
Theses and Dissertations

2013

Carbohydrate-protein interactions: structure, dynamics and free energy calculations

Sai Kumar Ramadugu
University of Iowa

Copyright 2013 Sai Kumar Ramadugu

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/1731>

Recommended Citation

Ramadugu, Sai Kumar. "Carbohydrate-protein interactions: structure, dynamics and free energy calculations." PhD (Doctor of Philosophy) thesis, University of Iowa, 2013.
<http://ir.uiowa.edu/etd/1731>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Chemistry Commons](#)

CARBOHYDRATE-PROTEIN INTERACTIONS: STRUCTURE, DYNAMICS
AND FREE ENERGY CALCULATIONS

by

Sai Kumar Ramadugu

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Chemistry
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Claudio J. Margulis

ABSTRACT

The current thesis presents work on the structure and dynamics of oligosaccharides and polysaccharides as well as the free energetics of carbohydrate-protein interactions. By applying various computational tools such as molecular dynamics simulations, our in-house fast sugar structure prediction software, replica exchange molecular dynamics, homology modeling, umbrella sampling, steered molecular dynamics as well as the thermodynamic integration formalism, we have been able to study the role of water on the surface of homopolysaccharides as well as complex oligosaccharides, we have been able to produce a prediction of the bound structure of triantennary oligosaccharide on the asialoglycoprotein receptor, we have been able to estimate the free energy of binding of $\text{Man}\alpha 1\rightarrow 2\text{Man}$ to the HIV-1 inactivating protein, Cyanovirin-N as well as the relative binding free energies of mutants of Cyanovirin-N to the same ligand.

Abstract Approved: _____
Thesis Supervisor

Title and Department

Date

CARBOHYDRATE-PROTEIN INTERACTIONS: STRUCTURE, DYNAMICS
AND FREE ENERGY CALCULATIONS

by

Sai Kumar Ramadugu

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Chemistry
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Claudio J. Margulis

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Sai Kumar Ramadugu

has been approved by the Examining Committee for the
thesis requirement for the Doctor of Philosophy degree
in Chemistry at the May 2013 graduation.

Thesis committee: _____

Claudio J. Margulis, Thesis Supervisor

Sarah C. Larsen

Christopher M. Cheatum

Amanda J. Haes

Ernesto J. Fuentes

ACKNOWLEDGEMENTS

I would like to thank God Almighty for everything.

I would like to thank my adviser, Prof Claudio J. Margulis. He is a great mentor, teacher and a caring person. I remember the days that we spent discussing research in his office, the long hours editing manuscripts, those barbecues at Lake MacBride, Coralville Lake and the get-togethers at his house. I learnt a great deal of science during my time in his lab.

I am also thankful to my parents, my wife Spandana, my brother and sister-in-law for believing in me and encouraging me to apply for PhD programs in the United States and particularly to the University of Iowa.

I would like to thank various past members of the Margulis lab, Dr. Hu, Dr. Annapureddy, Dr. Chung and post-doctoral fellows, Dr. Xia, Dr. Rahaman, Dr. Kashyap and all current members of the lab.

I would like to thank all the people in the department of chemistry, my friends in Iowa City (the list is humongous) and in India who made the journey to Ph.D. possible and life well lived.

ABSTRACT

The current thesis presents work on the structure and dynamics of oligosaccharides and polysaccharides as well as the free energetics of carbohydrate-protein interactions. By applying various computational tools such as molecular dynamics simulations, our in-house fast sugar structure prediction software, replica exchange molecular dynamics, homology modeling, umbrella sampling, steered molecular dynamics as well as the thermodynamic integration formalism, we have been able to study the role of water on the surface of homopolysaccharides as well as complex oligosaccharides, we have been able to produce a prediction of the bound structure of triantennary oligosaccharide on the asialoglycoprotein receptor, we have been able to estimate the free energy of binding of $\text{Man}\alpha 1\rightarrow 2\text{Man}$ to the HIV-1 inactivating protein, Cyanovirin-N as well as the relative binding free energies of mutants of Cyanovirin-N to the same ligand.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER	
1 INTRODUCTION	1
2 ROLE OF WATER ON THE SURFACE OF POLY- AND OLIGOSACCHARIDES	7
2.1 Introduction	7
2.2 Materials and Methods	11
2.3 Results and Discussion	16
2.3.1 The Effect of Saccharide Structure on the Surrounding Solvent Structure and Residence Time	19
2.3.2 Solvent Rotational and Translational Motion at the Saccharide Surface	23
2.4 Conclusions	33
3 IN SILICO PREDICTION OF THE 3-D STRUCTURE OF TRIMERIC ASIALOGLYCOPROTEIN RECEPTOR BOUND TO TRI-ANTENNARY OLIGOSACCHARIDE	35
3.1 Introduction	35
3.2 Methods	37
3.2.1 Fast Sugar Structure Prediction Software Studies	38
3.2.2 Replica Exchange Molecular Dynamics Simulations	39
3.2.3 Structural Alignment Studies	41
3.2.3.1 Identifying the Binding Site for H1	41
3.2.3.2 Alignment of Trimeric H1 of ASGP-R with TA	41
3.2.3.3 Homology Modeling of CRD of H2 subunit	42
3.2.4 Molecular Dynamics Studies of the Predicted Complex	44
3.3 Results and Discussion	47
3.3.1 FSPPS and REMD Studies of Free TA	47
3.3.2 Structural Alignment Studies	48
3.3.3 Simulated Annealing Studies of the H1-H1-H2-TA complex	56
3.3.4 Comparison of Structure and Dynamics of Homo- and Heterotrimeric Complexes	57

3.4	Conclusions	60
4	POTENTIAL OF MEAN FORCE CALCULATIONS OF DIMANNOSE BINDING TO CYANOVIRIN-N	61
4.1	Introduction	61
4.2	Methods	64
4.2.1	Model of Cyanovirin-N and Force Field Parameters	64
4.2.2	Steered Molecular Dynamics and Umbrella Sampling	66
4.2.2.1	AMBER11 Simulations	66
4.2.2.2	GROMACS Simulations	67
4.3	Results	68
4.3.1	Pulling Pathway in Gromacs: 1-D Pulling Along Z- or Y-Axis	68
4.3.2	Potential of Mean Force: 1-D Along Y- or Z-Axis	69
4.3.3	Pulling Pathway in AMBER11: 3D Pulling Along X,Y,Z-Axis	71
4.4	Conclusions	72
5	ROLE OF GLU41 IN BINDING OF DIMANNOSE TO P51G-M4-CVN	76
5.1	Introduction	76
5.2	Methods	77
5.2.1	System Preparation	77
5.2.2	Thermodynamic Integration	79
5.3	Results	83
5.3.1	Relative Binding Free Energy Calculations	83
5.4	Conclusions	84
6	SUMMARY AND FUTURE DIRECTIONS	95
	APPENDIX	97
	A AMBER99SB/GLYCAM06 PARAMETER CONVERSION TO GROMACS FORMAT	97
	A.1 Implementation of Mixed-Scaling	97
	REFERENCES	99

LIST OF TABLES

Table

2.1	Common names of the sugars	11
2.2	Characteristic residence times (τ values) for our model saccharides derived from simulation.	21
3.1	Dihedral angle increments for linkages in TA.	39
5.1	Thermodynamic Integration Analysis for Relative Binding Free Energies for the Glu41Ala Mutation.	85
5.2	Thermodynamic Integration Analysis for Relative Binding Free Energies for the Glu41Gly Mutation.	89

LIST OF FIGURES

Figure		
1.1	Complexity in carbohydrate linkage. Representative structures of digalactose are shown with different linkages and anomeric configurations. . . .	4
1.2	Triantennary Oligosaccharide: Sequence and Structure. Top panel, the sequence of triantennary oligosaccharide (TA), a natural ligand of the ASGP-R and bottom panel, the 3D structure of TA	5
1.3	Structure of the High Mannose Oligosaccharide, Man ₉ GlcNAc ₂ . Cyanovirin-N binds to the Man α 1 \rightarrow 2 moiety of D1, D2 or D3 arms of the high mannose oligosaccharide.	6
2.1	Some of the various oligosaccharide models used in the study. The Man ₉ oligosaccharide was manipulated to generate the different variants Man ₉ 1, Man ₉ 2 and Man ₉ dimer	12
2.2	Helical parameters n (number of residues per pitch) and h (advancement per monomer unit in Å) from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of α 1 \rightarrow 2, β 1 \rightarrow 2, α 1 \rightarrow 3, β 1 \rightarrow 3, α 1 \rightarrow 4, β 1 \rightarrow 4-glucans. Negative values of n represent a left-handed helix whereas positive values represent right-handed helix. These figures can be compared with French's work figures 2-6.	17
2.3	Representative structures from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of α 1 \rightarrow 2, β 1 \rightarrow 2, α 1 \rightarrow 3, β 1 \rightarrow 3, α 1 \rightarrow 4, β 1 \rightarrow 4-glucans. Mannose linear homopolymers are not shown in this figure but they have very similar secondary structures to the glucose analogues. In general, we find that linear homopolysaccharides with α 1 \rightarrow 2, β 1 \rightarrow 2, β 1 \rightarrow 3 and α 1 \rightarrow 4 linkages form wide helices while those with α 1 \rightarrow 3 and β 1 \rightarrow 4 linkages tend to form extended helices in solution.	18
2.4	Logarithm of the fraction of molecules with first passage time greater or equal to t as a function of t. The slopes of these plots correspond to the residence times.	20

2.5	Water density around saccharides with varied sequence, branching, linkage and secondary structure. From left to right top to bottom the saccharides are: $\alpha 1 \rightarrow 2$ -glucose, modified Man ₉ , $\alpha 1 \rightarrow 3$ -glucose, $\beta 1 \rightarrow 4$ -glucose and $\alpha 1 \rightarrow 3$ -mannose	24
2.6	Comparison of typical snapshots along simulation for $\alpha 1 \rightarrow 3$ -glucose (left) and $\alpha 1 \rightarrow 3$ -mannose (right) showing the loci of enhanced water trapping. The dashed lines indicate the distance between O6.....O2 and O2.....O4, showing the epimeric effect. In case of mannose the O2 is axial which makes the water molecule trapped in comparison to glucose. The epimeric difference between Man and Glu results in tighter solvent configurations in contact with the surface in the case of Mannose.	25
2.7	Water rotational correlation functions in the first solvation shell on the saccharide surface. Inset shows the complete decay of this function in the range from 0 to 10ps.	26
2.8	Comparison of rotational correlation functions for water on the first solvation shell of homopolysaccharides. (A) shows a comparison between different glucose nonamers, (B) shows a comparison between different mannose nonamers, (C) through (E) compare rotational correlation functions of water surrounding glucose and mannose nonamers with different linkages.	30
2.9	The mean square displacement of water around selected saccharide models chosen for this study.	31
2.10	A comparison of the the mean square displacement in units of Å ² of water around different nonamers models characterized by their linkage position, anomeric and epimeric configuration.	32
3.1	The sequence of a tri-antennary oligosaccharide (TA). The 1→6 linked arm is conventionally termed the 6' Arm, the 1→2 linked arm is the 6 Arm and the 1→4 linked arm the 8 Arm	45
3.2	Discrete Optimized Potential Energy (DOPE) score per residue for the template (H1 subunit of ASGP-R, PDB ID 1DV8) and target (H2 subunit of ASGP-R). The target (H2 subunit,) is shown in green and the template, (H1 subunit,) in red.	46
3.3	Dihedral angle map for TA. In red are the results from our REMD simulations and in green those from the FSPS for the 6' (a), 6 (b) and 8 (c) Arms.	49

3.4	Representative structures of TA derived from REMD simulations. The orientation of the 6', 6 and 8 Arms is indicated in the first snapshot and is the same for all other conformations.	50
3.5	Normalized probability distribution for distances between C1 of the terminal Gal in the 6', 6 and 8 Arms with respect to C1 of the terminal GlcNAc in the core region. Arrows indicate peaks referred to in the text. Inset shows the corresponding free energy functions derived from these probability distributions.	51
3.6	Structural Alignment of the H1 subunit of ASGP-R with the QPDWG mutant of MBP A. (a) All-atom alignment of the H1 subunit with the QPDWG mutant. The RMSD between the two proteins is 1.27 Å(all atom). Bound Gal is shown in stick format. (b) Zoom-in view of the active site. The labeled residues are those of H1.	53
3.7	Clockwise and counter-clockwise receptor-subunit arrangements. Attached to the 6 and 6' Arms are H1 subunits, and attached to the 8 Arm is an H2 subunit. H1 and H2 subunits are color coded and labeled to indicate the arm to which they are attached. Two types of arrangements, clockwise and counter-clockwise, can be derived from our studies. Since the protein subunits are large, it is unlikely that there is exchange between these conformations once the sugar is bound to the protein trimer.	55
3.8	The radii of gyration as a function of time for the four models of TA bound to H1-H1-H2 from our SAMD simulations. The first 10 ns (solid lines) correspond to SAMD whereas the subsequent 2.5 ns (circles) are unconstrained NPT simulations at 300K. CCW refers to the counter-clockwise model and CW to clockwise model that are described in Figure 3.7.	58
3.9	Comparison between TA-H1 subunit and TA-H2 subunit interactions at the 8 Arm. Red line represents the interaction energy between TA and H1 at the 8 Arm in the homo-oligomeric complex as a function of time during simulation. Green line represents the interaction energy between TA and H2 at the 8 Arm in the hetero-oligomeric complex. Blue line represents the interaction between ASP ²⁹⁰ of H2 and TA.	59
4.1	The pseudo-symmetry of Cyanovirin-N. Domain A is colored magenta and domain B is colored green. One can align domain A and domain B and both units	62
4.2	Snapshots and dissociation pathway along Y-axis. The overall pathway during the dissociation of dimannose along the Y-axis(a), snapshot at 5ns (b) and snapshot at 10ns (c).	70

4.3	Potential of Mean Force along Y- and Z-Axis for dimannose binding to Cyanovirin-N	71
4.4	Snapshots and dissociation pathway along 3D in AMBER11 Simulations. The overall pathway during the dissociation of dimannose is shown in (a), snapshot at 100ps (b) and snapshot at 400ps (c)	73
4.5	Potential of Mean Force obtained from AMBER11 Simulations. In the case of AMBER11, the pulling was in arbitrary directions	74
5.1	Trans configuration for the χ_1 angle of Glu 41 showing hydrogen bonding with 2'OH of the non-reducing end of dimannose (a). χ_1 of Glu41 in the gauche conformation centered around 60°.	78
5.2	Thermodynamic Cycle for Binding Free Energy Calculation	81
5.3	Three step approach for Thermodynamic Integration simulations, Trial I. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).	86
5.4	Three step approach for Thermodynamic Integration simulations, Trial II. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).	87
5.5	Three step approach for Thermodynamic Integration simulations, Trial III. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).	88
5.6	Three step approach for Thermodynamic Integration simulations, Trial I. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).	90
5.7	Three step approach for Thermodynamic Integration simulations, Trial II. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).	91
5.8	Three step approach for Thermodynamic Integration simulations, Trial III. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).	92
5.9	Torsional preferences of dimannose in solution (black), bound to E41-P51G-m4-CVN (red), bound to A41-P51G-m4-CVN (magenta) and bound to G41-P51G-m4-CVN (blue)	93

CHAPTER 1 INTRODUCTION

Along with proteins, lipids and vitamins, carbohydrates are an important class of biomolecules. Most functions within the cell are affected by carbohydrate-protein interactions and protein glycosylation is often a fundamentally important post-translational modification. In 2002, it was estimated that around 70% of the structures available in the protein data bank (PDB) included the sequence Asn-X-Ser/Thr, where a carbohydrate can link covalently to the Asn residue forming what is commonly referred to as the N-glycosylation.[146] Instead O-glycosylation, is defined as the covalent attachment of a carbohydrate moiety to the oxygen of a serine or threonine. A recent review article highlights evidence that any significant change in the O-GlcNAc leads to type II diabetes, cancer and neurodegenerative disease. [155]

Attempting to predict the structure of homo-polysaccharides, heteropolysaccharides or complex oligosaccharides is in itself a daunting task. Carbohydrates, unlike their protein counterparts, simply do not have a backbone and as opposed to proteins which are linear polymers, carbohydrates are often branched. Since these molecules are polyalcohols, there are various plausible linkage points such as 1-1, 1-2, 1-3, 1-4 and 1-6 [113](see Figure 1.1) and this is further complicated by the fact that monosaccharide residues exist in two anomeric configurations, namely α (where the OH group at C1 is axial) and β (where the OH group at C1 is equatorial) with respect to the 4C_1 chair conformation. In the past, our group has put significant effort in understanding the conformational preferences of oligosaccharies by developing and

using a fast sugar structure prediction software as well as by carrying out molecular dynamics simulation studies. [148, 149]

Biologically relevant carbohydrates must be studied as species in solution. Understanding the role of water on the surface of oligosaccharides is important since by often forming solvent-mediated hydrogen bonds, water is not a passive spectator in the binding of carbohydrate ligands to proteins. [35, 88, 84] Our own studies[112] show that for various types of carbohydrates, the residence times, diffusivities and local water number density as well as water rotational correlation times on the surface of carbohydrates depend on linkage type, anomeric configuration and sequence. Accordingly this could influence the binding ability of different carbohydrate ligands to proteins.

A significant number of bacterial, viral and fungal infections are at some stage associated with carbohydrates binding to proteins. For example, during HIV attachment to a host cell, glycoprotein-120 (gp120) on the virus capsid, which is highly decorated with high-mannose oligosaccharide [123] binds to receptors on the host cell. Likewise in the cell, many important carbohydrate interactions appear to involve a terminal sialic acid residue (N-acetylneuraminic acid, Neu5Ac or simply Sia). When the sialic acid binds to a receptor, corresponding signal transduction events occur. To recover proteins for degradation and reuse it is often the case that terminal sialic acid residues are chopped off by sialidases. In the case when the remaining carbohydrate motif is the galactose exposed triantennary oligosaccharide (see figure 1.2) the hepatic asialoglycoprotein receptor (ASGP-R)[7] recognizes it by binding to galactose

and inducing receptor mediated endocytosis.

Such multivalent interactions responsible for the high affinity binding of carbohydrate ligands to their respective proteins/receptors are common and the idea to utilize this concept to develop carbohydrate-based vaccines/drugs for effective therapies is not new. [78, 103] In order to fully understand the multivalent nature of carbohydrate based ligands, one must first unravel structural details and subsequently focus on the free energy of such interactions. Methodology and an attempt at the prediction of the structure of one such type of multivalent system is described in chapter 3 of this thesis where we study the triantennary oligosaccharide in complex with the asyialoglycoprotein receptor.

Once structural details of protein-carbohydrate interactions are established, the role of key amino acids at the binding site can be computationally evaluated in order to propose mutations that could potentially lead to enhanced binding affinities.

Among the very many systems that can be studied, one blue-green algal lectin, Cyanovirin-N is a promising candidate. This protein binds to the $\text{Man}\alpha 1 \rightarrow 2\text{Man}$ moiety of the D1 and D2 arms of high mannose (Man_9), see Figure 1.3. [26, 79] Several experimental [10, 13, 15, 74, 90, 94, 152, 26, 27, 55, 54] studies are available to date on various structural models of CVN and are complemented by computational studies.[93, 58, 143, 144, 57] However the significance of particular protein residues in binding and the possible enhancing role of mutations is not yet understood. This is the subject of chapters four and five in this thesis.

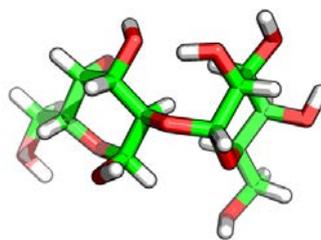
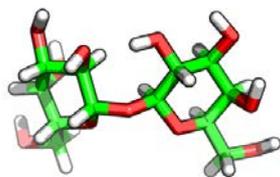
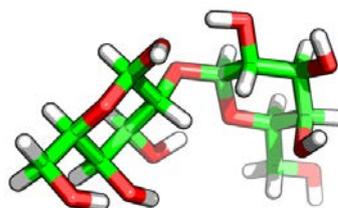
(a) Gal α 1→1Gal(b) Gal α 1→2Gal(c) Gal β 1→1Gal(d) Gal β 1→2Gal

Figure 1.1: Complexity in carbohydrate linkage. Representative structures of digalactose are shown with different linkages and anomeric configurations.

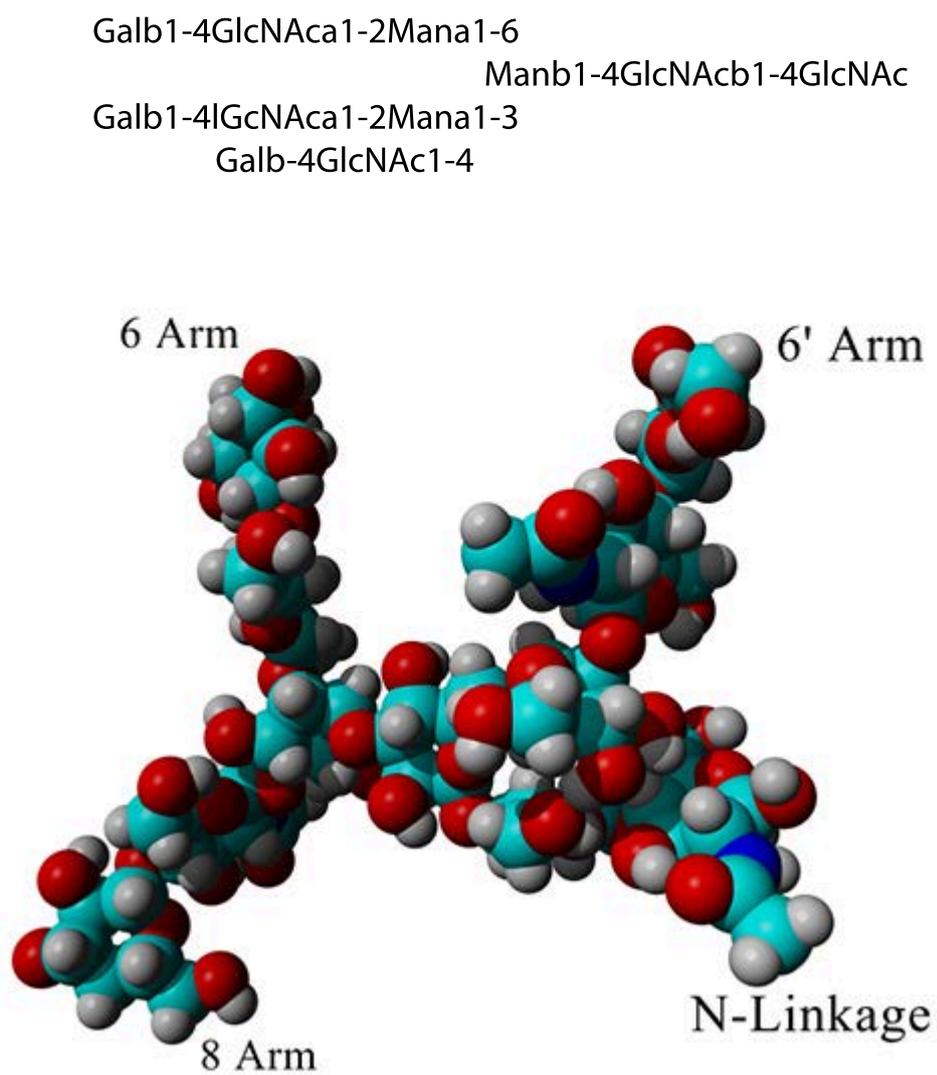


Figure 1.2: Triantennary Oligosaccharide: Sequence and Structure. Top panel, the sequence of triantennary oligosaccharide (TA), a natural ligand of the ASGP-R and bottom panel, the 3D structure of TA

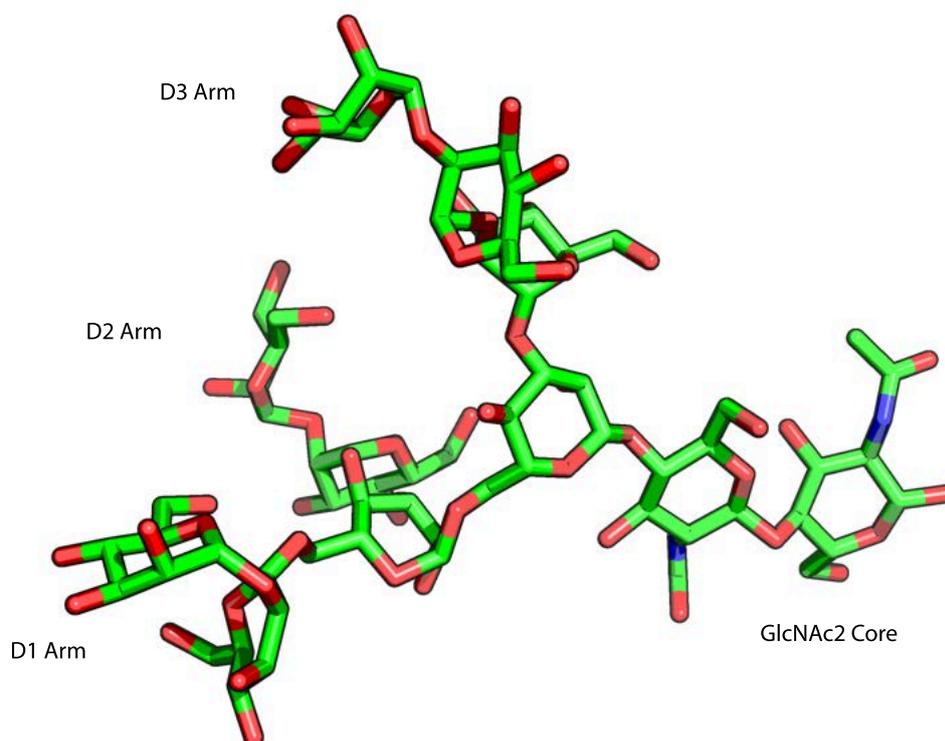


Figure 1.3: Structure of the High Mannose Oligosaccharide, Man₉GlcNAc₂. Cyanovirin-N binds to the Man α 1 \rightarrow 2 moiety of D1, D2 or D3 arms of the high mannose oligosaccharide.

CHAPTER 2

ROLE OF WATER ON THE SURFACE OF POLY- AND OLIGOSACCHARIDES

2.1 Introduction

Carbohydrates are an important class of biomolecules that can be found free in the cytoplasm, decorating the surface of proteins and as parts of glycolipids. They are involved in cell adhesion, immune responses, protein trafficking and signal processing. Understanding the structure and function of oligosaccharides is very important because these molecules are exquisite biological recognition agents. This uniqueness stems from the large number of chiral centers, the presence of branching and their conformational variability. As an example, in diseased states, glycans are expressed differently and act as biomarkers in cancer, AIDS and rheumatoid arthritis [141, 6, 46, 45].

Dashnau *et al* showed that the orientation of hydroxyl groups in axial and/or equatorial positions in aldohexopyranoses affects the water structuring in the first hydration shell. Aldohexopyranoses such as β -glucose, β -mannose and β -galactose have hydrophobic and hydrophilic hydration sites that play a role in aromatic interaction during carbohydrate-protein recognition [41, 132]. Resonance two photon ionization and ultraviolet and infrared ion-dip spectroscopy of hydrated mono- and disaccharides have shown that water on the surface of carbohydrates helps these biomolecules achieve the conformations that are recognized by proteins i.e., water is not a mute spectator but it actively participates in molecular recognition events [127] (see ref-

erence 7 and citations therein). In contrast with the large amount of information available for the role of water in contact with proteins and nucleic acids, no comprehensive study of the role of water on the surface of glycans is available. This is perhaps because of the topological complexity of sugars. It is therefore of crucial interest [135, 139] to shed light on the water structure patterns and the diffusive dynamics on the surface of carbohydrates as a function of the key elements present in carbohydrates but absent in proteins such as branching, linkage pattern and anomeric configuration [68, 43, 1].

Several experimental and theoretical studies have reported on the behavior of water at the interface with carbohydrates. Because of the complexity of these systems, most of these studies have been carried out on monosaccharides, disaccharides or small model oligosaccharides. Kirchner and Woods have performed high level quantum and molecular dynamics simulations and have shown that the conformational preferences for the 1→6 linkage are correctly reproduced only in the presence of water [77]. In exploring the role of water in the vicinity of simple monosaccharides such as α -d-glucopyranose and α -d-xylopyranose, Leroux *et al.* observed that the hydroxyl groups of the monosaccharide units align in such a way that they form hydrogen bonds with water instead of intra-molecular hydrogen bonds [89]. Consistent with these computational predictions, recent depolarized Rayleigh scattering (DRS) and low-frequency Raman Spectroscopy experiments performed on an aqueous glucose solution by Paolantoni *et al* showed that a solute with the ability to have multiple hydrogen bonds disrupts the tetrahedral geometry of water in its first hydration shell.

The loss of hydrogen bonding between water molecules is compensated by sugar-water hydrogen bonds leading to the denser water environment around the sugar [104].

Lee *et al.* have shown that disaccharides such as sucrose and trehalose not only disrupt the tetrahedral geometry of water in their vicinity but also its translational and rotational dynamics. Dynamics on the surface of these disaccharides is much slower than that on the surface of glucose [86]. Liu *et al* have shown that α,α -trehalose imposes a strong anisotropic structuring of solvent that extends up to three solvation shells away from the sugar due to the formation of water mediated H-bonds. In computational studies the self-diffusion coefficient of water in 87 μM α,α -trehalose solution was found to be 20% smaller than that in neat water simulations [91]. Englesen *et al* studied disaccharides maltose, sucrose and trehalose in dilute aqueous solutions. The H-bonding pattern, the solvent residence times and the solvent density around these disaccharides were observed in simulation to be different. Water surrounding trehalose displayed the longest residence times and was clearly more structured than in the vicinity of maltose and sucrose [48, 47].

Almond *et al* have elucidated the role of water on the surface of small oligosaccharides composed of α - and β -linkages of glucose and mannose. In the case of α -linkages, weak intramolecular hydrogen bond interactions along contiguous residues and many water mediated hydrogen bonds were observed contrary to the situation when β -linkages are present. In the case of β -linkages water mediated hydrogen bonds appeared not to be favored; instead strong direct hydrogen bonds with water were observed [2, 3].

Although the above studies are very important, as far as we know there have been no systemic and comprehensive analysis of the water structure and dynamics on the surface of moderate to large oligosaccharides consisting of various linkages and monosaccharide compositions. Several outstanding questions arise; how do water structure and dynamics differ around a linear homopolymer of glucose or mannose and across 1→2, 1→3 and 1→4 linkages? How does the anomeric configuration of the different monosaccharides affect surrounding water dynamics? How do branched oligosaccharides differ from linear oligosaccharides with the same monosaccharide composition in their ability to affect water structure and dynamics?

In order to address the effect of branching on a given type of oligosaccharides, we have studied several variations of the Man₉ sugar. In order to address issues related to the anomeric configuration, linkage point and monosaccharide identity, we have studied homopolymers of 9 monomeric units of 1→2, 1→3, and 1→4 linkages of α and β glucose and mannose sugars (Figure 2.1 and Table 2.1). The effect of polymer length is addressed by comparing the results for nonamers with two selected twenty-mers.

The Margulis group has previously written several articles [151, 148, 149, 93, 142] addressing the problem of sugar conformational variability in solution. The current article does not attempt to address this issue but instead is written from the point of view of the solvent around well defined sugar structures. In subsequent sections we will show that water structure and dynamics on the surface of sugars depends on the overall structure of the biomolecule, its anomeric configuration and

types of linkages. We will also show that some simple rules of thumb can be derived by systematically analyzing the results derived from our simulations.

2.2 Materials and Methods

The models chosen for this study are categorized into 4 groups. Homopolysaccharides of glucose, mannose, branched sugars and a trisaccharide. Within the homopolysaccharides we have studied 6 glucose and 6 mannose nonamers and 2 mannose twenty-mers. For easy reference, the branched structures and the glucose homo nonamers as well as the trimer are depicted in Figure 2.1. Common names for some of these molecules can be found in Table 2.1.

Common Name	residue	linkage
Amylose	Glucose	$\alpha 1 \rightarrow 4$
Cellulose	Glucose	$\beta 1 \rightarrow 4$
Crown Gall	Glucose	$\beta 1 \rightarrow 2$
Laminaran	Glucose	$\beta 1 \rightarrow 3$
Mannan	Mannose	$\beta 1 \rightarrow 4$

Table 2.1: Common names of the sugars

All oligosaccharide models were built using the xleap module in AMBER 9.0 [32]. The choice of the initial conformations was based on previous experimental information. [101, 121, 122, 133, 154, 4, 109]. We used the GLYCAM [77] force field for all our calculations. The total number of atoms in the system ranged from 5000-14000 depending on the model studied. All the oligosaccharides were solvated using

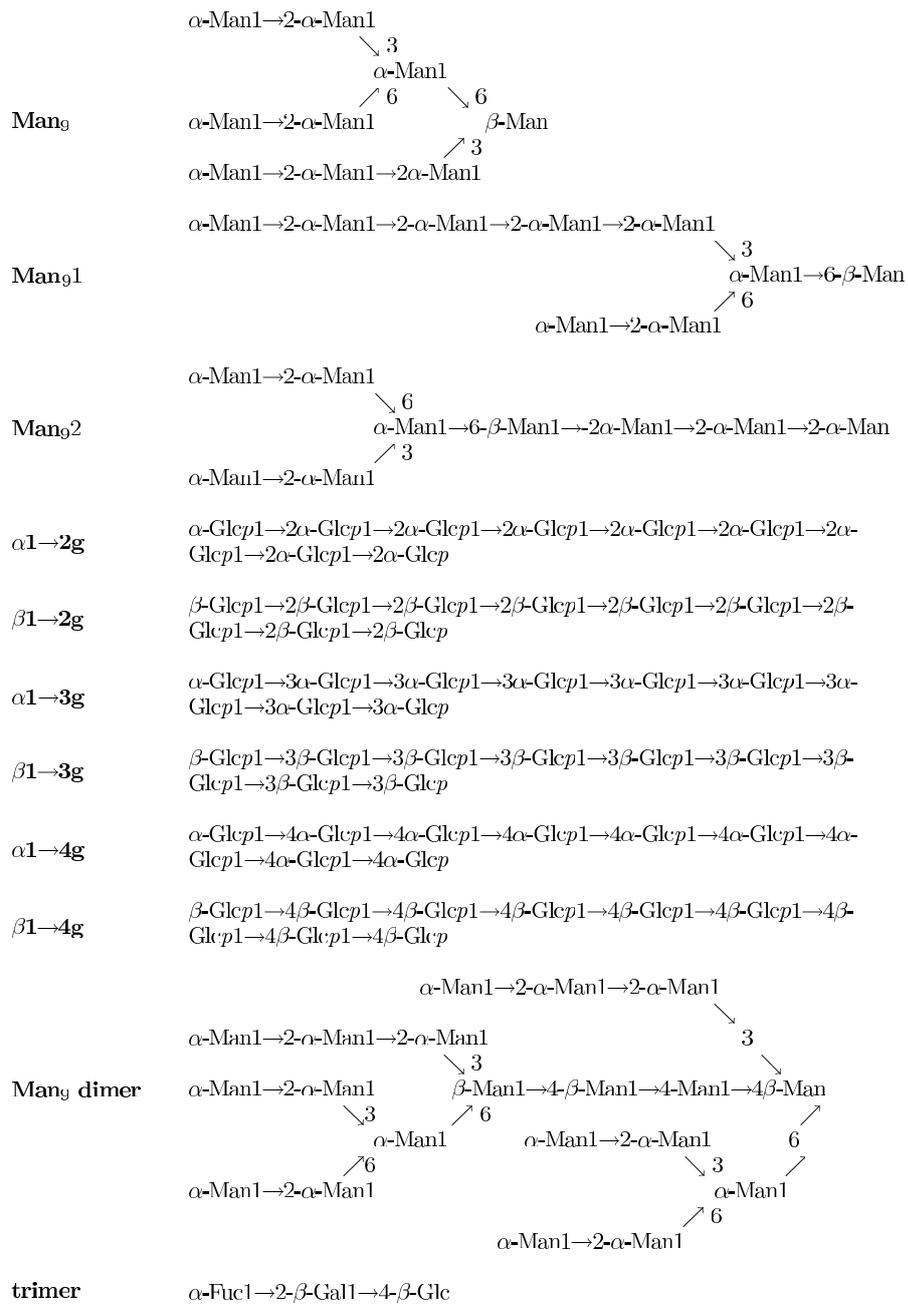


Figure 2.1: Some of the various oligosaccharide models used in the study. The Man₉ oligosaccharide was manipulated to generate the different variants Man₉1, Man₉2 and Man₉ dimer

SPC [71] water in truncated octahedron boxes. Long range electrostatic interactions were treated using the Particle Mesh Ewald summation [40]. Molecules were energy minimized using the steepest descent method followed by the conjugate gradient method. After constant pressure and temperature equilibrations for at least 250ps at ambient conditions, production molecular dynamics simulations in the NVE (constant energy, volume and number of atoms) ensemble of 5 ns in duration were carried out with the sander module of AMBER 9.0. Because of the time scale separation between solvent dynamics and sugar conformational dynamics 5 ns was enough to converge all correlation functions. Data was collected every 50 fs for subsequent analysis.

Chung and coworkers have previously developed theoretical methodology and a computational algorithm that significantly improve on the accuracy and reliability of residence time calculations on complex surfaces [34]. The problem with the proverbial methodology for the computation of residence times[69] at complex surfaces is not the theory but the actual implementation of this theory. Chung *et al* have demonstrated[34] that because of the way residence times are computed during finite simulations, solvent molecules that stay trapped on a surface for long periods of time significantly contribute to the survival probability time correlation function at short times and are the only ones contributing to this function at long times. The problem is that one has extremely bad statistics on solvent molecules that get trapped on a complex surface for long times since this is a rare event on typical computational time scales[34]! Because of statistical uncertainty, if one runs identical simulations with different initial random conditions the number of molecules that reside for long times on a surface will be different in each case [34]. Of course if molecular dy-

namics simulations were of infinite duration this would not be a problem, but in actuality the survival time correlation function of which the time constants define the residence time commonly differ significantly from simulation to simulation. The algorithm developed by Chung and coworkers [34] circumvents this problem by computing a different function (a normalized cumulative probability of passage times) that has been shown to be devoid of the aforementioned statistical instability and that is rigorously proven to be related in a differential way to Impey's [69] original survival probability time correlation function. When this function can be accurately fitted to an arbitrary sum of exponentials, the corresponding time constants are by definition identical to those derived from the proverbial algorithm but are much more statistically reliable. Equation 2.1 defines the distribution of first passage times.

$$\tilde{H}_{accum}(\geq m) = \sum_{j=1}^{N_{\text{sys}}} \sum_{m'=m}^{N_f-2} \frac{1}{N_f - m'} \sum_{n=1}^{N_f - m' - 1} [1 - A_j(n-1)][1 - A_j(n+m'+1)] \prod_{k=n}^{n+m'} A_j(k) \quad (2.1)$$

Here N_{sys} represents the total number of solvent molecules in the system. N_f is the total number of trajectory frames recorded and $A_j(k)$ takes a value of 1 if molecule j lies within the hydration shell (4\AA) at time frame k (i.e. at time $k * dt$); otherwise its value is 0. The product enforces that molecule j is found within the shell at all times from time frame n until time frame $(n+m')$. $[1 - A_j(n-1)] * A_j(n) = 1$ and $A_j(n+m') * [1 - A_j(n+m'+1)] = 1$ represent entering and leaving events. The fraction of molecules with first passage time greater or equal to $m * dt = \tau$ is

$$\tilde{f}(\geq \tau) = \frac{\tilde{H}_{accum}(\geq \tau)}{\tilde{H}_{accum}(\geq 0)} \quad (2.2)$$

and it can be proven[34] that

$$\tilde{f}(\geq \tau) = \frac{\frac{\partial f(\tau)}{\tau}}{\frac{\partial f(\tau)}{\tau}|_{\tau=0}} \quad (2.3)$$

where $f(\tau)$ is the proverbial survival time correlation function.

In order to compute solvent mean square displacements in this article, we used equations 2.4 and 2.5 from reference [34].

$$\langle \Delta r(t)^2 \rangle = \frac{1}{N_w(t)} \sum_{t_0}^{T_{\text{run}}-t} \sum_{j=1}^{N_w(t_0,t)} [r_j(t+t_0) - r_j(t_0)]^2, \quad (2.4)$$

where N_w is the number of water molecules inside the hydration shell at t_0 , which remained so until $t_0 + t$, T_{run} is the simulation time and

$$N_w(t) = \sum_{t_0}^{T_{\text{run}}-t} N_w(t_0, t). \quad (2.5)$$

In the analysis of the rotational dynamics of water, we calculate the autocorrelation function $C(t)$ defined in Eq. 2.6

$$C(t) = \frac{\langle \hat{v}(t_0 + t) \cdot \hat{v}(t_0) \rangle}{\langle \hat{v}(t_0) \cdot \hat{v}(t_0) \rangle} \quad (2.6)$$

Here $\hat{v}(t)$ is the unit vector directed along the dipole moment vector of water at time t . In the calculation of residence times, MSDs and rotational autocorrelation functions, the hydration shell is defined as the region within 4\AA from the solute.

3-D solvent occupancy plots were generated on typical 50ps fragments of our trajectories using the Chimera software from UCSF [110]. Grid cells of 1\AA^3 appear colored only when they are occupied by solvent molecules 10% of the time. This particular value was selected because it allows for optimal discrimination between our sugar models. At higher values, almost no occupancy can be detected for $\beta 1 \rightarrow 4$ homopolysaccharides.

In order to classify our linear saccharides, helical parameter, n (number of residues per pitch) and h (advancement per monomer unit) were calculated according to the def-

initiation in the following references. [113, 52, 115] According to Rees'[115] classification of perfectly periodic helices our studied sugars fall into the ribbon family $n=2\pm 4$ and $h=4-6\text{\AA}$ (this is what Rao's book[113] calls extended helices) and the hollow helix family $n=2\pm 10$ and $h=0-4\text{\AA}$ (this is what Rao's book[113] calls wide helices).

2.3 Results and Discussion

Several X-ray and conformational studies have been reported on homopolymers of $1\rightarrow 2$, $1\rightarrow 3$ and $1\rightarrow 4$ glucose and $1\rightarrow 4$ mannose. [101, 121, 122, 133, 154, 4, 109] In order to validate our simulation methodology we have computed torsional angles as a function of time for these sugars and found that both the minimum energy configurations as well as fluctuations are consistent with previously reported values [101, 121, 122, 133, 154, 4, 109]. We are therefore confident that sugar conformations are those experimentally reported and they do not significantly change on the time scale of our 5 ns simulations.

In order to classify the morphology of our saccharides, we computed helical parameters $n-h$ along simulation using the method described by French. [51, 52] Results are presented in Fig. 2.2. Though results are similar, the dispersion in our data is larger than that in reference [51] because our simulations do not force the saccharides to remain in a perfect helical configuration. According to reference [51] and our findings (see representative structures in Fig. 2.3), as the homopolysaccharide chain length is increased, these sugars form either wide helices (helices with large diameter that have a large number of residues per turn) or extended helices (helices with small diameter and low number of residues per turn). Based on the definitions in Section 2.2 and Fig. 2.2 we find that linear homopolysaccharides with linkages $\alpha 1\rightarrow 2$, $\beta 1\rightarrow 2$, $\alpha 1\rightarrow 4$ and $\beta 1\rightarrow 3$ tend to form wide helices and $\alpha 1\rightarrow 3$ and $\beta 1\rightarrow 4$ linkages tend to form extended helices

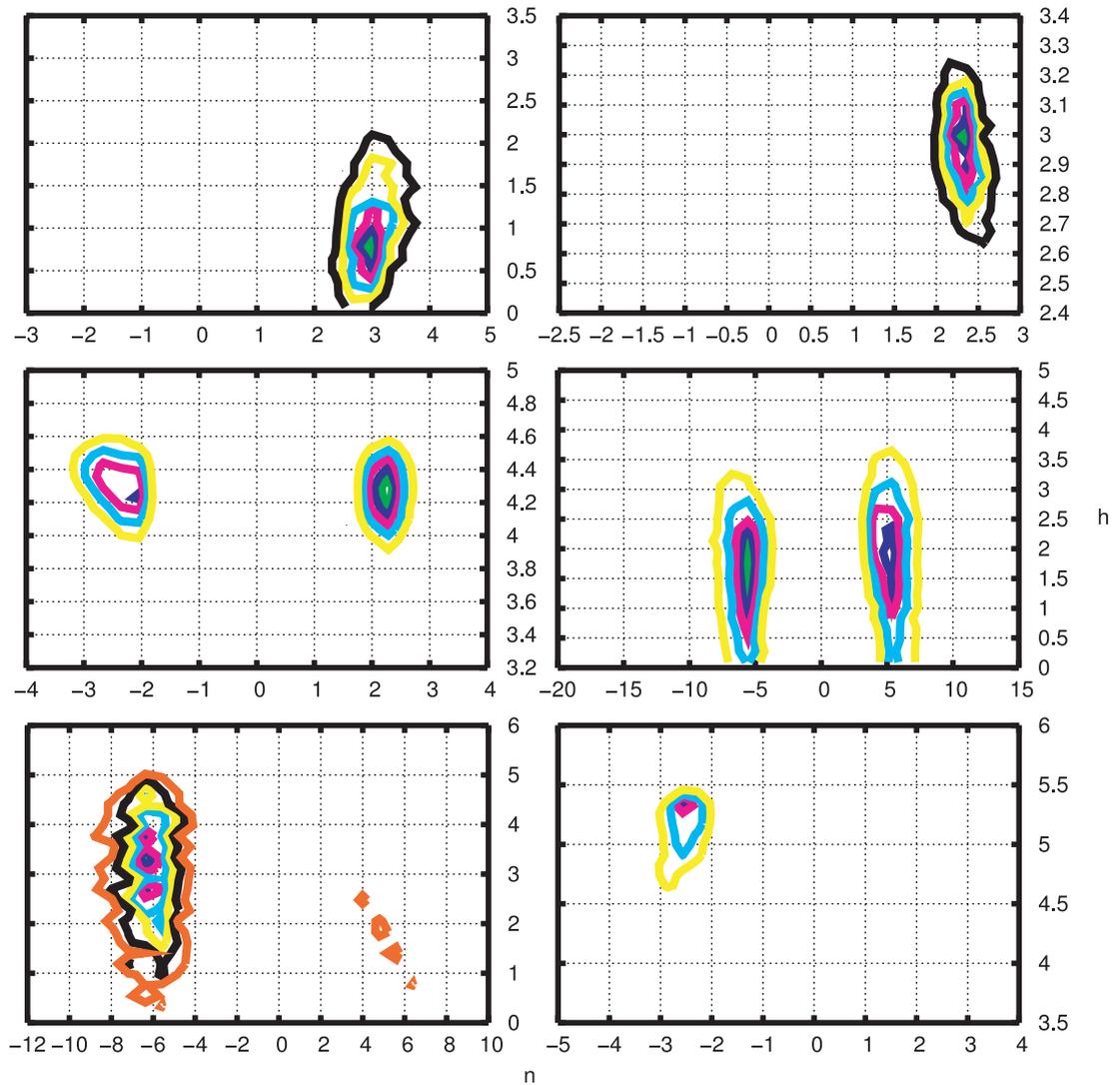


Figure 2.2: Helical parameters n (number of residues per pitch) and h (advancement per monomer unit in \AA) from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$, $\alpha 1 \rightarrow 3$, $\beta 1 \rightarrow 3$, $\alpha 1 \rightarrow 4$, $\beta 1 \rightarrow 4$ -glucans. Negative values of n represent a left-handed helix whereas positive values represent right-handed helix. These figures can be compared with French's work figures 2-6.

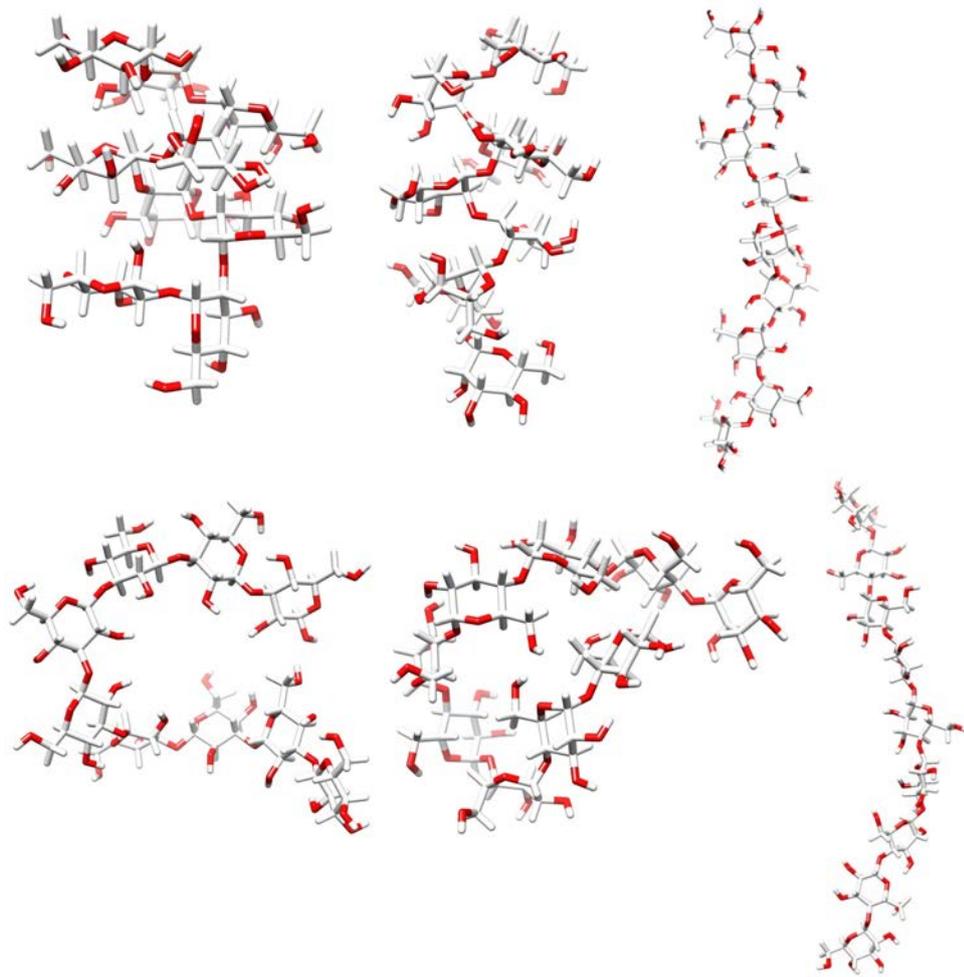


Figure 2.3: Representative structures from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$, $\alpha 1 \rightarrow 3$, $\beta 1 \rightarrow 3$, $\alpha 1 \rightarrow 4$, $\beta 1 \rightarrow 4$ -glucans. Mannose linear homopolymers are not shown in this figure but they have very similar secondary structures to the glucose analogues. In general, we find that linear homopolysaccharides with $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$, $\beta 1 \rightarrow 3$ and $\alpha 1 \rightarrow 4$ linkages form wide helices while those with $\alpha 1 \rightarrow 3$ and $\beta 1 \rightarrow 4$ linkages tend to form extended helices in solution.

2.3.1 The Effect of Saccharide Structure on the Surrounding Solvent Structure and Residence Time

The fraction of molecules with first passage time greater than or equal to t ($\tilde{f}(\geq t)$) is a measure of the behavior of water in contact with a surface and can be used to compare the nature of different sugar-solvent interfaces. In this study, we start by attempting to address the effect of linkage, anomeric configuration, secondary structure, branching, size and monosaccharide identity on $\tilde{f}(\geq t)$. Fig. 2.4 shows the logarithm of $\tilde{f}(\geq t)$ for a broad set of sugars with widely different structural characteristics ranging from a trisaccharide to a dimer of Man₉ and two structurally very different linear mannose twenty-mers. Some of these saccharides, are linear, others are highly branched and they have different linkage and anomeric configurations. For clarity this figure also shows typical conformations in solution for each of these molecules and surrounding solvent occupancy isosurfaces.

Since Fig. 2.4 is on a logarithmic scale, the slope of these curves correspond to the inverses of the characteristic water residence times on the surface of the different saccharides. Table 2.2 shows the value of these inverse slopes between 20 and 30ps. In broad terms, it is clear from Fig. 2.4 and Table 2.2 that branched or in general crowded sugars as well as sugars forming wide helices have longer solvent residence times. The residence time on the surface of extended helices is in general much shorter. Secondary structure plays a crucial role in defining the residence time of water on the sugar surface. Size appears to be less of a factor since for example Man₂₀ β 1 \rightarrow 4 and a β 1 \rightarrow 4 nonamer have almost identical residence times.

Heyden *et al* have studied solvation dynamics of water surrounding saccharides of different size using terahertz spectroscopy. In a comparison of a monosaccharide with two

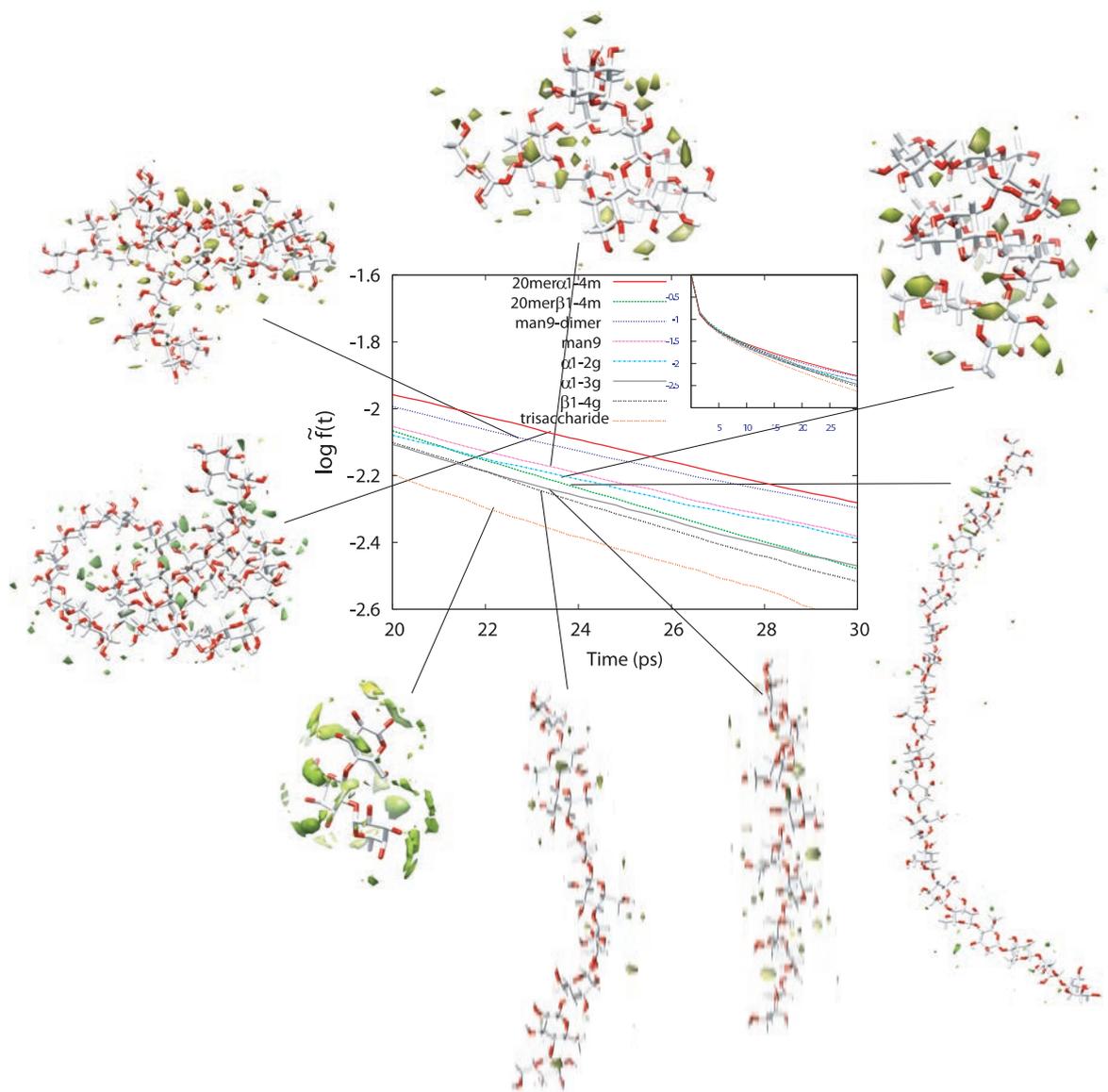


Figure 2.4: Logarithm of the fraction of molecules with first passage time greater or equal to t as a function of t . The slopes of these plots correspond to the residence times.

Model	τ (ps)	Secondary Structure
$\beta 1 \rightarrow 2m$	34	Wide Helix
man ₉ dimer	33	Branched
$\alpha 1 \rightarrow 3m$	33	Extended Helix †
$\alpha 1 \rightarrow 2g$	33	Wide Helix
$\beta 1 \rightarrow 3g$	32	Wide Helix
$\beta 1 \rightarrow 2g$	31	Wide Helix
Man ₉	31	Branched
$\alpha 1 \rightarrow 2m$	31	Wide Helix
Man20 $\alpha 1 \rightarrow 4$	30	Wide Helix
$\beta 1 \rightarrow 3m$	30	Wide Helix
$\alpha 1 \rightarrow 4g$	28	Wide Helix
$\alpha 1 \rightarrow 3g$	27	Extended Helix
Man20 $\beta 1 \rightarrow 4$	24	Extended Helix
$\alpha 1 \rightarrow 4m$	24	Wide Helix
$\beta 1 \rightarrow 4g$	23	Extended Helix
Trisaccharide	23	No Secondary Structure
$\beta 1 \rightarrow 4m$	23	Extended Helix

Table 2.2: Characteristic residence times (τ values) for our model saccharides derived from simulation.

disaccharides the authors observed that water hydrogen bond dynamics in the first solvation shell of the sugar was significantly less affected in the case of the monosaccharide [62]. Larger oligosaccharides with particular secondary structure patterns favor water mediated hydrogen bonds. In calculating the residence times of water in aqueous solution of sucrose, Englesen and Pérez [47] found that the presence of water mediated hydrogen bonds results in longer water residence times on the sugar surface. In particular, they observed that two water mediated hydrogen bonds, namely O-2g...Ow...O-3f and O-2g...Ow...O-1f were present in their simulations for more than 40% of time. More recently, Veluraja and Margulis found the same type of behavior in Sialyl LewisX containing oligosaccharides [142].

Among all the homopolysaccharide models selected for this study, the residence times of $\beta 1 \rightarrow 2$ -m and $\beta 1 \rightarrow 4$ -g or mannose constitute two extremes. Based on our distributions of first passage times, we are able to determine that residence times in the case of $\beta 1 \rightarrow 2$ -mannose (wide helix) is largest while those for $\beta 1 \rightarrow 4$ -glucose and mannose (extended helices) are shortest. The residence times for the solvent around all other nonamer homopolysaccharides fall in between these two limits. The distinction between wide and extended helices also explains the results in the case of our mannose twenty-mers in Fig. 2.4. The twenty-mer forming a wide helix has longer water residence time.

The effect of branching can be gauged by analyzing a group of different oligosaccharides belonging to the Man_9 family. The function of $\tilde{f}(\geq t)$ for all of these is fairly similar and the τ values are similar to that for Man_9 in table 2.2.

Most of our findings regarding residence times on the surface of homopolysaccharides of glucose and mannose can be explained in terms of their secondary structure and the extent to which they are able to trap water on their surface[91, 36, 2, 48]. As previously discussed, the $\alpha 1 \rightarrow 2$ -, $\beta 1 \rightarrow 2$ -glucose/mannose, $\beta 1 \rightarrow 3$ -[25] glucose/mannose and $\alpha 1 \rightarrow 4$ -glucose/mannose tend to form wide helices while $\alpha 1 \rightarrow 3$ -glucose and mannose and $\beta 1 \rightarrow 4$ -glucose [114] and mannose tend to form extended helices. Our analysis of local water occupancies appear to indicate that in wide helical conformations, the orientation of the pyranose rings is such that small pockets of water can get trapped giving rise to loci of high local density. On the contrary, in the case of extended helical conformation, solvent residence times are shorter, and for the most part solvent occupancies do not show appreciable water trapping though some interesting exceptions exist.

In Fig. 2.5 we show five characteristic examples of water occupancy isosurfaces in

contact with saccharides of different sequence, linkages and secondary structure. The top two saccharides $\alpha 1 \rightarrow 2$ -glucose, and a modified Man₉ correspond to a wide helix and a branched oligosaccharide respectively. In both cases significant water trapping and slow down occurs on the surface. From top to bottom, the following two structures, $\alpha 1 \rightarrow 3$ -glucose and $\beta 1 \rightarrow 4$ -glucose are extended helices; for these the residence time of water on the surface is shorter and water dynamics is faster. There are clear differences between the two extended sugars. In the case of $\alpha 1 \rightarrow 3$ -glucose (see Fig. 2.5) some water trapping can clearly be observed due to the particular structural pattern that facilitates the formation of water mediated hydrogen bonds between O6-O2-O4 in three consecutive residues. In the case of $\beta 1 \rightarrow 4$ -glucose almost no water trapping can be observed at the same iso-solvent occupancy contour level. A very interesting exception to the wide/extended rule appears to be $\alpha 1 \rightarrow 3$ -mannose (bottom of Fig. 2.5). Table 2.2 clearly shows that $\alpha 1 \rightarrow 3$ m has one of the largest τ values. This is due to a combination of secondary structure similar to that in $\alpha 1 \rightarrow 3$ -glucose and an epimeric effect that favors water trapping (see Figure 2.6).

2.3.2 Solvent Rotational and Translational Motion at the Saccharide Surface

While the distribution of first passage times and corresponding residence times provide information about how long a typical solvent molecule will remain in contact with the saccharide surface, it does not provide detailed information regarding its mobility. The calculation of mean square displacements and rotational correlation functions for water molecules on the saccharide surface convey information regarding the degree to which solvent motion is hindered and enables us to obtain deeper understanding of the similarities and differences between saccharide surfaces.

Figure 2.7 shows rotational correlation functions for water molecules in the first

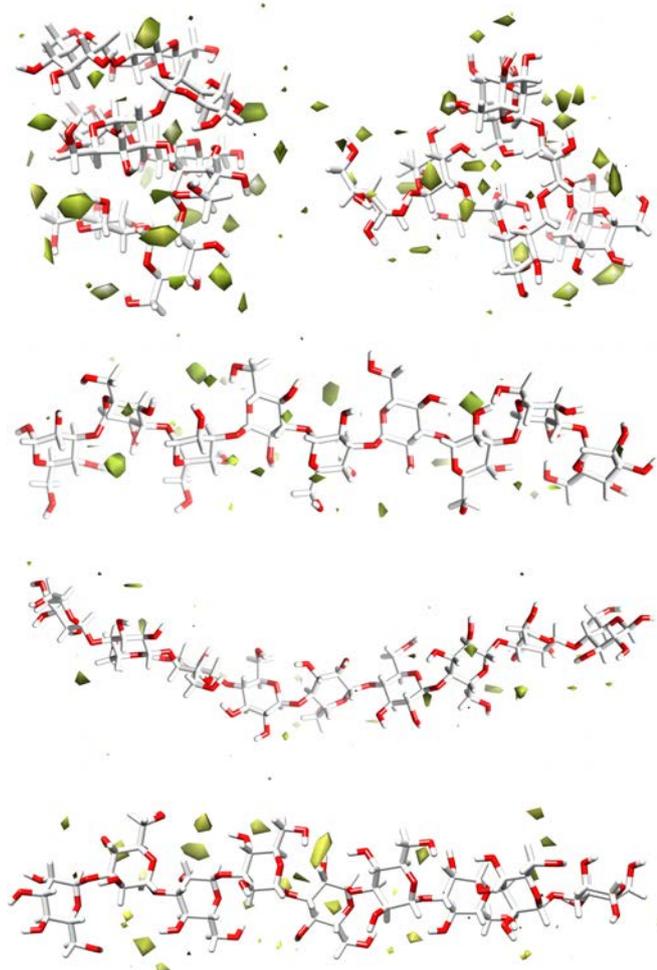


Figure 2.5: Water density around saccharides with varied sequence, branching, linkage and secondary structure. From left to right top to bottom the saccharides are: $\alpha 1 \rightarrow 2$ -glucose, modified Man_9 , $\alpha 1 \rightarrow 3$ -glucose, $\beta 1 \rightarrow 4$ -glucose and $\alpha 1 \rightarrow 3$ -mannose

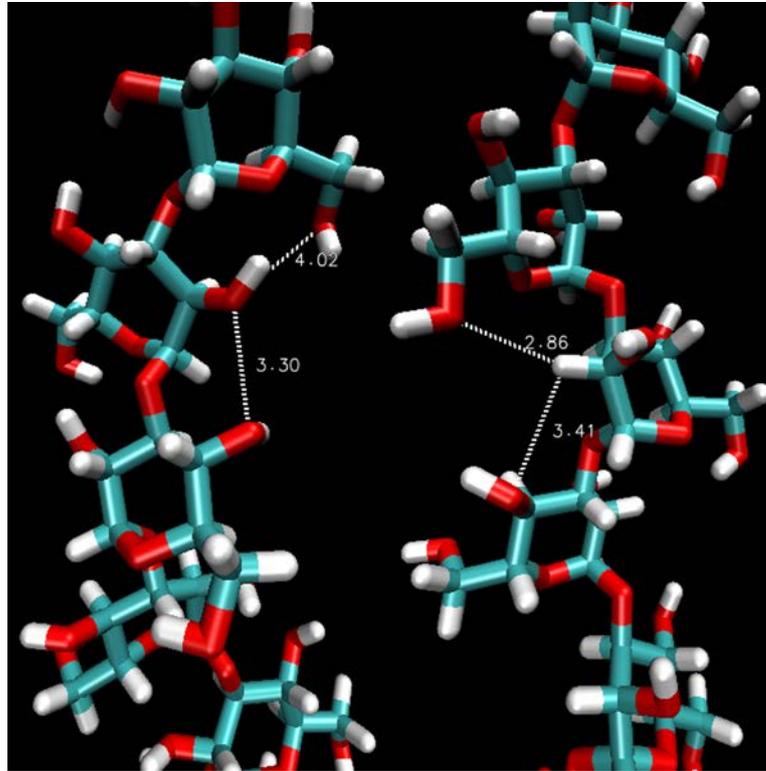


Figure 2.6: Comparison of typical snapshots along simulation for α 1 \rightarrow 3-glucose (left) and α 1 \rightarrow 3-mannose (right) showing the loci of enhanced water trapping. The dashed lines indicate the distance between O6.....O2 and O2.....O4, showing the epimeric effect. In case of mannose the O2 is axial which makes the water molecule trapped in comparison to glucose. The epimeric difference between Man and Glu results in tighter solvent configurations in contact with the surface in the case of Mannose.

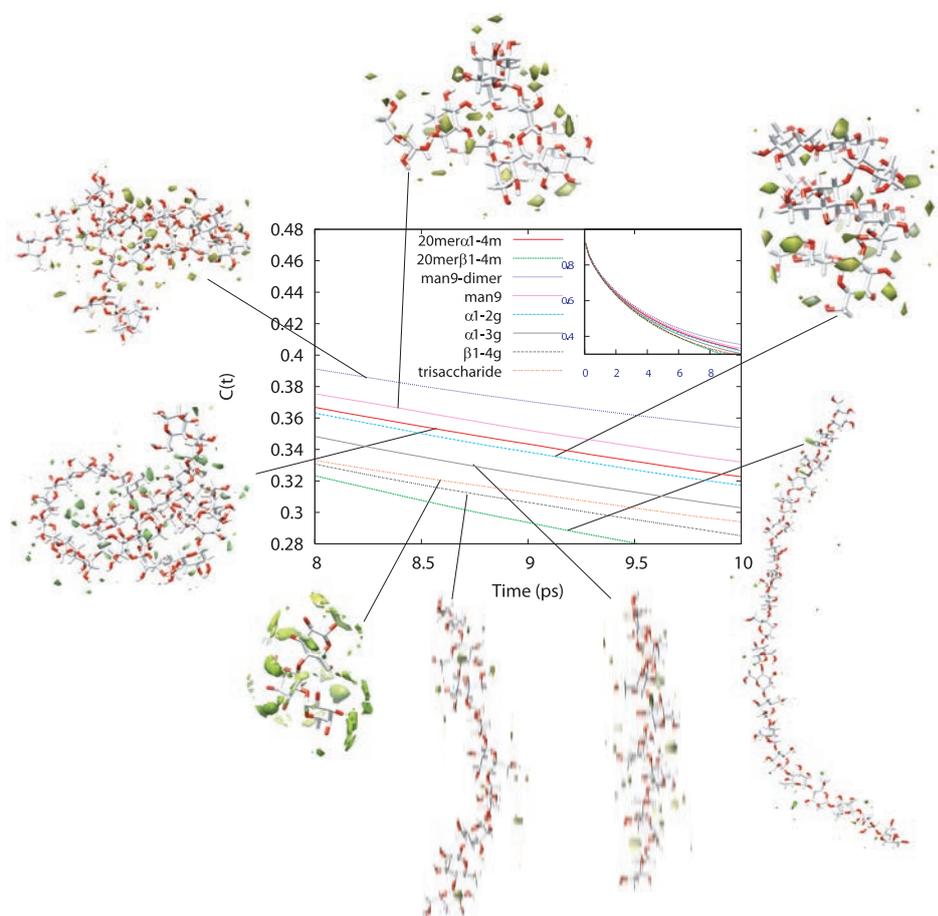


Figure 2.7: Water rotational correlation functions in the first solvation shell on the saccharide surface. Inset shows the complete decay of this function in the range from 0 to 10ps.

solvation shell around selected saccharides of different characteristics. Consistent with our previous analysis, the fastest relaxation occurs in the case of $\beta 1 \rightarrow 4$ -mannose twenty-mer followed by $\beta 1 \rightarrow 4$ -glucose nonamer. Fig. 2.7 clearly shows that these two are extended helices with modest ability for water trapping. Similarly fast decay of correlation is observed in the case of the trisaccharide. Intermediate behavior is observed in the case of $\alpha 1 \rightarrow 3$ -glucose an extended helix that shows moderate ability to trap water as demonstrated by solvent isosurfaces as well as residence time. The decay of correlation is slowest in the case of wide helices and complex branched oligosaccharides such as Man₉dimer and the $\alpha 1 \rightarrow 4$ -mannose twenty-mer.

A more detailed analysis of the rotational behavior of water surrounding homopolysaccharides can be obtained from Figure 2.8. Figure 2.8(A) and (B) compare the rotational behavior of glucose and mannose nonamers. The first notable difference between epimers can be seen in the case of $\alpha 1 \rightarrow 3$ and $\alpha 1 \rightarrow 4$ linkages. In the case of glucose, the water rotational autocorrelation functions at the surface of $\alpha 1 \rightarrow 3$ -glucan decays at an intermediate rate, faster than a typical wide helix such as $\beta 1 \rightarrow 2$ -glucose and slower than an extended helix such as $\beta 1 \rightarrow 4$ -glucose. In the case of mannose the water rotational correlation functions on the surface of $\alpha 1 \rightarrow 3$ -mannose shows uncommonly slow decay while that of $\alpha 1 \rightarrow 4$ -mannose displays a fast decay. The case of $\alpha 1 \rightarrow 3$ -m and $\alpha 1 \rightarrow 3$ -g are specially interesting because of the particular arrangement of atoms O6-O2-O4 in three consecutive rings that make water residence time on the surface unusually large compared to the case on the surface of other extended helices. The epimeric configuration at O2 in mannose makes this effect even more pronounced. The epimeric configuration of these two sugars can be clearly seen in Fig. 2.6 while the different local 3D solvent occupancies are shown in Fig. 2.5. The configuration of

O2 in the case of mannose facilitates the formation of a significantly rigid water structure that promotes water mediated hydrogen bonds between O6 and O4.

The case of $\alpha 1 \rightarrow 4$ -linkage is peculiar since it is known that amylose ($\alpha 1 \rightarrow 4$ -glucan) attains a wide range of helical conformations. [109] This may explain why water rotational correlation functions have dissimilar decays for glucose and mannose homopolymers. It is clear that 5ns is enough to study water structure and dynamics on systems with well defined helicity, however when the separation of time scales between water dynamics and saccharide conformational dynamics is smaller our results may reflect a combination of the water dynamics on two different saccharide secondary patterns.

Figures 2.8 (C) (D) and (E) compare the rotational correlation functions of water on the surface of α and β homopolymers at different linkage positions. It is clear as we have discussed in previous paragraphs that $\alpha 1 \rightarrow 3$ -m is special in the slow decay of water rotational correlation as can be appreciated from Figures 2.8 (C). The rest of the $1 \rightarrow 3$ homopolymers have water rotational correlation functions on their surface that decays at faster rates. In general, we find that regardless of whether they are α or β the $1 \rightarrow 2$ (Figures 2.8 (D)) homopolymers have water rotational correlation functions that decay slowly. This is because they all tend to form wide helices with many locations suitable for tight water mediated hydrogen bonds. Figures 2.8 (E) can also be easily understood. Except for $\alpha 1 \rightarrow 4$ -g which is in the wide helix configuration during simulation all other homopolymers form extended structures that are less suitable for water trapping.

The relationship between the ability of a sugar to form water mediated hydrogen bonds and the decay of surrounding solvent rotational correlation functions has been observed before. It has been shown earlier for the pentasaccharides of cellulose ($\beta 1 \rightarrow 4$ -glucose)

and maltose (α 1 \rightarrow 4-mannose) that solvent rotational correlation functions are slow decaying when compared to bulk water and also when compared to the decay on the surface of disaccharides [140].

The translationally diffusive behavior of water on the surface of saccharides can be studied using equation 2.4. We can get an accurate picture of the similarities and differences in the behavior of water on the surface of wide or extended helices as well as complex branched and small oligosaccharides by analyzing Fig. 2.9. The conclusions that can be derived from this plot are very similar to those obtained while studying rotational diffusion. In general complex branched sugars and wide helices which as we have demonstrated can trap significant amounts of water on their surface due to water mediated hydrogen bonds, show both the slowest decay of solvent rotational correlation and smallest solvent mean square displacements.

A detailed comparison of epimeric differences between mannose and glucose still show that α 1 \rightarrow 3-mannose has uncommonly slow water surface diffusion on its surface due to the particular arrangement of atoms displayed in Fig.2.6. In general, just as we found in the case of the rotational correlation functions, the MSD of water on the surface of nonasaccharides with 1 \rightarrow 4 linkages appear to be largest. In particular β 1 \rightarrow 4-mannose and glucose appear to be less able to slow down the motion of water in contact with their surface.

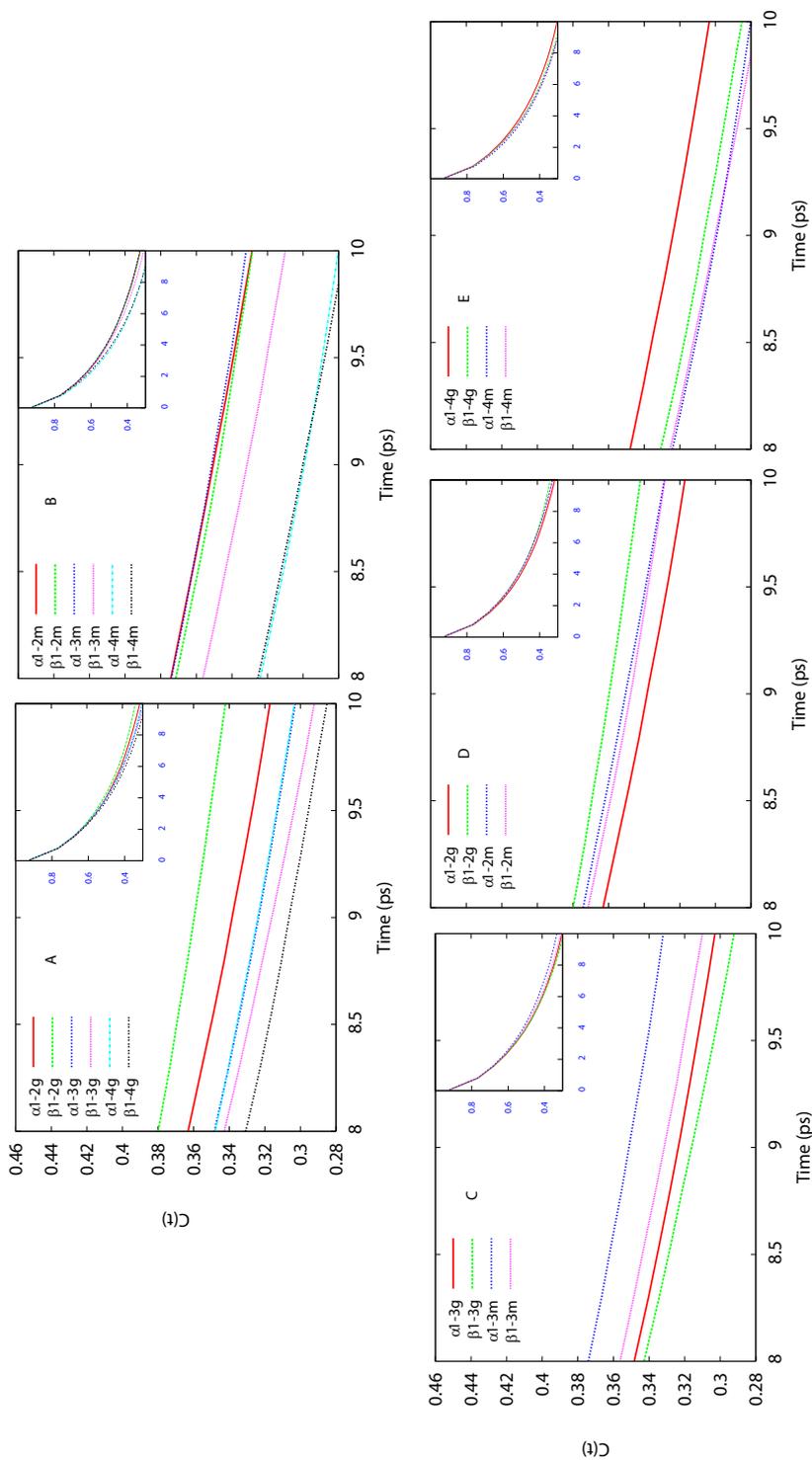


Figure 2.8: Comparison of rotational correlation functions for water on the first solvation shell of homopolysaccharides. (A) shows a comparison between different glucose monomers, (B) shows a comparison between different mannose monomers, (C) through (E) compare rotational correlation functions of water surrounding glucose and mannose monomers with different linkages.

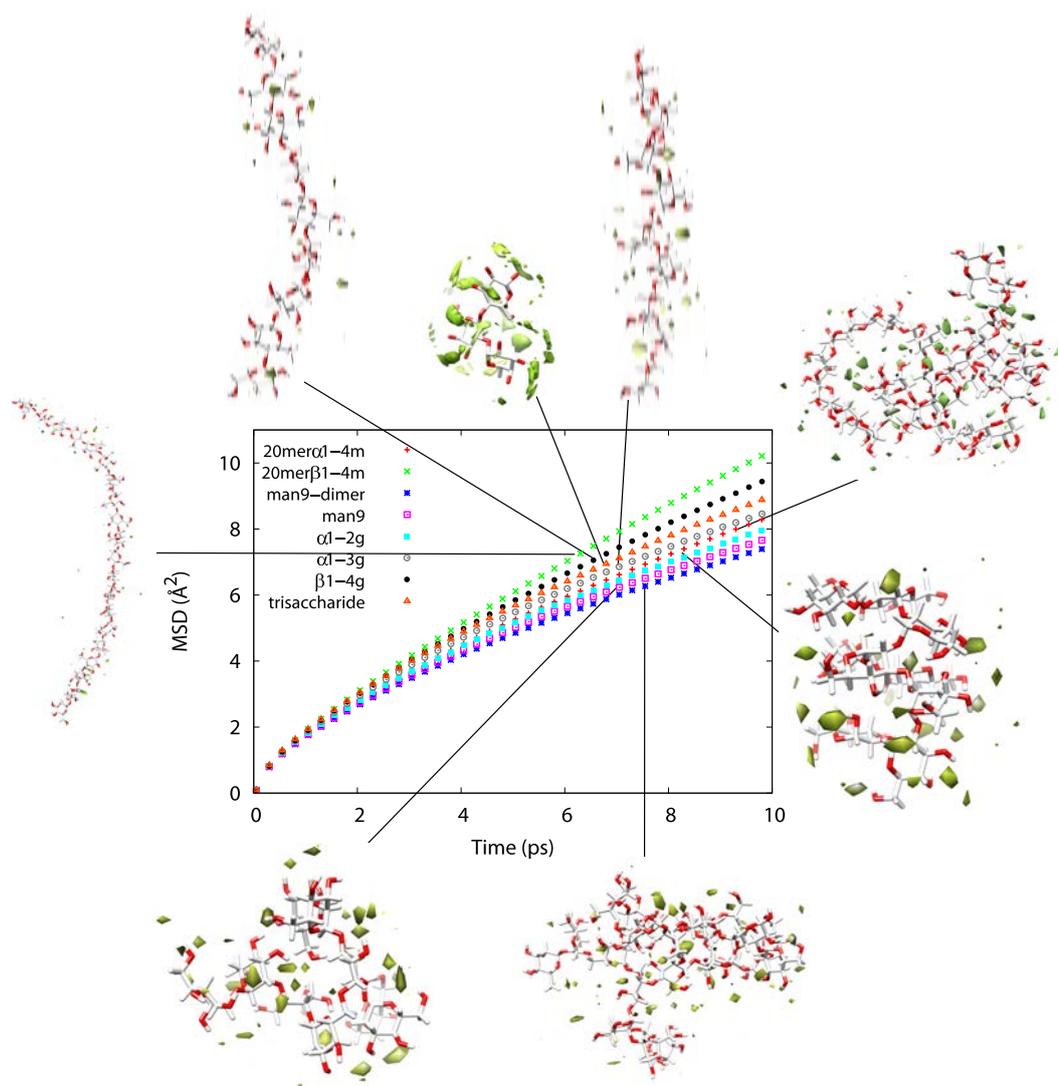


Figure 2.9: The mean square displacement of water around selected saccharide models chosen for this study.

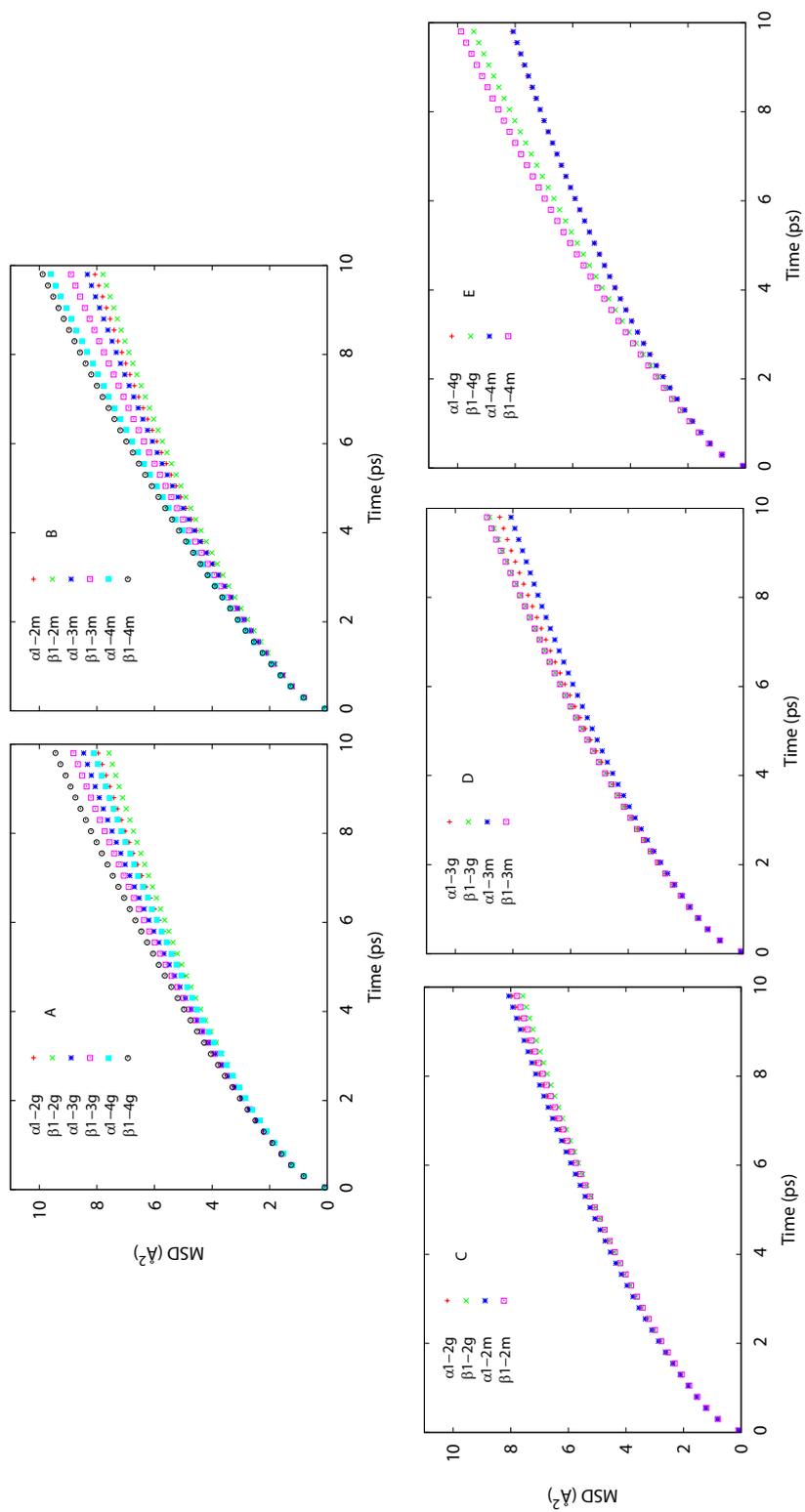


Figure 2.10: A comparison of the the mean square displacement in units of \AA^2 of water around different nonamers models characterized by their linkage position, anomeric and epimeric configuration.

2.4 Conclusions

As we have shown in previous subsections, geometric effects dramatically affect the behavior of water on the surface of sugars. Branched sugars are particularly well suited to trap water and so are wide helices. In general we find that $\alpha 1 \rightarrow 3$ and $\beta 1 \rightarrow 4$ homopolysaccharides of glucose and mannose form extended helices and it is likely that other homopolysaccharides with the same linkage and anomeric configuration as well as O4 of the second residue in equatorial position will form similar secondary patterns. Though we were not able to find in the literature the structure of homopolysaccharides of allose and altrose, these should form similar patterns to those found in mannose and glucose and based on table 2.2 we predict that they will have low or intermediate ability to trap water on their surface.

Glucose and mannose are epimers at O2; it is clear that all $\alpha 1 \rightarrow 2$ and $\beta 1 \rightarrow 2$ homopolysaccharides studied form wide helices which are able to trap water for long periods of time and slow down its rotational and translational diffusion. Therefore it appears that both $\alpha 1 \rightarrow 2$ and $\beta 1 \rightarrow 2$ configurations irrespective of whether the second residue is axial or equatorial will form such structures. We therefore predict that all aldohexoses that have a $1 \rightarrow 2$ linkage will form compact helices that are likely to strongly associate with water and slow down its dynamics.

Based on our studies, the $1 \rightarrow 3$ linkages are interesting since both extremes of very low and very high water trapping and slowing can be observed. While the $\beta 1 \rightarrow 3$ version of mannose and glucose homopolysaccharides form wide helical structures in solution with intermediate water trapping and slowing abilities the $\alpha 1 \rightarrow 3$ homopolysaccharides form well defined extended structures. Since mannose and glucose have equatorial O3 configurations,

we expect $\alpha 1 \rightarrow 3$ galactose and talose to also form well defined extended helices. Interestingly $\alpha 1 \rightarrow 3$ -g has moderate ability to trap and slow down surface water dynamics while $\alpha 1 \rightarrow 3$ -m is as highly effective as the branched sugars at modifying the behavior of water on its surface. This is because of the unique arrangement of O6, O2 and O4 in consecutive residues depicted in Fig. 2.6 which is only present in the case of mannose. It would be interesting to study the case of talose, which is an epimer of mannose in O4. Unfortunately this is an uncommon sugar for which a $\alpha 1 \rightarrow 3$ homopolysaccharide has not been studied. In the case of linear saccharides we find that beyond certain number of residues size does not significantly effect residence times as well as other properties.

CHAPTER 3

IN SILICO PREDICTION OF THE 3-D STRUCTURE OF TRIMERIC ASIALOGLYCOPROTEIN RECEPTOR BOUND TO TRI-ANTENNARY OLIGOSACCHARIDE

3.1 Introduction

The asialoglycoprotein receptor, identified by Ashwell and co-workers, functions as a transport receptor on the surface of hepatocytes. It binds to desialylated serum glycoproteins with exposed terminal galactose (Gal) and N-acetylgalactosamine (GalNAc) residues, eliciting receptor-mediated endocytosis operated via clathrine coated pits [7, 128, 44, 30]. The receptor consists of two polypeptide subunits, H1 and H2 (in human hepatic cells) or RHL1 and RHL2/3 (in rat hepatocytes), arranged as a hetero-oligomer [95, 126, 23, 24]. Several groups have proposed that the minimum ratio of the two polypeptides is 2:1, although some experiments have indicated ratios between H1 and H2 that are in the range 2-5:1-2 [92, 21]. It is thought that the presence of a hetero-oligomeric system confers high affinity to the receptor [20]. This receptor is known to be highly specific for the Gal or GalNAc residues. ASGP-Rs bind to Gal or GalNAc with high affinity when in the form of tri- and tetra-antennary oligosaccharides, whereas they bind with lower affinity when in the form of bi-antennary oligosaccharides. The K_d for tri- or tetra-antennary oligosaccharides to ASGP-R was found to be in the nM range, while that of bi-antennary was found to be in the μ M range [8, 87, 138]. In addition, when the terminal linkage was changed from β 1 \rightarrow 4 to β 1 \rightarrow 3, the binding affinity dropped significantly. This shows that ASGP-Rs are very sensitive to the geometry of their binding ligands [119, 117, 147, 118].

Many of the aforementioned binding studies were completed before the crystal structure of any carbohydrate recognition domain (CRD) of the ASGP-R had been determined.

Most C-type lectins such as the ASGP-R show significant sequence similarity in their CRDs. An important analysis of such CRDs was provided by Weiss and co-workers [80]. In this study the authors mutated the active site in the CRD of mannose binding protein A (MBP A, another C-type lectin) to that of ASGP-R and solved the structure of the protein bound to Gal. The modified protein was termed QPDWG because of mutations $\text{Glu}^{185} \rightarrow \text{Gln}^{185}$, $\text{Asn}^{187} \rightarrow \text{Asp}^{187}$, $\text{His}^{189} \rightarrow \text{Trp}^{189}$ and the insertion of a Gly rich loop. Gal and GalNAc were able to bind to QPDWG's CRD but mannose was not able to do so. The structure of the H1 CRD of ASGP-R was later determined by Meier *et al.* [96]. However, it was not co-crystallized with Gal or GalNAc. Since the overlay of the structures of the CRDs of the QPDWG mutant and H1 resulted in very low RMSD, Meier and coworkers reasonably concluded that the conformation of Gal-bound QPDWG found in Weiss' X-ray structure should be analogous to that of Gal-bound H1. Although H2 remains structurally uncharacterized, H1 and H2 share 68% sequence homology in the CRD and an overall 58% sequence similarity [129, 56].

Many studies have now provided important information on binding affinity of TA and the H1:H2 subunit ratio on the cell surface; however, none has addressed the 3-D arrangement of these subunits when bound to TA. In the study presented here, we predict the 3-D protein subunit arrangement on the cell membrane, as well as its conformational variability using computational techniques. Our novel approach takes advantage of the recently developed Fast Sugar Structure Prediction Software (FSPS) [148, 149, 150, 151] and the replica exchange method (REMD)[102, 66, 124, 131] to identify all conformations of TA that do not conflict with binding to the ASGP-R receptor. Our computational procedure also identifies the arrangements of H1 and H2 protein subunits in 3-D that are

compatible with TA binding.

3.2 Methods

In order to achieve the goal of predicting the 3-D arrangement of a multi-valent hetero-oligomeric transmembrane protein in ligand-bound conformation, we used several distinct and powerful computational techniques. Our in-silico protocol was based on using TA, a complex but computationally tractable molecule, to probe the 3-D arrangement of the much more complex protein receptor. In order to do this, several obstacles had to be overcome. First, a complete ensemble of conformational TA structures was obtained. Second, a reliable approximation to the Gal-H1/H2 binding geometry was obtained. Third, a homology model for H2 was constructed. Finally, because the predicted complexes should be stable in solution, explicit solvent simulations had to be run.

The TA ligand consists of three branches known as antennae, each of which has a terminal β -Gal residue that can bind to CRDs, such as those of H1 and H2 of the ASGP-R. The sequence of TA is shown in Figure 3.1. Assurance of full conformational sampling for a complex branched oligosaccharide is always challenging since most computational techniques are prone to fail when ergodicity problems are present. These ergodicity problems arise due to the coupled nature of dihedral rotations at branching points or crowded linkages. In order to guarantee that our sampling was reasonably complete we used two independent techniques: our in-house developed FSPS [148, 149, 150, 151] and the computationally expensive REMD method [102, 66, 124, 131].

3.2.1 Fast Sugar Structure Prediction Software Studies

In the case of TA, conformational flexibility is mainly due to glycosidic angle variability at the locations of branching. In order to better sample the conformations of TA we used FSPS to conduct a coarse-grained search in glycosidic space[148, 149, 150, 151]. Briefly, the protocol followed in a typical FSPS calculation consists of the following steps:

- Exhaustive linked coarse-grained search in ϕ - ψ space for all linkages, to discard sterically disallowed conformations.
- Energy minimizations of all allowed conformers in the gas phase or in an implicit solvent.
- Pooling of structures into “unique families” defined on the basis of their energetic and angular similarities. For example, in this work we required all members of a given family to have ϕ and ψ angular differences of $\leq 20^\circ$ and energy differences of ≤ 5 kcal/mol.
- Calculation of NMR observables (NOEs, RDCs, J-couplings) of a representative member of each family in order to rank its likelihood in comparison to experimental observables.

The initial 3-D model of TA was built using the YASARA software v 9.6.14 [82]. In this work, we were not interested in deriving the most likely structure of TA in solution, but instead were concerned with its conformational structure when bound to ASGP-R. Therefore the last two steps in the protocol are not needed. In order to balance accuracy and computational cost, we used the angular scanning increments shown in Table 3.1. A typical Ramachandran space basin of attraction has a radius of about 50° . We chose finer increments when close to branching points since these dihedral angles provide TA with most of its

conformational variability. This coarse graining resulted in 167792 allowed conformations without steric clashes. These conformations were subjected to gas-phase minimizations using the GROMACS 4.0 software [61] and the OPLS-AA force field [39]. Results derived using this method are compared against much more expensive REMD simulations in explicit solvent.

Linkage	Belongs To	Scanning Increment
α -Man-1 \rightarrow 6 β -Man	6' Arm	45°
β -GlcNAc-1 \rightarrow 2 α -Man	6 Arm	45°
β -GlcNAc-1 \rightarrow 4 α -Man	8 Arm	45°
α -Man-1 \rightarrow 3 β -Man	Core	60°
β -Gal-1 \rightarrow 4 β -GlcNAc	6', 6 and 8 Arms	180°
β -GlcNAc1 \rightarrow 4 β -GlcNAc	Core	120°
β -Man1 \rightarrow 4 β -GlcNAc	Core	120°
β -GlcNAc-1 \rightarrow 2 α -Man	6' Arm	120°

Table 3.1: Dihedral angle increments for linkages in TA.

3.2.2 Replica Exchange Molecular Dynamics Simulations

The REMD technique is based on parallel simulations of identical system replicas (M_1 to M_n) at different temperatures (T_1 to T_n)[102, 66, 124, 131]. Snapshots of the system throughout the simulation are exchanged based on a Monte Carlo acceptance scheme [98]. At higher temperature, the system is able to cross barriers that are inaccessible at lower temperatures. The protocol for exchange guarantees correct Boltzmann sampling at each of the studied temperatures for which free energies can be derived, if all ergodicity problems are surmounted. REMD simulations have been applied to the problem of protein folding

(see for example reference [59]) as well as to study many other systems such as DNA-Carbon nanotubes structures [70] and DNA-RNA dynamics [38].

As an initial equilibration step, we minimized TA in a cubic water box consisting of 4055 SPC (simple point charge) water molecules [71], using the steepest descent algorithm. Subsequently, we carried out an NPT (constant number of particles, pressure and temperature) molecular dynamics equilibration of the system for 1 ns at 300K. All simulations were carried out using the software GROMACS 4.0 [61] and the OPLS-AA force field [39]. In order to properly describe electrostatic interactions in a periodically replicated system we used the particle mesh Ewald (PME)[40] method, with a real space cutoff of 9 Å. The final structure from this equilibration run was used as the initial conditions for REMD simulations. Our REMD simulations consisted of forty eight molecular dynamics replicas ranging from 300K to 396K, in 3K increments. The choice of number of replicas and temperatures was generated using the protocol of Van der Spoel [107] (see also <http://folding.bmc.uu.se/remd/index.php>). Our explicit solvent REMD simulations were 35 ns in duration. This time scale was deemed appropriate based on the observed convergence of Ramachandran plots for the crowded linkage points at the branching of 6' Arm, 6 Arm and 8 Arm, which were almost unchanged after 15 ns. The integration time step for the MD and REMD simulations was set at 1 fs and 2 fs, respectively. Exchange of replicas was attempted every 1ps during the 35 ns simulation. We used the Nose-Hoover coupling scheme [100, 63] for temperature control with coupling τ value of 0.05 ps and the Parinello-Rahman coupling scheme [105] for pressure control with coupling parameter of 1.0 ps. In all simulations, the cut-off value for van der Waals interactions was set to 0.9 nm. This simulation resulted in 17500 models of TA saved from the lowest temperature replica

at a frequency of 2.5 ps.

3.2.3 Structural Alignment Studies

Our structural alignment studies consisted of three steps, described in subsections 3.2.3.1, 3.2.3.2 and 3.2.3.3. All alignments were performed using the PyMOL software DeLano Scientific, California [42].

3.2.3.1 Identifying the Binding Site for H1

The first predictive step involved deriving the orientation of Gal in the H1 binding site. Since the structure of the H1 subunit with its binding saccharide had not been solved [96], we used the x-ray structure of Gal bound to the QPDWG mutant [80] as a template to orient Gal with respect to H1. This approach is expected to result in only negligible error given the almost identical nature of the two binding sites (*vide infra*). To limit the complexity of the predictive problem, we have not attempted to predict variations on the well established binding mode for the interaction between the terminal Gal of TA and H1 or H2.

3.2.3.2 Alignment of Trimeric H1 of ASGP-R with TA

The second predictive step involved deriving all feasible 3-D arrangements of the ASGP-R protein subunits in the complex. The structure of H2 has not yet been determined. Since the H1 and H2 subunits are highly similar from a sequence and functional perspective, the structure of H2 is expected to be similar to that of H1. Therefore, we first used three H1 subunits instead of two H1 and one H2 subunit. We later replaced the H1 subunit connected at the 8 Arm by our homology model of H2 described in subsection 3.2.3.3. In order to derive all feasible 3-D arrangements of the three H1 subunits, we attached an ASGP-R

subunit to each of the three Gal terminated TA antennae using the orientation derived in subsection 3.2.3.1. In other words, each TA conformer derived from our REMD simulation was associated with a unique TA/ASGP-R 3-D arrangement that, in principle, can have steric clashes. Structures with steric clashes were subsequently discarded.

The following protocol was followed for all structural alignments:

- The terminal Gal residue of the 6' Arm of TA was aligned to the Gal binding site of H1 derived from our comparison, with QPDWG using the pairfit command in the PyMOL software. The coordinates of one H1 subunit and TA were saved to a pdb file for further alignment of the other two subunits.
- The terminal Gal binding site of the 6 Arm of TA in the monomeric complex was aligned with the Gal binding site of a second H1 and the dimeric complex was saved to a new pdb file.
- The terminal Gal residue of the 8 Arm of TA in the dimeric complex was aligned with a third H1 subunit and the final structure was saved in pdb format.

3.2.3.3 Homology Modeling of CRD of H2 subunit

The final predictive step involved deriving a homology model of the structurally uncharacterized H2 subunit. An attempt to derive the structure of H2 and H1 from homology to the QPDWG mutant was published[19] before the crystal structure of H1 was resolved. In contrast to this early study[19], our homology model used the known structure of H1 to derive that of H2. The residues coordinating Ca^{2+} in the binding pocket of H1 and H2 subunits are highly conserved [129]. Furthermore, amino acid sequence alignments of the ASGP-Rs from rat, mouse and humans show that the residues that ligate with Ca^{2+}

in each of these structures are identical. Thus it is highly probable that the H1 and H2 subunits interact with Gal in a very similar manner. These findings provide an excellent starting point for the theoretical derivation of a reasonable structure for the H2 subunit. Homology modeling of the H2 subunit was carried out using the MODELLER software, version 9v6 [49, 50].

The protocol consists of three steps:

- The MODELLER software has an internal database of related protein sequences with greater than 95% similarity. The sequence of H2 was used to query this database. Fourteen templates with sequence identity greater than 30% were selected for step two in the protocol.
- The second step is the use of a more accurate 2-D alignment program within MODELLER that takes into account structural information. All 14 template structures were considered. A sequence identity table and a clustering tree (dendrogram) expressing differences among closely related templates were generated. For the next step, three aspects were important: the resolution of the crystal structure, sequence similarity data (from the previous step) and dendrogram data from the present step. Based on these three criteria, the best choice for the next step (3-D model building) was the H1 subunit (resolution was 2.0 Å, sequence similarity was 68% and close relation to H2 through dendrogram data).
- The structure of H1 (PDB ID 1DV8) and the sequence of H2 were used to build a 3-D model for H2. Five candidates were proposed by the software. To evaluate these models we used a routine approach based on the objective score function and the Discrete Optimized Protein Energy (DOPE) score per residue. The DOPE score is

a statistical potential based on pair distribution functions of all atoms in the native structure [153]. A DOPE score value of 0.00 and above corresponds to high energy and negative values correspond to lower energy. This evaluative method is described in the MODELLER manual. The DOPE score for the template and best target are overlaid in Figure 3.2. The similarity in scores between the template and the target indicates high spatial homology except for moderate deviations at the H2 protein termini.

3.2.4 Molecular Dynamics Studies of the Predicted Complex

All 17500 trimeric H1-TA complex candidates were checked for inter-subunit steric clashes. Out of these 17500 candidate structures, only 22 had no inter-subunit clashes. If the core of the sugar is oriented towards the positive Z axis and one looks down along this axis, about half of the 22 structures have H1 subunits connected to the 6', 6 and 8 Arms arranged in a clockwise fashion. The remaining structures are arranged in a counter-clockwise fashion (the definition of clockwise and counter-clockwise is based on the arrangement of the arms see subsection 3.3.2). In order to further refine these two families of structures, the H1 subunit connected at the 8 Arm was replaced by our homology model of H2. The reason for the choice of the 8 Arm is based on experimental evidence from Rice *et al.* [118] showing that the 8 Arm Gal binds specifically to the RHL2/3 subunit (analogous to H2). Four out of the 22 complex candidates, two belonging to each orientation of three subunits (clockwise and counter-clockwise), were subjected to simulated annealing molecular dynamics (SAMD). These simulations were carried out in explicit solvent at constant volume, with a temperature range from 500K - 300K sampled every 10K for a total of 10 ns. All simulations were performed using the software GROMACS 4.0 [61] and the OPLS-AA force

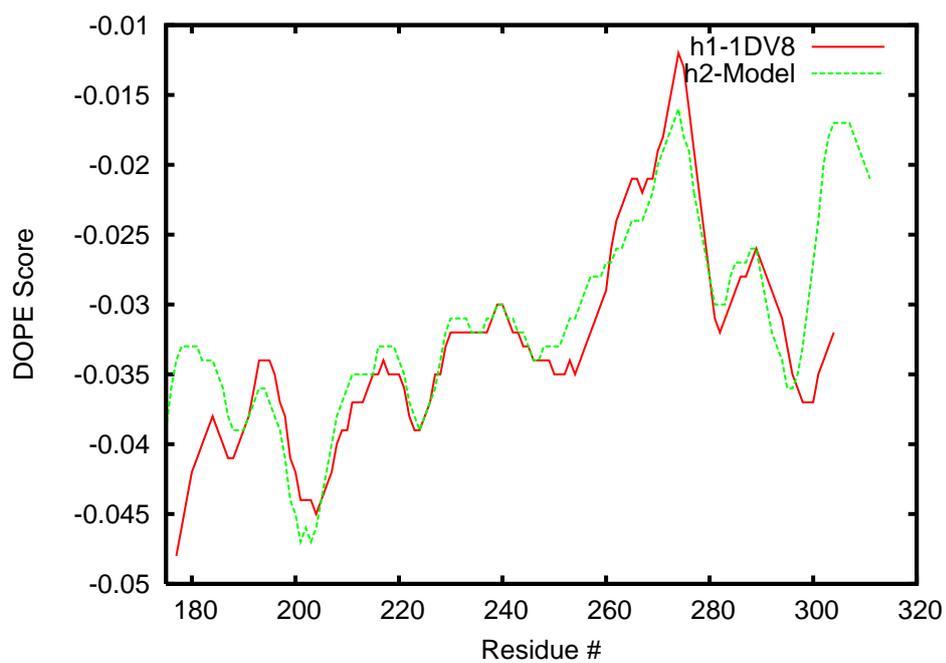


Figure 3.2: Discrete Optimized Potential Energy (DOPE) score per residue for the template (H1 subunit of ASGP-R, PDB ID 1DV8) and target (H2 subunit of ASGP-R). The target (H2 subunit,) is shown in green and the template, (H1 subunit,) in red.

3.3 Results and Discussion

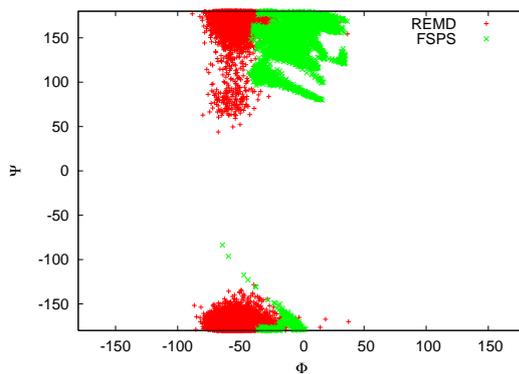
3.3.1 FSPS and REMD Studies of Free TA

A comparison of the fully coupled Ramachandran plots for the 1→6 linkage of the 6' Arm, the 1→2 linkage of the 6 Arm and the 1→4 linkage of the 8 Arm based on FSPS and REMD simulations are shown in Figure 3.3. These linkages are the most important to determine the overall structure of TA. We see from Figure 3.3 that both methods describe almost the same dihedral space, except for fairly small deviations arising from the fact that REMD is carried out in explicit solvent whereas FSPS is a gas phase procedure. In fact, FSPS samples a small region ($\phi = 30^\circ$ and $\psi = 170^\circ$), shown in Figure 3.3b, that was not sampled by REMD, possibly because this region may be of high energy. The overlap detected gives us confidence that we are indeed sampling all possible conformations of TA in solution. An important difference between the two approaches is that with FSPS it only takes three days and 10 processors to obtain these structures, whereas with REMD three months and 48 processors. Examples of the different conformations of TA derived from the REMD simulation are displayed in Figure 3.4. In this plot all conformers are displayed such that the 6 and 8 Arms are pointing in the same direction. Our plots of the explicit solvent REMD distance distribution at 300K and the corresponding free energies derived from them are shown in Figure 3.5. Distances are defined from the C1 atom in each terminal Gal to the C1 atom on the terminal GlcNAc in the core region. From the width of its probability distribution, it is clear that the 6' Arm displays the largest range of possible distances. Flexibility is a well known characteristic of 1→6 linkages [77] and is due to the presence of ω , the third dihedral angle. The probability distribution for the 6' Arm displays a major peak at 16 Å, and other less prominent peaks at 9 Å and

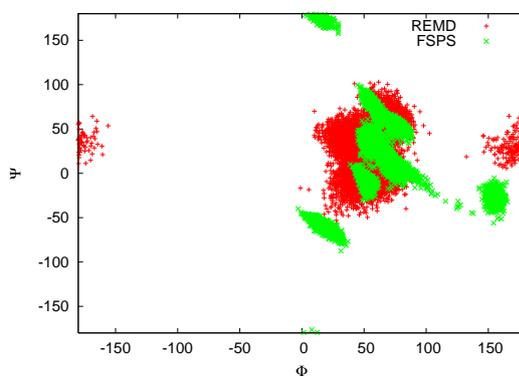
12 Å at about 0.6 and 1.7 kcal/mol above the free energy minimum, respectively. The probability distribution for the 6 Arm displays two peaks centered around 14 Å and 21 Å, with corresponding free energy differences of 1.8 kcal/mol. The 8 Arm shows a single peak centered around 22 Å. Based on these free energy profiles, transitions from the higher free energy conformational states to those at lower free energies appear to be almost barrierless. These results can be compared with those obtained in fluorescence resonance energy transfer (FRET) experiments performed by Rice *et. al.*, [119] in which a two-peaked distribution was observed for the 6' Arm, corresponding to our peaks at 9Å and 16Å, respectively. These represent folded and extended populations, consistent with previous gas phase molecular dynamics studies of TA by Balaji and coworkers [9]. The intermediate feature at 12Å in our probability distribution also appears to be important. It turns out that the four independent simulated-annealing molecular-dynamics simulations of the full ASGP-R-TA complex in explicit solvent gave rise to sugar structures that match extended, folded and intermediate 6' Arm configurations (*vide infra*). In agreement with our simulations, the FRET study [119] also identified a folded and extended set of populations for the 6 Arm, and only a single population in the case of the 8 Arm. These results and biochemical experiments in references [5, 116] appear to indicate that the inherent flexibility of TA is important for the high affinity binding of TA to ASGP-R.

3.3.2 Structural Alignment Studies

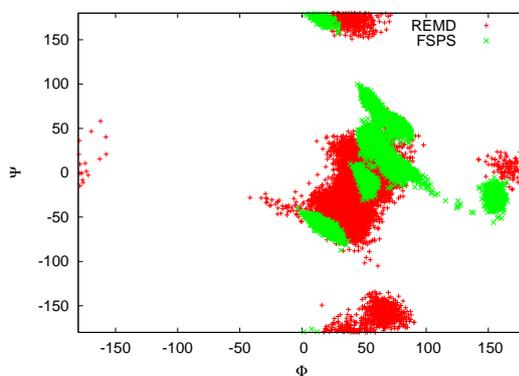
Probing the structure of the TA/ASGP-R complex experimentally has proven to be a daunting problem, so far the system has been refractory to high-resolution structural analysis. Thus, a number of important questions remain. What is the symmetry of the complex? What is the global 3-D arrangement of the protein subunits? Is there more



(a) 1→6 Linkage (6' Arm)



(b) 1→2 Linkage (6 Arm)



(c) 1→4 Linkage (8 Arm)

Figure 3.3: Dihedral angle map for TA. In red are the results from our REMD simulations and in green those from the FSPS for the 6' (a), 6 (b) and 8 (c) Arms.

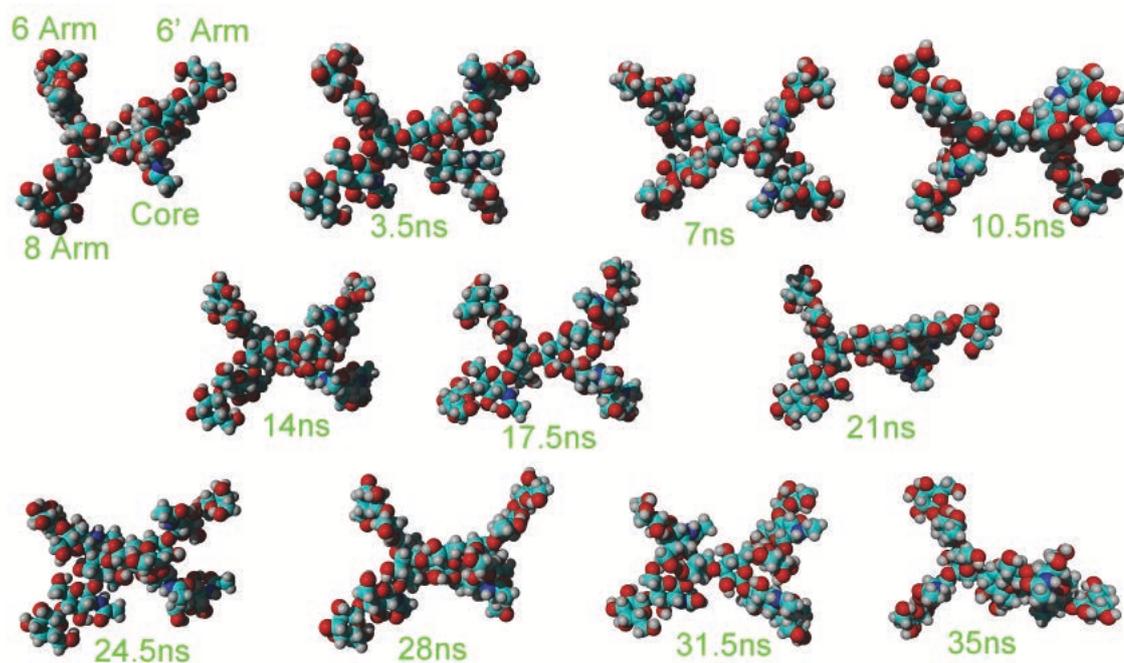


Figure 3.4: Representative structures of TA derived from REMD simulations. The orientation of the 6', 6 and 8 Arms is indicated in the first snapshot and is the same for all other conformations.

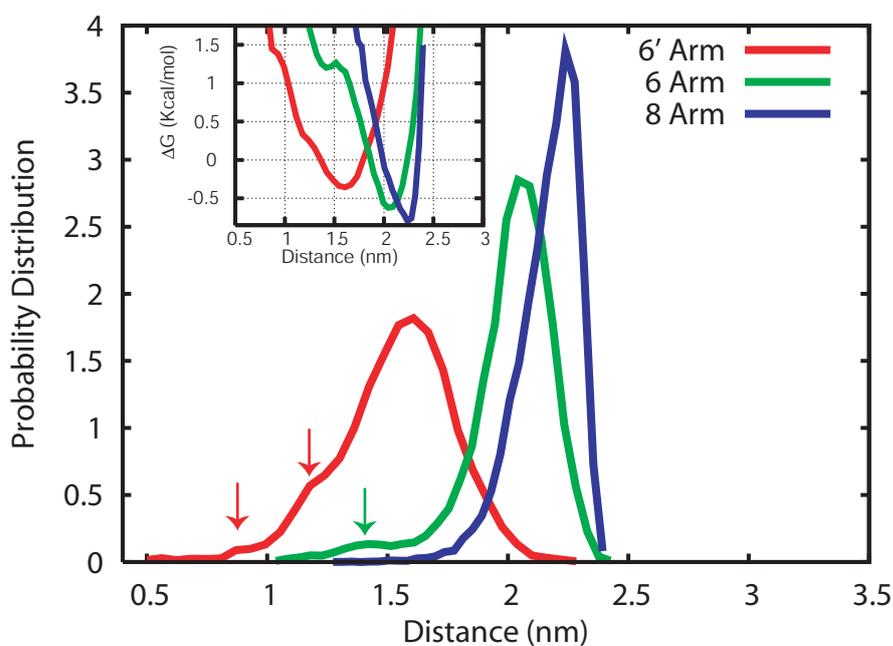
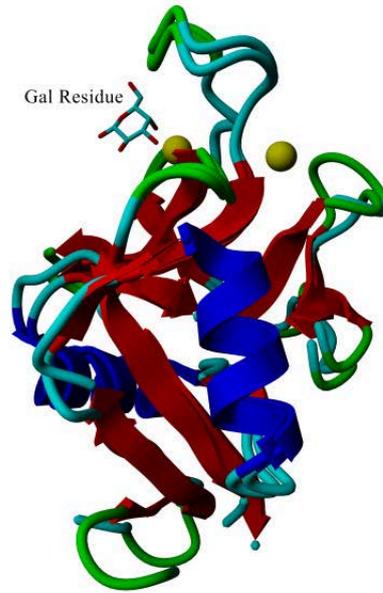


Figure 3.5: Normalized probability distribution for distances between C1 of the terminal Gal in the 6', 6 and 8 Arms with respect to C1 of the terminal GlcNAc in the core region. Arrows indicate peaks referred to in the text. Inset shows the corresponding free energy functions derived from these probability distributions.

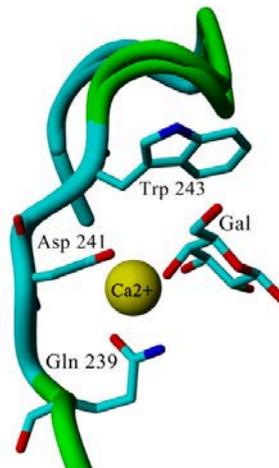
than one viable arrangement? Our goal was to predict the broad features of the complex, or set of possible complexes, computationally. Focusing on specific details of inter-protein subunit amino acid interactions would be unwise in this case, given that we derived the structure of the H2 subunit from homology. By taking advantage of experimentally derived structural data on the mode whereby Gal binds to the CRD of QPDWG, and having derived a complete pool of TA conformers, our alignment studies were able to predict the subset of these conformers that can bind three subunits of ASGP-R through terminal Gal without steric clashes.

Our pairwise structural alignment of the H1 subunit structure of ASGP-R to the QPDWG mutant of MBP with bound Gal is shown in Figure 3.6. This alignment resulted in a root mean square deviation (RMSD) of 1.27 Å (all atom, the C_{α} is around 0.94 Å). Given the almost perfect overlap of the CRDs in both proteins, we were able to construct 17500 models of TA bound to three H1 subunits by aligning each terminal Gal on TA to the known crystal structure of Gal bound to the QPDWG mutant. Only 22 out of the 17500 were viable structures for the complex, i.e. ones in which no inter protein subunit steric clashes were detected. These 22 viable complexes could be separated into two families distinguished by the “relative chirality” of protein subunit arrangement (see Figure 3.7). The chirality arises from the inherent flexibility of the three Arms of TA. At this point, we replaced the H1 subunit connected to 8 Arm with our model of the H2 subunit. Since the CRDs of H2 and H1 are highly similar to that of QPDWG, variations on the well established mode of binding between Gal and QPDWG are not expected. Because of the high level of similarity between the two protein structures, no clashes resulted from this substitution.

In the absence of further knowledge of the constraints imposed by the three membrane-



(a) Structural Alignment of H1 subunit of ASGP-R with QPDWG mutant of MBP A



(b) The Active Site of Binding

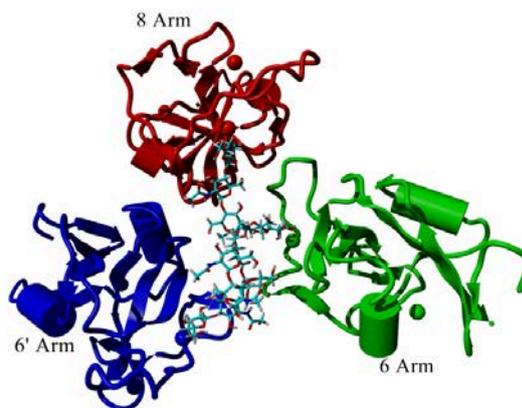
Figure 3.6: Structural Alignment of the H1 subunit of ASGP-R with the QPDWG mutant of MBP A. (a) All-atom alignment of the H1 subunit with the QPDWG mutant. The RMSD between the two proteins is 1.27 Å(all atom). Bound Gal is shown in stick format. (b) Zoom-in view of the active site. The labeled residues are those of H1.

bound protein stalk regions and the locations of glycosyl moieties on the H1 and H2 subunit surfaces, both of the families of complexes described above represent reasonable predictions of the trimeric complex. Not only do two H1 units and one H2 unit bind to the 6, 6' and 8 Arms, respectively, but the structures also accommodate a large glycoprotein attached to the core region of the sugar. This is a requirement for any predictive model, since in the biological system TA is attached to a variety of different proteins through its core.

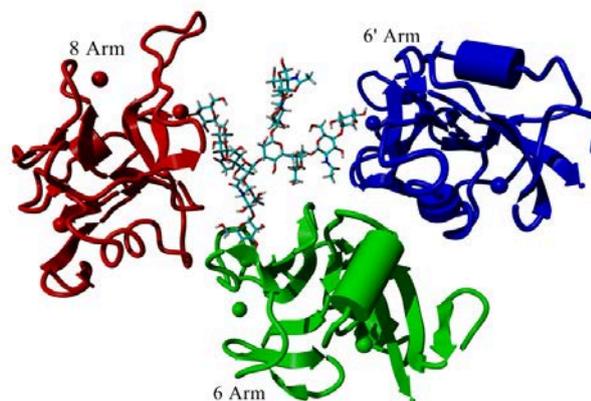
When bound to TA all candidate complexes are clearly asymmetric. The distances between terminal Gal residues are $20.9 \pm 1.3 \text{ \AA}$, $12.5 \pm 1.4 \text{ \AA}$ and $23.3 \pm 3.1 \text{ \AA}$ for the 6'-6, 6-8 and 6'-8 Arms, respectively. Clearly the largest variability is between 6' and 8 Arms, where some candidate complexes are separated by as much as 16 \AA - 25 \AA .

The asymmetry in inter Arm distances seen in each of the 22 candidate complexes is consistent with the prediction of asymmetry made by Lodish [92], based on simple geometric arguments using Rice's distance distributions of TA derived from FRET experiments. Interestingly, a symmetric structure of trimeric MBP A in the absence of oligosaccharide has been crystallized [125, 145]. If the results of our study apply more generally to other membrane-bound complexes such as trimeric MBP A, unless the ligand is highly symmetrical, the complex must be able to adapt in order to accept the ligand. A recent article by Menon and coworkers[97] on DCSIGN supports the latter prediction. Using force pulling experiments, these investigators showed that as the ligand approaches, it reaches a certain distance at which a significant conformational change occurs in the protein complex. This type of conformational rearrangement has also been observed in immunoglobulins [22].

To provide further details on the structure and dynamics of the full complex in solution, we carried out two different types of MD simulation studies. The first involved



(a) Clockwise Alignment



(b) Counter-clockwise Alignment

Figure 3.7: Clockwise and counter-clockwise receptor-subunit arrangements. Attached to the 6 and 6' Arms are H1 subunits, and attached to the 8 Arm is an H2 subunit. H1 and H2 subunits are color coded and labeled to indicate the arm to which they are attached. Two types of arrangements, clockwise and counter-clockwise, can be derived from our studies. Since the protein subunits are large, it is unlikely that there is exchange between these conformations once the sugar is bound to the protein trimer.

two constant-temperature and -pressure simulations of at least 10ns in duration, comparing changes in the H1, H1, H2 complex to those in the H1,H1,H1 homo-trimer analog. The second was a set of simulated-annealing molecular-dynamics simulations (SAMD) with a temperature ramp covering the range of 500K to 300K. The purpose of the second study was to measure the structural relaxation of the complex after a computational temperature jump. Two clockwise and two counter-clockwise arrangements of the complex were used as initial conditions for our four SAMD simulations.

3.3.3 Simulated Annealing Studies of the H1-H1-H2-TA complex

Simulated annealing (SA) is a technique in which the system is heated to high temperatures in order to disrupt structure and cross energy barriers and subsequently cooled to lower temperatures in order to allow the system to relax energetically. SA has been applied, for example, to predict the conformation of protein side chains [37, 85]. As described in subsection 3.2.4, distances between the ligand and binding receptors were constrained during the temperature ramp, but were released during the last 2.5 ns, during which the temperature of the system was kept constant at 300K.

The radius of gyration (R_g) as a function of time, for each of the models, is shown in Figure 3.8. From this figure it is clear that the three-subunit complex can potentially adopt a variety of conformations while still being compatible with the bound oligosaccharide. This flexibility is evident from the fact that some complexes are compact but others are not. Some of these conformations are much more likely to be physiologically relevant than others, but probing the relative probability of each conformer would require exhaustive simulation studies, both in the presence of the membrane and with the inclusion of the receptor stalk, whose conformation is still unknown. Furthermore, detailed attention would

have to be paid to glycosylation at the surface of the receptor proteins, and results may be specific to the particular protein attached to TA's core. Thus resolving these issues is beyond the ability of current computational technology. However, several fundamentally important conclusions can be drawn from our studies. Firstly, two different binding chiralities are likely to exist. Secondly, in all cases the complex is asymmetrical. Thirdly, the ASGP-R system appears to allow for significant fluidity which is likely required for binding to a highly flexible and multi-conformational sugar. This last point has important implications for our understanding of existing crystal structures of other sugar binding complexes. Specifically, although these complexes may appear highly symmetrical, fluidity and the potential for conformational variability may be key requirements for binding to asymmetric and highly flexible N-Glycans.

3.3.4 Comparison of Structure and Dynamics of Homo- and Hetero-trimeric Complexes

In order to gain a molecular-level understanding of differences in energetics between homo- and hetero-trimers, we subjected both systems to molecular dynamics simulations in explicit solvent at ambient pressure and temperature. In particular, we studied in detail the interactions between the subunit connected to 8 Arm and TA. Figure 3.9 shows the total interaction energy between TA and the protein subunit connected to the 8 Arm as a function of time. There are significantly larger fluctuations when H2 is connected to the 8 Arm. Interestingly, this apparent on-off behavior observed around 1 ns and again at 4ns in the case of H2 can be solely attributed to angular fluctuations of the side chain of ASP²⁹⁰ of H2, due to the interaction of the carboxylate with 8 Arm Gal. This is corroborated by the interaction energy between this residue and TA, which is shown as a function of time in

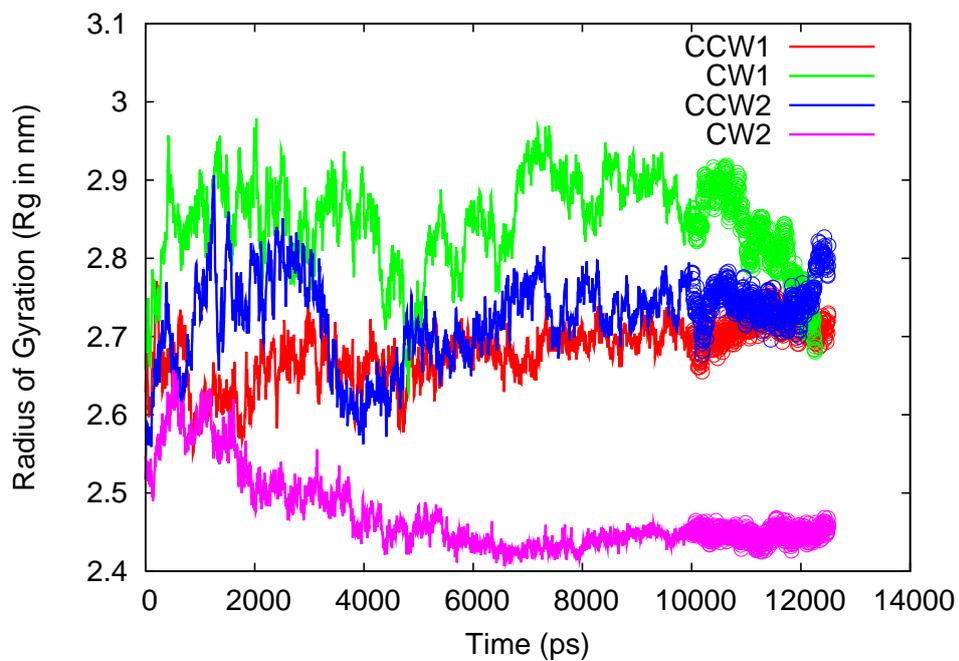


Figure 3.8: The radii of gyration as a function of time for the four models of TA bound to H1-H1-H2 from our SAMD simulations. The first 10 ns (solid lines) correspond to SAMD whereas the subsequent 2.5 ns (circles) are unconstrained NPT simulations at 300K. CCW refers to the counter-clockwise model and CW to clockwise model that are described in Figure 3.7.

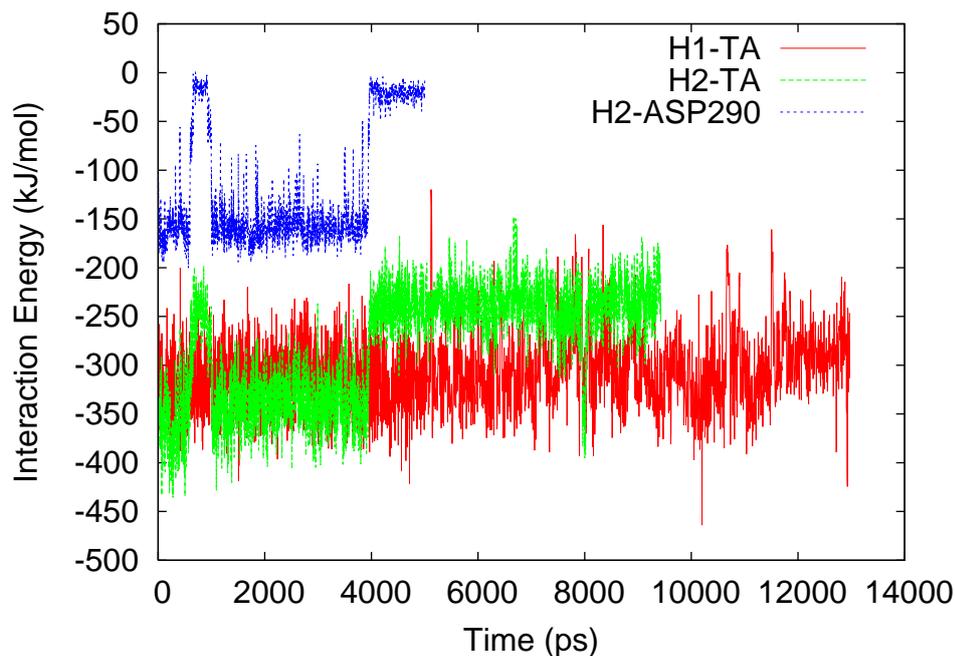


Figure 3.9: Comparison between TA-H1 subunit and TA-H2 subunit interactions at the 8 Arm. Red line represents the interaction energy between TA and H1 at the 8 Arm in the homo-oligomeric complex as a function of time during simulation. Green line represents the interaction energy between TA and H2 at the 8 Arm in the hetero-oligomeric complex. Blue line represents the interaction between ASP²⁹⁰ of H2 and TA.

Figure 3.9. Though the interaction energy between H2 and TA appears to fluctuate more, and its average appears to be lower, this does not necessarily mean that the free energy of binding to H2 is lower. For that, entropy and interactions with the solvent must also be considered.

3.4 Conclusions

In this study, an exhaustive set of conformations of TA in solution have been derived. These models display significant flexibility and are consistent with previous FRET experiments. Based on the crystal structure of QPDWG bound to Gal, a model of TA bound to H1 was derived. Although the H2 subunit in ASGP-R has never been crystallized, we have built a model for it based on homology, using H1 as a template. From the exhaustive set of conformations derived for TA in solution, all models that can accommodate two H1 subunits and one H2 subunit (attached at the 8 Arm), as well as a large protein attached to the core of the oligosaccharide, were constructed. All predicted complexes are asymmetric, and significant variation in the way TA can be presented to its binding partners appears to be possible based on our SA studies. If one assumes, as is the case for the recently crystallized MBP A, that in the absence of TA the tri-protein system in the membrane is symmetrical, then these proteins must undergo significant spatial rearrangements in order to bind TA, most likely in a stepwise process. This type of conformational rearrangement has been shown to occur in DCSIGN [97]. Our work is a first step to understanding a very complicated system. We hope that these findings will guide additional experiments. Finally, the approach described in this chapter is quite general and can be used to study a large set of very important but poorly understood multivalent transmembrane complexes. We are in the process of applying this procedure to study other receptors.

CHAPTER 4 POTENTIAL OF MEAN FORCE CALCULATIONS OF DIMANNOSE BINDING TO CYANOVIRIN-N

4.1 Introduction

Cyanovirin-N (CVN) is a carbohydrate binding protein belonging to the blue-green algae called cyanobacteria (*Nostoc ellipsosporum*). It has antiviral activity towards human (HIV) and simian immunodeficiency virus (SIV). These carbohydrates (particularly high mannose oligosaccharides) are present on the surface of glycoprotein-120 (gp120) of the HIV and SIV. [26] CVN binds to dimannose or oligomannose via two binding domains named as domain A and B. Domain B of wild type CVN binds to dimannose with high affinity ($K_d \ 10^{-7}$) and domain A binds with relatively low affinity ($K_d \ 10^{-6}$). [18] CVN has pseudo-symmetry with respect to these two domains (see Figure 4.1) and in solution two monomers of CVN dimerize to form a domain swapped dimer with two domains of A (residues 1-38 from one monomer and 90'-101' from other monomer) and two domains of B (39-50/51'-89' and 39'-50'/51-89). [29] Several studies exist that address binding of wild-type CVN to dimannose [18], trimannose [17] and oligomannose.[28] In order to avoid the domain-swapped dimer formation and to achieve the stability of the monomer and also to avoid the binding at the low-affinity site, the Ghirlanda group have engineered P51G-m4-CVN [55, 54], which binds to dimannose only at the high-affinity site.

Except for the studies by Vorontsov *et al*, [143, 144] all the previous experimental [10, 13, 15, 74, 90, 94, 152, 26, 27, 55, 54] and computational studies, [93, 58, 57] have been performed on the wild-type CVN. In the binding affinity experiments of wild-type CVN with dimannose, Bewley *et al* have delineated the amino acids that contribute the most

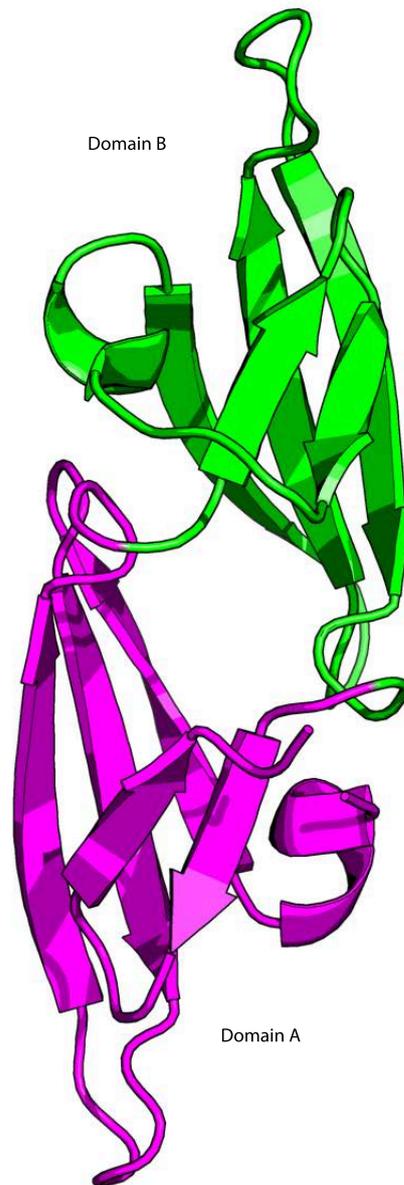


Figure 4.1: The pseudo-symmetry of Cyanovirin-N. Domain A is colored magenta and domain B is colored green. One can align domain A and domain B and both units

to the binding at the high-affinity site (domain B) and at the low-affinity site (domain A). The residues at the high-affinity site were identified as Glu41, Ser52, Asn53, Glu56, Thr57, Lys74, Thr75, Arg76 and Gln78. These residues interact either via hydrogen-bonds and/or electrostatic interactions.[14] Important amino acids in the low-affinity site were identified as Lys3, Gln6, Thr7, Glu23 and Thr25.[14]

Although the experimental dissociation constant for dimannose binding to the high-affinity site of P51G-m4-CVN has not yet been published, Vorontsov *et al* have investigated the binding of dimannose to this mutant using a MM/PB(GB)SA approach. Sandstorm *et al* have performed saturation transfer difference (STD) NMR experiments of dimannosides and trimannosides reminiscent of the larger Man₉ against another mutant of CVN in which the binding site at domain A was mutated to resemble that of the binding site at domain B. In their experiments, they found that the H2, H3 and H4 protons of the non-reducing end of dimannose receive the saturation transfer from the protein residues. They identified amino acids, Asn42, Ser52, Asn53 and Arg76 in the high affinity site as those interacting with the saccharides. [120, 11] According the studies by Vorontsov *et al* [143, 144], the key amino acids at the binding site of P51G-m4-CVN were identified to be Asn42, Asp44, Ser52, Asn53, Thr57, Lys74 and Thr75 based on the hydrogen bonding of these amino acids with the hydroxyl groups of dimannose. [143] Particularly, the 3'OH and 4'OH were cited as key to binding based on their subsequent studies on the binding of di-deoxy dimannose analogues to P51G-m4-CVN. [144]

Since so far, no dissociation studies are available for this system and to map the most important residues along the dissociation pathway we have carried umbrella sampling simulations to derive the potential of mean force as a function of different possible reaction

coordinates in the case of dimannose binding to P51G-m4-CVN.

4.2 Methods

4.2.1 Model of Cyanovirin-N and Force Field Parameters

As described in the introduction, we chose the P51G-m4-CVN model developed in the Ghirlanda group (PDB accession ID: 2RDK). P51G-m4-CVN has two sets of mutation introduced. One is the P51G mutation rendering the protein exclusively in the monomeric form. The second is a quadruple mutation at the low-affinity site (domain A). The four mutations at the low affinity site are K3N, T7A, E23I and N93A. In this model only the high-affinity binding site (or domain B) is active and binds to dimannose whereas the low-affinity site is non-functional and does not bind to dimannose[54].

For technical reasons the AMBER11 package[33] and the Gromacs package[61] provide facilities to use different reaction coordinates in pulling simulations. Therefore in order to use the same force fields in both packages a somewhat involved protocol had to be implemented.

In order to test for the usage of the AMBER ff99SB [64] and GLYCAM06 [76] force fields in the Gromacs package (version 4.5.x) [61], we converted the AMBER ff99SB/GLYCAM06 parameters to Gromacs force field parameters using `amb2gmx.pl` from Erin Sorin's website(see Appendix A for details). In AMBER11 the mixed scaling for 1-4 interactions is done separately in the case of proteins and carbohydrates. In Gromacs this is done via a so called fudge factor parameter. However in Gromacs the use of two different fudge factors is not implemented. In order to use in Gromacs a different fudge factor for proteins and for carbohydrates as required by the GLYCAM06 force field (0.5 for the Lennard-Jones interactions and 0.833 for the Coulomb interactions in the case of proteins and 1.0 for both

Lennard-Jones and Coulomb interactions in the case of carbohydrates), we utilized an approach by Nilmadhab *et al*[99]. Briefly, in this approach, the fudge factor for Lennard-Jones part is set to 1.0 and the fudge factor for Coulombic part is set to 0.16666666. The Lennard Jones ϵ is divided by 10 for protein atom types and the [pairs] section for proteins is replicated five times. The ϵ for carbohydrate atomtypes is divided by 6 and the [pairs] section replicated six times. This technical procedure achieves the desired mixed scaling factors. In order to show that this approach works and is compatible in Gromacs, we computed the energy of the protein, the carbohydrate and the complex with Gromacs and with AMBER11. The energy difference between the two codes for each term, i.e., bond, angle, dihedral, 1-4 coulomb, 1-4 van der Waals were less than 0.001%.

For the simulations using the AMBER11 package [33], the P51G-m4-CVN dimannose complex was solvated in an SPC water box of dimensions, 68 Å X 64 Å X 71 Å.

An initial energy minimization was performed for 5000 steps. The first 2500 steps were subjected to steepest descent method and the subsequent 2500 steps to the conjugate gradient method. A short, 20ps canonical ensemble (NVT) pre-equilibration run was performed followed by a 4ns equilibration run in the constant number of particles (N), pressure (P) and temperature (T) ensemble. Pressure was kept at 1 atm with isotropic position scaling, the reference temperature was maintained at 300 K by using a Langevin piston with collision frequency of 1 ps^{-1} . Long range electrostatics were treated with the Particle Mesh Ewald Summation. The cut-off distance for short range electrostatics was set at 10 Å. Data was saved every 10 ps. The last snapshot of this trajectory was used as input for steered molecular dynamics simulations (*vide infra*).

For the simulations using the Gromacs package the dimannose-CVN was solvated

in an SPC water box of dimensions 90 Å X 80 Å X 80 Å. The solvated complex was minimized for 5000 steps using the steepest descent method. After this we performed an initial pre-equilibration run for 100 ps run in the canonical (constant number of particles, N, volume, V and temperature, T) ensemble. This was followed by equilibration in the constant number of particles, N, pressure, P and temperature T ensemble for 200 ps. The final production run was in the NPT ensemble for 10 ns. A pressure of 1 atm was maintained using the Parinello-Rahman barostat, the reference temperature was maintained at 300 K using the V-rescale algorithm. The Particle mesh Ewald method was used to treat long range electrostatics and the cut-off for real space coulomb interactions was set at 10 Å. The final snapshot of this run was used as initial condition for steered molecular dynamics (SMD) simulations.

4.2.2 Steered Molecular Dynamics and Umbrella Sampling

4.2.2.1 AMBER11 Simulations

After preparing the system for steered MD the initial distance between glycosidic oxygen and C α of Lys74 was 8 Å. The ligand was pulled from the initial 8 Å distance to 28 Å, resulting in pulling of 20 Å along the reaction coordinate. The force constant applied for the SMD simulation was 50 kcal mol⁻¹ Å⁻² and at a pulling rate of 40 Å/ns. All other simulation parameters were kept the same as in the preparation run as discussed above. We will show below that the potential of mean force did not change beyond 14 Å, therefore 20 Å were deemed sufficient for our umbrella sampling study.

Positions along the pulling trajectory were used as initial conditions for our umbrella sampling calculations. For the umbrella sampling simulations, a restraining potential of 50 kcal mol⁻¹ Å⁻² was applied to each window and a total of 22 windows were used. Total

simulation time amounted to 220 ns. The analysis of the umbrella sampling simulations was done using the Weighted Histogram Analysis Method (WHAM) [83] implemented by Grossfield[60].

4.2.2.2 GROMACS Simulations

In the case of the simulations performed with Gromacs, the last structure from the preparation run was subjected to steered molecular dynamics simulations (SMD) or pulling simulations. During SMD, we applied a force constant of $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and a pull rate of 1 \AA/ns during 15 ns. This was done once along the Y direction and once along the Z direction.

In these simulations, dimannose was pulled away from the binding pocket and distance was measured between the glycosidic oxygen of dimannose and the N atom of Thr57. The maximum pulling distance in this case was 15 \AA . The force vs distance profile did not change beyond 12 \AA neither did the potential of mean force. In these simulations we applied restraints on $C\alpha$ atoms of the protein so that the rotation of the protein would be avoided during the pulling simulation. From each of the Y-axis pulling and Z-axis pulling simulations, snapshots were saved at regular intervals so that these can be used as starting structures for the umbrella sampling simulation[106, 136, 137] windows. The umbrella sampling simulations used a restraining potential which was applied to each window during the 10 ns of simulation time. 18 windows along the Y direction and 19 windows along the Z direction were used to generate the PMF curve. The total combined simulation time including both pulling directions was 400 ns. The analysis of the results obtained from these umbrella simulations was done using WHAM as coded in Gromacs[65].

4.3 Results

4.3.1 Pulling Pathway in Gromacs: 1-D Pulling Along Z- or Y-Axis

We converted the AMBER99SB protein force field and the GLYCAM06 carbohydrate force field to the Gromacs format and used the mixed scaling for protein and carbohydrates as described in the Methods section.

Representative snapshots along the pulling trajectories are shown in Figure 4.2. During the pulling simulation along the Z direction, the residues that appear to be important because they contribute to hydrogen bonding are Glu41, Ser52, Asn53, Thr57 and Lys74. Out of a total 15 ns, these residues interact with dimannose for almost 11.5 ns. At about 12 ns, the sugar breaks all the non-covalent bonding interactions with the protein and becomes free in solution. The hydrogen bond analysis for the trajectory was carried out with a facility implemented in the Visual Molecular Dynamics (VMD) software. [67]

Initially when dimannose is in its bound configuration, the residues that would appear to contribute to hydrogen bonding are backbone oxygen and side chain nitrogen of Asn42, backbone oxygen of Asn53 and backbone oxygen of Thr57. This close proximity of residues Asn42, Ser52, Asn53 and Lys74 was previously observed by Vorontsov *et al*[143] in their molecular dynamics simulations in solution. Also the CH₂OH of the non-reducing end of dimannose appears to show water mediated hydrogen bond with the CH₂OH of the reducing end of the sugar. As dimannose is pulled along the Z-Axis, new hydrogen bonds are formed. Among these carboxylate oxygen (O ϵ 1 and O ϵ 2) of Glu41, sidechain nitrogen (N ζ) of Lys74 along with carboxylate oxygen of Thr57 and Ser52 appear to be important.

The dissociation of dimannose from the binding site appears to take place in two steps. First the reducing end of dimannose breaks from the binding site and then the

non-reducing end breaks free. This is clearly visible in Figure 4.2. These residues were also highlighted in the NMR experiments by Sandstorm as previously described.[120] In this study the authors used STD-NMR experiments to determine that the protein transfers saturation to H2, H3 and H4 protons of the non-reducing end of the sugar.

In the pulling simulation along the Y-axis, the most important residues that interact with dimannose are carboxylate oxygen (O ϵ 1 and O ϵ 2) of Glu41, backbone oxygen of Asp44, backbone oxygen of Ser52, backbone oxygen of Asn53, backbone and side chain oxygen of Thr57 and backbone oxygen of Lys74. The main difference between the pulling along different directions appear to be the contacts with Asn42 vs Asp44 and Asn53. In both of these simulations, the salt bridge between Asp44 and Arg76 is maintained as observed earlier by Vorontsov *et al*[143] and Fujimoto *et al*[57]. The observation made by Vorontsov *et al* that Arg76 is not involved in hydrogen bonding with dimannose is indeed corroborated in our simulations using the AMBER ff99SB/GLYCAM06 force fields. This is not the case however with other force fields such as OPLS-AA.

4.3.2 Potential of Mean Force: 1-D Along Y- or Z-Axis

We used the g-wham facility implemented in Gromacs to compute the 1-D potential of mean force along the Y and Z directions and these are shown in figure 4.3. The fact that the dissociation of dimannose appears to occur in two stages can be appreciated both from figure 4.4 as well as from figure 4.3. For pulling along the Y-axis the reducing end of dimannose dissociates at around 9 Å and the non-reducing end of dimannose dissociates at around 11.8 Å. For the pulling along the Z-axis, the reducing end of dimannose dissociates at around 8.5 Å and the non-reducing end dissociates at around 10 Å. It can be said that whatever the pulling direction is, the dissociation is indeed a two step process and most of

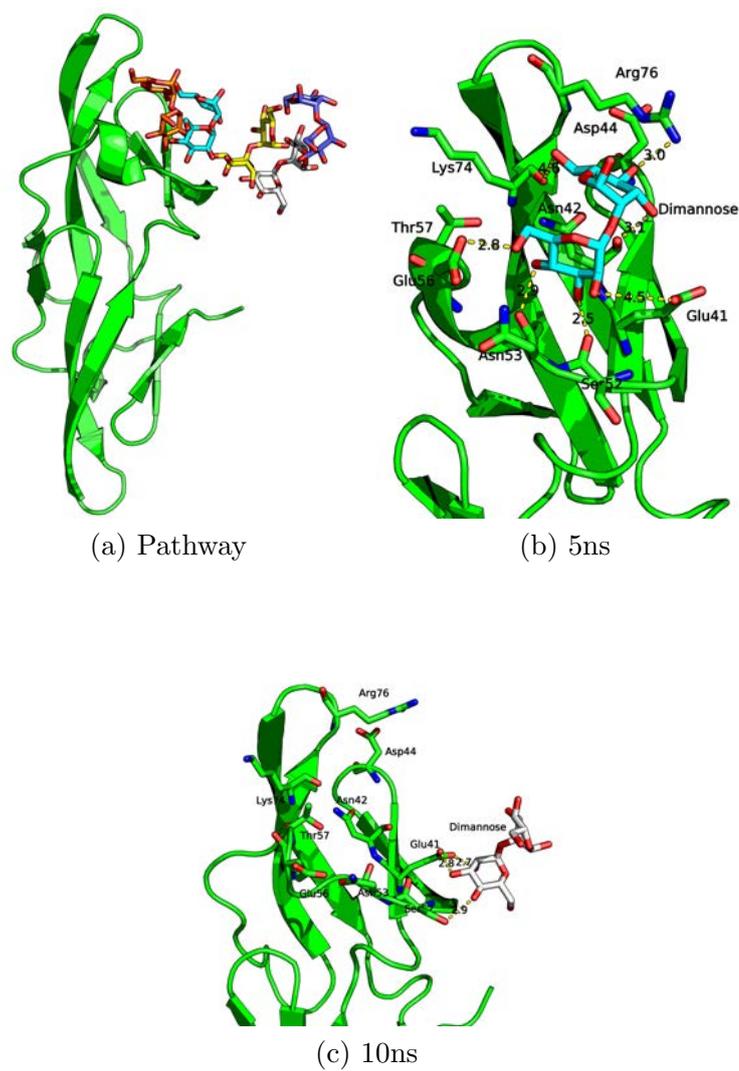


Figure 4.2: Snapshots and dissociation pathway along Y-axis. The overall pathway during the dissociation of dimannose along the Y-axis(a), snapshot at 5ns (b) and snapshot at 10ns (c).

Potential of Mean Force for Dimannose Binding to Cyanovirin-N

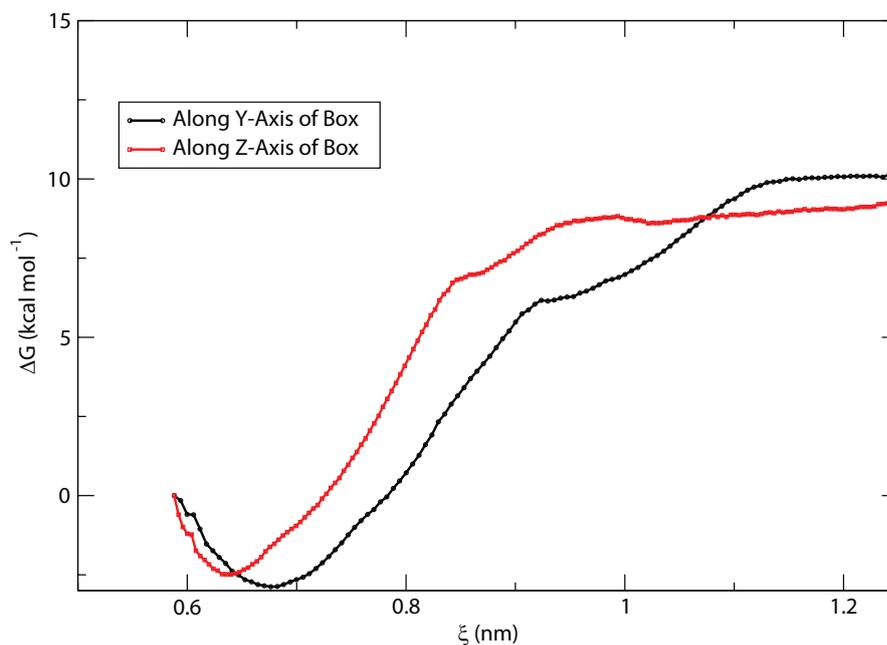


Figure 4.3: Potential of Mean Force along Y- and Z-Axis for dimannose binding to Cyanovirin-N

the energy is used in the dissociation of the reducing end of the sugar. The dissociation along the Y-axis requires 13.4 kcal mol⁻¹ whereas along the Y-axis it requires 12.5 kcal mol⁻¹.

4.3.3 Pulling Pathway in AMBER11: 3D Pulling Along X,Y,Z-Axis

Simulations using the AMBER 11 package were carried out because the sugar can be pulled without a constraining direction. In this case it is the distance between sugar and protein that change but the direction is arbitrary. The system was pulled apart at a

rate of 4 nm/ns with a constraining force constant of $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The total length of the simulation was 500 ps and structures were saved every 2 ps. The residues of CVN that interact with dimannose are sidechain carboxylate oxygen (O ϵ 1 and O ϵ 2), backbone carboxylate of Asn42, backbone carboxylate of Ser52 and backbone carboxylate oxygen of Asn53. Clearly in the AMBER11 simulations the dissociation of dimannose from CVN is also a two step process. Representative structures along the pulling trajectory using AMBER11 are shown in Figure 4.4.

Figure 4.5 shows the PMF obtained from AMBER11 simulations. Whereas in this case the dissociation is also a two step process, it is the detachment of the non-reducing end that is the most energetically expensive step. This appears to be different from what is observed when pulling along specific directions. The difference in free energy that can be estimated from the AMBER11 simulations is $12.5 \text{ kcal mol}^{-1}$. This number is similar to that obtained along specific pulling directions.

4.4 Conclusions

We have performed umbrella sampling simulations using the AMBER99SB/GLYCAM06 force fields to study the unbinding of dimannose from P51G-m4-CVN. In all of our pulling simulations we found that dissociation appears to be a two step process. The most important residues contributing to binding appear to be Glu41, Asn41, Asp44, Ser52, Asn53, Thr57 and Lys74. As previously observed by Vorontsov and coworkers[143], using the AMBER99SB/GLYCAM06 force fields, Arg76 forms a salt bridge with the Asp44 side chain but it rarely interacts with dimannose. Free energies differences estimated using the different potentials of mean force computed in our simulations vary from 12.5 kcal/mol to 13.4 kcal/mol depending on the pulling direction. These 1D estimations do not necessarily reflect

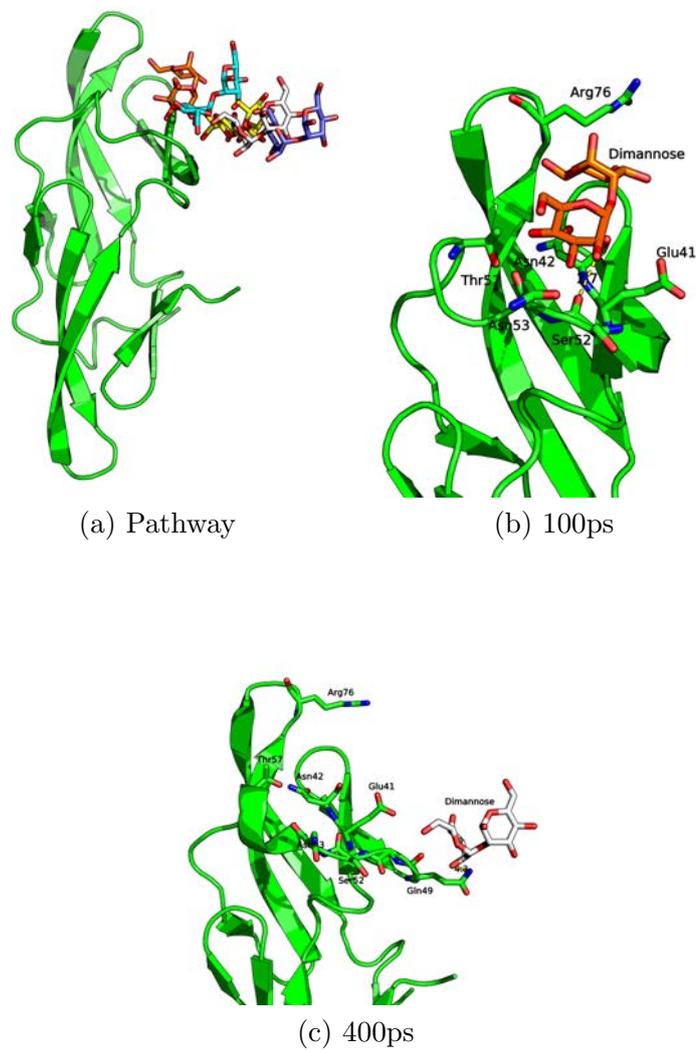


Figure 4.4: Snapshots and dissociation pathway along 3D in AMBER11 Simulations.

The overall pathway during the dissociation of dimannose is shown in (a), snapshot at 100ps (b) and snapshot at 400ps (c)

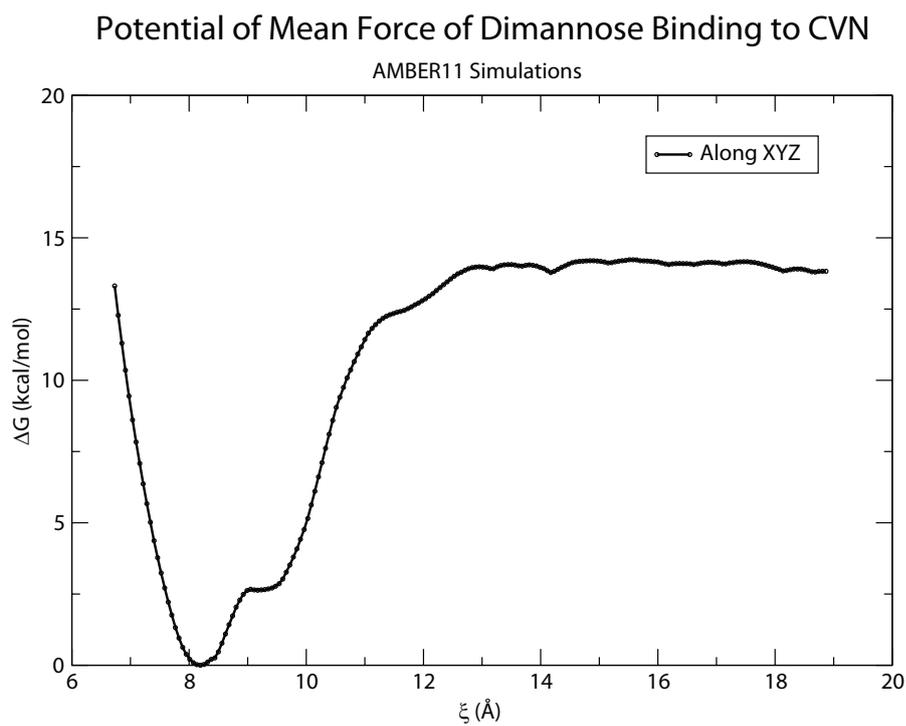


Figure 4.5: Potential of Mean Force obtained from AMBER11 Simulations. In the case of AMBER11, the pulling was in arbitrary directions

the experimental free energy difference between bound and unbound states and are perhaps reminiscent to AFM pulling experiments from which insight into mechanistic details may be obtained. The experimental free energy of binding of dimannose to P51G-m4-CVN has not yet been published however the number reported for wild type CVN[18] is 9.35 kcal mol⁻¹.

CHAPTER 5

ROLE OF GLU41 IN BINDING OF DIMANNOSE TO P51G-M4-CVN

5.1 Introduction

Cyanovirin-N a potentially very important anti-HIV protein was introduced in the previous chapter. Building on previous knowledge about binding affinities and the tendency of wild type CVN to form domain swapped dimers the Ghirlanda group has recently developed a mutant that enforces the monomeric structure (P51G mutation) and that abolishes the activity of the low affinity site (m4). The four mutations at the low affinity site in this case are K3N, T7A, E23I and N93A. The structure of P51G-m4-CVN was solved using X-ray crystallography. As expected, the protein was found to be in the monomeric form and it binds dimannose only at the high-affinity site.[55]

Vorontsov *et. al.* were the first to carry out computer studies on this system in complex with dimannose.[143, 144] Some of the main conclusions of this study were that 3'OH and 4'OH of dimannose were crucial to binding and that Asn42, Asp44, Ser52, Asn53, Thr57, Lys74 and Thr75 but not Arg76 or Glu41 were important to binding. These findings are quite controversial since in particular Glu41 has been highlighted as potentially very important by several experimental [14, 55, 54] and simulation studies[93, 58, 57] at least in the case of wild type CVN. NMR studies by Bewley and coworkers demonstrated that Glu41 interacts strongly with 2'OH of dimannose[14]. In fact this appears to be the case as well in the structure of the P51G-m4-CVN. [54]

Interestingly Glu41 appears to have rotameric configurations that are distinctly different in the case of monomeric CVN structures and domain-swap structures. In the case of monomeric structures χ_1 is in the trans configuration.[16, 14, 55, 54] Instead most

domain swapped dimers have Glu41 in the gauche conformation.

In the gauche χ_1 conformation carboxylate groups of Glu41 are unable to interact with the ligand whereas in the trans conformation they are ideally suited for hydrogen bonding of 2'OH of the non-reducing end of dimannose. This is depicted in Figure 5.1. This is likely the reason for the discrepancy between the simulations by Vorontsov and coworkers[143, 144] when compared to previous experimental and simulation studies. We found that the ff99SB force field for proteins[64] used in these studies highly favors the gauche configuration, however the OPLS-AA force field[72, 73] favors instead the trans configuration. Since the protein monomer is expected to have χ_1 of Glu41 in the trans configuration we have chosen to use the OPLS-AA force field for our in depth study of the relevance of Glu41 to dimannose binding to the P51G-m4-CVN mutant.

5.2 Methods

5.2.1 System Preparation

Two mutants of P51G-m4-CVN, E41A-P51G-m4-CVN and E41G-P51G-m4-CVN were generated by modifying the structure of (PDB ID 2RDK)[54]. 15103 water molecules were added to each of these models. We chose the TIP3P water model[71] for our simulations. Initial system dimensions were 78 X 78 X 78 Å³. For reasons described in the previous subsection the OPLS-AA force field was used to describe the proteins[72, 73] whereas the OPLS force field for carbohydrates [39] was chosen to represent dimannose. All simulations were carried out using the software Gromacs version 4.5.x [61]. Small changes to the force field for sugars had to be implemented since OPLS only has parameters for monomeric hexopyranoses.

Three K^+ counterions were included to balance the charge of P51G-m4-CVN. The

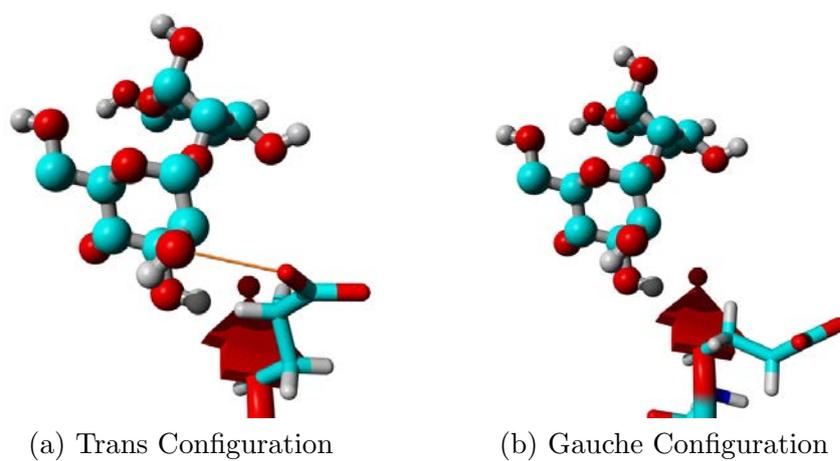


Figure 5.1: Trans configuration for the χ_1 angle of Glu 41 showing hydrogen bonding with 2'OH of the non-reducing end of dimannose (a). χ_1 of Glu41 in the gauche conformation centered around 60° .

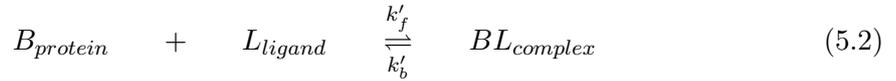
two mutants of this study have A41 and G41 rendering the protein with a charge of -2 electron units. We added 2 K^+ ions for these systems to maintain the electroneutrality. In a full thermodynamic cycle this would result in an imbalance of ions. (The cycle starts with three cations and ends with two) Another option was to use three counter ions in all simulations, but then in the case of the A41 and G41 mutants simulations would be overall positive. Both approaches were tested yielding results that are to within computational error.

Complexes initially underwent minimization for 5000 steps followed by constant volume and temperature simulations at 300K for 100 ps. After this initial pre-equilibration 5ns of equilibration were run in the NPT ensemble where volume is allowed to fluctuate while the system is kept at ambient pressure. All simulations used the Particle-Mesh Ewald (PME)[40] summation with a real space cutoff of 10 Å. Identical simulation protocol was used for the systems without dimannose.

5.2.2 Thermodynamic Integration

Consider a reaction between ligand L and a protein A in solution. Once the ligand reaches the protein, it binds to the protein at the binding/active site. Now consider a single point mutation in protein A resulting in protein B and the same ligand (L) binding to the mutated protein B. The relative free energy difference between a ligand L when bound to protein A as compared to mutant B can be computed using the thermodynamic integration (TI) approach or the free energy perturbation (FEP) method. The foundations for both methods were laid by Kirkwood[75] and Zwanzig [156]. For the application of this method in molecular dynamics or Monte Carlo simulations, an artificial parameter λ is defined that couples two thermodynamic states in an alchemical mutation such as ligand L bound to

protein A and ligand L bound to protein B.[12, 130, 81]



The two end states (complexes) in the equations 5.1 and 5.2 can be coupled using the so called λ parameter through a number of intermediate λ values. These intermediate λ values are artificial constructs. The λ dependence on the potential function can be described using a linear relationship such as

$$V(\lambda) = (1 - \lambda)V_{AL} - \lambda V_{BL} \quad (5.3)$$

The two end states can be recovered when $\lambda = 0$ (protein A bound to ligand L, eqn. 5.1) and $\lambda = 1$ (protein B bound to ligand L, eqn. 5.2). The same holds true for the transformation of protein A into protein B in the absence of ligand. For applying this method to the study of binding free energies, a thermodynamic cycle (see Figure 5.2) is utilized. [53] In figure 5.2, P51G-m4-CVN is represented in the lower left corner by the letter “A”, E41A-P51G-m4-CVN or E41G-P51G-m4-CVN are represented by “B”, dimannose is the ligand “L”. From Figure 5.2, it is clear that $\Delta G_1 - \Delta G_2 - \Delta G_3 + \Delta G_4 = 0$ which implies $\Delta G_3 - \Delta G_4 = \Delta G_1 - \Delta G_2$. The quantity $\Delta G_1 - \Delta G_2$ involves the transformation of complex AL into complex BL (ΔG_1) and protein A to protein B (ΔG_2). This is the preferred way to indirectly obtain $\Delta G_3 - \Delta G_4$. The other path requires simulating the removal of the ligand and is computationally more challenging.

Further, ΔG_1 and ΔG_2 are split into three processes as shown below.

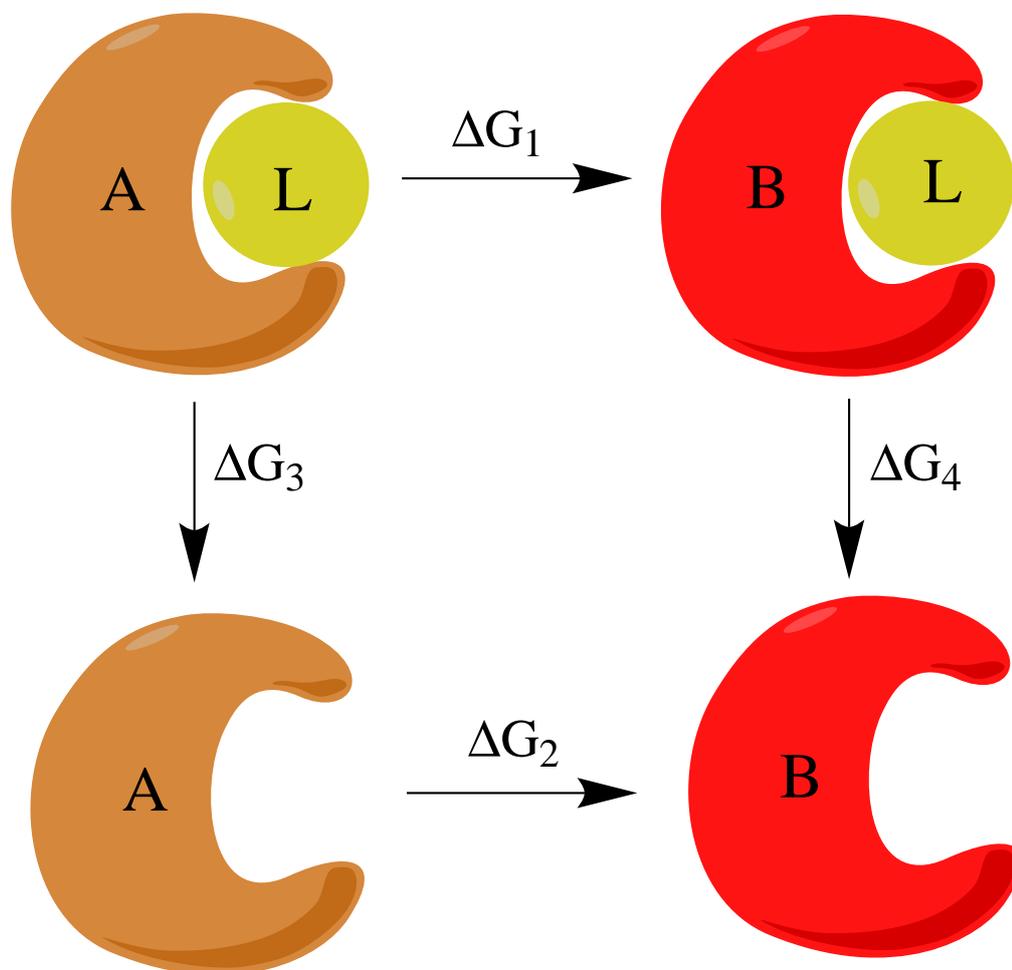


Figure 5.2: Thermodynamic Cycle for Binding Free Energy Calculation

1. Remove the charges from glutamate residue (charge removal)
2. Convert the sidechain of glutamate to alanine or glycine (Lennard-Jones transformation)
3. Addition of charge on alanine residue (charge addition)

For step 3, we calculate the free energy of charging of A41 and G41 as the negative of the discharging of the respective amino acid. In most practical applications of the thermodynamic integration approach, either a single topology or a dual topology approach is used [108]. In our study we used the single topology approach. For the Lennard-Jones (LJ) transformation we employed soft-core potentials[31, 134, 111] with default parameter values as coded in GROMACS version 4.5.x.

For the charge transformation, we used 11 λ values starting from 0 to 1 at intervals of 0.1.

Since the Lennard-Jones transformation is more prone to fluctuations that result in poor convergence particularly in the regions close to $\lambda = 0$ and $\lambda = 1$ we set λ intervals of 0.02 from $\lambda = 0$ to $\lambda = 0.1$ and also from $\lambda = 0.9$ to $\lambda = 1$. In between these values intervals were set at 0.1 as in the case of the Coulomb transformation.

The protocol for each λ simulation was as follows: first a minimization of 5000 steps, then 100 ps of NVT simulation followed by 250 ps in the NPT ensemble followed by 2 ns NPT used for production. The three step approach, in the presence and absence of dimannose resulted in 0.85 μ s of simulation.

5.3 Results

5.3.1 Relative Binding Free Energy Calculations

In order to estimate the relative binding free energy difference for dimannose in complex with P51G-m4-CVN and two of its mutants E41A and E41G, we used the thermodynamic integration approach. In order to gauge the error in simulation we performed three trials for each mutation. As previously discussed in section 5.2 and as shown in Figure 5.2, each of these trials has three steps and each of the steps has to be carried out both in the presence and absence of dimannose. Furthermore, each step consists of λ values ranging from 0 to 1. In all we have 82 simulations for each trial for the E41A mutation and each λ simulation runs for 2 ns resulting in 492 ns. For the E41G mutation not all simulations needed to be repeated since the step corresponding to the discharge of Glu41 is the same for both mutations. 60 new simulations are needed in the case of the E41G mutation. This resulted in extra 360 ns of simulation time.

The below table summarizes the results obtained for each step in the presence and absence of dimannose. From these three trials we see that both in the presence and absence of dimannose discharging of glutamate results in a large positive free energy penalty as shown in Figures 5.3a, 5.6a, 5.4a, 5.7a, 5.5a and 5.7a. The integral of $\langle \partial V / \partial \lambda \rangle$ for this step yields an average value 741.7 ± 0.4 kJ mol⁻¹ in the presence of ligand and 733.2 ± 0.6 kJ mol⁻¹ in the absence of the ligand (see Tables 5.1 and 5.2). In the case of the discharging of Ala41 or Gly41 this penalty is close to a third of this value (see tables 5.1 and 5.2) and is almost the same in the presence and absence of the ligand. This can be expected as Glu41 interacts strongly with dimannose or the solvent via hydrogen bonding whereas side chains of Ala41 or Gly41 do not form hydrogen bonds either with dimannose or solvent. The

relative free energy for Coulombic discharge in the presence and absence of ligand results in 7.6 ± 0.8 kJ mol⁻¹ for E41A and 7.9 ± 1.0 kJ mol⁻¹ for E41G.

The relative free energy difference in the case of Lennard-Jones transformation of Glu41 to Ala41 in the presence and absence of dimannose result is -6.4 ± 0.9 kJ mol⁻¹ (see Figures 5.3b, 5.4b and 5.4b and table 5.1). In the case of Glu41 to Gly41, the relative free energy difference for this step is -8.0 ± 1.2 kJ mol⁻¹ (see Figures 5.6b, 5.7b and 5.8b and table 5.2). The relative size ordering of the side chains of Gly, Ala and Glu is Gly smaller than Ala which in turn is smaller than Glu. The Van Der Waals interactions appear to follow the same trend, namely the smaller the amino acid the more favorable the mutation.

The overall relative free energy difference when Coulombic and Lennard-Jones parts are added together comes out to be 1.3 ± 1.2 kJ mol⁻¹ for E41A-P51G-m4-CVN and -0.1 ± 2.1 kJ mol⁻¹ for E41G-P51G-m4-CVN.

In order to check whether our different λ simulations properly sampled the conformational space of dimannose, we analyzed the torsional space of dimannose free in solution and bound to P51G-m4-CVN and the two mutants. Dimannose exists in two minima when free in solution as seen in Figure 5.9 and it exists only in one dominant conformation when bound to P51G-m4-CVN or its mutants. At least in our 10 ns simulations there were no conformational changes observed in dihedral space.

5.4 Conclusions

Glu41 binds very strongly to 2'OH of the non-reducing end of dimannose. This however does not mean that the importance of this residue towards the binding free energy is high. We believe this is quite an interesting observation. A surface charged residue interacts strongly with a hydrogen bonded ligand, but it also is expected to interact strongly with

	Discharging of Glu41 (kJ/mol)	LJ Transformation (kJ/mol)	Discharging of Ala41 (kJ/mol)	Total ΔG (kJ/mol)
Trial I + Dimannose	742.6 (0.5)	-5.4 (0.6)	-275.2 (0.1)	
Trial II + Dimannose	741.8 (0.4)	-4.1 (0.6)	-275.4 (0.2)	
Trial III + Dimannose	740.8 (0.4)	-6.0 (0.5)	-275.5 (0.1)	
AVG ΔG Dimannose	741.7 (0.4)	-5.2 (0.6)	-275.4 (0.1)	461.2 (0.7)
Trial I - Dimannose	733.2 (0.7)	0.6 (0.7)	-273.8 (0.2)	
Trial II - Dimannose	733.1 (0.7)	2.0 (0.2)	-274.7 (0.3)	
Trial III - Dimannose	733.1 (1.1)	0.9 (1.0)	-274.7 (0.3)	
AVG ΔG NoDimannose	733.2 (0.6)	1.2 (0.7)	-274.4 (0.3)	459.1 (1.0)
$\Delta\Delta G$	8.6 (0.8)	-6.4 (0.9)	-0.9 (0.3)	1.3 (1.2)

Table 5.1: Thermodynamic Integration Analysis for Relative Binding Free Energies for the Glu41Ala Mutation.

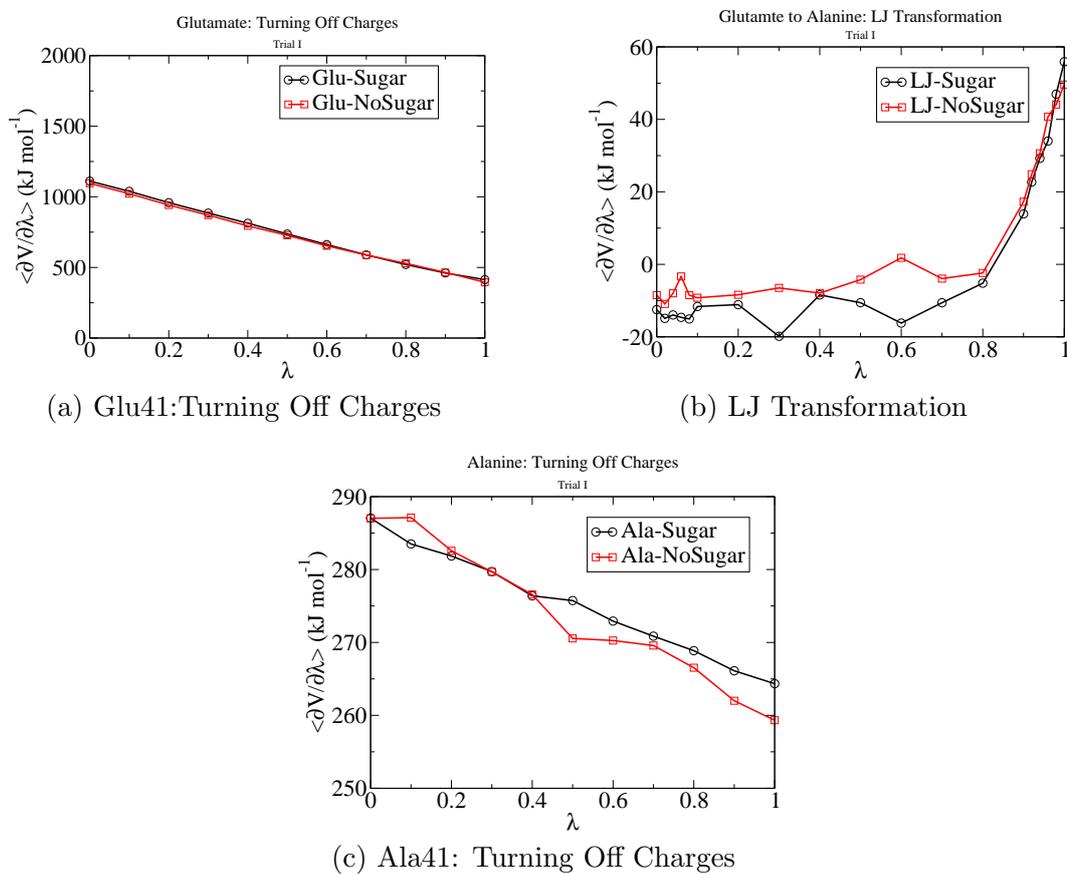


Figure 5.3: Three step approach for Thermodynamic Integration simulations, Trial I. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).

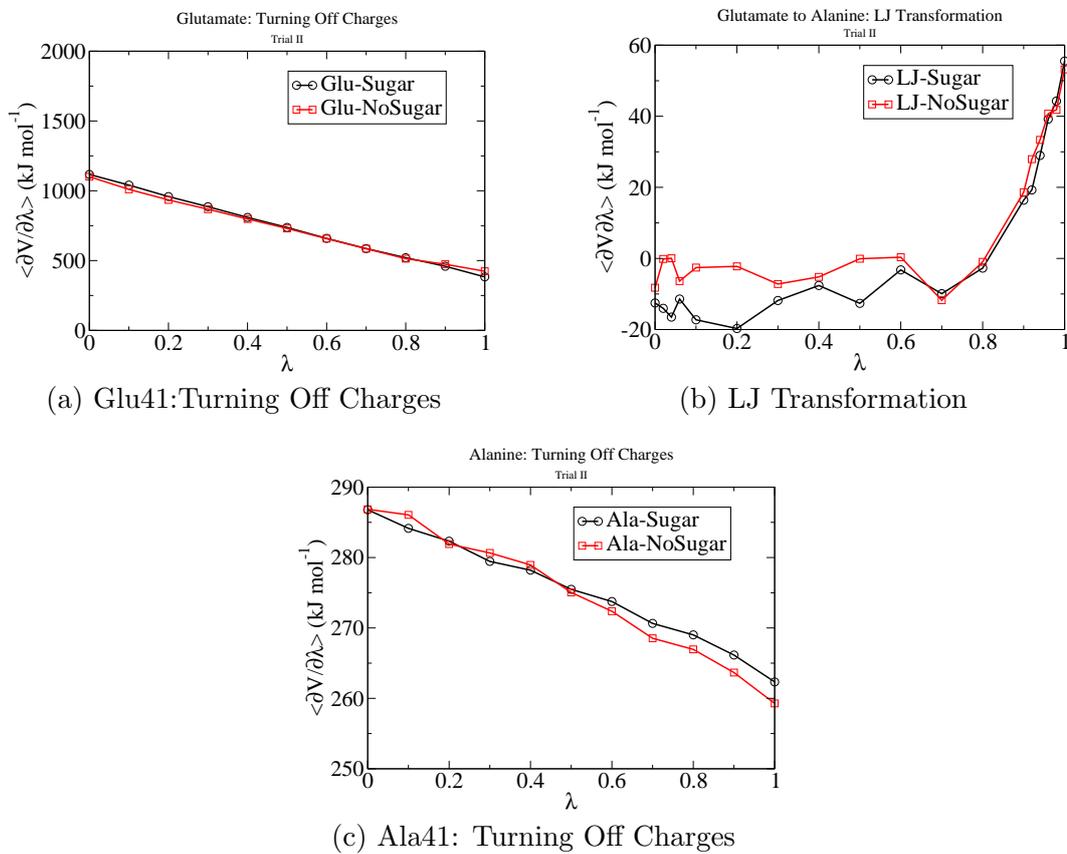


Figure 5.4: Three step approach for Thermodynamic Integration simulations, Trial II. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).

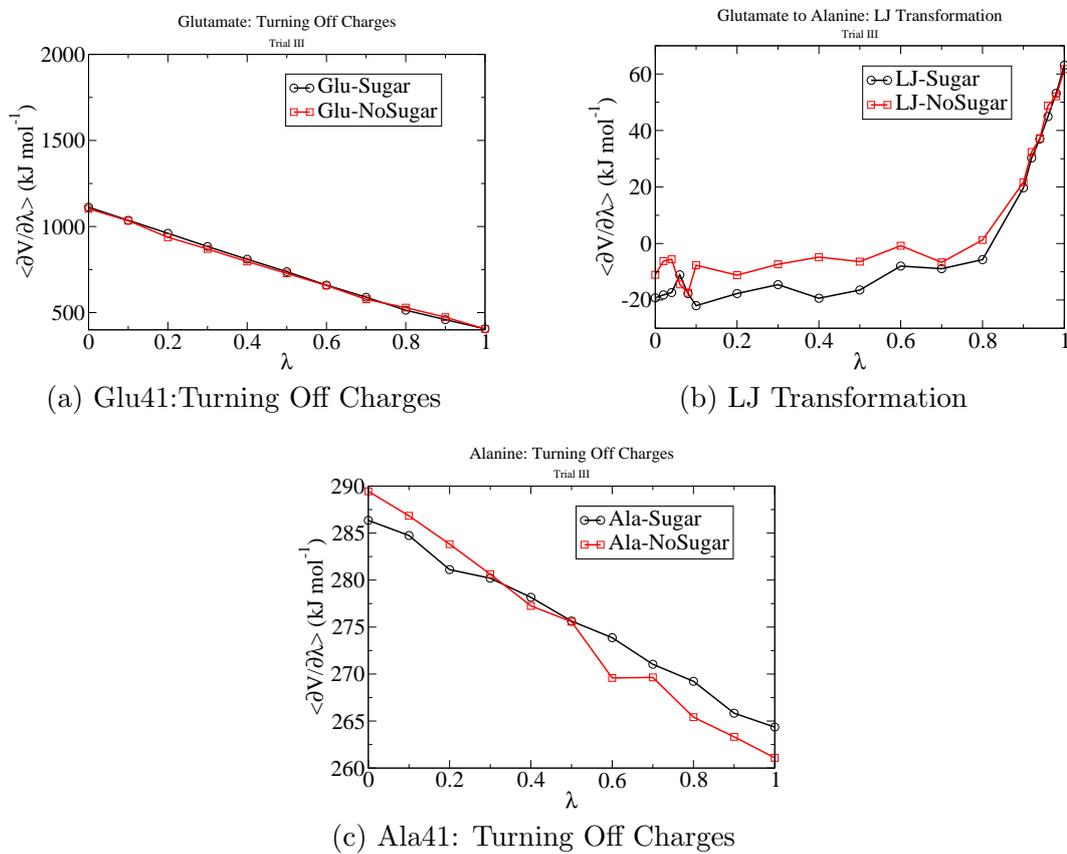


Figure 5.5: Three step approach for Thermodynamic Integration simulations, Trial III. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Alanine (c).

	Discharging of Glu41 (kJ/mol)	LJ Transformation (kJ/mol)	Discharging of Ala41 (kJ/mol)	Total ΔG (kJ/mol)
Trial I + Dimannose	742.6 (0.5)	-48.7 (1.2)	-270.5 (0.1)	
Trial II + Dimannose	741.8 (0.4)	-47.7 (0.9)	-270.2 (0.1)	
Trial III + Dimannose	740.8 (0.4)	-48.0 (0.6)	-270.6 (0.2)	
AVG ΔG Dimannose	741.7 (0.4)	-48.1 (0.9)	-270.4 (0.1)	423.2 (1.0)
Trial I - Dimannose	733.2 (0.7)	-40.6 (0.4)	-269.6 (0.1)	
Trial II - Dimannose	733.1 (0.7)	-40.6 (1.1)	-269.6 (0.2)	
Trial III - Dimannose	733.1 (1.1)	-39.1 (0.6)	-270.0 (0.2)	
AVG ΔG NoDimannose	733.2 (0.6)	-40.1 (0.8)	-269.7 (0.2)	423.3 (1.8)
$\Delta\Delta G$	8.6 (0.8)	-8.0 (1.2)	-0.7 (0.2)	-0.1 (2.1)

Table 5.2: Thermodynamic Integration Analysis for Relative Binding Free Energies for the Glu41Gly Mutation.

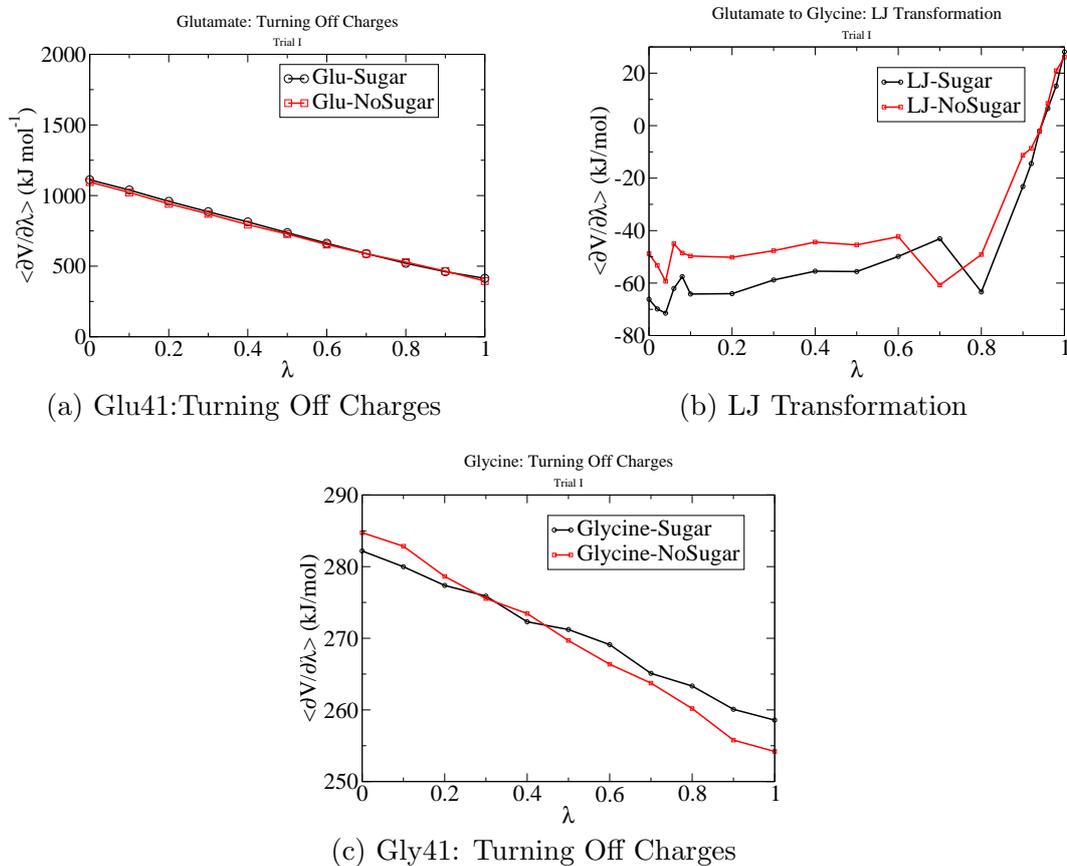


Figure 5.6: Three step approach for Thermodynamic Integration simulations, Trial I. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).

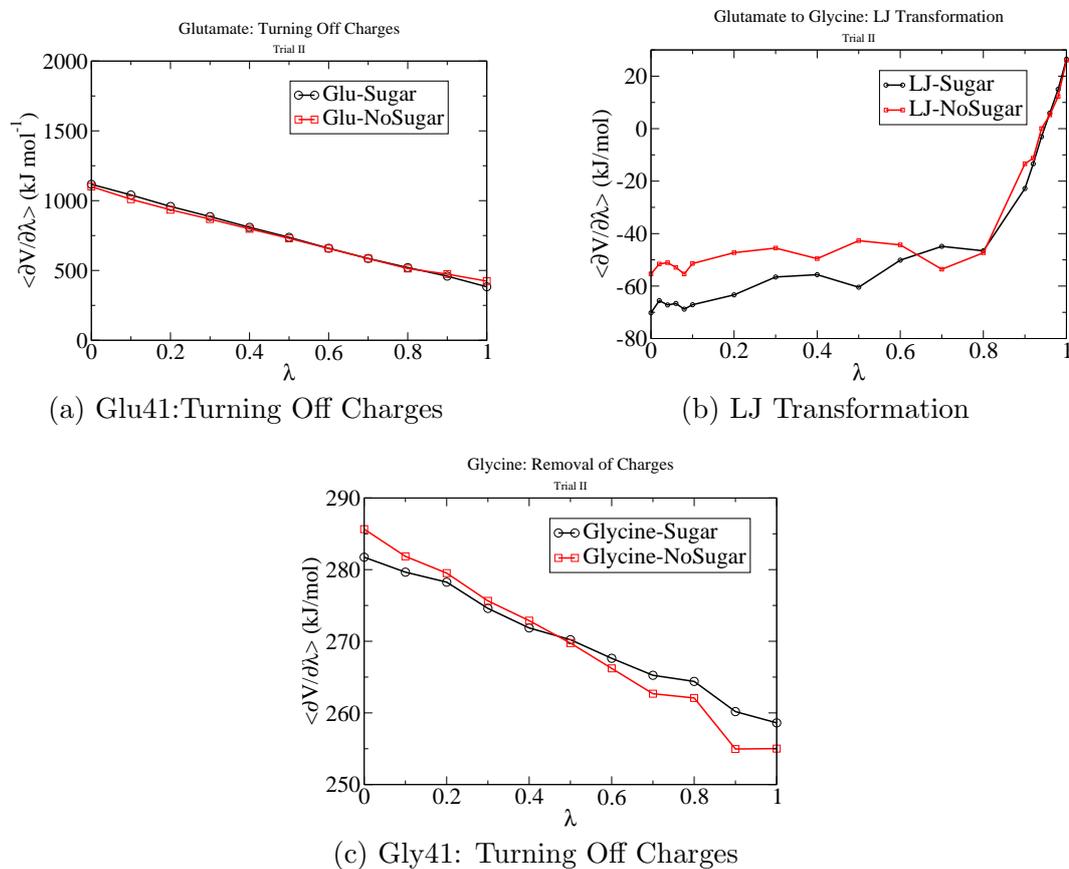


Figure 5.7: Three step approach for Thermodynamic Integration simulations, Trial II. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).

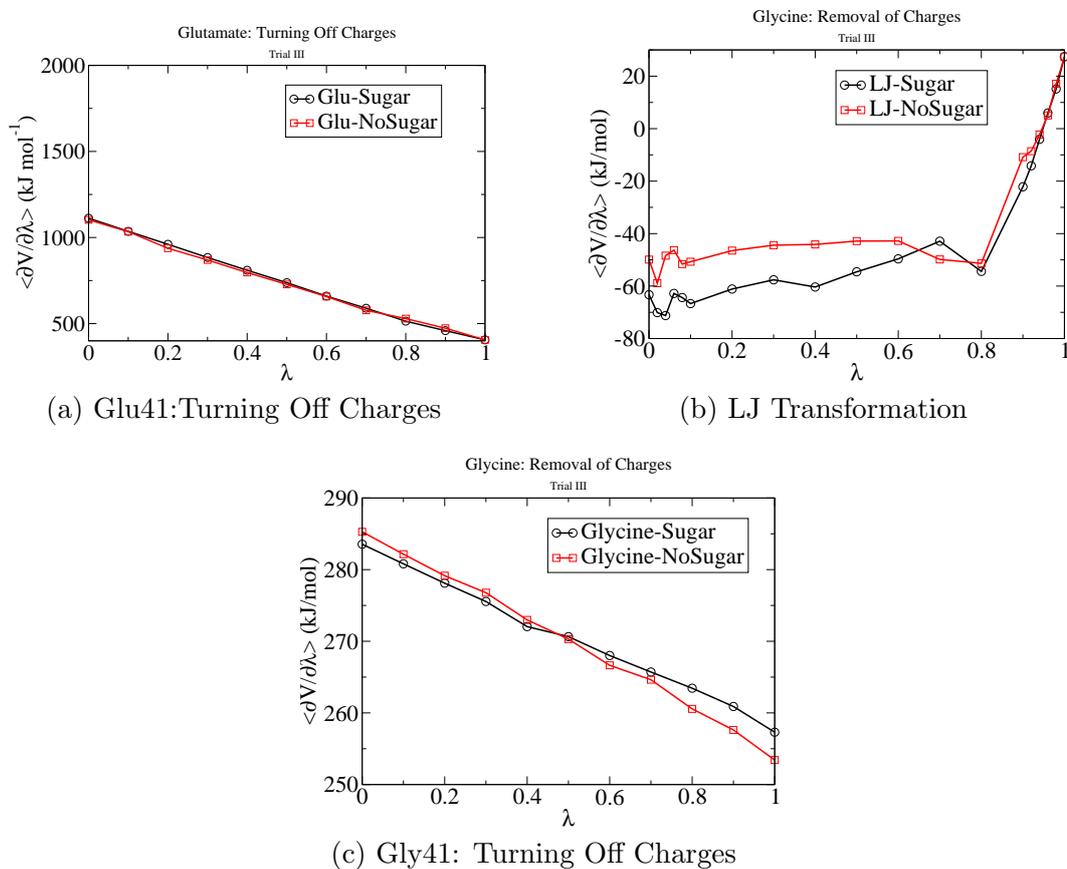


Figure 5.8: Three step approach for Thermodynamic Integration simulations, Trial III. Turning off charges in Glutamate(a), LJ transformation (b) and Turning off charges in Glycine (c).

Dihedral Angle Preferences of Dimannose

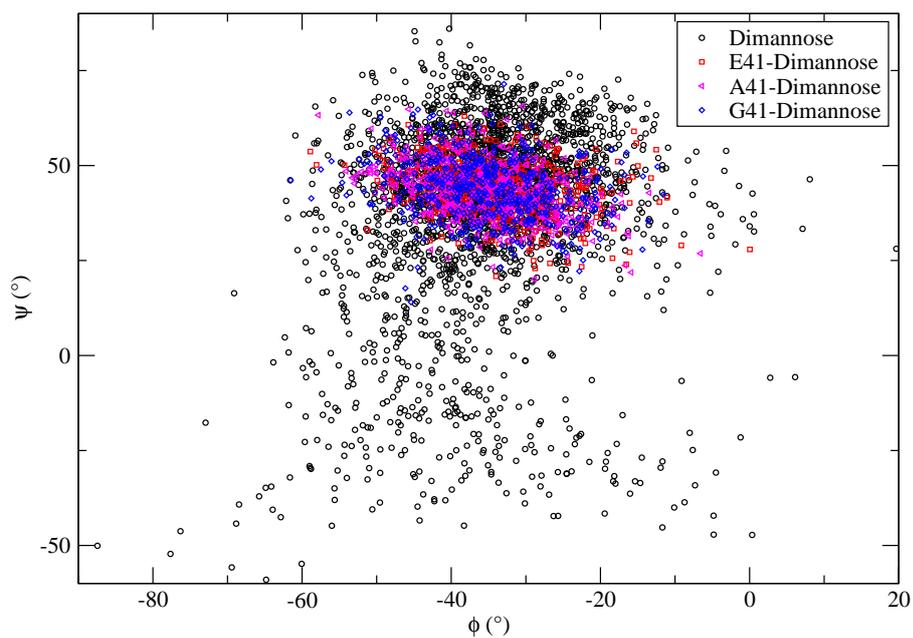


Figure 5.9: Torsional preferences of dimannose in solution (black), bound to E41-P51G-m4-CVN (red), bound to A41-P51G-m4-CVN (magenta) and bound to G41-P51G-m4-CVN (blue)

solvent in the absence of the ligand. Absolute binding free energies include the difference between both of these large numbers and in cases this difference may be modest. Interestingly in the case of dimannose the overall penalty for discharging Glu41 in the presence of the sugar as compared to discharging the same residue in its absence but surrounded by solvent is on the order of 8 kJ/mol. Although this number is significant it is not large enough to significantly overcome the gain in the Lennard Jones transformation into residues with smaller side chains. This has important implications for our understanding of the CVN family of systems. Whereas structural information derived from experiments may point to a residue having close contact hydrogen bonding with a ligand and simulations may point to the large energy of interaction between ligand and such residue, neither of these is conclusive evidence for the significance of this residue in terms of free energies. Such analysis only focuses on the bound state and does not take into account the free energy of the unbound state.

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

In this thesis work we have investigated various structural and dynamical aspects of the nature of carbohydrates as well as carbohydrate-protein interactions. These include the role of water on the surface of oligosaccharides and the binding free energies of a disaccharide in complex with CVN. From our study of the role of water on the surface of poly- and oligosaccharides we learned about the different types of solvent behaviors that can be expected as the anomeric configuration, type of linkage between monosaccharides or branching points are modified. This work was one of the most exhaustive studies so far on the behavior of water on the surface of carbohydrates.

Previous work in the Margulis group focused on the development of software tools to study conformational sampling of sugars free in solution and in the context of carbohydrate-protein complexes. Building on these concepts, we took advantage of the Fast Sugar structure Prediction Software to derive conformations of a complex N-glycan involved in the degradation of glycoproteins. By deriving the conformations of the oligosaccharide and using these conformers as probes, we were able to predict possible structures of the heterotrimeric asialoglycoprotein receptor bound to triantennary oligosaccharide. The general methodology developed in this project can be applied to any type II membrane bound multimeric receptors.

In an ongoing project with the Ghirlanda group at Arizona State University, we are working on the binding free energies of dimannose in complex to various mutants of Cyanovirin-N. Previous work in the Margulis group has shown that certain amino acids at the binding site interact strongly with dimannose due to hydrogen bonding and electrostatic

interactions. The current binding free energy calculations have shown that in the estimation of free energies one has to consider not only the interactions of the protein with the ligand but also those of the protein with solvent in the absence of the ligand.

To better understand the free energetics of CVN and its complexes and in order to propose transformations that will result in better binding mutants, more mutation studies must be pursued. Furthermore single mutations may not be sufficient because of possible cooperativity effects. In such cases multiple mutations may need to be performed. This is quite a daunting computational task which we plan to undertake. We plan to initially pursue mutations of Asn53Ser and Lys74Glu.

APPENDIX A AMBER99SB/GLYCAM06 PARAMETER CONVERSION TO GROMACS FORMAT

A.1 Implementation of Mixed-Scaling

The original `amb2gmx.pl` script was written by Eric Sorin in the Pande Lab at Stanford University. This script works well for the conversion of topology and coordinate files in standard AMBER protein force fields into Gromacs format. But when the AMBER ff99SB protein force field or its variants are used in conjunction with the GLYCAM06 force field for carbohydrates the issue of mixed-scaling has to be considered.

For the proteins, when the ff99SB force field is used, the non-bonded 1-4 interactions are scaled with factors of 1/2 in the case of Lennard Jones interactions and 1/1.2 in the case of Coulomb interactions. For carbohydrates, the GLYCAM force field is parameterized using 1.0 as scaling factor for both types of non-bonded interactions. The use of different scaling factors for different species in simulation is well implemented in the AMBER11 software but is not implemented in Gromacs 4.5.x.

In order to achieve the mixed scaling in the case of the Gromacs package these are the steps that need to be followed:

1. Set the `fudgeLJ` to 1.0 and `fudgeQQ` to 0.16666666
2. Calculate each possible carbohydrate [pairtypes] interactions according to the combination rules, using Lorentz-Berthelot rules. For σ_{ij} this is the arithmetic average and for ϵ_{ij} the geometric average. Then ϵ_{ij} needs to be divided by 6 and the carbohydrate [pairs] section needs to be replicated in the topology file *six* times.
3. Calculate each possible protein [pairtypes] interactions according to the Lorentz-

Berthelot combination rules and divide the ϵ_{ij} value by 10 and repeat the [pairs] section in the topology file *five* times.

4. Verify that you have not broken anything by combining carbohydrate and protein itp files and recompute 1-4 interactions. Since these are intramolecular you should obtain the same thing as in the sum of each of the individual molecules.

REFERENCES

- [1] P Adhikari, K Bachhawat-Sikder, C Thomas, R Ravishankar, A Jeyaprakash, Vivek Sharma, M Vijayan, and A Surolia. Mutational analysis at asn-41 in peanut agglutinin. a residue critical for the binding of the tumor-associated thomsen-friedenreich antigen. *J. Biol. Chem.*, 276:40734, 2001.
- [2] A Almond. Towards understanding the interaction between oligosaccharides and water molecules. *Carbohydr. Res.*, 340:907–920, 2005.
- [3] A Almond, A Brass, and JK Sheehan. Oligosaccharides as model systems for understanding water-biopolymer interaction: Hydrated dynamics of a hyaluronan decamer. *J. Phys. Chem. B*, 104:5634–5640, 2000.
- [4] Andrew Almond and John Sheehan. Predicting the molecular shape of polysaccharides from dynamic interactions with water. *Glycobiology*, 13:255, 2003.
- [5] S Andre, S Kojima, G Gundel, R Russwurm, X Schrott, C Unverzagt, and HJ Gabius. Branching mode in complex-type triantennary n-glycans as regulatory element of their ligand properties. *Biochem. Biophys. Acta*, 1760(5):768–782, May 2006.
- [6] James N Arnold, Mark R Wormald, Robert B Sim, Pauline M Rudd, and Raymond A Dwek. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.*, 25:400–500, 2007.
- [7] G Ashwell and J Harford. Carbohydrate-specific receptors of the liver. *Annu. Rev. Biochem.*, 51:531–554, 1982.
- [8] J. U. Baenziger and D. Fiete. Galactose and n-acetylgalactosamine-specific endocytosis of glycopeptides by isolated rat hepatocytes. *Cell*, 22(2):611–620, 1980.
- [9] P V Balaji, P K Qasba, and V S Rao. Molecular dynamics simulations of asialoglycoprotein receptor ligands. *Biochemistry*, 32(47):12599–611, Nov 1993.
- [10] Laura G. Barrientos, John M. Louis, Istvan Botos, Toshiyuki Mori, Zhaozhong Han, Barry R. O’Keefe, Michael R. Boyd, Alexander Wlodawer, and Angela M. Gronenborn. The domain-swapped dimer of cyanovirin-n is in a metastable folded state: Reconciliation of x-ray and nmr structures. *Structure*, 10(5):673–686, 2002. doi: 10.1016/S0969-2126(02)00758-X.

- [11] Laura G. Barrientos, Elena Matei, Fatima Lasala, Rafael Delgado, and Angela M. Gronenborn. Dissecting carbohydrate cyanovirin-n binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Engineering, Design and Selection*, 19(12):525–535, 2006. doi: 10.1093/protein/gzl040.
- [12] D. L. Beveridge and F. M. Dicapua. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.*, 18:431–492, 1989.
- [13] Carole A. Bewley. Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-hiv protein cyanovirin-n using residual dipolar couplings. *Journal of the American Chemical Society*, 123(5):1014–1015, 2001. doi: 10.1021/ja005714o.
- [14] Carole A. Bewley. Solution structure of a cyanovirin-n:man1-2man complex: Structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure*, 9(10):931–940, 2001. doi: 10.1016/S0969-2126(01)00653-0.
- [15] Carole A. Bewley and G. Marius Clore. Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-n in solution using dipolar couplings and rigid body minimization. *Journal of the American Chemical Society*, 122(25):6009–6016, 2000. doi: 10.1021/ja000858o.
- [16] Carole A. Bewley, Kirk R. Gustafson, Michael R. Boyd, David G. Covell, Ad Bax, G. Marius Clore, and Angela M. Gronenborn. Solution structure of cyanovirin-n, a potent hiv-inactivating protein. *Nat. Struct. Biol.*, 5(7):571–578, 1998. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 1998:463471(Journal) 10.1038/828.
- [17] Carole A. Bewley, Shigeki Kiyonaka, and Itaru Hamachi. Site-specific discrimination by cyanovirin-n for α -linked trisaccharides comprising the three arms of man8 and man9. *Journal of Molecular Biology*, 322(4):881–889, 2002. doi: 10.1016/S0022-2836(02)00842-2.
- [18] Carole A. Bewley and Sarah Otero-Quintero. The potent anti-hiv protein cyanovirin-n contains two novel carbohydrate binding sites that selectively bind to man8 d1d3 and man9 with nanomolar affinity: Implications for binding to the hiv envelope protein gp120. *Journal of the American Chemical Society*, 123(17):3892–3902, 2001. doi: 10.1021/ja004040e.
- [19] A M Bianucci and F Chiellini. A 3d model for the human hepatic asialoglycoprotein receptor (asgp-r). *J. Biomol. Struct. Dyn.*, 18(3):435–51, Dec 2000.

- [20] M D Bider and M Spiess. Ligand-induced endocytosis of the asialoglycoprotein receptor: evidence for heterogeneity in subunit oligomerization. *FEBS Lett*, 434(1-2):37–41, Aug 1998.
- [21] M D Bider, J M Wahlberg, R A Kammerer, and M Spiess. The oligomerization domain of the asialoglycoprotein receptor preferentially forms 2:2 heterotetramers in vitro. *J. Biol. Chem.*, 271(50):31996–2001, Dec 1996.
- [22] Başar Bilgiçer, Samuel W Thomas, 3rd, Bryan F Shaw, George K Kaufman, Vijay M Krishnamurthy, Lara A Estroff, Jerry Yang, and George M Whitesides. A non-chromatographic method for the purification of a bivalently active monoclonal igg antibody from biological fluids. *J. Am. Chem. Soc.*, 131(26):9361–7, Jul 2009.
- [23] J Bischoff, S Libresco, M A Shia, and H F Lodish. The h1 and h2 polypeptides associate to form the asialoglycoprotein receptor in human hepatoma cells. *J. Cell Biol.*, 106(4):1067–74, Apr 1988.
- [24] J Bischoff and H F Lodish. Two asialoglycoprotein receptor polypeptides in human hepatoma cells. *J. Biol. Chem.*, 262(24):11825–32, Aug 1987.
- [25] T L Blumh and A Sarko. The triple helical structure of lentinan, a linear β -(1 \rightarrow 3)-d-glucan. *Can. J. Chem.*, 55:400–500, 1977.
- [26] I. Botos and A. Wlodawer. Cyanovirin-n: a sugar-binding antiviral protein with a new twist. *Cell. Mol. Life Sci.*, 60(2):277–287, 2003. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 2003:761707(Journal; General Review) 10.1007/s000180300023.
- [27] Istvan Botos, Toshiyuki Mori, Laura K. Cartner, Michael R. Boyd, and Alexander Wlodawer. Domain-swapped structure of a mutant of cyanovirin-n. *Biochem. Biophys. Res. Commun.*, 294(1):184–190, 2002. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 2002:482288(Journal) 10.1016/S0006-291X(02)00455-2.
- [28] Istvan Botos, Barry R. O’Keefe, Shilpa R. Shenoy, Laura K. Cartner, Daniel M. Ratner, Peter H. Seeberger, Michael R. Boyd, and Alexander Wlodawer. Structures of the complexes of a potent anti-hiv protein cyanovirin-n and high mannose oligosaccharides. *Journal of Biological Chemistry*, 277(37):34336–34342, 2002.

- [29] M. R. Boyd, K. R. Gustafson, J. B. McMahon, R. H. Shoemaker, B. R. O’Keefe, T. Mori, R. J. Gulakowski, L. Wu, M. I. Rivera, C. M. Laurencot, M. J. Currens, J. H. Cardellina, R. W. Buckheit, P. L. Nara, L. K. Pannell, R. C. Sowder, and L. E. Henderson. Discovery of cyanovirin-n, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrobial Agents and Chemotherapy*, 41(7):1521–30, 1997.
- [30] P P Breitfeld, C F Simmons, Jr, G J Strous, H J Geuze, and A L Schwartz. Cell biology of the asialoglycoprotein receptor system: a model of receptor-mediated endocytosis. *Int. Rev. Cytol.*, 97:47–95, 1985.
- [31] Beutler T. C., Mark A. E., van Schaik R. C., Gerber P. R., and van Gunsteren WF. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physical Letters*, 222:529–539, 2006. doi: 10.1093/protein/gzl040.
- [32] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, , and P.A. Kollman. *AMBER 9*. University of California, San Fransisco, San Fransisco, USA, 2006.
- [33] D.A. Case, T.A. Darden, III T.E. Cheathama, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossvry, K.F. Wong, F. Paesani, R.M. Wolf J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. *AMBER 11*. University of California, San Francisco, 11 edition, 2010.
- [34] Ying-Hua Chung, Junchao Xia, and Claudio J Margulis. Diffusion and residence time of hydrogen peroxide and water in crowded protein environments. *J Phys. Chem. B*, 111:13336–44, 2007.
- [35] C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley, and G-J. Boons. Involvement of water in carbohydrateprotein binding. *J. Am. Chem. Soc.*, 123:12238–12247, 2001.
- [36] L Cordone, G Cottone, and S Giuffrida. Role of residual water hydrogen bonding in sugar/water/biomolecule systems: a possible explanation for trehalose peculiarity. *J. Phys.: Condens. Matter*, 19:205110, 2007.

- [37] Paul E Correa. Prediction of protein side-chain conformation by packing optimization. *Proteins*, 7(4):366–377, 1990.
- [38] Jeremy Curuksu and Martin Zacharias. Enhanced conformational sampling of nucleic acids by a new hamiltonian replica exchange molecular dynamics approach. *J. Chem. Phys.*, 130(10):104110, Mar 2009.
- [39] Wolfgang Damm, Antonio Frontera, Julian Tirado-Rives, and William L. Jorgensen. Opls all-atom force field for carbohydrates. *Journal of Computational Chemistry*, 18(16):1955–1970, 1997.
- [40] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [41] JL Dashnau, KA Sharp, and JM Vanderkooi. Carbohydrate intramolecular hydrogen bonding cooperativity and its effect on water structure. *J. Phys. Chem. B*, 109:24152–24159, 2005.
- [42] W. L. DeLano. *The PyMOL User's Manual*. DeLano Scientific, Palo Alto, CA, 2002.
- [43] Mari L DeMarco and Robert J Woods. Structural glycobiology: a game of snakes and ladders. *Glycobiology*, 18:426–40, 2008.
- [44] K Drickamer and M E Taylor. Biology of animal lectins. *Annu. Rev. Cell Biol.*, 9:237–64, 1993.
- [45] Danielle H Dube and Carolyn R Bertozzi. Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discovery*, 4:477–88, 2005.
- [46] R Dwek. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.*, 96:683–720, 1996.
- [47] S Engelsen and S Pérez. The hydration of sucrose. *Carbohydr. Res.*, 292:21–38, 1996.
- [48] S B Engelsen, C Monteiro, C Hervé de Penhoat, and S Pérez. The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and nmr spectroscopy. *Biophys. Chem.*, 93:103–27, 2001.

- [49] Narayanan Eswar, David Eramian, Ben Webb, Min-Yi Shen, and Andrej Sali. Protein structure modeling with modeller. *Methods Mol. Biol.*, 426:145–59, 2008.
- [50] Narayanan Eswar, Ben Webb, Marc A Marti-Renom, M S Madhusudhan, David Eramian, Min-Yi Shen, Ursula Pieper, and Andrej Sali. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinformatics*, Chapter 5:Unit 5.6, Oct 2006.
- [51] A. D. French. Conformational accessibility of some simple polyglucosides. *ACS Symposium Series*, 260:43–59, 1984.
- [52] A. D. French and V. G. Murphy. The effects of changes in ring geometry on computer models of amylose. *Carbohydr. Res.*, 27(2), 1973.
- [53] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*. Academic Press, San Diego, California 92101, USA, 2 edition, 2002.
- [54] Raimund Fromme, Zivile Katiliene, Petra Fromme, and Giovanna Ghirlanda. Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 resolution. *Protein Sci.*, 17(5):939–944, 2008. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 2008:553563(Journal) 10.1110/ps.083472808.
- [55] Raimund Fromme, Zivile Katiliene, Barbara Giomarelli, Federica Bogani, James Mc Mahon, Toshiyuki Mori, Petra Fromme, and Giovanna Ghirlanda. A monovalent mutant of cyanovirin-n provides insight into the role of multiple interactions with gp120 for antiviral activity,. *Biochemistry*, 46(32):9199–9207, 2007. doi: 10.1021/bi700666m.
- [56] C Fuhrer, I Geffen, K Huggel, and M Spiess. The two subunits of the asialoglycoprotein receptor contain different sorting information. *J. Biol. Chem.*, 269(5):3277–82, Feb 1994.
- [57] Y. K. Fujimoto and D. F. Green. Carbohydrate recognition by the antiviral lectin cyanovirin-n. *J. Am. Chem. Soc.*, 134:19639–19651, 2012.
- [58] Y. K. Fujimoto, R. N. Terbush, V. Patsalo, and D. F. Green. Computational models explain the oligosaccharide specificity of cyanovirin-n. *Protein Sci.*, 17(11):2008–14, 2008. Fujimoto, Yukiji K Terbush, Ryan N Patsalo, Vadim Green, David F Research Support, Non-U.S. Gov’t United States Protein science : a publication of the Protein Society *Protein Sci.* 2008 Nov;17(11):2008-14. Epub 2008 Sep 22.

- [59] A E García and K Y Sanbonmatsu. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins*, 42(3):345–54, Feb 2001.
- [60] Alan Grossfield. *WHAM: the weighted histogram analysis method*, volume 2.0.6. University of Rochester.
- [61] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindhal. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.
- [62] M Heyden, E Bründermann, U Heugen, G Niehues, D M Leitner, and M Havenith. Long-range influence of carbohydrates on the solvation dynamics of water—answers from terahertz absorption measurements and molecular modeling simulations. *J. Am. Chem. Soc.*, 130:5773–9, 2008.
- [63] WG Hoover. Canonical dynamics - equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.
- [64] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [65] J. S. Hub, B. L. de Groot, and D. v. d. Spoel. g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.*, 6(12):3713–3720, 2010.
- [66] Koji Hukushima and Koji Nemoto. Exchange monte carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.*, 65(6):1604–1608, Jun 1996.
- [67] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [68] A Imberty and S Pérez. Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem. Rev.*, 100:4567–88, 2000.
- [69] R. W. Impey, P. A. Madden, and I. R. McDonald. Hydration and mobility of ions in solution. *J. Phys. Chem.*, 87:5071–5083, 1983.
- [70] Robert R Johnson, Axel Kohlmeyer, A T Charlie Johnson, and Michael L Klein. Free energy landscape of a dna-carbon nanotube hybrid using replica exchange molecular dynamics. *Nano Lett*, 9(2):537–41, Feb 2009.

- [71] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffery D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926–935, 1983.
- [72] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of American Chemical Society*, 118(45):11225–11236, 1996.
- [73] G. A. Kaminski, Friesner, R. A., J. Tirado-Rives, and William L. Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*, 105(28):64746487, 2001.
- [74] Brendan S. Kelley, Leng Chee Chang, and Carole A. Bewley. Engineering an obligate domain-swapped dimer of cyanovirin-n with enhanced anti-hiv activity. *Journal of the American Chemical Society*, 124(13):3210–3211, 2002. doi: 10.1021/ja025537m.
- [75] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.
- [76] Karl N. Kirschner, Austin B. Yongye, Jorge Tschampel, Sarah M. Gonzalez-Outeirino, Charlisa R. Daniels, Lachele B. Foley, and Robert J. Woods. Glycam06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry*, 29(4):1–34, 2008. doi: 10.1002/jcc.20820.
- [77] KN Kirschner and RJ Woods. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.*, 98:10541–10545, 2001.
- [78] P. I. Kitov, G. Sadowska, J. M. Mulvey, G. D. Armstrong, H. Ling, N. S. Pannu, R. J. Read, and D. R. Bundle. Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature*, 403:669–672, 2000.
- [79] Leonardus M. I. Koharudin and Angela M. Gronenborn. Sweet entanglementsprotein: Glycan interactions in two hiv-inactivating lectin families. *Biopolymers*, 99(3):196–202, 2012.
- [80] A R Kolatkar and W I Weis. Structural basis of galactose recognition by c-type animal lectins. *J. Biol. Chem.*, 271(12):6679–85, Mar 1996.
- [81] P. A. Kollman. The evolutionarily conserved family of cyanovirin-n homologs: Structures and carbohydrate specificity. *Structure*, 93:23952417, 1993.

- [82] E Krieger, G Koraimann, and G Vriend. Increasing the precision of comparative models with yasara nova - a self-parameterizing force field. *Proteins*, 47(3):393–402, May 2002.
- [83] S Kumar, J. M. Rosenberg, D Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comp. Chem.*, 8(13):1011–1021, 1992.
- [84] J. E. Ladbury. Just add water!: the effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.*, 3:973–980, 1996.
- [85] Christopher Lee and S Subbiah. Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.*, 217(2):373–388, 1991.
- [86] Sau Lawrence Lee, Pablo G Debenedetti, and Jeffrey R Errington. A computational study of hydration, solution structure, and dynamics in dilute carbohydrate solutions. *J. Chem. Phys.*, 122:204511, 2005.
- [87] Y. C. Lee, R. R. Townsend, M. R. Hardy, J. Lonngren, J. Arnarp, M. Haraldsson, and H. Lonn. Binding of synthetic oligosaccharides to the hepatic gal galnac lectin - dependence on fine-structural features. *Journal of Biological Chemistry*, 258(1):199–202, 1983.
- [88] R. U. Lemieux. How water provides the impetus for molecular recognition in aqueous solution. *Accounts of Chemical Research*, 29:373–380, 1996.
- [89] B Leroux, H Bizot, J Brady, and V Tran. Water structuring around complex solutes: theoretical modeling of -d-glucopyranose. *Chem. Phys.*, 216:349–363, 1997.
- [90] Lin Liu, In-Ja L. Byeon, Ivet Bahar, and Angela M. Gronenborn. Domain swapping proceeds via complete unfolding: A 19f- and 1h-nmr study of the cyanovirin-n protein. *Journal of the American Chemical Society*, 134(9):4229–4235, 2012. doi: 10.1021/ja210118w.
- [91] Q Liu, R Schmidt, B Teo, P Karplus, and J Brady. Molecular dynamics studies of the hydration of , α , α -trehalose. *J. Am. Chem. Soc.*, 119:7851–7862, 1997.
- [92] H F Lodish. Recognition of complex oligosaccharides by the multi-subunit asialoglycoprotein receptor. *Trends Biochem. Sci.*, 16(10):374–7, Oct 1991.

- [93] C. J. Margulis. Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-hiv virucidal protein cyanovirin. *The Journal of Physical Chemistry B*, 109(8):3639–3647, 2005. doi: 10.1021/jp0406971.
- [94] Elena Matei, Andrew Zheng, William Furey, Jeremy Rose, Christopher Aiken, and Angela M. Gronenborn. Anti-hiv activity of defective cyanovirin-n mutants is restored by dimerization. *Journal of Biological Chemistry*, 285(17):13057–13065, 2010.
- [95] M McPhaul and P Berg. Formation of functional asialoglycoprotein receptor after transfection with cdnas encoding the receptor proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 83(23):8863–7, Dec 1986.
- [96] M Meier, M D Bider, V N Malashkevich, M Spiess, and P Burkhard. Crystal structure of the carbohydrate recognition domain of the h1 subunit of the asialoglycoprotein receptor. *J. Mol. Biol.*, 300(4):857–65, Jul 2000.
- [97] Sindhu Menon, Kenneth Rosenberg, Sarah A Graham, Eliot M Ward, Maureen E Taylor, Kurt Drickamer, and Deborah E Leckband. Binding-site geometry and flexibility in dc-sign demonstrated with surface force measurements. *Proc. Natl. Acad. Sci. U. S. A.*, 106(28):11524–9, Jul 2009.
- [98] A Mitsutake, Y Sugita, and Y Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, 60(2):96–123, 2001.
- [99] Chakrabarti Nilmadhab, Chris Neale, Jian Payandeh, Oaum Emil F., and Pomès Régis. An iris-like mechanism of pore dilation in the cora magnesium transport system. *Biphsical Journal*, 98(5):784–792, 2010.
- [100] S Nose. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.
- [101] K. Ogawa, K. Okamura, and A. Sarko. Molecular and crystal structure of the regenerated form of (1→3) α -d-glucan. *Int. J. Biol. Macromol.*, 3:31–36, 1981.
- [102] T Okabe, M Kawata, Y Okamoto, and M Mikami. Replica-exchange monte carlo method for the isobaric-isothermal ensemble. *Chem. Phys. Lett.*, 335(5-6):435–439, March 2001.
- [103] M. K. O’Reilly, B. E. Collins, S. Han, L. Liao, C. Rillahan, P. I. Kitov, D. R. Bundle, and J. C. Paulson. Bifunctional cd22 ligands use multimeric immunoglobulins as protein scaffolds in assembly of immune complexes on b cells. *J. Am. Chem. Soc.*, 130(24):7736–7745, 2008.

- [104] Marco Paolantoni, Paola Sassi, Assunta Morresi, and Sergio Santini. Hydrogen bond dynamics and water structure in glucose-water solutions by depolarized rayleigh scattering and low-frequency raman spectroscopy. *J. Chem. Phys.*, 127:024504, 2007.
- [105] M Parrinello and A Rahman. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.
- [106] G. N. Patey and J. P. Valleau. The free energy of spheres with dipoles: Monte carlo with multistage sampling. *Chem. Phys. Lett*, 21(2):297–300, 1973.
- [107] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, 10(15):2073–2077, 2008.
- [108] D. A. Pearlman. A comparison of alternative approaches to free energy calculations. *J. Phys. Chem.*, 98(5):1487–1493, 1994.
- [109] S. Pérez, A. Imberty, and R. P. Scaringe. Modeling of interactions of polysaccharide chains. *ACS Symposium Series*, 430:281–299, 1990.
- [110] EF Pettersen, TD Goddard, CC Huang, GS Couch, DM Greenblatt, EC Meng, and TE Ferrin. Ucsf chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25:1605–1612, 2004.
- [111] Shirts M. R. and Pande V. S. Solvation free energies of amino acid side chains for common molecular mechanics water models. *J. Chem. Phys.*, 122(13):134508–134520, 2005.
- [112] S. K. Ramadugu, Chung, Y-H, J Xia, and C. J. Margulis. When sugars get wet. a comprehensive study of the behavior of water on the surface of oligosaccharides. *J. Phys. Chem. B*, 113(31):11003–11015, Jul 2009.
- [113] V. S. R. Rao, P. K. Qasba, P. V. Balaji, and R. Chandrashekar. *Conformation of Carbohydrates*. Harwood Academic Publishers, Amsterdam, The Netherlands, 1998.
- [114] S Raymond, B Henrissat, D T Qui, A Kvick, and H Chanzy. The crystal structure of methyl -cellotrioxide monohydrate 0.25 ethanolate and its relationship to cellulose ii. *Carbohydr. Res.*, 277:209–229, 1995.
- [115] D. A. Rees. *Polysaccharide Shapes*. Chapman and Hall, London, United Kingdom, 1977.

- [116] K G Rice and Y C Lee. Oligosaccharide valency and conformations in determining binding to the asialoglycoprotein receptor of rat hepatocytes. *Adv. Enzymol.*, 66:41–83, 1993.
- [117] KG Rice and YC Lee. Modification of triantennary glycopeptide into probes for the asialoglycoprotein receptor of hepatocytes. *J. Biol. Chem.*, 265(30):18423, Oct 1990.
- [118] KG Rice, OA Weisz, T Barthel, RT Lee, and YC Lee. Defined geometry of binding between triantennary glycopeptide and the asialoglycoprotein receptor of rat hepatocytes. *J. Biol. Chem.*, 265(30):18429, Oct 1990.
- [119] KG Rice, P Wu, L Brand, and YC Lee. Interterminal distance and flexibility of a triantennary glycopeptide as measured by resonance energy transfer. *Biochemistry*, 30(27):6646–6655, 1991.
- [120] Corine Sandstorm, Olivier Berteau, Emiliano Gemma, Stefan Oscarson, Lennart Kenne, and Angela M. Gronenborn. Atomic mapping of the interactions between the antiviral agent cyanovirin-n and oligomannosides by saturation-transfer difference nmr. *Biochemistry*, 43(44):13926–13931, 2004. doi: 10.1021/bi048676k.
- [121] B. K. Sathyanaraya and V. S. R. Rao. Conformational studies of β -glucans. *Biopolymers*, 10:1605–1615, 1971.
- [122] B. K. Sathyanaraya and V. S. R. Rao. Conformational studies of α -glucans. *Biopolymers*, 11:1379–1394, 1972.
- [123] Christopher N. Scanlan, John Offer, Nicole Zitzmann, and Raymond A. Dwek. Exploiting the defensive sugars of hiv-1 for drug and vaccine design. *Nature*, 446:1038–1045, 2007.
- [124] M Marvin Seibert, Alexandra Patriksson, Berk Hess, and David Van Der Spoel. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J. Mol. Biol.*, 354(1):173–83, Nov 2005.
- [125] S Sheriff, C Y Chang, and R A Ezekowitz. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat. Struct. Biol.*, 1(11):789–94, Nov 1994.
- [126] M A Shia and H F Lodish. The two subunits of the human asialoglycoprotein receptor have different fates when expressed alone in fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.*, 86(4):1158–62, Feb 1989.

- [127] JP Simons, RA Jockusch, P Carcabal, I Hung, RT Kroemer, NA Macleod, and LC Snoek. Sugars in the gas phase. spectroscopy, conformation, hydration, co-operativity and selectivity. *Intl. Rev. Phys. Chem.*, 24:489–531, 2005.
- [128] M Spiess. The asialoglycoprotein receptor: a model for endocytic transport receptors. *Biochemistry*, 29(43):10009–18, Oct 1990.
- [129] M Spiess and H F Lodish. Sequence of a second human asialoglycoprotein receptor: conservation of two receptor genes during evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 82(19):6465–9, Oct 1985.
- [130] T. P. Straatsma and J. A. McCammon. Computational alchemy. *Annual Review of Physical Chemistry*, 43:407–435, 1992.
- [131] Y Sugita and Y Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, November 1999.
- [132] Mannargudi Sujatha, Yellamraju Sasidhar, and Petety Balaji. Energetics of galactose- and glucose-aromatic amino acid interactions: Implications for binding in galactose-specific proteins. *Protein Sci.*, 13:2502, 2004.
- [133] P. R. Sundararajan and V. S. R. Rao. Conformational studies of linear β -d-1,4 mannan and galactan. *Biopolymers*, 9:1239–1247, 1970.
- [134] Steinbrecher T., Mobley D. L., and Case D. A. Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *Journal of Chemical Physics*, 127(21):214108–214120, 2007.
- [135] E J Toone. Structure and energetics of protein carbohydrate complexes. *Curr. Opin. Struc. Biol.*, 4:719–728, 1994.
- [136] G. M. Torrie and J. P. Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chem. Phys. Lett*, 28(4):578–581, 1975.
- [137] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187199, 1977.
- [138] R. R. Townsend, M. R. Hardy, T. C. Wong, and Y. C. Lee. Binding of n-linked bovine fetuin glycopeptides to isolated rabbit hepatocytes - gal/galnac hepatic lectin discrimination between galbeta(1,4)glnac and galbeta(1,3)glnac in a triantennary structure. *Biochemistry*, 25(19):5716–5725, 1986.

- [139] S.M Tschampel and R.J Woods. Quantifying the role of water in protein-carbohydrate interactions. *J. Phys. Chem. A*, 107:9175–9181, 2003.
- [140] M Umemura, S Hayashi, T Nakagawa, H Urakawa, and K kajiwara. Structure of water molecules in aqueous solutions of di-and penta-d-glucopyranoses using molecular dynamics simulation. *J. Mol. Struct.*, 639:69–86, 2003.
- [141] Ajit Varki. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3:97, 1993.
- [142] K Veluraja and C J Margulis. Conformational dynamics of sialyl lewisx in aqueous solution and its interaction with selectine. a study by molecular dynamics. *J. Biomol. Struct. Dyn.*, 23:101–112, 2005.
- [143] Ivan I. Vorontsov and Osamu Miyashita. Solution and crystal molecular dynamics simulation study of m4-cyanovirin-n mutants complexed with di-mannose. *Biophys. J.*, 97(9):2532–2540, 2009. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 2009:1440415(Journal) 10.1016/j.bpj.2009.08.011.
- [144] Ivan I. Vorontsov and Osamu Miyashita. Crystal molecular dynamics simulations to speed up mm/pb(gb)sa evaluation of binding free energies of di-mannose deoxy analogs with p51g-m4-cyanovirin-n. *J. Comput. Chem.*, 32(6):1043–1053, 2011. Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved. CAPLUS AN 2011:285029(Journal) 10.1002/jcc.21683.
- [145] W I Weis and K Drickamer. Trimeric structure of a c-type mannose-binding protein. *Structure*, 2(12):1227–40, Dec 1994.
- [146] Mark R. Wormald, Andrei J. Petrescu, Pao Ya-Lan, Ann Glithero, Tim Elliott, and Raymond A. Dwek. Conformational studies of oligosaccharides and glycopeptides: Complementarity of nmr, x-ray crystallography, and molecular modelling. *Chemical Reviews*, 102(2):371–386, 2002.
- [147] P G Wu, K G Rice, L Brand, and Y C Lee. Differential flexibilities in three branches of an n-linked triantennary glycopeptide. *Proc. Natl. Acad. Sci. U. S. A.*, 88(20):9355–9, Oct 1991.
- [148] J Xia, R Daly, F Chuang, L Parker, and J Jensen. Sugar folding: A novel structural prediction tool for oligosaccharides and polysaccharides 1. *J. Chem. Theory Comput*, 3:1620–1628, 2007.

- [149] J Xia, R Daly, F Chuang, L Parker, and J Jensen. Sugar folding: A novel structural prediction tool for oligosaccharides and polysaccharides 2. *J. Chem. Theory Comput*, 3:1629–1642, 2007.
- [150] Junchao Xia and Claudio Margulis. A tool for the prediction of structures of complex sugars. *J. Biomol. NMR*, 42(4):241–56, Dec 2008.
- [151] Junchao Xia and Claudio Margulis. Computational study of the conformational structures of saccharides in solution based on j couplings and the fast sugar structure prediction software. *Biomacromolecules*, 10(11):3081–3088, 2009.
- [152] Fan Yang, Carole A. Bewley, John M. Louis, Kirk R. Gustafson, Michael R. Boyd, Angela M. Gronenborn, G. Marius Clore, and Alexander Wlodawer. Crystal structure of cyanovirin-n, a potent hiv-inactivating protein, shows unexpected domain swapping. *Journal of Molecular Biology*, 288(3):403–412, 1999. doi: 10.1006/jmbi.1999.2693.
- [153] Min yi Shen and Andrej Sali. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 15(11):2507–2524, November 2006.
- [154] T Yui, K Ogawa, and A Sarko. Molecular and crystal structure of konjac glucomannan in the mannan ii polymorphic form. *Carbohydr. Res.*, 229:41–55, 1992.
- [155] Natasha E. Zachara and Gerald W. Hart. The emerging significance of o-glcnaC in cellular regulation. *Chemical Reviews*, 102(2):371–386, 2002.
- [156] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *Journal of Chemical Physics*, 22(8):1420–1426, 1954.