Stony Brook University

Using Computational Models to Understand Specific Carbohydrate Recognition by Cyanovirin-N

A Thesis Presented

by

Yukiji Karen Fujimoto

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Chemistry

Stony Brook University

August 2009

Stony Brook University

The Graduate School

Yukiji Karen Fujimoto

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

David F. Green – Thesis Advisor
Assistant Professor of Department of Applied Mathematics and Statistics

Carlos Simmerling – Chairperson of Defense
Professor of Department of Chemistry

Isaac Carrico – Third Member
Assistant Professor of Department of Chemistry

This thesis is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Thesis

Using Computational Models to Understand Specific Carbohydrate Recognition by Cyanovirin-N

by

Yukiji Karen Fujimoto

Master of Science

in

Chemistry

Stony Brook University

2009

As an initial step in HIV infection of T-cells, the viral envelope glycoprotein gp120 must bind to CD4 receptors on the target cell surface. Over the past several years, proteins isolated from various prokaryotes have been shown to inhibit HIV cell entry by binding to gp120 and thus blocking the association with CD4. Unlike the majority of inhibitors, these proteins bind to the carbohydrate moiety of the glycoprotein, not the protein component. The binding of one of these proteins, cyanovirin-N, to a range of carbohydrate targets has been well-characterized, both structurally and thermodynamically, and thus provides an ideal system for the development of computational methods that address the problem of protein-carbohydrate binding. Using a range of computational approaches, we have dissected the energetic contribution of each protein and carbohydrate functional group to the binding affinity. Studies have shown cyanovirin-N contains two symmetry–related carbohydrate–binding sites with different affinities that recognize the high mannose oligosaccharides, and that the minimum structure required for high-affinity binding is the disaccharide mannose-α(1-2)-mannose. Our calculations begin with analyzing the disaccharide and understanding how it is bound to the binding sites of cyanovirin-N. We continue to build up the sugar one monomer at a time and analyze each conformation that is built. Understanding how diverse carbohydrate-binding proteins bind their carbohydrate ligands with high affinity can expand our knowledge of specific carbohydrate recognition.

Table of Contents

Acknowledgments

First I would like to thank my advisor Dr. David Green for his intellectual support, patience, and continual encouragement through my studies. I appreciate the opportunity to present results of my work at various conferences, such as Protein Society, Protein Design and Engineering (sponsored by the New York Academy of Sciences), annual meetings sponsored by NIGMS, and many others.

I would also like to thank Dr. Bruce Tidor from MIT for making available the MultigridPBE software. I would like to mention my deepest thanks to my other research advisors from years past: Dr. Mali Yin, and Dr. Adele L. Boskey. Professor Yin served as a good mentor and gave me the encouragement to continue my interest in science. I want to express my gratitude to Dr. Boskey for giving me the opportunity to join her lab before I entered graduate school. She gave me the first of many advices and recommendations for starting my graduate study.

Next, I would like to thank all the current and past members of the Green lab. In particular, I would like to thank Vadim Patsalo and Jonathan Cheng. It was a pleasure to share the first office the Green lab had. We shared not only nice discussions during work, but also nice times during conferences. Ryan Terbush also deserves a special acknowledgment that began the work contained in this thesis. I am also grateful for all the support from Noel Carrascal, Tao Jiang, Yulin Huang, and the rest of Dr. Robert Rizzo's lab.

Lastly, I want to thank my family for encouraging me to pursue my goals. I especially wish to express my thanks to Chris Marraccino for his endless support, love, patience, and understanding.

Chapter 1

Introductory Background

Aims

This work involves the development of comprehensive set of computational methods for analysis and design in glycobiology, and the application of these tools to the glycobiology of HIV-1 infection. The long-term aim is to understand the role of carbohydrates on the Env glycoproteins both in recognition of cellular targets and in recognition by the immune system. It is shown how the continuum electrostatic models were used and how molecular dynamics simulations allowed us to dissect the protein-carbohydrate interactions. Using this method allows the separation of the individual contributions of various parts of the complex with their energies, which is inaccessible by experiments.

1.1 An introduction to protein-carbohydrate interactions

Protein-carbohydrate interactions play an important role in various cellular processes, including viral and microbial pathogenesis, inflammation, and fertilization, as

shown in Figure 1. These interactions play a key role in a variety of cell adhesion events, including the binding of parasites, fungi, bacteria, and viruses to their host cells, which is the initial step in infection [14]. One can think of this type of binding as the "lock and key" analogy seen with an enzyme and a substrate, where the "lock" is the enzyme and the "key" is the substrate. In protein-carbohydrate interactions the "key" represents the complex carbohydrates and the "lock" symbolizes the carbohydrate binding proteins, also known as lectins [39]. Complex carbohydrates are commonly found on the cell surface and interact with lectins in solution or on the surfaces of other cells. There are many various types of lectins that differ in size and structure [38].

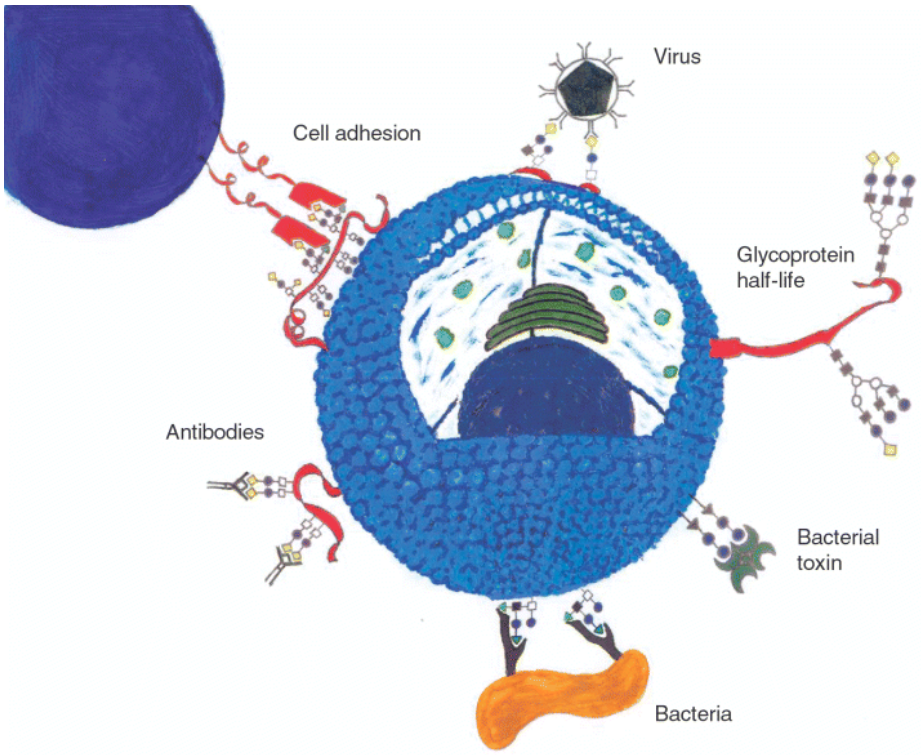

Figure 1: **Schematic diagram portraying protein-carbohydrate interactions at the cell surface.** Red ribbons indicate that the sugar chains can be linked to proteins or anchored in the plasma membrane via a lipid. [19]

Information about lectins and their ligands can come from an assortment of sources, including X-ray crystallography, binding experiments with sugars, site-directed mutagenesis, NMR experiments, and molecular modeling. Currently, over 300 complexes of lectins with carbohydrates [10] have been solved mainly by X-ray crystallography; most of these are from bacterial or viral sources. Table 1 shows several examples of these types of complexes.

Source and name of lectin	Ligand	Resolution (Å)	PDB code
Clostridium botulinum Botulinum neurotoxin B	NeuAcα2,3Galβ4Glc	2.60	1F31
Clostridium tetani Tetanus toxin	Galβ4Glc	1.80	1DLL
Escherichia coli FimH	Man	2.79	1KLF
Influenza virus Hemagglutinin	NeuAcα2,3βGalβ3-GlcNAcβ3Galβ4Glc	2.50	IRV0
Ralstonia solanacearum RS-IIL	Man	1.70	1UQX
Staphylococcus aureus Enterotoxin B	Galβ4Glc	1.90	1SE4
Vibrio cholerae Cholera toxin	NPαGal	2.00	1EEI

Table 1: **Examples of three-dimensional complexes of bacterial and viral lectins**. The source and name of the lectin, ligand, structure resolution, and PDB codes are listed [10].

Experimental studies have shown that the binding regions of carbohydrate–lectin complexes are mostly in the form of shallow clefts on the surface of the protein, where typically one or two segments of the ligand are bound [39]. Lectins can interact by way of hydrogen bonds, hydrophobic, electrostatic, and water-mediated interactions [30]. Hydrogen bonds are involved in providing affinity and specificity to protein–

carbohydrate interactions [36]. They depend largely on interactions involving the hydroxyls of the carbohydrate. A sugar hydroxyl has the ability to interact with a protein both as a hydrogen bond donor and as an acceptor. As a hydrogen bond donor, the hydroxyl has rotational freedom around the C–OH torsional angle, and thus can often attain a strong linear bond with an acceptor group. When each of two adjacent hydroxyls of a monosaccharide interact with different atoms of the same amino acid (*e.g.* two oxygens from the carboxylate of glutamic or aspartic acid), it forms bidentate hydrogen bonds [36]. Even though carbohydrates are highly polar molecules, the position of the hydroxyl groups can create hydrophobic regions on their surfaces, which can form contacts with hydrophobic side chains of the protein [6]. Sometimes contacts between the protein and ligand are mediated by water bridges [31]. These water–mediated interactions can consist of a single water molecule or a chain of many water molecules which may be important for ligand recognition.

Chapter 2

Biological context: Human Immunodeficiency Virus (HIV)

2.1 History and Epidemiology

One leading example of a protein-carbohydrate interaction comes from the human immunodeficiency virus, also known as HIV. The HIV infection is one of the most destructive epidemics in history. Usually after several years, a patient who has been infected with HIV is likely to become ill more often as the number of immune system cells left in their body drops below a particular point; the patient is said to have the acquired immunodeficiency syndrome (AIDS). HIV is a member of the retrovirus family, and it is different from most other viruses because it attacks the immune system. HIV finds and destroys a type of white blood cell called T cells that the immune system must have to fight disease. HIV is believed to have originated in primates in sub-Saharan Africa and transferred to humans early in the 20th century [16]. As of December 2007, the Joint United Nations Program on HIV/AIDS and the World Health Organization

estimated 30 to 36 million people worldwide were living with HIV. Of those, approximately 2.2 to 3.2 million people became newly infected and 1.8 to 2.3 million lost their lives to AIDS [45]. In 2007, The World Health Organization has estimated 1.2 million adults (15 and over) and children in the United States were infected with HIV [45].

2.2 Transmission

HIV is a virus that is sensitive to its environment, and cannot live very long outside of the body. As a result, the virus cannot be transmitted through everyday activities (*e.g.* shaking hands, hugging, drinking from a water fountain, etc.). HIV is primarily found in blood, semen, and vaginal fluid from an infected person. HIV infection has been identified to occur via three main transmission routes—unprotected sexual relations with an infected person, sharing needles and/or syringes with an infected person, or being exposed to HIV from mother to child *in utereo* or during breast feeding as an infant.

2.3 Structure

HIV belongs to the class of retroviruses, which are viruses that contain RNA (ribonucleic acid) as their genetic material. Within the retrovirus family, HIV belongs to a subgroup known as lentiviruses, which are “slow" viruses due to their characteristics of having a long period of time between initial infection and the beginning of serious symptoms. This is why many people who are unaware of being infected with HIV can unfortunately spread the virus to others. Outside a human cell, HIV is roughly spherical

with a diameter of about 120nm, which is smaller than a red blood cell, but large for a virus. It consists of two basic components: the viral envelope and the viral core [34], as shown in Figure 1. The viral envelope is the outer layer of the virus consisting of a lipid bilayer. Embedded around the viral envelope are proteins or "spikes", known as Env. The Env consists of a cap made of three molecules called glycoprotein 120 (gp120), and a stem consisting of three molecules called glycoprotein 41 (gp41) that anchors the structure in the viral envelope. Information about the envelope spike structure is limited, and it is not clear how many spikes there are on each virus or how they are distributed [9]. The second component — the viral core — is found within the viral envelope. It is a bullet-like shaped structure, also known as a capsid, which is made up of 2,000 copies of the viral protein p24. The capsid surrounds two single strands of HIV RNA, each of which has a complete copy of the viral genome. HIV has three structural genes (gag, pol, and env) that contain information needed to make structural proteins for new virus particles. The env gene codes for a protein called gp160 that is broken down by a viral enzyme to form gp120 and gp41. HIV has six regulatory genes (tat, rev, nef, vif, vpr, and vpu) that contain information needed to produce proteins that control the ability of HIV to infect a cell, produce new copies of virus, or cause disease. The protein encoded by nef is necessary for the virus to replicate efficiently, the vpu-encoded protein influences the release of new virus particles from infected cells, and vif interacts with an antiviral defense protein in host cells, causing inactivation of the antiviral effect and enhancing HIV replication. This interaction may serve as a new target for antiviral drugs. The ends

of each strand of HIV RNA contain an RNA sequence called the long terminal repeat (LTR). Regions in the LTR act as switches to control production of new viruses and can be triggered by proteins from either HIV or the host cell. The viral core also includes a protein called p7, the HIV nucleocapsid protein [34]. Three enzymes carry out later steps in the viral life cycle: reverse transcriptase, integrase, and protease. Reverse transcriptase reads the sequence of viral RNA that has entered the host cell and transcribes the sequence into a complementary DNA sequence. Without it, the viral genome could not become incorporated into the host cell, and could not reproduce [34]. Afterwards, the DNA can be inserted into the DNA of the lymphocyte. The virus has its own enzyme called integrase that facilitates incorporation of the viral DNA into the host cells DNA. The integrated DNA is called a provirus [34].

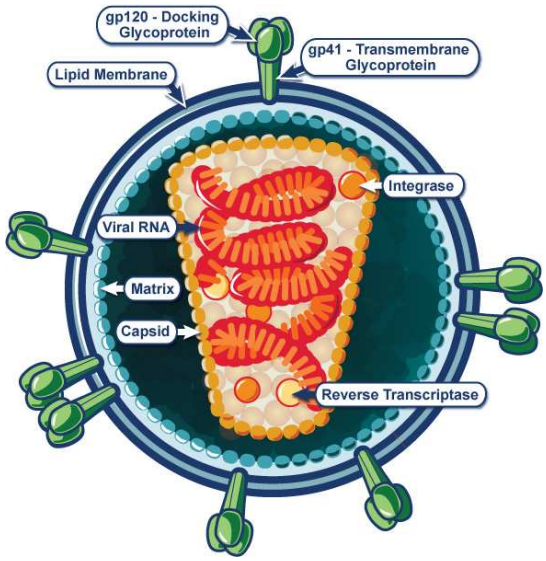

Figure 1: **Schematic image of the HIV virion structure.** The viral envelope and viral core are shown. [33]

2.4 Replication cycle

Cell membranes are usually barriers that keep out harmful chemicals and microbes. However, HIV has little trouble penetrating the membrane of its target cells. The HIV life cycle involves multiple steps. HIV begins its life cycle when it binds to a CD4 receptor and one of the two co-receptors (CCR5 or CXCR4 receptor) on the surface of a $CD4^+$ T-lymphocyte. The HIV virus will fuse with its host cell, and after fusion the virus will release RNA into the host cell. Reverse transcriptase converts the single stranded HIV RNA to double stranded HIV DNA. As DNA enters the host cell's nucleus, an enzyme called an integrase “hides” the HIV DNA within the host's own DNA. This is known as a provirus. The provirus may remain inactive for several years, producing few or no copies of HIV. However, when the host cell receives a signal to become active, the shorter strands of messenger RNA (mRNA) form templates to make long chains of HIV proteins. Protease then cuts the long chains of HIV proteins into smaller protein strands. As the smaller HIV proteins come together with copies of RNA genetic material, a new virus particle is assembled. During this time, it buds out from the host cell. The new virus steals part of the cell's outer envelope, which acts as a covering and it is studded with viral glycoproteins. These glycoproteins are necessary for the virus to bind CD4 and its co-receptors, and for fusion of the viral and cellular membranes. The new copies of HIV can now move on to infect other cells [34]. Figure below is a schematic drawing of the replication cycle.

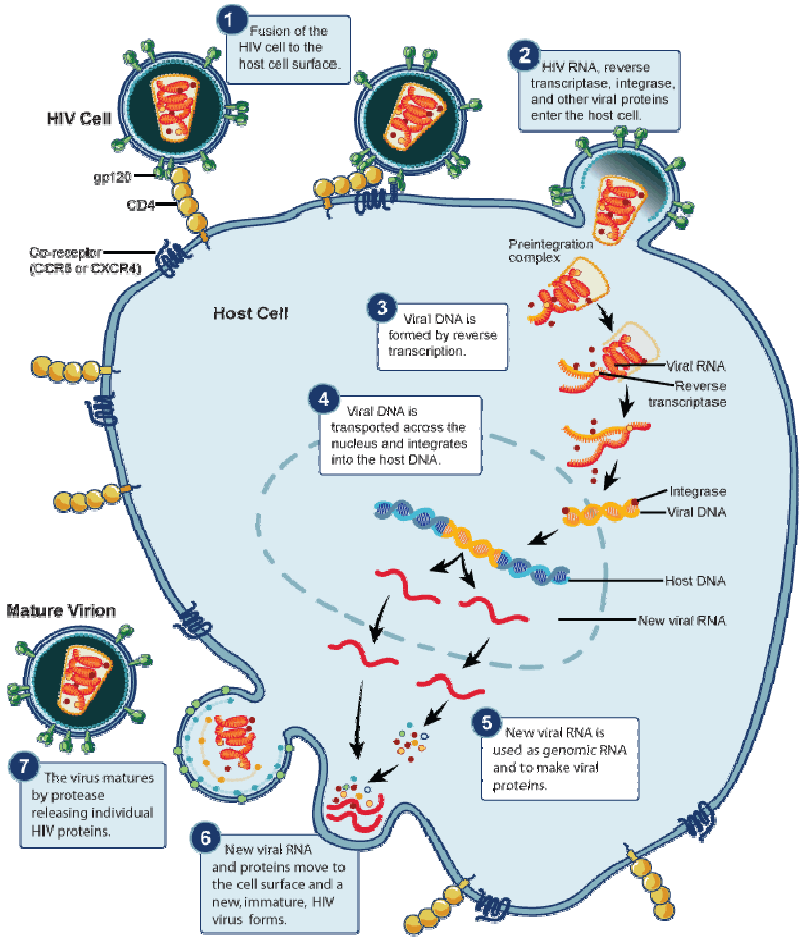

Figure 2: **Steps in the HIV Replication Cycle.** In the first step, there is a fusion of the HIV cell to the host cell surface. Second, the HIV RNA, reverse transcriptase, integrase, and other viral proteins enter the host cell. Third, the viral DNA is formed by reverse transcription; Fourth, the viral DNA is transported across the nucleus and integrates into the host DNA. Next, the new viral RNA is used as genomic RNA and to make viral proteins. This is followed by the new viral RNA and proteins moving into cell surface and a new, immature, HIV virus forms. Lastly, the virus matures by protease releasing individual HIV proteins. [34]

2.5 Carbohydrate-binding proteins inhibit HIV-1 cell entry

Due to the high rate of mutation, HIV is able to optimize its interactions with various host proteins and pathways, thereby multiplying more quickly. The virus ensures

that the host cell survives until the replication cycle is completed, and — possibly even more damaging — may establish a stable latent form that supports the chronic nature of the infection. The complete suppression of the virus appears unlikely until effective methods are developed to purge these latent viral forms. More testing is needed to solve the mysteries of viral latency and replication.

Recently, there has been a growing interest in developing anti-viral microbicides. Microbicides are substances that protect the body from infection by microorganisms such as bacteria, viruses, and fungi. They work by either destroying the microbes or preventing them from establishing an infection. Cyanovirin-N (CVN) is one of the microbicides being studied today. It was originally isolated from cultures of the cyanobacterium (blue green algae), *Nostoc ellipsosporum* [7]. CVN — a HIV−cell fusion blocker — was discovered in a National Cancer Institute (NCI) screening program for natural anti-HIV agents [12, 13]. CVN binds to the sugars attached to the HIV envelope protein, and thus is an inhibitor of all strains of HIV. While CVN is being studied for the prevention of HIV infection, it does not cure HIV or AIDS.

As seen in numerous experiments, it is a known fact that the glycosylation of proteins plays a key role in human health and disease; however, the details of these effects continue to be poorly understood. For that reason, an important area of research is the study of how structural and energetic properties of proteins are affected by carbohydrates. We are focusing on non-bonded protein−carbohydrate interactions, using

the association of virucidal proteins with the oligosaccharides of HIV gp120 as a model system.

Chapter 3

Molecular Modelling Techniques

The range of systems that can be considered in molecular modeling is extremely broad, from simple isolated molecules to polymers and biological macromolecules (proteins and DNA). Computational studies of biological systems can play an important role complementary to experimental studies. These studies can separate individual contributions and energies of various parts of the complex in a way inaccessible to experiments.

In theory, quantum mechanical treatments should be the tool for a reliable description of a complex in a system. However, this is not feasible especially for large macromolecules. A large number of particles must be considered and the calculations are time consuming. To overcome these obstacles, there are two alternative methods that have been particularly successful — molecular mechanics and continuum electrostatics.

3.1 Molecular Mechanics

Molecular mechanics considers molecules as a collection of point particles, each with its own mass and charge, and linked together with classical chemical bonds. The forces acting on each particle are evaluated as combinations of classical bond-stretching, bond angle, bond torsion, and non-bonded interactions, and in this manner the energy of a given atomic configuration is calculated using a molecular mechanics force field. One form of this force field that can be used for a single molecule or an ensemble of molecules is:

$$\nu = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angle} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right)$$
$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

where ν represents the potential energy [27]. The contributions that are listed in the equation are schematically represented in Figure 1.

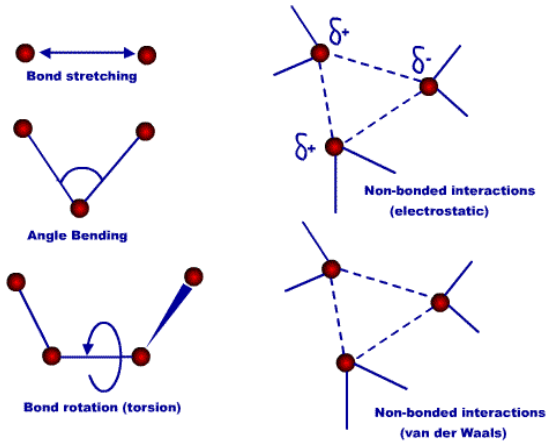

Figure 1. **Schematic representations of four key contributions to molecular mechanics.** These include: bond stretching, angle bending, torsional, and non-bonded interactions. [27]

The first term comes from the interaction between pairs of bonded atoms by a harmonic potential that increases the energy as the bond length, l_i, moves away from the reference, l_{i0}. The second term comes from the angle bending in a molecule between three atoms. The third term considers the bond rotation, and the final term is the non-bonded term. This is a calculation between all pairs of atoms (*i* and *j*) that are in different molecules or that are in the same molecule but separated by three or more bonds. It is usually modeled using a Coulomb potential for electrostatic interactions and a Lennard−Jones potential for van der Waals interactions.

3.2 Molecular Dynamics

To complement experimental data, molecular dynamics can be used to understand more about complex molecular systems. The static view of a molecule from an X-ray crystal or NMR structure is just not enough to fully explain its biological role. Molecules are not frozen; the atoms continuously interact amongst themselves and environment. The motions can explain more about their structure and function. Molecular dynamic simulations involve a number of steps, including: obtaining a three-dimensional structure from X-ray crystallography or NMR data; identifying each atom and residue by assigning each atom with its force field parameters; placing the protein in a pre-equilibrated solvent environment; running molecular dynamics at a certain temperature and density; collecting data from the output files; and analyzing the data over the period of time.

3.2.1 Boundary Conditions

The proper treatment of boundaries is important in simulation methods. Without any boundary conditions, the outermost atoms may not be able to interact with surrounding atoms. To overcome this effect, a method has been used called periodic boundary conditions in which a modeled system is placed in a unit cell that has an infinite number of images. If a molecule leaves the system from one cell it immediately enters at the same velocity at the opposite boundary [27].

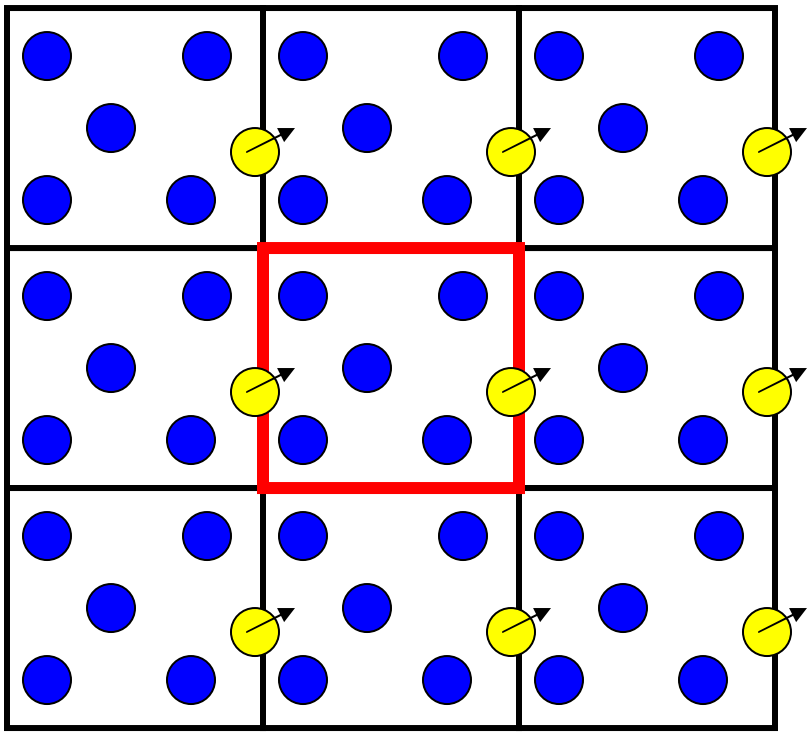

Figure 3. **A representation of periodic boundary conditions.** The primary region (in red square) contains a solute (in yellow circle). The nearest-neighbor repeats this region (in black squares). The solute avoids interacting with its own image by having a cutoff of all non-bonded interactions be smaller than the distance between the edge of the solute and the nearest edge of its nearest image.

3.2.2 Thermodynamic Features

Molecular dynamics is normally performed in various thermodynamic ensembles. The two most common ensembles are the constant *NVT* (constant volume–constant temperature) and the constant *NPT* (constant pressure–constant temperature) ensembles. A constant temperature simulation may be required to compare another system where it changes temperature (*e.g.* unfolding of a protein). Running simulations under constant pressure is important because one can test the effects of any phase transitions with an induced pressure [27].

3.2.3 Water Models

Systems are normally simulated in a setting that represents its biological environment. For many cases, the molecular complex can be simulated in a solvent of water with sodium chloride ions or with a lipid bilayer. Molecular dynamics simulations can be done with explicit solvent molecules or by implicitly accounting for the effects of solvent. There are a variety of models for explicit water including TIP3P, TIP4P, TIP5P, SSD, SPC, and PPC [23]. Figure 2 illustrates various structures these models can adopt.

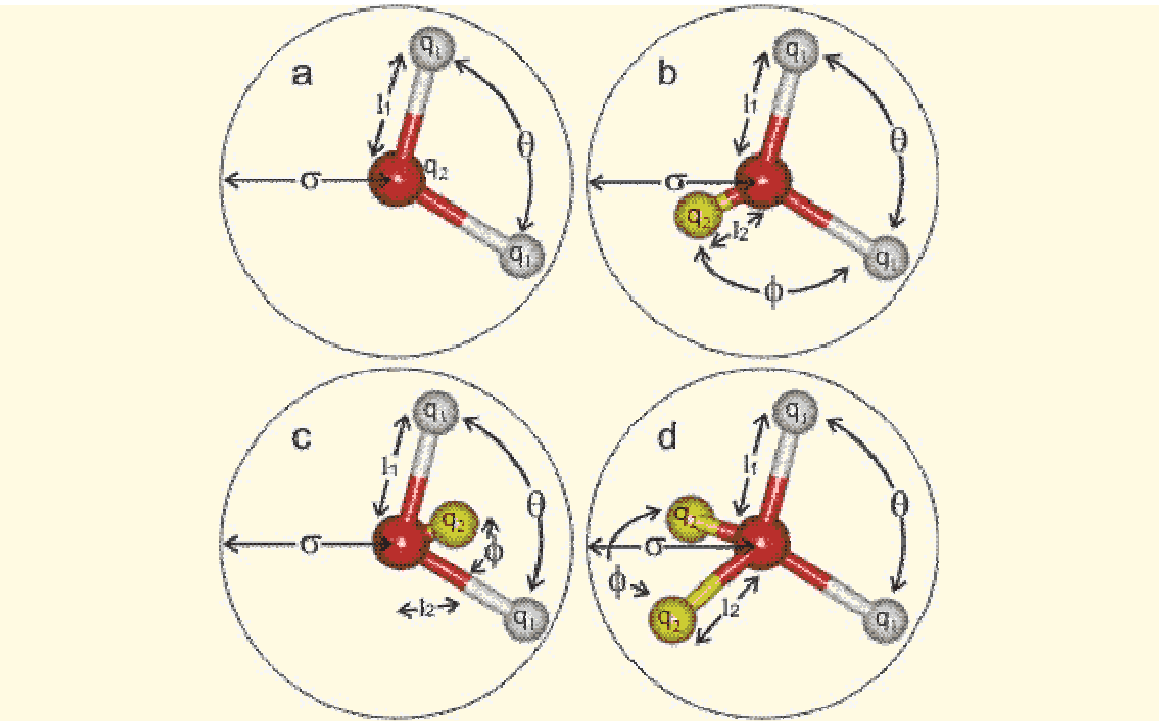

Figure 2. **Representations of simple explicit water models.** Models differ in the quantity, positions, and magnitudes of partial charges used to represent the effects of water. Image extracted from [10]

3.3 The continuum electrostatics model

To treat solvent explicitly by placing the system of interest in a bath of individual solvent molecules is computationally costly because it is expensive to calculate the free energy of solvent. To address this issue, the second method that is applied to biological systems is the continuum electrostatic model [17]. It is used to compute the electrostatic effects in binding of proteins. In this approach, molecules are described as a set of point charges located at the center of the atoms in a region of low dielectric constant and the

solvent treated as a region of high dielectric. The electrostatic potential can then described by the Poisson−Boltzmann equation:

$$\vec{\nabla} \cdot \left[\varepsilon \left(\vec{r} \right) \vec{\nabla} \phi \left(\vec{r} \right) \right] - \varepsilon \left(\vec{r} \right) \cdot \kappa^2 \left(\vec{r} \right) \sinh \left[\phi \left(\vec{r} \right) \right] = -4\pi\rho \left(\vec{r} \right)$$

where the variables of ϕ, ε, κ, and ρ are all functions of the position $\vec{r}$. The ϕ describes the electrostatic potential, ε describes the dielectric constant, ρ describes the charge density, and $\kappa^2 = \frac{8\pi z^2 \mathrm{I}}{ekT}$ describes the effect of mobile ions using the Debye−Hückel model. The electrostatic potential produced by the system can be obtained by the linearized Poisson−Boltzmann equation by replacing the hyperbolic sine with the first term in the series expansion and yields:

$$\vec{\nabla} \cdot \left[\varepsilon \left(\vec{r} \right) \vec{\nabla} \phi \left(\vec{r} \right) \right] - \varepsilon \left(\vec{r} \right) \cdot \kappa^2 \left(\vec{r} \right) \phi \left(\vec{r} \right) = -4\pi\rho \left(\vec{r} \right)$$

By using the continuum electrostatic model, the solvent is considered as a bulk entity. It is reasonably fast in computing the electrostatic energy of a given conformation of the solute. Additionally, since the dielectric constants represents electrostatic properties averaged over many configurations of the solvent, it prevents the need to sample all the solvent configurations.

Since all charges act independently within the linearized form, the contributions to the electrostatic free energy from various parts of the system are separable. Thus, the binding energies can easily be divided into various parts of the molecule or functional group.

Chapter 4

Trajectory Analysis

After a molecular dynamics simulation is completed, an enormous amount of data is generated to be analyzed. Individual pieces of data (i.e. bond energies, atomic positions, temperature, etc) provide little insight to the behavior of the biomolecule. A visual aide (such as Visual Molecular Dynamics (VMD) software) is used to load a molecular dynamics trajectory to visualize the trajectory [21]. Even though visual analysis is important, it is not enough to understand the system. Thermodynamic information can be obtained through mathematical analysis.

4.1 Binding Free Energy Calculation

The binding energy is calculated by the difference in Gibbs Free Energy (*G*) between the bound and unbound states of a receptor–ligand conformation.

$$\Delta G_{binding} = G_{complex} - \left(G_{protein} + G_{ligand}\right)$$

The binding free energy can be broken down into various terms. First, there are van der Waals and electrostatic interactions in the bound state. Additionally, there is a solvation term that is comprised of polar and non-polar terms, each calculated separately:

$$G_{solv} = G_{polar} + G_{nonpolar}$$

It incorporates the van der Waals interaction between solute and solvent, the electrostatic interactions between the solvent and molecules on binding, and the energy cost of creating a cavity within the solvent. The non-polar terms can be approximated from the solvent accessible surface area:

$$G_{nonpolar} = \gamma A + b$$

The values for γ and b comes from experimentally determined free energies [41].

The binding energy calculations done in this work involve a semi-rigid approximation, which means the structures of the proteins do not change significantly upon complex formation, and since we wish to study the energetics of binding essentially due to the contact of protein surfaces rather than to the conformational change of the main chains of the protein, the internal degrees of freedom can be ignored. When this is ignored, the only terms to be considered are the direct Coulombic interactions between the binding partners, the solvent screening of this interaction, and the desolvation of each binding partner.

4.2 Calculation of relative free energies of binding

Calculating free energies has been widely used with researchers working with molecular simulations. However, calculating free energies has been troublesome from

the fact that accurate free energy results can be obtained if the contributions of all the populated states (including configurational, conformational, and vibrational states) are included in the calculation. To deal with such problems, relative free energies of rigid binding energies are calculated for a set of molecules that are computed by comparing the average ensemble energies extracted from bound state molecular dynamics and those of related conformations from the bound state.

4.3 Electrostatic component analysis

In many cases, electrostatic interactions play a vital role in a complex. In order to understand electrostatics in more detail, it is possible to break down the electrostatic interactions from various parts of the molecule. For proteins, every residue is divided into a side chain, backbone carbonyl, and backbone amino group. For carbohydrates, we can define each functional group. Three types of calculations can be determined for every group. One term is the desolvation energy of an individual group, the second term can be defined as the solvent-screened Coulombic interactions between two groups in the bound state (direct interaction), and lastly, the difference in the solvent screening that intramolecular interactions experience in the bound and unbound states (indirect interactions). All three terms can be summed up to give the electrostatic contributions to binding.

Within a group, the total of the desolvation, direct, and indirect terms gives an energy called the mutation free energy. It corresponds to the binding energy difference of the native complex and that of a complex with a specific group of interest that is

substituted by a hydrophobic isostere. The positive or negative values indicate the unfavorable and favorable contributions to the binding affinity. The electrostatic component analysis method is a good way to determine the influences of single residues on the binding affinity and pinpoint hotspot residues. This information can be used as a guide for design.

4.4 Calculation of ligand strain energies

There are multiple conformations a prospective ligand can adopt in a given binding site [17] and in unbound state. The ligand strain energy is the energetic cost to conform the ligand into the binding conformation. It can be computed by comparing the total ensemble-averaged energies of the ligand conformations extracted from the bound state simulations and for the unbound simulations. The energies computed for the ligand included the sum of all bonded terms (bonds, angles, and dihedrals), intramolecular van der Waals, and intramolecular Coulombic interactions, hydrophobic solvation free energies, and electrostatic solvation free energies.

Chapter 5

Implementation and Results – Using restrained solvent model [a]

5.1 Introduction

To date, numerous lectins from diverse sources have been identified as having virucidal activity against HIV, but among the best characterized is cyanovirin-N (CVN). Cyanovirin-N is a 101-amino acid protein that contains two pseudosymmetrical carbohydrate binding sites with different affinities (as shown in Figure 1).

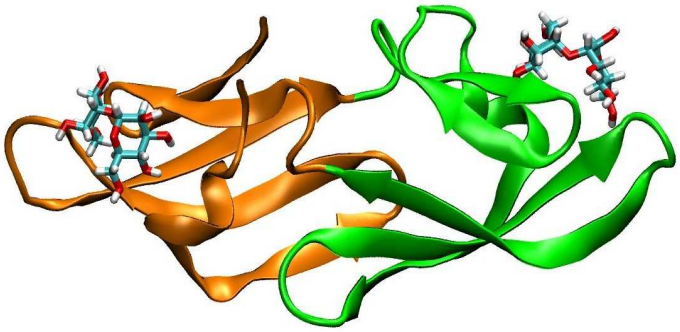

Figure 1. **Ribbon diagram of CVN bound with two dimannoses showing the pseudosymmetry of the protein (PDB 1IIY).** Domain A is colored in orange and domain B is colored in green. Image generated with VMD [20]

[a] *Details in this chapter appear in Prot Sci.: Fujimoto, Y.K., et al. Computational models explain the oligosaccharide specificity of cyanovirin-N. Prot. Sci., 11: 2008-2014 (2008).*

The "back" of the protein contains two adjacent triple stranded anti-parallel β sheets and the "front" comprises of two opposing β hairpins. One 3_{10}-helical turn directs each of these structural elements, and a four residue linker connects the sequential domain. CVN binds to the sugars attached to the HIV envelope glycoprotein, gp120, and thus is an inhibitor of all strains of HIV. It has been established that gp120 is glycosylated with an abundance of N-linked carbohydrates, and that CVN selectively binds with nanomolar affinity to the D1D3 isomer of Man_8 and to Man_9 [1].

A number of enlightening studies on carbohydrate binding to CVN have been carried out by Dr. Carole Bewley and her collaborators. The solution NMR structure in complex with Manα(1-2)Man (PDB: 1IIY) revealed two sugar binding sites located on opposite ends of the protein. Isothermal titration calorimetry experiments showed the sites to differ in affinity for the disaccharides by a factor of 10 (data shown in Table 1).

Carbohydrate	**Site**	**K_A (M^{-1})**	**ΔG @ 298K (kcal mol^{-1})**
Manα(1-2)Man	Site A	6.8 (±4) x 10^5	-8.0
	Site B	7.2 (±4) x 10^6	-9.4
Manα(1-2)Manα(1-2)Man	Site A	6.6(±0.7) x 10^6	-9.3
	Site B	3.7(±0.3) x 10^5	-7.6
Manα(1-2)Manα(1-3)Man	Site A	8.1(±0.8) x 10^3	-5.3
	Site B	1.7(±0.5) x 10^5	-7.1
Manα(1-2)Manα(1-6)Man	Site A	7.1(±0.9)x10^4	-6.6
	Site B	2.8(±0.7)x10^5	-7.4

Table 1. **Isothermal titration calorimetry results.** These experiments show the different affinities for the carbohydrates in each site. [3, 4]

Domain B is comprised of a deep pocket, while domain A appears to have a reduced-size pocket. Along the carbohydrate binding sites, there are a number of polar and charged amino acid residues and thus a side chain is within hydrogen bonding distance of almost all the disaccharide hydroxyl groups, except for the C1 hydroxyl group on the reducing mannopyranose ring [3].

Following the disaccharide studies, experimentalists determined the affects of affinity and specificity of CVN to larger mannose oligosaccharides. Once again, using isothermal titration calorimetry, the specificity and affinity was evaluated for CVN binding to several synthetic trisaccharides (as indicated in Table 1), methyl Manα(1-2)Manα(1-2)Man, methyl Manα(1-2)Manα(1-3)Man, and methyl Manα(1-2)Manα(1-6)Man), which represent the D1, D2, D3 arms of the Man_9 structure. An illustrated image of the Man_9 structure is depicted in Figure 2. The results of the binding of CVN to the trisaccharides were unanticipated. Site A was able to discriminate between the three trisaccharides with K_A values over three orders of magnitude; however, site B showed all three trisaccharides binding with similar K_A values as provided in Table 1 [3].

Despite the apparent wealth of data for this system, there remain open problems regarding the structure and energetics of specific sugar recognition. The crystal structures of CVN bound to each of two high-mannose oligosaccharides, Man_6 and Man_9, have been solved [5], but the carbohydrates are not fully resolved. One of the two binding sites is occupied, and the oligosaccharide structures deviate strongly from that expected; several mannoses are in the β configuration (where α anomers are expected),

and numerous rings are in disfavored ring conformations. Overall, only the small portion of the oligosaccharide (Manα(1-2)Manα(1-2)Man) making the most intimate contact with the protein is particularly well structured.

A solution structure with both sites bound to the disaccharide Manα(1–2)Man is available [1]. In this structure, however, few NOE constraints were available to accurately define the lower affinity site. Here, we present computationally refined models of α(1–2)-linked di- (Man_2) and trimannose (Man_3), representative of the D1 arm of Man_9, bound in both sites. These models, combined with molecular dynamics and continuum electrostatic analysis, capture the observed specificity of binding with semi-quantitative accuracy.

5.2 Methods

5.2.1 Preparation of structures.

The initial structure was the solution NMR structure of CVN bound to Manα(1–2)Man (PDB 1IIY) [1]. CVN has two pseudo-symmetric binding domains. The backbone atoms of equivalent residues in each site were superimposed by an RMSD fit and the coordinates of the sugar in site A were then replaced with those from the superimposed structure to construct increased symmetrical binding models. The next step is to understand the effects of the trimannose structures with CVN which represent the three arms of Man9. As there are no structures solved for these models, we have computationally built them by extending the newly constructed dimannose by one unit. Two forms of Manα1-2Manα1-2Man were generated—a glycosidic bond can be formed

either with OH2 of dimannose “Y” or with OH1 of dimannose “X”. The crystal structure of Man_9 bound in domain A indicates the first of these configurations is preferred for that region [5]. A schematic figure of this is shown (Figure 3).

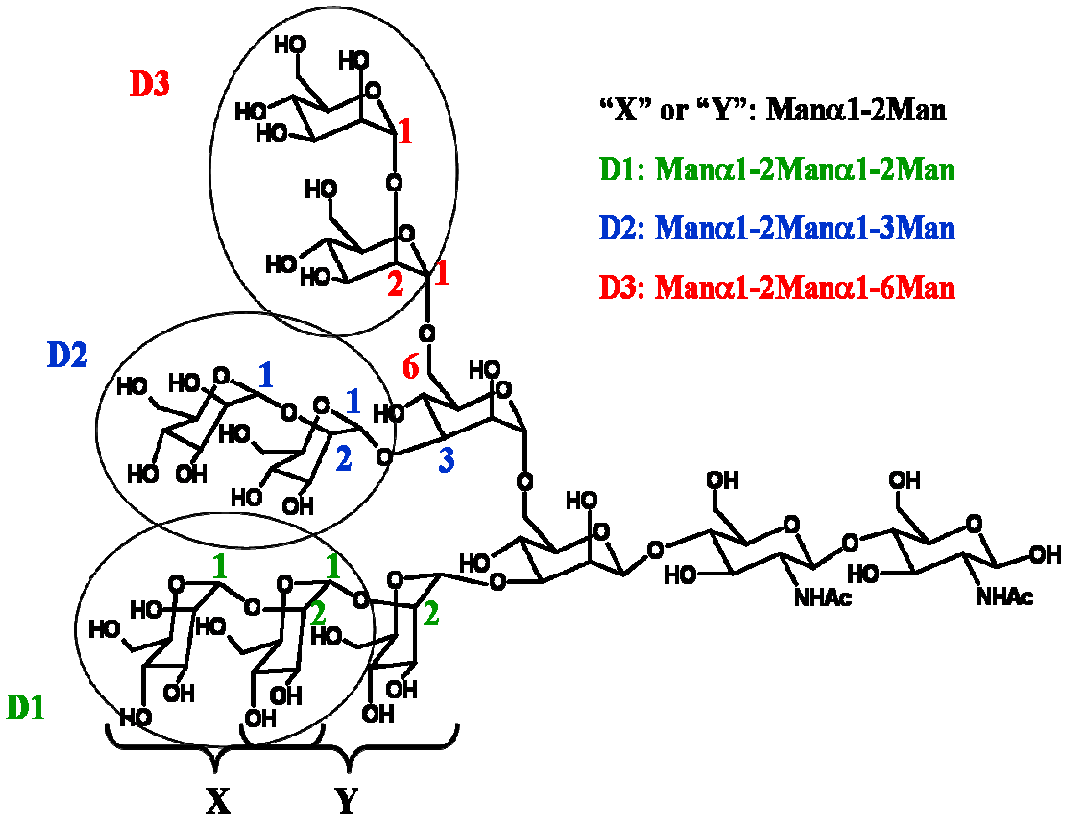

Figure 2. **Three arms of Man_9** are labelled as: D1, D2, and D3. The carbon atoms of the glycosidic bonds are labelled by their linkages. The dimannose (circle) represents the terminal ends of all three arms. Trimannose structures that represent the three arms were computationally built by extending the newly constructed dimannose by one unit. Two forms of Manα1-2Manα1-2Man were generated—a glycosidic bond can be formed either with OH2 of dimannose “Y” or with OH1 of dimannose “X”. The crystal structure of Man_9 bound in domain A indicates the first of these configurations is preferred [5] for that region.

Visual analysis indicated no major steric clashes, and this initial placement was subjected to a short minimization with all protein residues >4.0A from the sugar held fixed to alleviate any other clashes. Manipulations were done using the CHARMM software.

5.2.2 Explicit-solvent molecular dynamics.

Molecular dynamics simulations were performed with CHARMM. Hydrogen-atoms were positioned using the HBUILD facility of the CHARMM computer program. The SHAKE algorithm was used to fix all bonds involving hydrogens. Sugars were surrounded by 15 Å radius spheres of water, with all atoms outside this sphere held fixed. The simulations ran for 20ns after 2ns of equilibration, using Langevin dynamics. All calculations used the PARAM 22 protein parameter set [32], the TIP3P water model [23], and the Carbohydrate Solution Force Field (CSFF) sugar parameter set [25].

5.2.3 Calculating binding free energies.

Binding free energies were computed with an MD/PBSA model. The trajectories were sampled every 20ps for a total of one thousand frames. Rigid body binding free energies were calculated for every snapshot and averaged over for all frames. The linearized Poisson–Boltzmann equation [17] was solved using a multigrid finite-difference solver and distributed with the Integrated Continuum Electrostatic (ICE) package [19]. Charges were taken from the PARAM 22 and CSFF sugar parameter sets, radii optimized for use in continuum models were used (protein radii [34], sugar radii from [18]). An internal dielectric constant of 2 and an external dielectric constant of 80 were used, and the ionic strength was set to 145mM. The dielectric boundary was chosen

based on the molecular surface with a 1.4 Å radius probe, and a 2.0Å ion exclusion layer were used. Boundary conditions were computed using a 3-step focusing procedure on a 129x129x129 unit cubic grid, with the molecule occupying first 23%, then 92%, and 184% of the grid, with Debye–Hückel boundary potentials used for the lowest level. The total binding energies for each snapshot were computed as the sum of the intermolecular van der Waals energy, electrostatic contribution, and the term related to the solvent accessible surface area buried in binding. The surface area was calculated using a 1.4A radius probe and the contribution was given as: $\Delta G^{np} = 0.005\Delta A + 0.86 \frac{\text{kcal}}{\text{mol}}$ [41].

5.2.4 Calculating sugar strain energies.

Sugar strain energies were computed by comparing the total ensemble-averaged energies of sugar conformations extracted from bound-state molecular dynamics and those from unbound simulations. Electrostatic solvation free energies were computed with a Poisson–Boltzmann model, as the difference between a system with solvent dielectric constant of 80 (ionic strength of 0.145 M) and solute dielectric constant of 2, and a system of uniform dielectric constant of 2 (zero ionic strength). For these calculations, the largest grid contained the whole sugar (92% of the longest dimension). Hydrophobic solvation free energies were estimated with a term proportional to the total solute surface area, as described above. These energies were added to the molecular-mechanics energy of the solute, including all bonded terms (bonds, angles, dihedrals), intramolecular van der Waals, and intramolecular Coulombic (in uniform dielectric of 2) interactions. The

sugar strain energy is given by the difference in the ensemble average of the total energy of the sugar in conformations extracted from the complex simulation and the similar average for conformations from a simulation of the free sugar. Thus, these values correspond to the energetic cost of perturbing the unbound conformational ensemble into the ensemble that is capable of binding, in a fully solvated context.

5.3 Discussion

In the initial explicit-solvent molecular dynamics simulations with the solution structure in complex with two mannose disaccharides (PDB 1IIY) [1], it quickly revealed instability in the lower affinity binding site (Site A). The disaccharide began to dissociate from cyanovirin-N within the first 500 ps of simulation (see Figure 3). In the higher affinity site (Site B), the sugar remains stably bound during the 20 ns simulation.

The difference between the two sites in affinity for Man_2 is approximately a 10-fold difference in K_A, as measured by the calorimetric data [4]; thus, this behavior is not expected. The two binding sites of CVN are pseudo-symmetric, but the orientations of the sugar in each site of the structure are different; the sugar in the low-affinity site makes less intimate contact with some binding-site residues (see Figure 4, left). A higher symmetry model of dimannose binding in the low-affinity site was thus considered. The protein backbone of the high affinity binding site was superimposed on that of the lower affinity site, and the sugar from the high affinity site placed in the lower affinity pocket. For a short time the structure was minimized while all protein residues further than 4.0A from the sugar were held fixed. Molecular dynamics simulation from this starting

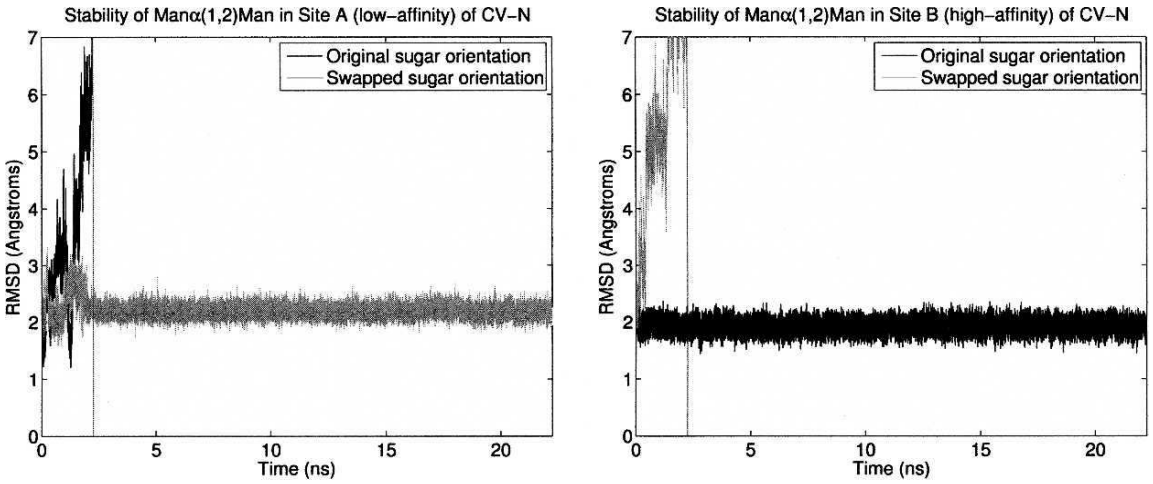

Figure 3. **Stability of Man2 in each CV-N binding site.** The root mean-square deviation (RMSD) of sugar-heavy atoms, relative to the initial coordinates, is shown over 22 ns of explicit-solvent molecular dynamics. Trajectories beginning from the original structure are in black; those beginning with the sugar placed in the newly found orientation in the other site are shown in gray. A vertical bar indicates the end of the equilibration phase; all analysis was done on frames following this point. [15]

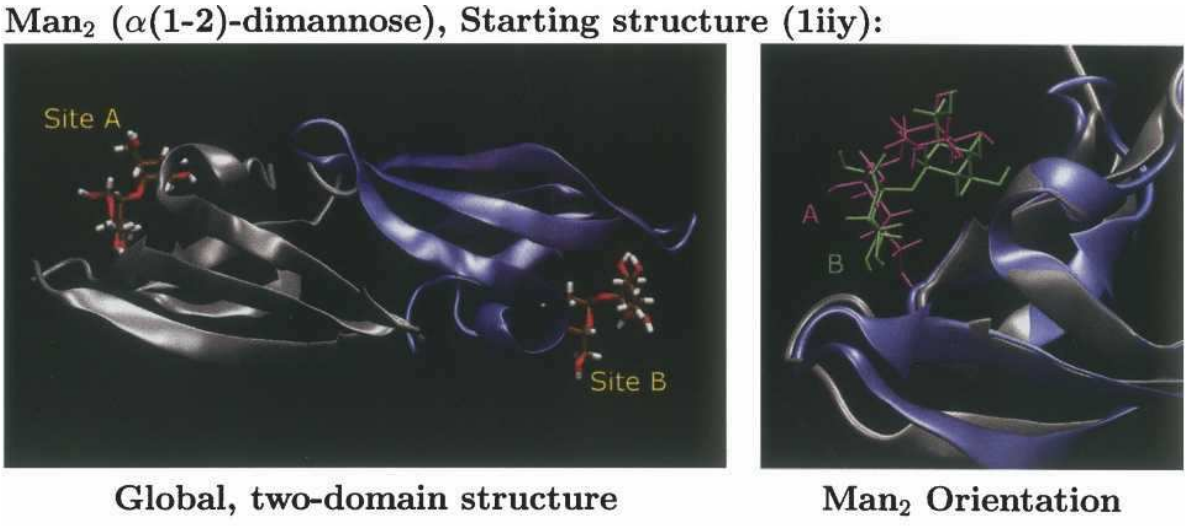

Figure 4. **Structures of dimannose bound to CV-N.** (Left) The overall structure of CV-N, showing two sugar binding sites in pseudosymmetric domains (domain A in gray, domain B in blue). (Right) A superimposed image of the two binding sites of CV-N; showing the difference in orientation of the original Man_2 model in the low-affinity site (pink/gray) from the structure of the high-affinity site (green/blue). [15]

structure shows similar stability to that of the high-affinity site. As an additional test, this procedure was repeated in reverse—the low-affinity site was superimposed on the high affinity site, and the low-affinity sugar orientation placed in the high-affinity pocket. The binding site was briefly relaxed, and then subjected to molecular dynamics. This model shows the same instability as was first observed in the lower affinity site. These data are displayed in Figure 3 (right).

Fewer NOEs were observed experimentally for the lower affinity site than for the higher affinity, and thus the published model was not uniquely determined by experimental constraints [1]. To further test the validity of the new model, we tracked the distances of all atoms involved in observed intermolecular NOEs throughout the first 10 ns of molecular-dynamics simulation. As seen in Table 2, all experimentally observed contacts remain within 6.0A for the majority of the simulation, and most remain within 5.0A or less. In comparison, several contacts in the published model are beyond the largest constraint distance (6.0A). While the timescale of the simulation is much lower than that observed in the NOESY experiment, these data demonstrate the consistency of the refined model with available data.

Sugar atom	CVN atom	d[a]	Occ.[b]	τ[c]	CVN atom	d[a]	Occ.[b]	τ[c]
Man1:								
H2	I94-$C_\delta H_3$	6.0(6.8)	1.70	1.37				
H4	T7-$C_\gamma H_3$	5.0(5.2)	1.43	0.92				
H5	T7-$C_\gamma H_3$	5.0(2.3)	0.96	3.56	T25-$C_\alpha H_1$	4.0(2.4)	1.00	3333
H6b	T25-$C_\alpha H_1$	4.0(2.9)	0.98	12.5				
Man2:								
H2	A92-$C_\beta H_3$	4.0(5.4)	1.05	3.81	I94-$C_\delta H_3$	4.0(5.0)	1.16	0.90
H3	A92-$C_\beta H_3$	5.0(5.5)	0.97	3.66	I94-$C_\delta H_3$	6.0(5.2)	1.12	0.90
H4	K3-$C_\delta H_3$	4.0(3.4)	1.11	17.2	Q6-$C_\beta H_3$	5.0(4.2)	1.45	1.77
	T7-$C_\gamma H_1$	5.0(4.5)	1.40	1.06				
H5	T7-$C_\beta H_3$	5.0(5.2)	1.00	175	T7-$C_\gamma H_3$	4.0(3.0)	2.60	2.16
H6a	K3-$C_\delta H_3$	6.0(6.7)	0.70	0.95	Q6-$C_\beta H_2$	4.0(7.2)	1.29	1.93
	Q6-$C_\gamma H_3$	5.0(5.3)	1.38	1.85	T7-$C_\beta H_1$	4.0(6.4)	0.62	4.00
	T7-$C_\gamma H_1$	4.0(4.4)	1.98	1.97				
H6b	K3-$C_\delta H_3$	6.0(5.5)	0.59	0.96	Q6-$C_\beta H_2$	5.0(5.7)	1.48	4.78
	T7-$C_\beta H_3$	5.0(7.3)	0.82	1.32	T7-$C_\gamma H_3$	4.0(5.1)	2.33	2.17

[a] Persistant contact cutoff from MD (distance in published structure [bewley 2001]) in Å.
[b] Fraction of simulation in which contact is made (sum of all hydrogens in group).
[c] Average lifetime of contact, in ps.

Table 2. **Observed NOEs in simulation of refined structure.**

Binding energies of each complex were calculated using three energetic terms: the intermolecular van der Waals interactions made in the bound state (ΔG^{vdw}), the hydrophobic solvation energy ($\Delta G^{h\phi}$), and the electrostatic contribution (ΔG^{elec}). The sum of all these terms gives the "semi-rigid" binding free energy. The energies are averaged over multiple frames. An additional term was calculated into the binding energy. The strain energy was computed for the carbohydrates to obtain the energetic cost of a structure found in the bound and unbound states. Strain energies were calculated by running additional simulations of just the unbound structure. The difference in total energy given from internal energies (bond, angle, and dihedral), intramolecular van der Waals, Coulombic interactions, and the solute-solvent interactions between the unbound state and the bound state is the strain (ΔG^{str}).

The binding free energy methodology can be used to calculate the relative binding affinities of the two domains from each carbohydrate model (presented in Table 3); and these are computed with reasonable accuracy for the dimannose model and Manα(1-2)Manα(1-2)Man trimannose model, further supporting the validity of the structural models. In these cases, the trends are computed correctly, both trimer/dimer specificity in each site and the relative affinities of either ligand in the two sites. However, the magnitude of the differences is somewhat overestimated. Considering each energetic term suggests a dominant role for electrostatic interactions in defining the specificity at each site. This is consistent with the structural analysis, in which all major differences involved electrostatic interactions.

	ΔG^{vdw}	$\Delta G^{h\phi}$	ΔG^{elec}	ΔG^{str}	ΔG^{comp}	ΔG^{expt}
ΔΔG Site A – Site B						
Man_2	–0.9 (0.1)	–0.1 (0.0)	–2.3 (0.1)	+0.4 (0.2)	–2.8 (0.2)	–1.5
Man_3	–3.8 (0.1)	+0.3 (0.0)	+8.2 (0.1)	+2.1 (0.2)	+6.9 (0.2)	+1.7
ΔΔG Man_3 – Man_2						
Site A	–1.3 (0.1)	–0.7 (0.0)	–3.7 (0.0)	–0.7 (0.3)	–6.4 (0.3)	–1.4
Site B	–4.2 (0.1)	–0.2 (0.0)	+6.5 (0.1)	+1.0 (0.3)	+3.3 (0.4)	+1.8

Table 3. **Relative free energies of binding.** Van der Waals, hydrophobic surface burial, electrostatics, carbohydrate strain, total computed, and experimental (Bewley and Otero-Quintero 2001; Bewley et al. 2002), all in kcal/mol. Errors are the standard error of the mean for the ensemble averages.

Using the component analysis method these structures began to explain the key determinants of oligosaccharide specificity in the two sites of CV-N, as shown in Figure 5 and 6.

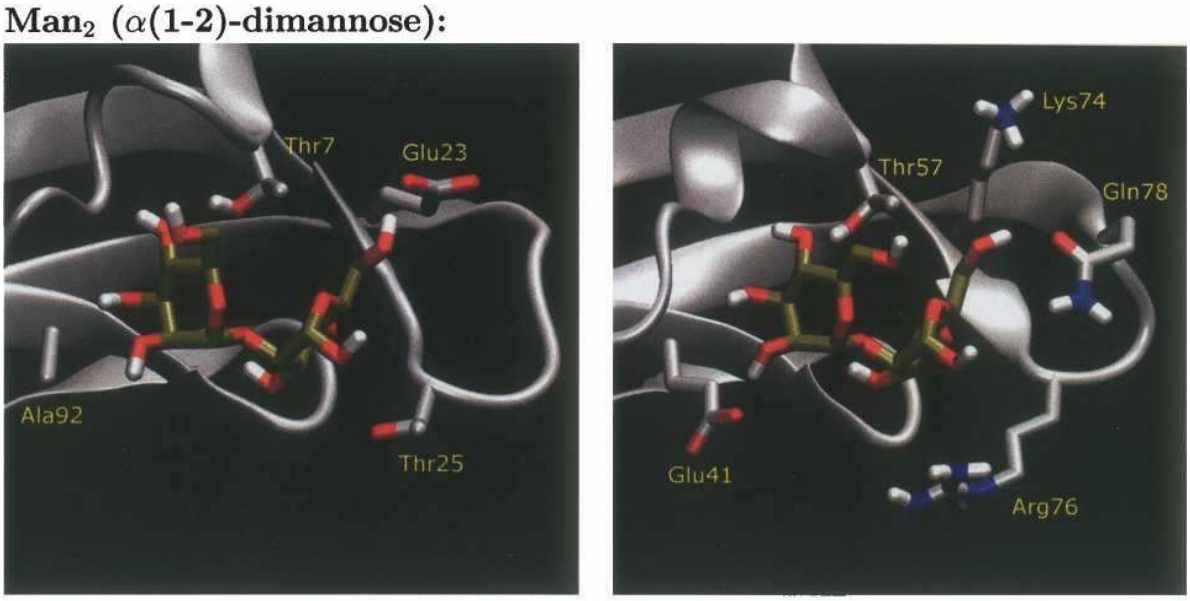

Figure 5. **Structures of diimannose bound to CV-N.** A representative snapshot from each dynamic simulation is displayed, with protein in gray and sugar in bronze. All amino acid side chains involved in significant electrostatic interactions are shown. Figures generated with VMD [20].

There is a strong electrostatic interaction in the high-affinity site (between Glu41 and Man2 OH2) which is absent in the low-affinity site, where the corresponding residue is alanine 92. There are other ways in which the two binding sites vary, but the overall differences in interactions are less prominent. A hydrogen bond is made between Glu23 and Man1 OH6 in domain A but its corresponding residue, Lys74, does not. Instead of Lys74 there is a similar interaction with Gln78 in domain B; these residues are not in equivalent positions, but make equivalent interactions. Thr25 in domain A is replaced by Arg76 in domain B, but neither makes close, specific interactions with the sugar. The Glu41/Ala92 variation found in the disaccharide model seems to be an important feature when looking at the CVN bound to the Manα(1-2)Manα(1-2)Man trimannose. In this

trimannose model it binds to these sites with a specificity reversed from that of the dimannose. Domain A (low affinity for dimannose) binds the trimannose with higher affinity than does domain B (high affinity for dimannose). The favorable hydrogen bond made by Glu41 with Man2 in domain B is lost when there is an additional α(1–2)-linkage extending from Man2. It now becomes part of the glycosidic bond. The lack of an interaction with this combination explains the change in affinity in the sites. Figure 6 shows the structure of the Manα(1-2)Manα(1-2)Man trimannose model.

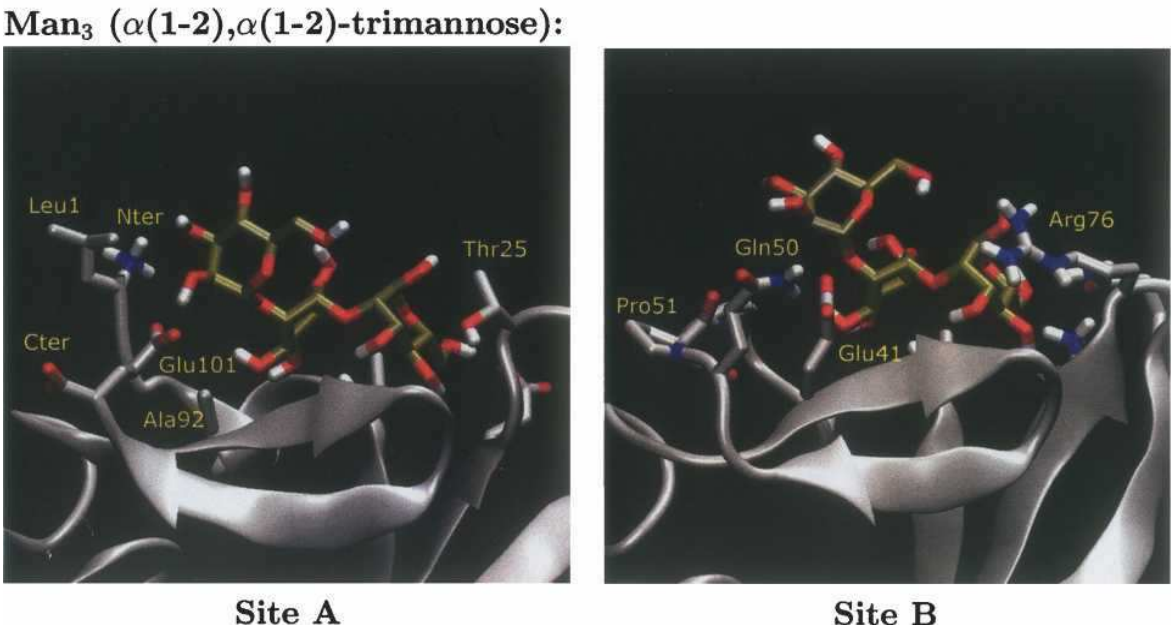

Figure 6. **Structures of Manα(1-2)Manα(1-2)Man trimannose bound to CV-N.** A representative frame from each dynamic simulation is displayed, with protein in gray and sugar in bronze. All amino acid side chains involved in significant electrostatic interactions are shown. Figures generated with VMD.

Another set of interactions are made in domain A, explaining the increased affinity of this site for this trisaccharide: Glu101 (the C-terminal residue) receives a hydrogen bond from

OH2 of the third sugar; and the charged N-terminus donates a hydrogen bond to OH3. In domain B, neither of these interactions is made.

Unfortunately, simulating the other trimannose models did not show consistencies in affinities with experimental data as seen in Table 4.

	$\Delta\Delta G^{vdw}$	$\Delta\Delta G^{h\phi}$	$\Delta\Delta G^{elec}$	$\Delta\Delta G^{str}$	$\Delta\Delta G^{comp}$	$\Delta\Delta G^{exp}$
ΔΔG Site B – Site A						
Man_{3b}	–4.1 (0.4)	–0.2 (0.0)	–4.7 (0.4)	+0.6 (1.0)	–8.4 (1.1)	+1.7
Man_{3c}	–3.1 (0.4)	–0.2 (0.0)	–1.9 (0.4)	–0.6 (1.3)	–5.8 (1.4)	–1.8
Man_{3d}	–5.5 (0.4)	–0.5 (0.0)	–4.2 (0.4)	–1.2 (1.3)	–11.4 (1.4)	–0.8
ΔΔG Man_{3b} – Man_2 :						
Site A	–0.9 (0.3)	–0.3 (0.0)	+0.1 (0.2)	+0.6 (1.0)	–0.5 (1.1)	–1.3
Site B	–4.1 (0.3)	–0.4 (0.0)	–2.3 (0.7)	+0.7 (1.0)	–6.1 (1.1)	+1.8
ΔΔG Man_{3c} – Man_2 :						
Site A	–1.2 (0.3)	–0.2 (0.0)	+0.3 (0.2)	+0.4 (0.9)	–0.7 (1.0)	+2.7
Site B	–3.4 (0.3)	–0.3 (0.0)	+0.7 (0.3)	–0.7 (1.0)	–3.7 (1.1)	+2.3
ΔΔG Man_{3d} – Man_2 :						
Site A	–0.3 (0.3)	–0.1 (0.0)	+0.6 (0.2)	–0.2 (1.0)	0.0 (1.1)	+1.4
Site B	–4.9 (0.3)	–0.5 (0.0)	–1.3 (0.3)	–1.9 (0.9)	–8.6 (1.0)	+2.0

Table 4. **Relative free energies of binding of other carbohydrates.** Van der Waals, hydrophobic surface burial, electrostatics, carbohydrate strain, total computed and experimental data, all in kcal/mol. Errors printed are the standard error of the mean for the ensemble averages.

The inconsistencies in this data can be due to many reasons. One may be how the complex was solvated. The possibility of restraining the system in a sphere of water might not accurately calculate the binding affinities for these complexes. Discussion about this will be further analyzed in the next chapter.

5.4 Conclusion

The results validate the use of the MD/PBSA approach in the study of carbohydrate-binding proteins. While this method has been used extensively in the study of protein–protein and protein–small molecule interactions, to date there have been few applications

to carbohydrates. This work strongly motivates the pursuit of future studies on protein–carbohydrate recognition by these approaches.

Chapter 6

Using unrestrained simulations to understand carbohydrate-protein interactions

Previously it was discussed how the dimannose and Manα1-2Manα1-2Man interacted in the binding pockets. The other trimannose models were not discussed in great detail because the trends for the computed relative energies did not correspond well to experimental data. One possible reason for this issue was the restraint put on the binding modes by using a sphere of water. In the new method, CVN and its carbohydrate ligands were simulated in a box of water, allowing the entire complex to be mobile.

6.1 Methods

6.1.1 Construction of CVN Complex.

Sixteen separate models of CVN complexes were simulated with the following carbohydrates: Manα(1–2)Man, two binding modes of Manα(1–2)Manα(1–2)Man, Manα(1–2)Manα(1–3)Man, and Manα(1–2)Manα(1–6)Man. Trimannose structures that represent the three arms of Man_9 were

computationally built using the same method as in the previous chapter. Five simulations represented each of the carbohydrates bound in both binding domains (doubly bound). Another set had the same five carbohydrates but with domain A bound (singly bound – domain A), another five simulations had them bound to domain B (singly bound – domain B), and lastly an unbound structure of CVN where no sugars were bound was considered. Molecular dynamics simulations and visual analysis revealed a problem in domain B, in which the sugar often dissociated from its binding pocket. Ten simulations had sugars bound in domain B, including the doubly bound and singly bound forms in domain B; and from that group, five simulations showed the carbohydrates dissociating from its binding pocket in domain B (doubly bound Manα1-2Manα1-6Man, singly bound Manα1-2Man, both conformations of Manα1-2Manα1-2Man in the singly bound complexes, and the singly bound Manα1-2Manα1-6Man). After further looking at the structure, there was a disallowed backbone dihedral angle between S52 (ϕ 173 and ψ 13) and N53 (ϕ 59 and ψ 60). Molecular dynamics revealed alternate structures that relieved that clash. Carbohydrates were rebuilt using the new backbone structure and re-simulated. Manipulations were done using the CHARMM software package.

6.1.2 Molecular Dynamics

All CVN complexes were solvated in a box of TIP3P water, 145mM NaCl and simulated under NPT ensemble conditions with periodic boundary conditions and PME electrostatics for an overall simulation time of 200ns using a 2fs time step. All calculations used the PARAM 22 protein parameter set, the TIP3P water model, and the

CSFF sugar parameter set. Binding energies were calculated when simulations were converged (after 50ns) as the sum of van der Waals, solvent-accessible surface area ($\Delta G = 0.005A+0.86$ kcal/mol), and the continuum electrostatic terms, and were averaged over 1500 snapshots. Simulations were done using the NAMD 2.6 software.

6.2 Discussion

Many different simulations were performed with the various CVN complexes. One group of simulations had carbohydrates bound in both domains and in other sets the carbohydrates were bound to domain A while domain B was left empty, and vice versa. An additional simulation was done where no carbohydrates were bound. After the simulations were completed and analyzed, domain A showed structural and energetic consistencies in the doubly and singly bound models; however, domain B revealed issues in its structure. Each region in domain B was further analyzed and the inconsistencies were found in a particular hinge region from residues 50 through 57 which is shown in a circle in Figure 1.

There was an unfavorable backbone dihedral angle around Ser52 (ϕ 173 and ψ 13) and Asn53 (ϕ 59 and ψ 60) shown in Figure 2A. After simulating the complex, two possible favorable conformations were discovered that relieved this structure (Figures 2B and 2C). Figure 2A shows the protein undergoing a crank-shaft motion of the peptide bond between S52 and N53 (highlighted by yellow arrows) and Figure 2B shows the final structure it conforms to (the region that was changed is highlighted in yellow circle). In

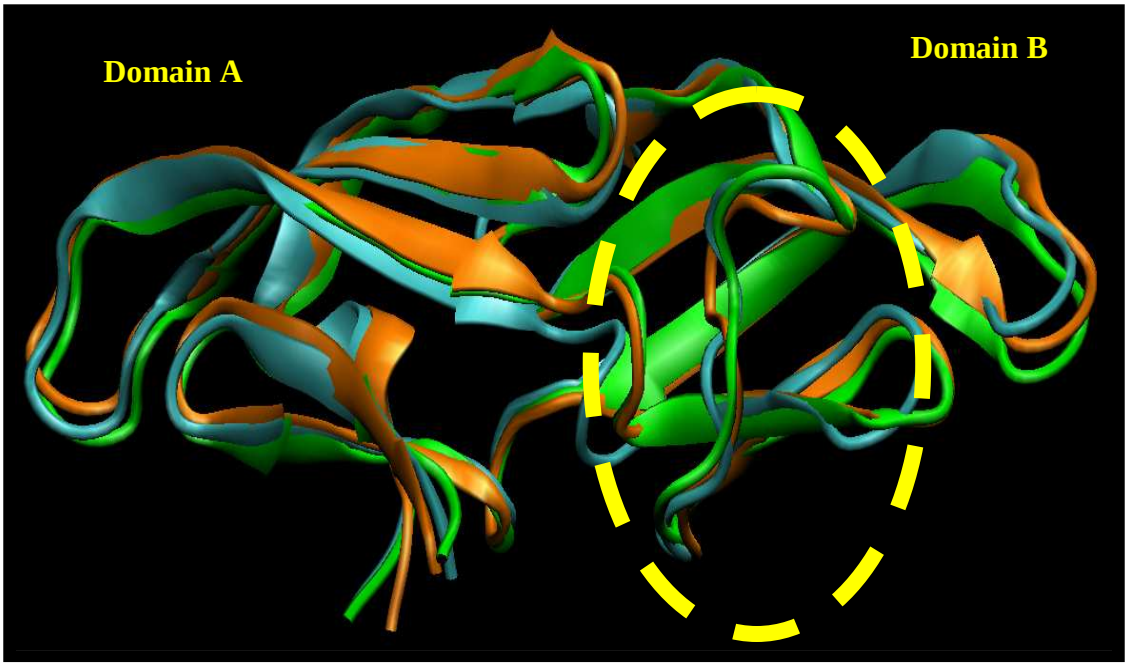

Figure 1. **Overlay image of three snapshots of CVN** (no sugar is present in image). Orange: doubly bound CVN with Manα(1–2)Man and unbound form, cyan: singly bound CVN with Manα(1–2)Man, green: unbound form.

another case, it undergoes a cis peptide conformation shown in a red arrow in Figure 2A by rotating the bond between the carbon atom on P51 and the nitrogen on S52; and the conformation looks like the structure shown in Figure 2C. All doubly bound models and singly bound–domain B models (except singly bound–Manα(1–2)Man) showed the crank-shaft motion; and only the doubly bound Manα1-2Manα1-3Man simulation showed the cis peptide conformation. To further analyze the movement of the dihedral angles around S52 and N53 during the time of the simulation, Ramachandran plots of these dihedral angle were generated (Figure 3).

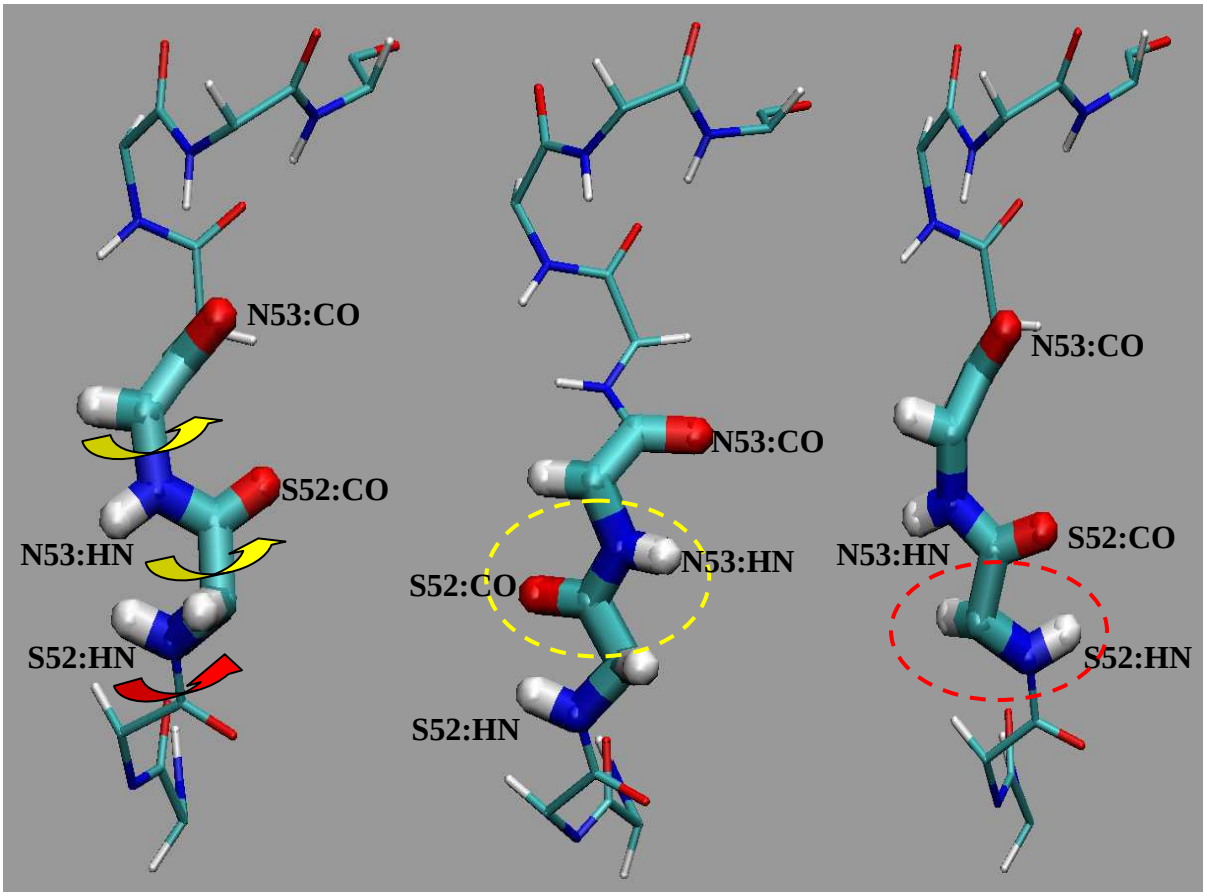

Figure 2. Backbone hinge region highlighting S52 & N53 (A) Monomeric NMR structure of wt CVN (PDB: 1IIY) (B) Crank-shaft movement of the middle region (C) Cis peptide conformation movement in the lower region.

In Figure 3, the left-hand side of the plots represent phi and psi angle of serine 52 and the right-hand side of the plots represent phi and psi angles of asparagines 53. The y-axis represents the angles ranging from 0 to 360 degrees and the x-axis represents the time of the simulation. The two plots on top show the movement when the original NMR structure was the starting structure. Around 60ps that structure quickly changed into the

crank-shaft structure. However, it did not remain at that structure for the entire simulation. At around 70 to 80 ns it converts back into its native structure. After 80ns it changes into the crank-shaft conformation and stays there for the remainder of the simulation. The middle two plots show the movement using the crank-shaft structure further extended from the previous model. Although the movement in that region did not change from the crank-shaft like motion to the original NMR structure for another 100ns, those residues still showed a lot of movement allowing the sugar in that domain to be unstable. Since the carboxyl group from S52 faced away from the sugar that lose of interaction allowed more movement in the sugars and even having the dimannose move away from the protein causing dissociation from its binding pocket. Calculating the binding energies from the other conformations was difficult since there was sugar dissociation. The bottom plots represent the cis peptide structure. By using this as the starting structure, it was observed that the sugar stayed more tightly bound to its binding pocket because the carboxyl groups from S52 and N53 had strong interactions with the hydroxyls on the sugars. That strong interaction kept the motion of the dihedral angle more stable as seen in the bottom two plots. When calculating the binding energies when using the last model, it correlated more to the experimental results.

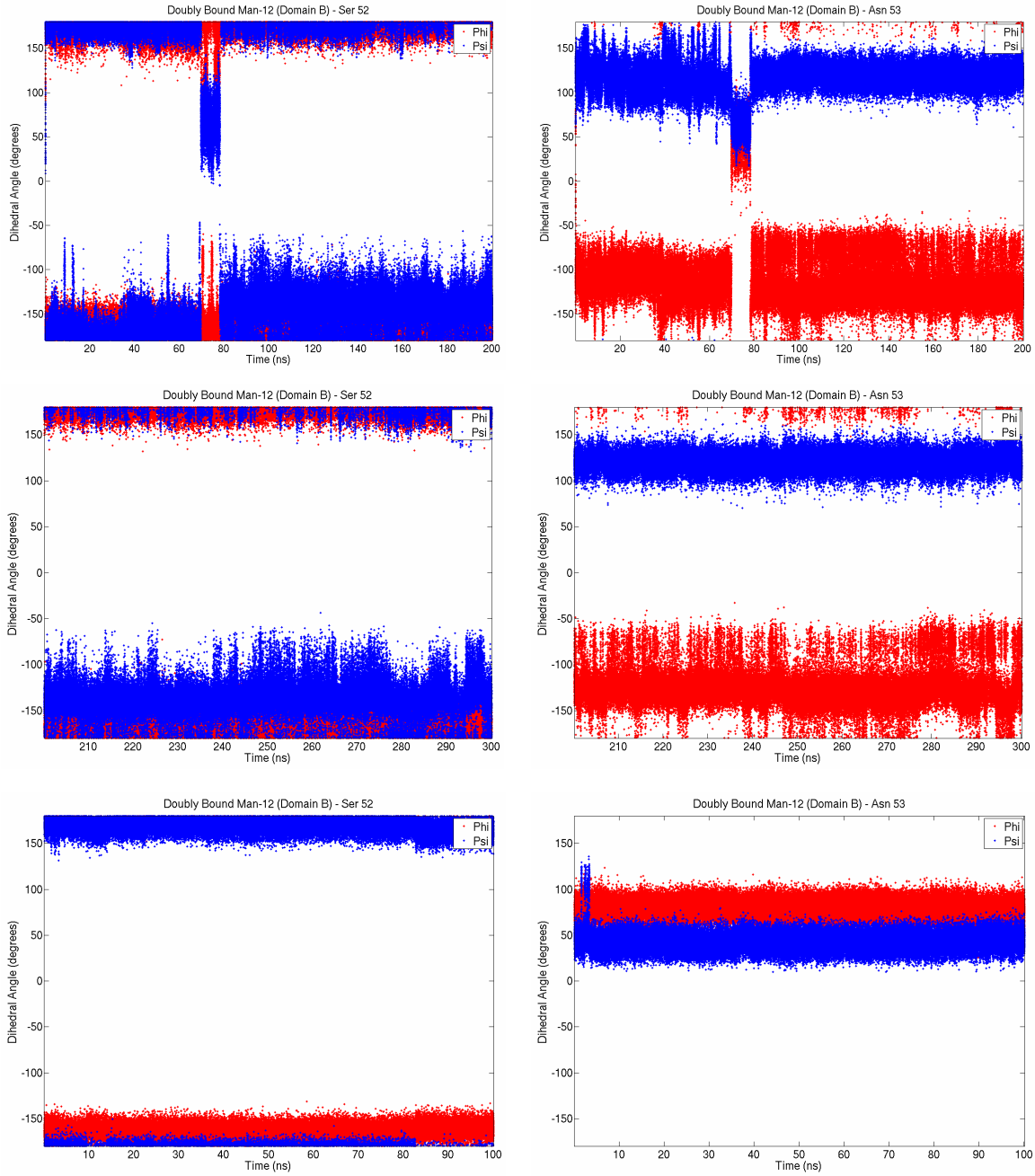

Figure 3. **Backbone dihedral angle movement from Ser52 and Asn53 versus time.** (Top) Simulation from original NMR structure (Middle) Simulation from the crank-shaft conformation (Bottom) Simulation from the cis peptide conformation.

In the other CVN complexes, both conformations were simulated, and in many cases carbohydrates dissociated from the results from the crank-shaft motion.

To further validate the cis peptide structure, other monomeric structures of CVN were observed. Crystal structures and an NMR structure showed a similar representation of this hinge region to the computed model. In experimental cases Pro51 was mutated to a glycine to relieve that unfavorable dihedral angle. Since the computed simulation originally used the NMR structure, the other NMR structure (PDB: 2RP3) with the P51G mutation was used to compare structures. The P51G mutation resembled the simulated model shown in Figure 2C. A representation of this similarity is shown in Figure 4. The backbone regions of S52 and N53 in both models face the same direction, as shown in a dotted yellow circle. The only difference between these models (shown in a yellow square) comes from the backbone region of residue 51. As long as the carboxyl groups face the sugar, those interactions can keep the sugar stably bound.

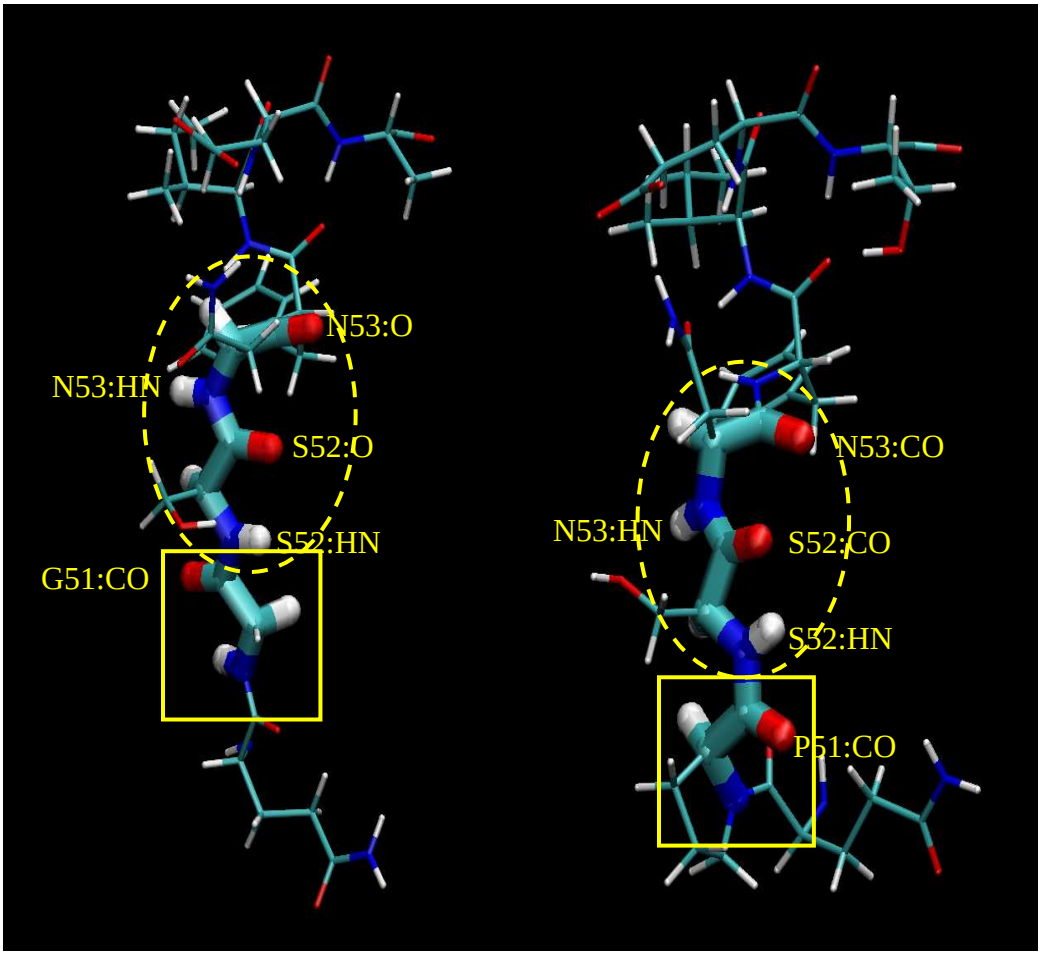

Figure 4. **Backbone hinge region of residues 50 to 57.** Residues 51 to 53 are highlighted in licorice Left: Monomeric NMR structure of P51G mutant (PDB: 2RP3) Right: Simulated structure showing how wt P51 resembles P51G mutation

Initial observations gave support the use of the structure shown as the cis peptide structure as the ideal conformation.

To further investigate this model, binding affinities for both sites in all the doubly bound simulations were calculated. Although the magnitude of binding free energies were overestimated compared to the experimental results, the overall trends were

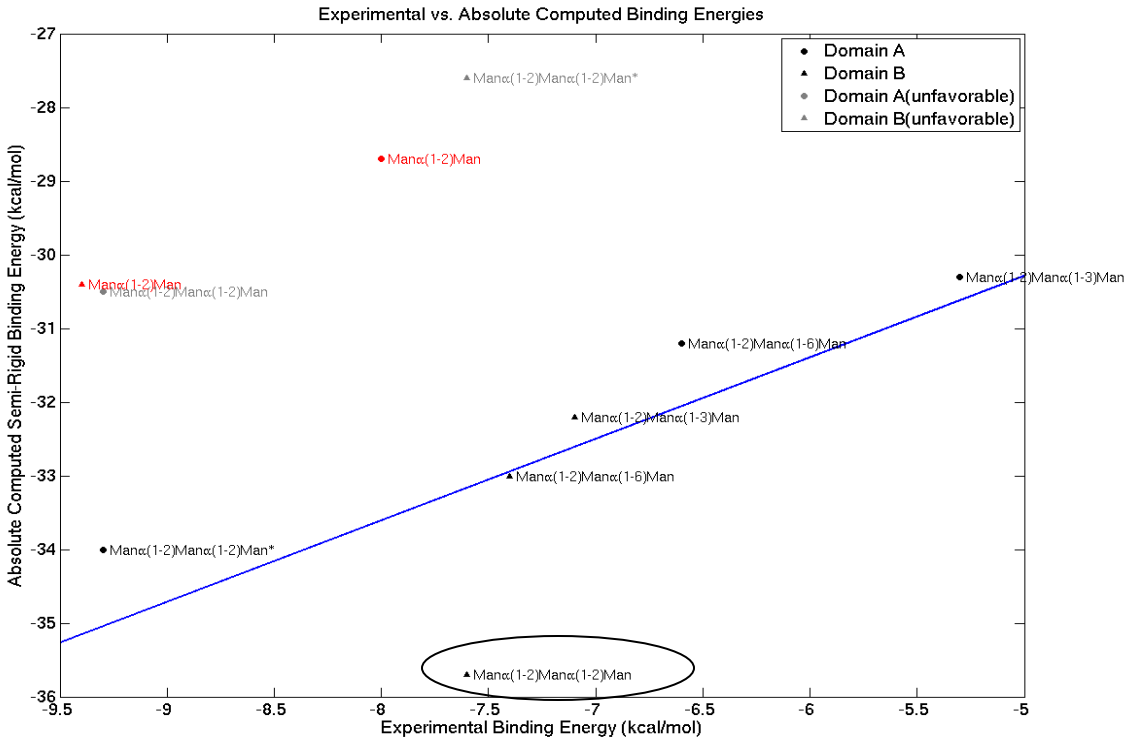

Figure 5. **Comparison of experimental and computed binding energies.** Computed binding energies were calculated as the sum of the intermolecular van der Waals energy, electrostatic contribution, and the term related to the solvent accessible surface area buried in binding. Plots colored in red represent the dimannose model, plots colored in gray are the unfavorable Manα(1-2)Manα(1-2)Man structures, the plots colored in black represent the favorable trimannose models.

computed with reasonable accuracy further supporting the validity of the models (Figure 5). The binding free energies were overestimated because not all energetic terms (*e.g.* entropy and strain) were considered. A linear fit was plotted using the favorable trimannose structures from both domains. This line was used as a reference to compare the other carbohydrate binding energies. Previously, it was stated two forms of Manα(1-2)Manα(1-2)Man structures were built. The unfavorable Manα(1-2)Manα(1-2)Man

corresponding to its domain is highlighted in grey. As expected these complexes showed a large destabilization as compared to the other complexes. Experimentally, the dimannose model showed very good binding affinities compared to the other trimannose models. However, the computed results were not able to match that data (as shown in red data points). There was a significant destabilization in the dimannose models for both domains compared to all the favorable trimannose models. It appears that an additional sugar can increase the affinity by a significant amount (-1.6 to -5.3 kcal/mol). The dimannose models should be considered as a category on its own, and unable to compare it with the trimannose models. A breakdown of the energies is found is Table 1. If one compares the relative energies within each domain of the diamannose model, the correct trend is present. Experimental data shows -1.4kcal/mol difference between domain B and domain A, and the computed data shows a -1.7 difference. The trends from the data of the trimannose models in domain A (black filled circles) showed perfect correlation to experimental data. Both experimental and computed data showed the least favorable binding energy from Manα(1-2)Manα(1-3)Man and the best from Manα(1-2)Manα(1-2)Man. Although the magnitude of the computed calculations are overestimated compared to the experimental results, the trend is still present. In domain A, if the favorable Manα(1-2)Manα(1-2)Man was used as a reference point for all the other trimannose models in this domain, the computed difference between Manα(1-2)Manα(1-2)Man and Manα(1-2)Manα(1-3)Man computed is +3.7 and experimentally it is +4.0 and

the computed difference with Manα(1-2)Manα(1-6)Man is +2.8, while experimentally it is +2.7.

Carbohydrate Model	ΔG^{vdw}	$\Delta G^{h\phi}$	ΔG^{elec}	ΔG^{bind}	ΔG^{exp}
Domain A					
Manα(1-2)Man	-18.9 (0.1)	-2.2 (0.0)	-7.6 (0.1)	-28.7 (0.1)	-8.0
Manα(1-2)Manα(1-2)Man*	-20.7 (0.1)	-2.7 (0.0)	-10.6 (0.1)	-34.0 (0.1)	-9.3
Manα(1-2)Manα(1-2)Man	-21.5 (0.1)	-2.5 (0.0)	-6.5 (0.1)	-30.5 (0.1)	ND
Manα(1-2)Manα(1-3)Man	-21.0 (0.1)	-2.4 (0.0)	-6.9 (0.1)	-30.3 (0.1)	-5.3
Manα(1-2)Manα(1-6)Man	-21.5 (0.1)	-2.5 (0.0)	-7.2 (0.1)	-31.2 (0.1)	-6.6
Domain B					
Manα(1-2)Man	-20.0 (0.1)	-2.1 (0.0)	-8.3 (0.1)	-30.4 (0.1)	-9.4
Manα(1-2)Manα(1-2)Man	-25.2 (0.1)	-2.6 (0.0)	+0.2 (0.1)	-27.6 (0.1)	ND
Manα(1-2)Manα(1-2)Man*	-23.7 (0.1)	-2.6 (0.0)	-9.4 (0.1)	-35.7 (0.1)	-7.6
Manα(1-2)Manα(1-3)Man	-22.2 (0.1)	-2.4 (0.0)	-7.6 (0.1)	-32.2 (0.1)	-7.1
Manα(1-2)Manα(1-6)Man	-22.2 (0.1)	-2.6 (0.0)	-8.2 (0.1)	-33.0 (0.1)	-7.4

Table 1. **Semi-rigid Binding Energies.** Average energies calculated over 1500 snapshots. Energies are in kcal/mol with its standard error of the mean.

As for data in domain B the experimental and computed data were not consistent with one another. The major feature that contributed most to the affinity differences was the electrostatic energy. An electrostatic component analysis was done to determine which interactions contributed most significantly to the binding affinity, and why there was such a difference in affinity between experimental and computational data. By using the component analysis we are able to distinguish where the differences are coming from. Figure 6 shows snapshots of the dimannose models and the preferred Manα1-2Manα1-2Man model illustrating the same interaction seen from the older model. The dimannose model has an interaction in domain B between E41 and Man2 OH2 which is absent in the corresponding site (A92). Previously, there was a strong interaction with Glu23 which is

still seen in this simulation. However, by not restraining the solvent in a sphere of water the corresponding nearby residues in the other domain do not make that same interaction. The preferred Manα1-2Manα1-2Man interacts between E101 (C-term) and the OH2 of the third sugar; and the charged N-term donates a hydrogen bond to OH3. Neither of these interactions is made with Q50 in domain B.

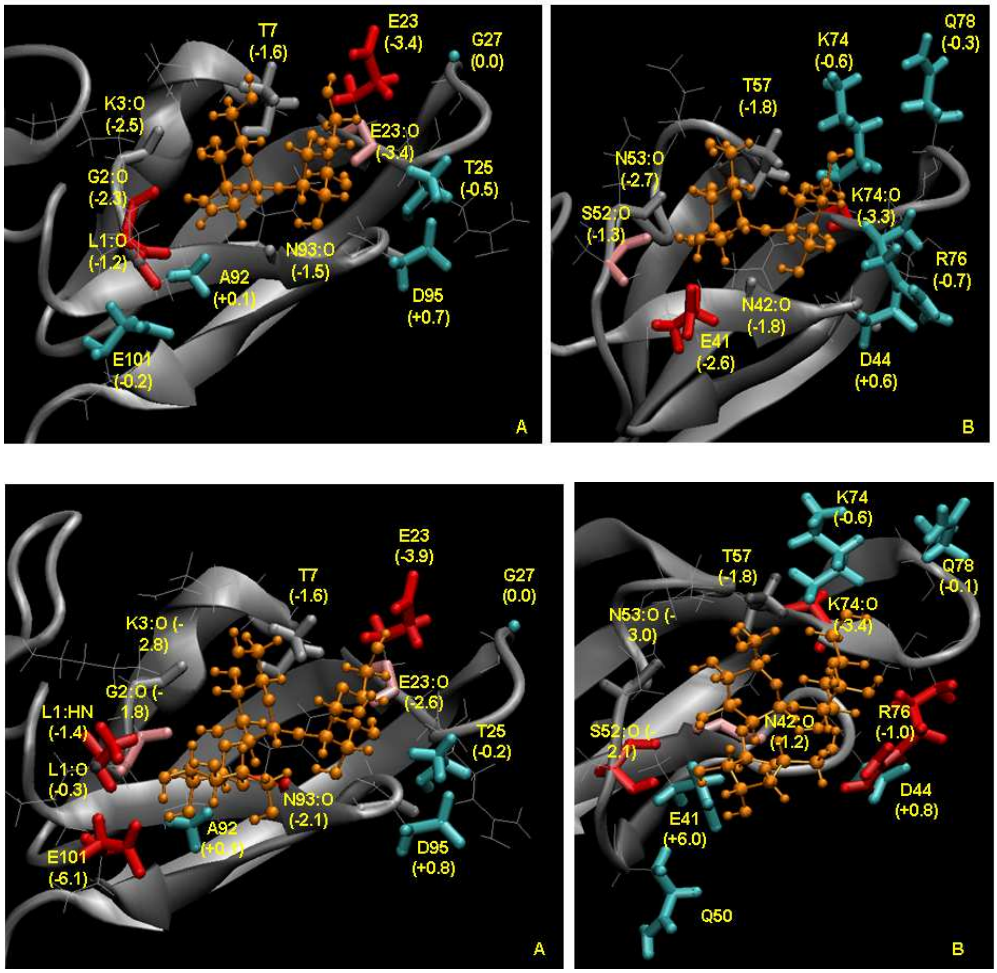

Figure 6. **Structures of CVN bound to dimannose (top) and preferred Manα1-2Manα1-2Man for domain A (bottom).** Values represented next to each residue correspond to the mutation energy. (Left images: Domain A and Right images: Domain B). Red residue: Strong interaction, Pink residue: Moderate interaction, Cyan: Poor interaction, Gray: Similar interaction in corresponding domains

For the other favorable trimannose structures in domain B computed binding affinity was more favorable than in domain A, including the preferred Manα1-2Manα1-2Man. The modeling studies suggested a critical role of Glu 41 in the tight binding of the sugar and possibly in the selectivity for Manα(1–2)Man that is common in all three arms of the trimannose structures in domain B. In the simulations, Arg 76 undergoes a conformational change that brings the side chain from an unlocked position, far from the ligand, to a locked position in which two direct hydrogen bonds to the ligand are observed. This change happens longer for the Manα1-2Manα1-2Man than the other trimannose models in domain B which explains for the offset in the linear fit (as shown in Figure 7). A combination of E41 and R76 can explain the favorable affinity for domain B. The corresponding resides A92 and T25 do not contribute to the binding affinity.

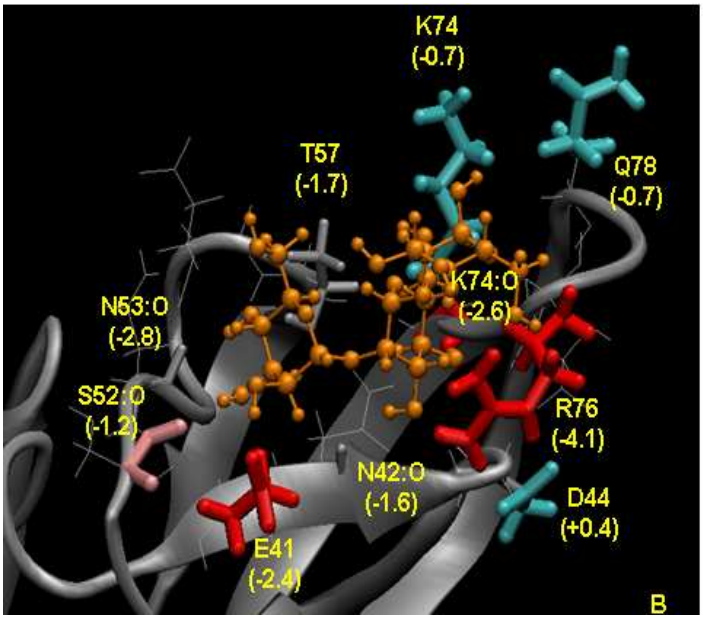

Figure 7. **Structure of the favorable Manα(1-2)α(1-2)Man for domain B.** All values represented next to each residue correspond to the mutation term in kcal/mol. Red residue: Strong interaction, Pink residue: Moderate interaction, Cyan: Poor interaction, Gray: Similar interaction in corresponding domains

Chapter 7

General Conclusions

The role of electrostatic energies and binding energies of the various CVN complexes were investigated in this thesis. A lot of information was analyzed by changing the environment of the simulation. By unrestraining the complex by placing it in a box of water and allowing the complex to be mobile, the calculations seemed to be better correlated. We were able to get a better trend for the other trimannose models.

Molecular dynamics was able to reveal a lot of information about the structure of the complex. Multiple molecular dynamic simulations were performed on several forms of cyanovirin-N bound to different carbohydrate models. We were able to capture different conformations of the hinge region from domain B that experiments were not able to detect from the wild type structure. By using this technique, we were able to improve the model and capture the trend of the experimental results.

The observed variations in binding affinity were mainly due to electrostatic effects. The electrostatic analysis showed important information for affinity and

specificity for the various CVN complexes. Although the computational approach to calculating the binding free energies were significantly larger than the experimental results, the observed trends were present. The electrostatic analysis for all preferred models in domain B revealed strong electrostatic effects from Glu 41, which is known to be important for specificity. The electrostatic interactions from the N and C termini of the preferred Manα1-2Manα1-2Man in domain A gave a favorable affinity over the other models in that domain.

In general, molecules can change conformation upon binding; thus, the present method can be extended to allow other terms (*e.g.* entropy, strain, etc). By adding these terms, the current overestimated values may get closer to the experimental results. Additionally, water-mediated interactions should be determined to see if any interactions might have enhanced the affinities. For example, since water is removed when the energies are calculated, we are not fully capturing the effects of the binding affinities. Maybe there is more information that can emerge from domain B by studying these effects. Also, in the future, we can take the information from the electrostatic analysis and apply it towards the design of mutants with altered affinity and specificity.

Bibliography

[1] Bewley, C.A. Solution structure of a cyanovirin-N:Manα1–2Manαcomplex: Structural basis for high-affinity carbohydrate-mediated binding. *Structure* **9:** 931–940, 2001.

[2] Bewley, C.A., K.R. Gustafson, M.R. Boyd, D.G. Covell, A. Bax, G.M. Clore, and A.M. Gronenborn. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat. Struct. Biol.* **5:** 571–578, 1998.

[3] Bewley, C.A., S. Kiyonaka, and I. Hamachi. Site-specific discrimination by cyanovirin-N for a-linked trisaccharides comprising the three arms of Man8 and Man9. *J. Mol. Biol.* **322:** 881–889, 2002.

[4] Bewley, C.A. and S. Otero-Quintero. The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man8D1D3 and Man9 with nanomolar affinity: Implications for binding to the HIV envelope protein gp120. *J. Am. Chem. Soc.* **123:** 3892–3902, 2001.

[5] Botos, I., B.R. O'Keefe, S.R. Shenoy, L.K. Cartner, D.M. Ratner, P.H. Seeberger, M.R. Boyd, and A. Wlodawer. Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J. Biol. Chem.* **277:** 34336-34342, 2002.

[6] Bourne, Y., P. Rouge, and C. Cambillaun. X-ray Structure of a Biantennary Oligosaccharide-Lectin Complex Refined at 2.3-A Resolution. *J. Biol. Chem.*, **267**: 197-203, 1992.

[7] Boyd, M.R., K.R. Gustafson, J.B. McMahon, R.H. Shoemaker, B.R. O'Keefe, R.J. Gulakowski, L. Wu, M.I. Rivera, C.M. Laurencot, M.J. Currens, J.H. Cardellina 2nd, R.W. Buckheit Jr., P.L. Nara, L.K. Pannell, R.C. Sowder 2nd, 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. *Antimicrob. Agents Chemother*. **41:** 1521–1530, 1997.

[8] Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, , D.J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4:** 187–217, 1983.

[9] Burton, D.R. Structural Biology: Images from the surface of HIV. *Nature*. **441**: 817-818, 2006.

[10] Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique (Cermav-CNRS), National Center for Scientific Research. http://www.cermav.cnrs.fr/lectines/.

[11] Chaplin, M. Water Models. http://www.lsbu.ac.uk/water/models.html.

[12] Currens, M. J., R. J. Gulakowski, J. M. Mariner, R. A. Moran, R. W. Buckheit, Jr., K. R. Gustafson, J. B. McMahon, M. R. Boyd. Antiviral activity and mechanism of action of calanolide A against the human immunodeficiency virus. *J. Pharmacol. Exp. Ther.* **279,** 645-651, 1996.

[13] Currens, M. J., J. M. Mariner, R. A. Moran, J. B. McMahon, M. R. Boyd. Kinetic analysis of inhibition of HIV-1 reverse transcriptase by calanolide A. *J. Pharmacol. Exp. Ther.* **279,** 652-661, 1996.

[14] Disney, M.D. and P.H. Seeberger. The Use of Carbohydrate Microarrays to Study Carbohydrate-Cell Interactions and to Detect Pathogens. *Chem & Biol*., **11**: 1701–1707, 2004.

[15] Fujimoto, Y.K., R.N. TerBush, V. Patsalo, D.F. Green. Computational models explain the oligosaccharide specificity of cyanovirin-N. *Prot. Sci*., **11:** 2008-2014.

[16] Gao, F., E. Bailes, D.L. Robertson, Y. Chen, C.M. Rodenburg, S.F. Michael, L.B. Cumminsk, L.O. Arthur, M. Peeters, G.M. Shaw, P.M. Sharp, and B.H. Hahn. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature.* **397:** 436-441, 1999.

[17] Gilson, M.K. and B. Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins Struct. Funct. Genet.* **4:** 7–18, 1988.

[18] Green, D.F. Optimized parameters for continuum solvation calculations with carbohydrates. *J. Phys. Chem. B* **112:** 5238–5249, 2008.

[19] Green, D.F. and B. Tidor. Evaluation of electrostatic interactions. *In Current protocols in bioinformatics* (ed. G.E. Petsko), chapter 8.3. John Wiley & Sons, Inc., New York, 2003.

[20] Holgersson, J., A. Gustafsson, and M.E. Breimer. Characteristics of protein-carbohydrate interactions as a basis for developing novel carbohydrate-based antirejection therapies. *Immunol. and Cell Biol.*, **83**: 694–708, 2005.

[21] Humphrey, W., A. Dalke, and K. Schulten. VMD—visual molecular dynamics. J. Mol. Graph. **14:** 33–38, 1996.

[22] Im, W., M.S. Lee, and C.L. Brooks III. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **24:** 1691–1702, 2003.

[23] Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79:** 926–935, 1983.

[24] Karplus, M. and J.N. Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14:** 325–332, 1981.

[25] Kuttel, M., J.W. Brady, and K.J. Naidoo. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J. Comput. Chem.* **23:**1236–1243, 2002.

[26] Lazaridis, T. Binding affinity and specificity from computational studies. *Curr. Org. Chem.* **6:** 1319–1332, 2002.

[27] Leach, A.R. *Molecular Modelling Principles & Applications,* 2nd edition; Pearson Education Limited: 2001;

[28] Leonard, C.K., M.W. Spellman, L. Riddle, R.J. Harris, J.N. Thomas, and T.J. Gregory. Assignment of interchain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **265:** 10373–10382, 1990.

[29] Lifson, J., S. Coutre´, E. Huang, and E. Engleman. Role of envelope glycoprotein carbohydrate in human immunodeficiency virus (HIV) infectivity and virus-induced cell fusion. *J. Exp. Med.* **164:** 2101–2106, 1986.

[30] Lis, Halina and N. Sharon. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev*., **98**: 637-674, 1998.

[31] Loris, Remy, S.P.G. Philippe, and L. Wyns. Conserved Waters in Legume Lectin Crystal Structures. *J. Biol. Chem*., **269**: 26722-26733, 1994.

[32] MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102:** 3586–3616, 1998.

[33] Margulis, C.J. Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-HIV virucidal protein cyanovirin. *J. Phys. Chem. B* **109:** 3639–3647, 2005.

[34] National Institute of Allergy and Infectious Diseases (NIAID). http://www3.niaid.nih.gov/.

[35] Nina, M., D. Beglov, and B. Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **101:** 5239–5248, 1997.

[36] Quiocho, Florante A. Probing the atomic interactions between proteins and carbohydrates. *Biochem. Soc. Trans.*, **21**: 442-448, 1993.

[37] Scanlan, C.N., J. Offer, N. Zitzmann, and R.A. Dwek. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature* **446:** 1038–1045, 2007.

[38] Sharon, N. Lectin-carbohydrate complexes of plants and animals: an atomic view. *Trends Biochem. Sci.*, **246**: 221-226, 1993.

[39] Sharon, Nathan and H. Lis. Lectins as Cell Recognition Molecules. *Science*, **246**: 227-234, 1989.

[40] Sheinerman, F.B., R. Norel, and B. Honig. Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **10:** 153–159, 2000.

[41] Sitkoff, D., K.A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98:** 1978–1988, 1994.

[42] Srinivasan, J., T.E. Cheatham, P. Cieplak, P.A. Kollman, and D.A. Case. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **120:** 9401–9409, 1998.

[43] Tsai, C.C., P. Emau, Y. Jiang, B.P. Tian, W.R. Morton, K.R. Gustafson, and M.R. Boyd. Cyanovirin-N as a topical microbicide prevents rectal transmission of SHIV89.6P in macaques. *AIDS Res. Hum. Retroviruses* **19:** 535–541, 2003.

[44] Tsai, C.C., P. Emau, Y. Jiang, M.B. Agy, R.J. Shattock, A. Schmidt, W.R. Morton, K.R. Gustafson, and M.R. Boyd. Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res. Hum. Retroviruses* **20:** 11–18, 2004.

[45] World Health Organization (WHO). http://www.who.int/hiv/data/en/.

[46] Yang, F., C.A. Bewley, J.M. Louis, K.R. Gustafson, M.R. Boyd, A.M. Gronenborn, G.M. Clore, and A. Wlodawer. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J. Mol. Biol.* **288:** 403–412, 1999.