon *J. Chem. Phys.* **1993**, 98(2), 1358-1371.

(80) S. S. Xantheas and T. H. Dunning Jr. Ab initio studies of cyclic water clusters (h2o)n, n=1-6. I. optimal structures and vibrational spectra. *J. Chem. Phys.* **1993**, 99(11), 8774-8792; S. S. Xantheas and T. H. Dunning Jr. The structure of the water trimer from ab initio calculations *J. Chem. Phys.* **1993**, 98(10), 8037-8040.

(81) M. R. Viant, J. D. Cruzan*, et al.* pseudorotation in water trimer isotopomers using terahertz laser spectroscopy *J. Phys. Chem. A* **1997**, 101, 9032-9041; J. D. Cruzan, M. R. Viant*, et al.* Terahertz Laser Vibration-Rotation Tunneling Spectroscopy of the Water Tetramer *J. Phys. Chem. A* **1997**, 101, 9022-9031; J. D. Cruzan, L. B. Braly*, et al.* Quantifying Hydrogen Bond Cooperativity in Water: VRT Spectroscopy of the Water Tetramer *Science* **1996**, 271, 59-62; D. J.

Wales and R. R. Walsh Theoretical study of the water tetramer *J. Chem. Phys.* **106**, 106(17), 7193-7207; S. Graf and S. Leutwyler An ab initio derived torsional potential energy surface for the cyclic water tetramer *J. Chem. Phys.* **1998**, 109(13), 5393-5403; I. M. B. Nielsen, E. T. Seidl*, et al.* Accurate structures and binding energies for small water clusters: The water trimer *J. Chem. Phys.* **1999**, 110(19), 9435-9442; A. Khan theoretical studies of large water clusters: (h2o)28, (h2o)29, (h2o)30, and (h2o)31 hexakaidecahedral structures *J. Chem. Phys.* **1997**, 106(13), 5537-5540; H. M. Lee, S. B. Suh*, et al.* Structures, energies, vibrational spectra and electronic properties of water monomer to decamer *J. Chem. Phys.* **2000**, 112(22), 9759-9772.

(82) F. N. Keutsch and R. J. Saykally Water clusters: Untangling the mysteries of the liquid, one molecule at a time *Proc. Natl. Acad. Sci., USA* **2001**, 98(19), 10533-10540.

(83) W. Klopper, J. G. C. M. van Duijneveldt-van de Rijdt*, et al.* Computational determination of equilibrium geometry and dissociation of the water dimer *Phys. Chem. Chem. Phys.* **2000**, 2(10), 2227-2234.

(84) K. N. Kirschner and R. J. Woods Quantum Mechanical Study of the Non-Bonded Forces in Water-Methanol Complexes *J. Phys. Chem. A* **2001**, 105(16), 4150-4155.

(85) K. A. Peterson and T. H. Dunning Jr. *J. Chem. Phys.* **1995**, 102, 2032-2041.

(86) S. S. Xantheas, C. J. Burnham*, et al.* Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles *J. Chem. Phys.* **2002**, 116(4), 1493-1499.

(87) L. Gonzalez, O. Mo*, et al.* High level ab initio and density functional theory studies on methanol-water dimers and cyclic methanol(water)2 trimer *J. Chem. Phys.* **1998**, 109(1), 139-150.

(88) O. Mo, M. Yanez*, et al.* Cooperative (nonpairwise) effects in water trimers: An ab initio molecular orbital study *J. Chem. Phys.* **1992**, 97(9), 6628-6638.

(89) S. S. Xantheas Ab initio studies of cyclic water clusters (h2o)n, n=1-6. II. Analysis of many-body interactions *J. Chem. Phys.* **1994**, 100(10), 7523-7534.

(90) M. Brandl, M. Meyer*, et al.* Water-Mediated Base Pairs in RNA: A Quantum-Chemical Study *J. Phys. Chem. A* **2000**, 104, 11177-11187.

(91) J. E. Del Bene and M. J. T. Jordan What a difference a decade makes: progress in ab initio studies of the hydrogen bond *J. Mol. Struct.* **2001**, 573, 11-23.

(92) M. Lozynski, D. Rusinska-Roszak*, et al.* Hydrogen Bonding and Density Functional Calculations: The B3LYP Approach as the Shortest Way to MP2 Results *J. Phys. Chem. A* **1998**, 102, 2899-2903; S. S. Xantheas Ab initio studies of cyclic water

clusters (h2o)n, n=1-6. III. Comparison of density functional with MP2 results *J. Chem. Phys.* **1995**, 102(11), 4505-4517.

(93) A. Dkhissi, R. Ramaekers*, et al.* dft/b3lyp study of the hydrogen-bonding cooperativity: application to (2-pyridone)2, 2-pyridone-h2o, 2-pyridone-ch3oh and 2-pyridone-ch3och3 *Chem. Phys. Lett.* **2000**, 331, 553-560.

(94) A. Milet, T. Korona*, et al.* anisotropic intermolecular interactions in van der waals and hydrogen-bonded complexes: what can we get from density functional calculations? *J. Chem. Phys.* **1999**, 111(17), 7727-7735; Y. S. Wang, H. C. Chang*, et al.* structures and isomeric transitions of nh4+(h2o)3-6: from single to double rings *J. Am. Chem. Soc.* **1998**, 120, 8777-8788.

(95) J. E. Del Bene, W. B. Person*, et al.* Properties of Hydrogen-Bonded Complexes obtained from the B3LYP functional with 6-31G(d,p) and 6-31+G(d,p) basis sets: comparison with MP2/6-31+G(d,p) results and experimental data *J. Phys. Chem.* **1995**, 99, 10705-10707.

(96) E. E. Smith and I. J. Goldstein Protein-carbohydrate interaction V. *Arch. Biochem. Biophys.* **1967**, 121(1), 88.

(97) G. M. Bradbrook, T. Gleichmann*, et al.* x-ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes *J. Chem. Soc., Faraday Trans.* **1998**, 94, 1603-1611.

(98) J. H. Naismith, C. Emmerich*, et al.* Refined structure of concanavalin A complexed with methyl alpha-d-mannopyranoside at 2.0A resolution and comparison with the saccharide-free structure *Acta Cryst. D* **1994**, 50, 847-858.

(99) J. H. Naismith and R. A. Field Structural Basis of Trimannoside Recognition by Concanavalin A *J. Biol. Chem.* **1996**, 271(2), 972-976.

(100) C. P. Swaminathan, N. Surolia*, et al.* Role of water in the specific binding of mannose and mannooligosaccharides to concanavalin A *J. Am. Chem. Soc.* **1998**, 120, 5153-5159.

(101) D. K. Mandal, N. Kishore*, et al.* Thermodynamics of Lectin-Carbohydrate Interactions. Titration Microcalorimetry Measurements of the binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A *Biochemistry* **1994**, 33, 1149-1156; R. A. Bryce, I. H. Hillier*, et al.* Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy analysis of Oligosaccharide Binding to concanavalin A *Biophysical Journal* **2001**, 81, 1373-1388; F. P. Schwarz, K. D. Puri*, et al.* Thermodynamics of monosaccharide binding to concanavalin A, Pea (Pisum sativum) Lectin, and Lentil (Lens culinaris) Lectin *J. Biol. Chem.* **1993**, 268(11), 7668-7677.

(102) D. K. Mandal, L. Bhattacharyya*, et al.* Studies of the Binding Specificity of Concanavalin A. nature of the Extended Binding Site for Asparagine-Linked Carbohydrates *Biochemistry* **1994**, 33, 1157-1162.

(103) A. K. Soper and J. L. Finney Hydration of Methanol in Aqueous Solution *Phys. Rev. Lett.* **1993**, 71(26), 4346-4349.

(104) J. R. Carney and T. S. Zwier Infrared and Ultraviolet Spectroscopy of Water-Containing Clusters of Indole, 1-Methylindole, and 3-Methylindole *J. Phys. Chem. A* **1999**, 103, 9943-9957.

(105) S. F. Boys and F. Bernardi The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Mol. Phys.* **1970**, 19(4), 553-566.

(106) S. S. Xantheas On the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy *J. Chem. Phys.* **1996**, 104(21), 8821-8824.

(107) S. S. Xantheas Cooperativity and hydrogen bonding network in water clusters *Chem. Phys.* **2000**, 258, 225-231.

(108) R. D. Parra, M. Furukawa*, et al.* Energetics and cooperativity in three-center hydrogen bonding interactions. I. diacetamide-X dimers (X=HCN, CH3OH) *J. Chem. Phys.* **2001**, 115(13), 6030-6035.

(109) K.-Y. Kim, U.-I. Cho*, et al.* Ab initio Study on Structures, Energies and Vibrations of Methylammonium-(water)n(n=1-3) Complexes *Bull. Korean Chem. Soc.* **2001**, 22(6), 597-604.

(110) C. Chaudhuri, Y. S. Wang*, et al.* Infrared spectra and isomeric structures of hydroxide ion - water clusters OH-(h2o)1-5 : a comparison with H3O+(h2o)1-5 *Mol. Phys.* **2001**, 99(14), 1161-1173; J. C. Jiang, Y. S. Wang*, et al.* Infrared Spectra of H+(h2O)5-8 clusters: evidence for symmetric proton hydration *J. Am. Chem. Soc.* **2000**, 122, 1398-1410.

(111) E. Garcia-Hernandez and A. Hernandez-Arana Structural bases of lectin-carbohydrate affinities: Comparison with protein-folding energetics *Protein Science* **1999**, 8, 1075-1086.

(112) S. Scheiner In Hydrogen bonding a theoretical perspective; D. G. Truhlar, Ed.; Oxford University Press: New York, 1997, p 23-26.

(113) L. A. Curtiss and M. Blander Thermodynamic properties of gas-phase hydrogen-bonded complexes *Chem. Rev.* **1988**, 88(6), 827-841.

(114) A. Almond Towards understanding the interaction between oligosaccharides and water molecules *Carbohydr. Res.* **2005**, 340(5), 907-920; J. W. Brady and R. K.

Schmidt The Role of Hydrogen-Bonding in Carbohydrates - Molecular-Dynamics Simulations of Maltose in Aqueous-Solution *J. Phys. Chem.* **1993**, 97(4), 958-966.

(115) S. M. Tschampel and R. J. Woods Quantifying the role of water in protein-carbohydrate interactions *J. Phys. Chem. A* **2003**, 107(43), 9175-9181.

(116) C. Clarke, R. J. Woods*, et al.* involvement of water in carbohydrate-protein binding *J. Am. Chem. Soc.* **2001**, 123, 12238-12247.

(117) R. U. Lemieux, L. T. J. Delbaere*, et al.* Involvement of Water in Host-Guest Interactions *Ciba Foundation Symposia* **1991**, 158, 231-248.

(118) F. Corzana, M. S. Motawia*, et al.* A hydration study of (1 -> 4) and (1 -> 6) linked alpha-glucans by comparative 10 ns molecular dynamics simulations and 500-MHz NMR *J. Comput. Chem.* **2004**, 25(4), 573-586.

(119) A. K. Soper Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water *Mol. Phys.* **2001**, 99(17), 1503-1516.

(120) G. M. Brown and H. A. Levy Alpha-Delta-Glucose - Further Refinement Based on Neutron-Diffraction Data *Acta Cryst. B* **1979**, 35(MAR), 656-659.

(121) G. A. Jeffrey, R. K. McMullan*, et al.* Neutron-Diffraction Study of Hydrogen-Bonding in Crystal-Structures of Methyl Alpha-D-Mannopyranoside and Methyl Alpha-D-Glucopyranoside *Acta Cryst. B* **1977**, 33(MAR15), 728-737.

(122) S. Takagi and G. A. Jeffrey Neutron-Diffraction Refinement of the Crystal-Structures of Methyl Alpha-D-Galactopyranoside Monohydrate and Methyl Beta-D-Galactopyranoside *Acta Cryst. B* **1979**, 35(APR), 902-906.

(123) E. Hough, S. Neidle*, et al.* Crystal-Structure of Alpha-D-Glucose Monohydrate *Acta Cryst. B* **1973**, B 29(FEB15), 365-367.

(124) F. Mo and L. H. Jensen Refined Model for N-Acetyl-Alpha-D-Glucosamine *Acta Cryst. B* **1975**, 31(DEC15), 2867-2873.

(125) D. A. Case, T. A. Darden*, et al.*; University of California: San Francisco, 2004.

(126) M. L. C. E. Kouwijzer, B. P. van Eijck*, et al.* Comparison of two force fields by molecular dynamics simiultaion of glucose crystals: effect of using ewald sums *J. Comput. Chem.* **1993**, 14(11), 1281-1289.

(127) B. Roux The Calculation of the Potential of Mean Force Using Computer-Simulations *Comput. Phys. Commun.* **1995**, 91(1-3), 275-282; S. Kumar, D. Bouzida*, et al.* The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method *J. Comput. Chem.* **1992**, 13(8),

1011-1021; S. Kumar, J. M. Rosenberg*, et al.* Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method *J. Comput. Chem.* **1995**, 16(11), 1339-1350.

(128) Q. Liu and J. W. Brady Anisotropic solvent structuring in aqueous sugar solutions *J. Am. Chem. Soc.* **1996**, 118(49), 12276-12286.

(129) P. E. Mason, G. W. Neilson*, et al.* Neutron diffraction studies on aqueous solutions of glucose *J. Chem. Phys.* **2003**, 119(6), 3347-3353.

(130) D. J. Price and C. L. Brooks A modified TIP3P water potential for simulation with Ewald summation *J. Chem. Phys.* **2004**, 121(20), 10096-10103.

(131) W. L. Jorgensen, J. D. Madura*, et al.* Optimized Intermolecular Potential Functions for Liquid Hydrocarbons *J. Am. Chem. Soc.* **1984**, 106(22), 6638-6646.

(132) N. L. Allinger, Y. H. Yuh*, et al.* Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .1 *J. Am. Chem. Soc.* **1989**, 111(23), 8551-8566; N. L. Allinger, M. Rahman*, et al.* A Molecular Mechanics Force-Field (Mm3) for Alcohols and Ethers *J. Am. Chem. Soc.* **1990**, 112(23), 8293-8307.

(133) A. D. French, D. P. Miller*, et al.* Miniature Crystal Models of Cellulose Polymorphs and Other Carbohydrates *Int. J. Biol. Macromol.* **1993**, 15(1), 30-36.

(134) J. D. Cox, D. D. Wagman*, et al.* CODATA Key Values for Thermodynamics; Hemisphere Publishing Corp.: New York, 1984; V. V. Ponomarev and T. A. Alekseeva A massive microcalorimeter for determining heats of combustion *Russ. J. Phys. Chem. (Engl. Transl.)* **1961**, 35, 800-803; J. D. Cox and G. Pilcher Thermochemistry of Organic and Organometallic Compounds; Academic Press: New York, 1970; S. M. Skuratov, A. A. Strepikheev*, et al.* The Reactivity of 5-Member and 6-Member Heterocyclic Compounds *Doklady Akademii Nauk Sssr* **1957**, 117(3), 452-454.

(135) M. Kraszni, Z. Szakacs*, et al.* Determination of rotamer populations and related parameters from NMR coupling constants: a critical review *Analytical and Bioanalytical Chemistry* **2004**, 378(6), 1449-1463.

(136) Y. Nishida, H. Hori*, et al.* H-1-Nmr Analyses of Rotameric Distribution of C5-C6 Bonds of D-Glucopyranoses in Solution *J. Carbohydr. Chem.* **1988**, 7(1), 239-250; Y. Nishida, H. Ohrui*, et al.* H-NMR studies of 6R and 6S deuterated D-hexoses: assignment of the preferred rotamers about C5-C6 bond of D-glucose and D-galactose derivatives in solution *Tetrahedron Lett.* **1984**, 25(15), 1575-1578.

(137) M. Kuttel, J. W. Brady*, et al.* Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations *J. Comput. Chem.* **2002**, 23(13), 1236-1243.

(138) S. M. Tschampel, M. R. Kennerty*, et al.* A TIP5P compatible carbohydrate force field **2005**.

(139) Q. Liu and J. W. Brady Model dependence of the anisotropic structuring of solvent water around sugars in molecular dynamics simulations *J. Phys. Chem. B* **1997**, 101(8), 1317-1321; H. Beierbeck and R. U. Lemieux Computer-Simulation Studies of the Solvation of Oligosaccharides *Can J Chem* **1990**, 68(6), 820-827.

(140) R. U. Lemieux How Proteins Recognize and Bind Oligosaccharides *Acs Symposium Series* **1993**, 519, 5-18; R. U. Lemieux How water provides the impetus for molecular recognition in aqueous solution *Acc. Chem. Res.* **1996**, 29(8), 373-380.

(141) A. K. Soper The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa *Chem. Phys.* **2000**, 258(2-3), 121-137; T. Head-Gordon and G. Hura Water structure from scattering experiments and simulation *Chem. Rev.* **2002**, 102(8), 2651-2669.

(142) I. Tvaroska, M. Hricovini*, et al.* An Attempt to Derive a New Karplus-Type Equation of Vicinal Proton Carbon Coupling-Constants for C-O-C-H Segments of Bonded Atoms *Carbohydr. Res.* **1989**, 189, 359-362.

(143) J. Cavanaugh and N. J. Skelton Protein NMR Spectroscopy; Elsevier Science: San Diego, 1995.

(144) S. N. Ha, A. Giammona*, et al.* A revised potential-energy surface for molecular mechanics studies of carbohydrates *Carbohydr. Res.* **1988**, 180(2), 207-221.

(145) P. Wernet, D. Nordlund*, et al.* The structure of the first coordination shell in liquid water *Science* **2004**, 304(5673), 995-999.

(146) A. Nilsson, P. Wernet*, et al.* Comment on "Energetics of hydrogen bond network: Rearrangements in liquid water" *Science* **2005**, 308(5723).

(147) S. Woutersen, U. Emmerichs*, et al.* Femtosecond mid-IR pump-probe spectroscopy of liquid water: Evidence for a two-component structure *Science* **1997**, 278(5338), 658-660.

(148) G. A. Jeffrey and J. Mitra The Hydrogen-Bonding Patterns in the Pyranose and Pyranoside Crystal-Structures *Acta Cryst. B* **1983**, 39(AUG), 469-480.

(149) P. S. Vermersch, J. J. G. Tesmer*, et al.* Protein Ligand Energetics Assessed Using Deoxy and Fluorodeoxy Sugars in Equilibrium Binding and High-Resolution Crystallographic Studies *Journal of Molecular Biology* **1992**, 226(4), 923-929; F. A. Quiocho and N. K. Vyas Novel Stereospecificity of the L-Arabinose-Binding Protein *Nature* **1984**, 310(5976), 381-386; R. Loris, P. P. G. Stas*, et al.* Conserved waters in legume lectin crystal structures: The importance of bound water for the sequence-structure relationship within the legume lectin family. *J. Biol. Chem.*

1994, 269(43), 26722-26733; A. H. Daranas, H. Shimizu*, et al.* Thermodynamics of binding of D-galactose and deoxy derivatives thereof to the L-arabinose-binding protein *J. Am. Chem. Soc.* **2004**, 126(38), 11870-11876.

(150) F. A. Quiocho, D. K. Wilson*, et al.* Substrate-Specificity and Affinity of a Protein Modulated by Bound Water-Molecules *Nature* **1989**, 340(6232), 404-407.

(151) A. Wlodawer, J. Walter*, et al.* Structure of Bovine Pancreatic Trypsin-Inhibitor - Results of Joint Neutron and X-Ray Refinement of Crystal Form-Ii *Journal of Molecular Biology* **1984**, 180(2), 301-329.

(152) G. Otting, E. Liepinsh*, et al.* Protein Hydration in Aqueous-Solution *Science* **1991**, 254(5034), 974-980; G. Otting and K. Wuthrich Studies of Protein Hydration in Aqueous-Solution by Direct Nmr Observation of Individual Protein-Bound Water-Molecules *J. Am. Chem. Soc.* **1989**, 111(5), 1871-1875.

(153) A. Varki, R. Cummings*, et al.*, Eds. Essentials of Glycobiology; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1999.

(154) T. K. Dam and C. F. Brewer Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry *Chem. Rev.* **2002**, 102(2), 387-429; H. Fukada, J. M. Sturtevant*, et al.* Thermodynamics of the Binding of L-Arabinose and of D-Galactose to the L-Arabinose-Binding Protein of Escherichia-Coli *J. Biol. Chem.* **1983**, 258(21), 3193-3198.

(155) U. Spohr, E. Paszkiewiczhnatiw*, et al.* Molecular Recognition .11. The Synthesis of Extensively Deoxygenated Derivatives of the H-Type-2 Human Blood-Group Determinant and Their Binding by an Anti-H-Type-2 Monoclonal-Antibody and the Lectin-1 of Ulex-Europaeus *Can J Chem* **1992**, 70(1), 254-271.

(156) J. B. Sumner *J. Biol. Chem.* **1919**, 37, 137-142.

(157) Z. Derewenda, J. Yariv*, et al.* The Structure of the Saccharide-Binding Site of Concanavalin-A *Embo Journal* **1989**, 8(8), 2189-2193.

(158) J. Greer, H. W. Kaufman*, et al.* An X-Ray Crystallographic Study of Concanavalin-A *Journal of Molecular Biology* **1970**, 48(2), 365-&.

(159) R. Kadirvelraj *personal communication* **2005**.

(160) J. Wang, P. Cieplak*, et al.* how well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, 21, 1049-1074.

(161) C. L. Brooks Protein folding - With a little help *Nature* **2002**, 420(6911), 33-34.

(162) Y. Duan and P. A. Kollman Computational protein folding: From lattice to all-atom *Ibm Systems Journal* **2001**, 40(2), 297-309; C. L. Brooks and D. A. Case

Theory and simulation - The control and timescale of structure and reactivity in biological systems: from peptide folding to cellular networks - Editorial overview *Curr. Opin. Struct. Biol.* **2003**, 13(2), 143-145; W. C. Swope, J. W. Pitera*, et al.* Describing protein folding kinetics by molecular dynamics simulations. 1. Theory *J. Phys. Chem. B* **2004**, 108(21), 6571-6581; C. D. Snow, N. Nguyen*, et al.* Absolute comparison of simulated and experimental protein-folding dynamics *Nature* **2002**, 420(6911), 102-106.

(163) J. W. Pitera and W. Swope Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins *Proc. Natl. Acad. Sci. USA* **2003**, 100(13), 7587-7592.

(164) J. W. Neidigh, R. M. Fesinmeyer*, et al.* Designing a 20-residue protein *Nat. Struct. Biol.* **2002**, 9(6), 425-430.

(165) C. Simmerling, B. Strockbine*, et al.* All-Atom structure prediction and folding simulations of a stable protein *J. Am. Chem. Soc.* **2002**, 124, 11258-11259.

(166) C. D. Snow, B. Zagrovic*, et al.* The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations *J. Am. Chem. Soc.* **2002**, 124(49), 14548-14549.

(167) A. Okur, B. Strockbine*, et al.* Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins *J. Comput. Chem.* **2003**, 24(1), 21-31.

(168) H. Hu, M. Elstner*, et al.* Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution *Proteins* **2003**, 50(3), 451-463.

(169) S. C. Lovell, I. W. Davis*, et al.* Structure validation by C alpha geometry: phi,psi and C beta deviation *Proteins* **2003**, 50(3), 437-450.

(170) A. E. Garcia and K. Y. Sanbonmatsu Exploring the energy landscape of a beta hairpin in explicit solvent *Proteins* **2001**, 42(3), 345-354.

(171) B. Zagrovic, E. J. Sorin*, et al.* beta-hairpin folding simulations in atomistic detail using an implicit solvent model *Journal of Molecular Biology* **2001**, 313(1), 151-169.

(172) R. H. Zhou, B. J. Berne*, et al.* The free energy landscape for beta hairpin folding in explicit water *Proc. Natl. Acad. Sci. USA* **2001**, 98(26), 14931-14936.

(173) R. H. Zhou and B. J. Berne Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proc. Natl. Acad. Sci. USA* **2002**, 99(20), 12777-12782.

(174) F. J. Blanco, G. Rivas, *et al.* A Short Linear Peptide That Folds into a Native Stable Beta-Hairpin in Aqueous-Solution *Nat. Struct. Biol.* **1994**, 1(9), 584-590.

(175) E. Dealba, F. J. Blanco, *et al.* Interactions Responsible for the Ph-Dependence of the Beta-Hairpin Conformational Population Formed by a Designed Linear Peptide *European Journal of Biochemistry* **1995**, 233(1), 283-292.

(176) V. Munoz, P. A. Thompson, *et al.* Folding dynamics and mechanism of beta-hairpin formation *Nature* **1997**, 390(6656), 196-199.

(177) V. Munoz, E. R. Henry, *et al.* A statistical mechanical model for beta-hairpin kinetics *Proc. Natl. Acad. Sci. USA* **1998**, 95(11), 5872-5879.

(178) A. G. Cochran, N. J. Skelton, *et al.* Tryptophan zippers: Stable, monomeric beta-hairpins (vol 98, pg 5578, 2001) *Proc. Natl. Acad. Sci. USA* **2002**, 99(13), 9081-9081.

(179) X. Daura, W. F. vanGunsteren, *et al.* Studying the stability of a helical beta-heptapeptide by molecular dynamics simulations *Chem-Eur J* **1997**, 3(9), 1410-1417.

(180) X. Daura, B. Jaun, *et al.* Reversible peptide folding in solution by molecular dynamics simulation *Journal of Molecular Biology* **1998**, 280(5), 925-932.

(181) X. Daura, W. F. van Gunsteren, *et al.* Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations *Proteins* **1999**, 34(3), 269-280.

(182) W. Damm and W. E. Van Gunsteren Reversible peptide folding: Dependence on molecular force field used *J. Comput. Chem.* **2000**, 21(9), 774-787.

(183) K. Gademann, B. Jaun, *et al.* Temperature-dependent NMR and CD spectra of beta-peptides: On the thermal stability of beta-peptide helices - Is the folding process of beta-peptides non-cooperative? *Helvetica Chimica Acta* **1999**, 82(1), 1-11.

(184) L. L. Qiu, S. A. Pabit, *et al.* Smaller and faster: The 20-residue Trp-cage protein folds in 4 mu s *J. Am. Chem. Soc.* **2002**, 124(44), 12952-12953.

(185) P. A. Kollman, R. Dixon, *et al.* In Computer Simulations of Biomolecular Systems; W. F. van Gunsteren, P. K. Weiner and A. J. Wilkinson, Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997, p 83-96.

(186) I. R. Gould, W. D. Cornell, *et al.* A quantum-mechanical investigation of the conformational energetics of the alanine and glycine dipeptides in the gas-phase and in aqueous-solution *J. Am. Chem. Soc.* **1994**, 116(20), 9250-9256.

(187) M. D. Beachy, D. Chasman*, et al.* Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields *J. Am. Chem. Soc.* **1997**, 119(25), 5908-5920.

(188) A. E. Garcia and K. Y. Sanbonmatsu alpha-Helical stabilization by side chain shielding of backbone hydrogen bonds *Proc. Natl. Acad. Sci. USA* **2002**, 99(5), 2782-2787.

(189) S. Gnanakaran and A. E. Garcia Validation of an all-atom protein force field: From dipeptides to larger peptides *J. Phys. Chem. B* **2003**, 107(46), 12555-12557.

(190) Y. Duan, C. Wu*, et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations *J. Comput. Chem.* **2003**, 24(16), 1999-2012.

(191) A. D. MacKerell, M. Feig*, et al.* Improved treatment of the protein backbone in empirical force fields *J. Am. Chem. Soc.* **2004**, 126(3), 698-699; A. D. Mackerell, M. Feig*, et al.* Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations *J. Comput. Chem.* **2004**, 25(11), 1400-1415.

(192) A. D. Mackerell Empirical force fields for biological macromolecules: Overview and issues *J. Comput. Chem.* **2004**, 25(13), 1584-1604.

(193) C. J. Bosques, S. M. Tschampel*, et al.* Effects of glycosylation on peptide conformation: A synergistic experimental and computational study *J. Am. Chem. Soc.* **2004**, 126(27), 8421-8425.

(194) B. D. Bursulaya and C. L. Brooks Comparative study of the folding free energy landscape of a three-stranded beta-sheet protein with explicit and implicit solvent models *J. Phys. Chem. B* **2000**, 104(51), 12378-12383.

(195) D. Elking. Athens, GA, 2004.

(196) G. A. Kaminski, R. A. Friesner*, et al.* Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides *J. Phys. Chem. B* **2001**, 105(28), 6474-6487.

(197) J. D. Dyekjaer *private communication* **2005**.

(198) R. Fletterick, C. C. Tsai*, et al.* Crystal and Molecular Structure of L-Alanyl-L-Alanine *J. Phys. Chem.* **1971**, 75(7), 918-&.

(199) R. Vargas, J. Garza*, et al.* Conformational study of the alanine dipeptide at the MP2 and DFT levels *J. Phys. Chem. A* **2002**, 106(13), 3213-3218.

(200) K. J. Jalkanen and S. Suhai N-acetyl-L-alanine N'-methylamide: A density functional analysis of the vibrational absorption and vibrational circular dichroism spectra *Chem. Phys.* **1996**, 208(1), 81-116.

(201) J. Gonzalez-Outeirino, K. N. Kirschner*, et al. Journal of Canadian Chemistry* **2005**.

(202) E. B. Nielsen and Schellma.Ja Rotatory Properties of Molecules Containing 2 Peptide Groups - Experimental and Theoretical Results *Biopolymers* **1971**, 10(9), 1559-&.

(203) J. Neel Experimental study of the influence of specific intramolecular interactions on the conformation of model molecules. (Peptides and oligopeptides) *Pure Appl. Chem.* **1972**, 31, 201-225.

(204) G. M. Crippen and J. T. Yang Conformation of N-Acetyl-L-Alanine-N'-Methylamide in 1,2-Dichloroethane by Circular-Dichroism and Optical-Rotatory Dispersion *J. Phys. Chem.* **1974**, 78(11), 1127-1130.

(205) V. Madison and K. D. Kopple Solvent-Dependent Conformational Distributions of Some Dipeptides *J. Am. Chem. Soc.* **1980**, 102(15), 4855-4863.

(206) C. F. Weise and J. C. Weisshaar Conformational analysis of alanine dipeptide from dipolar couplings in a water-based liquid crystal *J. Phys. Chem. B* **2003**, 107(14), 3265-3277.

(207) M. A. Mehta, E. A. Fry*, et al.* Structure of the alanine dipeptide in condensed phases determined by C-13 NMR *J. Phys. Chem. B* **2004**, 108(9), 2777-2780.

(208) T. Takekiyo, T. Imai*, et al.* Temperature and pressure effects on conformational equilibria of alanine dipeptide in aqueous solution *Biopolymers* **2004**, 73(2), 283-290.

10. ABBREVIATIONS

ΔH_{vap}	enthalpy of vaporization
ΔH_f	enthalpy of formation
ε_r	dielectric constant
ε_{ij}	well depth in the Lennard-Jones potential energy function
η_h	number of waters in the hydration shell of a given entity
θ_i	equilibrium angle
r_i	equilibrium bond length
τ_c	correlation time: time required to reorient a molecule in solution
τ_f	folding time
τ_m	melting temperature
χ	error between quantum and classical molecular electrostatic potentials
ω-angle	torsion angle defined by the atomic sequence O5-C5-C6-O6 in hexopyranoses
1CVN	PDB ID for the X-ray structure of Concanavalin – A complexed with Man*p*-α-(1→3)[Man*p*-α- (1→6)]-Man*p*
AMBER	*Assisted Model Building and Refinement:* molecular modeling package employed for molecular mechanics energy minimizations, molecular dynamics simulations, and analysis (versions: AMBER5, AMBER7, and AMBER8)
AMBER*	carbohydrate parameter set developed for use in the AMBER force field with implicit solvent
B3LYP	hybrid density functional theory implementing the Becke-3 correlation and the Lee, Yang, and Parr exchange functionals
BSSE	*Basis-Set Superposition Error*
CBS	*Complete Basis Set*
CCSD(T)	Coupled-Cluster level of theory with single (S) and double (D) substitutions, with triple (T) excitations included explicitly
CD	*circular dichroism* spectroscopy
CHARMM	*Chemistry at HARvard Molecular Mechanics*
CHELP	*Charges from Electrostatic Potentials*
CHelpG	grid-based approach to the CHELP partial charge derivation algorithm
Con A	*Concanavalin – A*
Cooperativity	3-body interaction energy/non-additive energetic contributions
CSFF	*Carbohydrate Solution Force Field*: carbohydrate parameter set for use in the CHARMM force field
DFT	*Density Functional Theory*
EP	*Extra Point* (Same as LP)

Gal*p*	Galactopyranoside
Gal*p*-OMe	methyl galactopyranoside
gg	hexopyranosides conformation in which the O5-C5-C6-O6 angle adopts the *gauche* conformation (-60°) and the C4-C5-C6-O6 angle adopts the *gauche* conformation (60°)
Glc*p*	Glucopyranoside
GLYCAM	*Glycosides and Glycoconjugates in AMBER*
2000a	carbohydrate parameter set developed for use with the AMBER PARM94, 96, 98, and 99 protein parameter sets in the AMBER force field
2000b	modification of OH van der Waals from 2000a parameter set
04 (04)	independent carbohydrate parameter set that can be utilized with the AMBER parameters in the AMBER force field
04-LP (04-LP)	modification of the 04 parameter set to include lone-pairs
GROMOS	*GROningen MOlecular Simulations*
GB	*Generalized Born* implicit solvent model
gt	hexopyranosides conformation in which the O5-C5-C6-O6 angle adopts the *gauche* conformation (60°) and the C4-C5-C6-O6 angle adopts the *trans* conformation (180°)
HF	*Hartree-Fock*
HGFB	standard carbohydrate parameter set implemented in the CHARMM force field
IR	*infrared* spectroscopy
k_B	*Boltzman constant*
K_r	bond force constant
K_θ	angle force constant
LP	*Lone-Pair* (Same as EP)
lectin	generally defined as a carbohydrate – binding protein
LMP2	*Local 2nd Order Møller Plesset pertubation theory*
Man*p*	Mannopyranoside
MEP	*Molecular Electrostatic Potential*
methyl α-maltoside	α-Glc*p*-(1→4)-α-Glc*p*-OMe
methyl α-isomaltoside	α-Glc*p*-(1→6)-α-Glc*p*-OMe
MK	*Merz-Kollman* partial charge fitting scheme
MM3	*Molecular Mechanics Program, Version 3*
MP2	*2nd Order Møller Plesset pertubation theory*
[*nφ-γ*]	n – multiplicity, φ – torsion angle, and γ – phase shift

NOE	*Nuclear Overhauser Effect*
NPT	constant number of particles (N), pressure (P), and temperature (T)
OPLS	*Optimized Potentials for Liquid Simulations*
OPLS – AA	*OPLS – All Atom*
ORD	*Optical Rotary Dispersion*
PARM	parameter sets developed for implementation in the AMBER force field for proteins and nucleic acids
PARM94/98	all-atom parameter set that utilized alanine dipeptide and glycine dipeptide conformers for derivation of the backbone torsion parameters
PARM96	modification of protein backbone torsion parameters from PARM94 in which more points along the φ, ψ potential energy surfaced of the alanine dipeptide were included over PARM94
PARM99	modification of the backbone torsion parameters from PARM94 in which more points along the φ, ψ potential energy surface of the alanine dipeptide were included along with ala_4 conformers
PARM02	PARM99 parameter set including atomic polarizabilities
PARM02EP	PARM99 parameter set including lone-pairs (EPs) and atomic polarizabilities
PME	*Particle Mesh Ewald* algorithm
PMF	*Potential of Mean Force*
q	atomic partial charge
QM	*Quantum Mechanical*
RDF	*Radial Distribution Function*
RESP	*Restrained ElectroStatic Potential*
R_{ij}	distance between atom *i* and atom *j*
R_{i*}	van der Waals radius of atom i
R_{ij*}	internuclear separation at the minimum energy position, $R_{i*} + R_{j*}$
SANDER	*Simulated Annealing with NMR-derived Energy Restraints*
SPC	*Simple Point Charge* Water Model
SPC/E	*Extended Simple Point Charge* Water Model
T_1 relaxation	spin-lattice or longitudinal relaxation time
TBD	X-ray structure of Concanavalin – A complexed with Man*p*-α-(1→3)[Man*p*-α- (1→6)]-2-deoxy-2-ethylhydroxy-Man*p*-OMe
tg	hexopyranosides conformation in which the O5-C5-C6-O6 angle adopts the *trans* conformation (180°) and the C4-C5-C6-O6 angle adopts the *gauche* conformation (-60°)
TIP3P	*Transferable Intermolecular Potential, 3 – Point* Water Model
TIP4P	*Transferable Intermolecular Potential, 4 – Point* Water Model
TIP4P-EW	modification of TIP4P for use with the Ewald algorithm for non-

	bonded interactions
TIP5P	*Transferable Intermolecular Potential, 5 – Point* Water Model
TIP5P-EW	modification of TIP5P for use with the Ewald algorithm for non-bonded interactions
TMD	*Temperature at Maximum Density*
WHAM	*Weighted Histogram Analysis Method*
Xyl*p*	Xylopyranoside
ZPVE	*Zero-Point Vibrational Energy*

11. APPENDIX

Perl Program: AddLP.pl

```
# This program was written to add lone-pairs to a list of coordinates.
# List of xyz coordinates (FS=",", no spaces, $1=atomic #, in Bohrs (RESP consistent))
# Based on the distance of nearby atoms – if two atoms are within 1.5 Å of O atom, then
# LPs added in sp3 fashion, i.e. hydroxyl group.  If only one nearest neighbor then sp2.
# Remove division factor for Angstroms.

# ***** DEFINE ALL SUBROUTINES *****

sub normalize {
      @listcc = @_;
      $length = sqrt($listcc[0]**2 + $listcc[1]**2 + $listcc[2]**2);
      @listnormal=($listcc[0]/$length,$listcc[1]/$length,$listcc[2]/$length);
      return @listnormal};

sub scalarproduct {
    @listcc = @_;
    @list1cc = ($listcc[0],$listcc[1],$listcc[2]);
    @list2cc = ($listcc[3],$listcc[4],$listcc[5]);
    $scalar = (($list1cc[0]*$list2cc[0])+
               ($list1cc[1]*$list2cc[1])+
               ($list1cc[2]*$list2cc[2]));
    return $scalar};

sub crossproduct {
    @listcc = @_;
    @list1cc = ($listcc[0],$listcc[1],$listcc[2]);
    @list2cc = ($listcc[3],$listcc[4],$listcc[5]);
    @vector[0] = ($list1cc[1]*$list2cc[2]-$list1cc[2]*$list2cc[1]);
    @vector[1] = ($list1cc[2]*$list2cc[0]-$list1cc[0]*$list2cc[2]);
    @vector[2] = ($list1cc[0]*$list2cc[1]-$list1cc[1]*$list2cc[0]);
    return @vector};

sub distance {
    @listcc = @_;
    @connector=($listcc[0]-$listcc[3],$listcc[1]-$listcc[4],$listcc[2]-$listcc[5],
                $listcc[0]-$listcc[3],$listcc[1]-$listcc[4],$listcc[2]-$listcc[5]);
    $distant = sqrt (scalarproduct (@connector));
    return $distant};

# ***** SPECIFY INPUT FILE *****

open(FILE, "<input.xyz") || die print " Can't open input!";
@input=<FILE>;
$numinput=@input;
for ($p=0; $p<$numinput; $p++) {
        @temp=split(",", $input[$p]);
        for ($q=0; $q<=3; $q++) {
                @matrix[$p]=([$temp[0], $temp[1], $temp[2], $temp[3]])}};

# ***** set ALPHA to the correct sp3 (default 109.54) or sp2 (default 120) value *****

$pi=3.141592654;
@alpha_sp3 = (sqrt(1/3),sqrt(2/3));
@alpha_sp2 = (0.5,sqrt(3/4));

# *****  Sort atoms with respect to distance from the oxygen atom  *****

$num=@matrix ;
for ($i=0; $i<$num; $i++) {
        if ($matrix[$i][0]==8) {
                @oxygen=($matrix[$i][1],$matrix[$i][2],$matrix[$i][3]);
                for ($t=0; $t<$num; $t++) {
```

```
                        $distances[$t] = distance($oxygen[0],$oxygen[1],$oxygen[2],
                                        $matrix[$t][1],$matrix[$t][2],$matrix[$t][3])};
                @distance2 = sort { $a <=> $b } @distances;
                for ($t=0; $t<$num; $t++) {
                        if ($distances[$t]==$distance2[1] ) {$close[0]=$t}};
                for ($t=0; $t<$num; $t++) {
                        if ($distances[$t]==$distance2[2] and $close[0]!=$t)
                        {$close[1]=$t}};

# ***** if third closest is within 1.5 A (can change this) *****
# ***** It is an sp3 hybridized oxygen *****

        if ($distance2[2] < (1.5/.5291772)) {
                print " It's an sp3  distance : $distance2[2]    atom#  : $i\n";
               @closeatom = ([$matrix[$close[0]][1],$matrix[$close[0]][2],$matrix[$close[0]][3]],
                             [$matrix[$close[1]][1],$matrix[$close[1]][2],$matrix[$close[1]][3]]);
                @vector=([normalize($oxygen[0]-$closeatom[0][0],$oxygen[1]$closeatom[0][1],
                             $oxygen[2]-$closeatom[0][2])],[normalize($oxygen[0]-$closeatom[1][0],
                              $oxygen[1]-$closeatom[1][1],$oxygen[2]-$closeatom[1][2])]);
                        @gaa = normalize($vector[0][0],$vector[0][1],$vector[0][2]);
                        @gab = normalize($vector[1][0],$vector[1][1],$vector[1][2]);
                        @gamma = normalize ($gaa[0]+$gab[0],$gaa[1]+$gab[1],$gaa[2]+$gab[2]);
                        @zeta = normalize (crossproduct($vector[0][0],$vector[0][1],$vector[0][2],
                                        $vector[1][0],$vector[1][1],$vector[1][2]));
                        @alpha  = @alpha_sp3};

# ***** if third closest is farther than 1.5 A *****
# ***** It is an sp2 hybridized oxygen *****

         if ($distance2[2] >= (1.5/.5291772) )   {
         print " It's an sp2   distance : $distance2[2]      atom#  : $i\n";
        @oxygen=($matrix[$i][1],$matrix[$i][2],$matrix[$i][3]);
        @carbon=($matrix[$close[0]][1],$matrix[$close[0]][2],$matrix[$close[0]][3]);
                for ($t=0; $t<$num; $t++) {
                        $distances[$t] = distance($carbon[0],$carbon[1],$carbon[2],
                                        $matrix[$t][1], $matrix[$t][2],$matrix[$t][3])};
        @distance3 = sort { $a <=> $b } @distances;
        if ($distance3[1]==$distance2[1]) {$closeatom2 = 2};
        if ($distance3[1]!=$distance2[1]) {$closeatom2 = 1};
        for ($t=0; $t<$num; $t++) {
                if ($distances[$t]==$distance3[$closeatom2]) {$close[0]=$t}};
        @gab= normalize($matrix[$close[0]][1]-$carbon[0],$matrix[$close[0]][2]-$carbon[1],
                        $matrix[$close[0]][3]-$carbon[2]);
        @gamma = normalize($oxygen[0]-$carbon[0],$oxygen[1]-$carbon[1],
                        $oxygen[2]-$carbon[2]);
        @eta = normalize(crossproduct($gamma[0],$gamma[1],$gamma[2],
                        $gab[0],$gab[1],$gab[2]));
        @zeta = normalize(crossproduct($eta[0],$eta[1],$eta[2],
                        $gamma[0],$gamma[1],$gamma[2]));
        @alpha  = @alpha_sp2};

# *****  Set the LP-oxygen distance *****
# *****  Get the final LP location  and Print Them *****

        for ($lplength1=(0.70/.5291772)) {
                @lp1a=($alpha[0]*$gamma[0]+$alpha[1]*$zeta[0],$alpha[0]*$gamma[1]
                        +$alpha[1]*$zeta[1],$alpha[0]*$gamma[2]+$alpha[1]*$zeta[2]);
                @lp2a=($alpha[0]*$gamma[0]-$alpha[1]*$zeta[0],
                        $alpha[0]*$gamma[1]-$alpha[1]*$zeta[1],
                        $alpha[0]*$gamma[2]-$alpha[1]*$zeta[2]);
                @lp1=(($lplength1*$lp1a[0])+$oxygen[0],($lplength1*$lp1a[1])+$oxygen[1],
                        ($lplength1*$lp1a[2])+$oxygen[2]);
                @lp2=(($lplength1*$lp2a[0])+$oxygen[0],($lplength1*$lp2a[1])+$oxygen[1],
                        ($lplength1*$lp2a[2])+$oxygen[2]);
        open(APPEND1, ">> output") || die print "Can't open newfile! ";
        printf APPEND1 ("%32.7E %15.7E %15.7E\n", $lp1[0], $lp1[1], $lp1[2]);
        printf APPEND1 ("%32.7E %15.7E %15.7E\n", $lp2[0], $lp2[1], $lp2[2])};
}};
```

Modifications to the AMBER7 source code for GLYCAM04-LP

i. Filename “unit.c” in the LEAP subprogram to treat LP identically to EP

Original file lines 280-284

```
            /* Don't include torsions relating to extra points:  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "EP" ) == 0 || strcmp( sAtom4, "EP" ) == 0 )
            continue;
        }
```

Modified file

```
            /* Don't include torsions relating to extra points:  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "EP" ) == 0 || strcmp( sAtom4, "EP" ) == 0 )
            continue;
        }
            /* Don't include torsions relating to lone pairs (SMW):  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "LP" ) == 0 || strcmp( sAtom4, "LP" ) == 0 )
            continue;
        }
```

Original file lines 421-424

```
/*  delete all angles related to extra points    */
        if( GDefaults.iDeleteExtraPointAngles ){
        if ( strcmp( sA, "EP" ) == 0 || strcmp( sC, "EP" ) == 0 ) return(TRUE);
        }
```

Modified file

```
/*  delete all angles related to extra points    */
        if( GDefaults.iDeleteExtraPointAngles ){
        if ( strcmp( sA, "EP" ) == 0 || strcmp( sC, "EP" ) == 0 ) return(TRUE);
        if ( strcmp( sA, "LP" ) == 0 || strcmp( sC, "LP" ) == 0 ) return(TRUE);
        }
```

ii. Filename “unitio.c” in the LEAP subprogram to treat LP identically to EP

Original file lines 1853-1857

```
        /* skip if this torsion refers to an extra point:  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "EP" ) == 0 || strcmp( sAtom4, "EP" ) == 0 )
            continue;
        }
```

Modified file

```
        /* skip if this torsion refers to an extra point:  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "EP" ) == 0 || strcmp( sAtom4, "EP" ) == 0 )
            continue;
        }
        /* skip if this torsion refers to an extra point:  */
        if( GDefaults.iDeleteExtraPointAngles ){
            if( strcmp( sAtom1, "LP" ) == 0 || strcmp( sAtom4, "LP" ) == 0 )
            continue;
        }
```

Original file lines 4332-4338

```
    /*NUMEXTRA*/
    iNumExtra = 0;
    for ( i=0; i<iVarArrayElementCount(uUnit->vaAtoms); i++ ) {
                cPTemp = sAtomType( PVAI(uUnit->vaAtoms, SAVEATOMt, i )->aAtom );
                if( !strncmp( cPTemp, "EP", 2 )) iNumExtra++;
    }
```

Modified file

```
    /*NUMEXTRA*/
    iNumExtra = 0;
    for ( i=0; i<iVarArrayElementCount(uUnit->vaAtoms); i++ ) {
                cPTemp = sAtomType( PVAI(uUnit->vaAtoms, SAVEATOMt, i )->aAtom );
                if( !strncmp( cPTemp, "EP", 2 ) || !strncmp( cPTemp, "LP", 2 ) )
iNumExtra++;
    }
```

Original file lines 6344-6350

```
    /*NUMEXTRA*/
    iNumExtra = 0;
    for ( i=0; i<iVarArrayElementCount(uUnit->vaAtoms); i++ ) {
                cPTemp = sAtomType( PVAI(uUnit->vaAtoms, SAVEATOMt, i )->aAtom );
                if( !strncmp( cPTemp, "EP", 2 )) iNumExtra++;
    }
```

Modified file

```
    MESSAGE(( "Writing NumExtra\n" ));
    FortranFormat( 5, LBLFORMAT );
    iNumExtra = 0;
    for ( i=0; i<iVarArrayElementCount(uUnit->vaAtoms); i++ ) {
                cPTemp = sAtomType( PVAI(uUnit->vaAtoms, SAVEATOMt, i )->aAtom );
                if( !strncmp( cPTemp, "EP", 2 ) || !strncmp( cPTemp, "LP", 2 ) )
iNumExtra++;{
        }
    }
```

iii. Filename “extra_pts.f” in the SANDER subprogram to increase the number of lone-pairs per molecule and treat LP identically to EP

Original file

line 122
```
integer ep,blank,n,max14
```

line 129
```
max14 = 12*(nphih+nphia+ndper)
```

lines 160-161
```
integer ep,blank,n,numextra_test
data ep/4HEP  /
```

line 172
```
 if ( isymbl(n) .eq. ep )then
```

line 209
```
+    natom,isymbl,ep,
```

line 286
```
+    numatoms,isymbl,ep,
```

line 291
```
+    isymbl(*),numatoms,ep
```

line 311
```
 if ( isymbl(ii) .eq. ep ) then
```

line 316
```
 else if ( isymbl(jj) .eq. ep ) then
```

Modified file

line 122
```
integer ep,lp,blank,n,max14
```

line 129
```
max14 = 24*(nphih+nphia+ndper)
```

lines 160-161
```
integer ep,lp,blank,n,numextra_test
data ep/4HEP  /, lp/4HLP  /
```

line 172
```
 if ((isymbl(n) .eq. ep) .or. (isymbl(n) .eq. lp))then
```

line 209
```
+    natom,isymbl,ep,lp,
```

line 286
```
+    numatoms,isymbl,ep,lp,
```

line 291
```
+    isymbl(*),numatoms,ep,lp
```

line 311
```
 if (( isymbl(ii) .eq. ep) .or. ( isymbl(ii) .eq. lp)) then
```

line 316
```
 else if (( isymbl(jj).eq.ep).or.(isymbl(jj) .eq. lp)) then
```