राष्ट्रीय प्रतिरक्षाविज्ञान संस्थान
NATIONAL INSTITUTE OF IMMUNOLOGY

CERTIFICATE

This is to certify that the thesis entitled "**Computational analysis of the structural stability and substrate specificity of legume lectins**" submitted by Sandeep Kaushik in partial fulfillment of Ph.D. degree of Jawaharlal Nehru University, New Delhi comprises the work done by the scholar under my guidance at National Institute of Immunology, New Delhi. The work is original and has not been submitted before, in part or full for any degree or diploma of any university.

Debasisa Mohanty
Dr. Debasisa Mohanty
Thesis Supervisor
National Institute of Immunology
New Delhi

Avadhesha Surolia
Prof. Avadhesha Surolia
Thesis Co-Supervisor
National Institute of Immunology
New Delhi

अरूणा आसफ अली मार्ग, नई दिल्ली - 110067 फैक्स : 91-011-26742125, 26742626 तार : IMMUNOLOGY
Aruna Asaf Ali Marg, New Delhi- 110067 Fax : 91-011-26742125, 26742626 Grams : IMMUNOLOGY
दूरभाष (Telephones) EPABX : 26717121 - 45 & 26717010 - 19

Computational analysis of the structural stability and substrate specificity of legume lectins

Thesis submitted to Jawaharlal Nehru University for the degree of

Doctor of Philosophy

May 2011

Sandeep Kaushik

National Institute of Immunology

Aruna Asaf Ali Marg, New Delhi, INDIA

Dedicated to

My *Parents* and My Sunshine,

My Son *Anand*

Acknowledgements

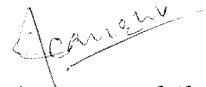
"God .... Even if He doesn't exist in solid form, He makes His presence felt every single day, in every color, in everything that exists."

I acknowledge Him as the One and only, responsible for everything, not only on this paper but in my heart, every single moment I breathe. I look back and I ask myself, how can I acknowledge Him when I have never really revered Him the way He should be? I always tried to defy Him, thinking of His existence as any fool would, getting unruly every day. But, then came days when I realized that my attitude generated nothing but troubles for everybody around me. And yet I still had things which made me feel happy! And so I realize that the day has come when I acknowledge Him for whatever I am and I have.

He gave me two great gems of my life whom anybody would hold close to their heart, my parents. My father seemingly a lean and frail man, is as strong as granite. He is the only machine that has the power to keep going on, overcoming all the obstacles. While he is still around, I can live without dying inside but when he will be gone he will leave his fire inside me which will burn forever until I fizzle out. I am so sorry that have been yet another portal of problems for him. I could not fathom what I brought upon both of them. When my children would do the same to me will I truly realize the cureless pain that I gave to my parents.

The other great blessing that God gave me was in the form of my mother. She is the one lady I truly love, who is worth love itself. She is like gold, all the fire (of problems) that came her way, made her shine more and more. She has fought for us like any lioness would fight for her cubs. She crossed the desert yet keeping us safe from every hot wind and every hot grain of sand that leapt for us. I am born of her and to God I say that every drop of my blood is grateful to my mother. She is worth every sacrifice I could ever make. If today I am writing my thesis, it's because of for her efforts and her love.

All of my existence thanks God to have given me a life full of people who have been helpful in making me what I am, especially, my wife. She has been instrumental in turning me from that boy to this man. I would like to thank my dearest friends at home like Yogesh, Vinod and Radha along with my mates at work namely, Deepak (Sharma), Bhushan, Swadha, Garima (Gary), Nikhil (Niks), Shradha (Shreds) and Deepak (Mr. Photogenic) for bearing my idiosyncrasies and madness. I would like to thank Dr. Mohanty and Prof. Surolia, and even Dr. Vidya, for bearing the insolent attitude of a bonehead like me and for pushing me in a direction they thought was right for me. I wish I could do better for them and for everybody else. Finally, I wish this thesis could be of any good to somebody.



Sandeep Kaushik
May 2011

Table of Contents

Abbreviations

ABP	Abrin derived peptide
AGP	Agglutinin derived peptide
AMBER	Assisted Model Building with Energy Refinement
AMML	*Astragalus mongholicus lectin*
APF	Atomic Positional Fluctuations
ASA	(solvent) Accessible Surface Area
BF	B-factors
CancerLectinDB	Cancer Lectin Database
CHARMm	Chemistry at Harvard Molecular Mechanics
CINC	Cytokine-Induced Neutrophil Chemoattractant
COM	Centre of Mass
ConA	Concanavalin A
DNA	Deoxyribonucleic acid
DrosL	*Dioclea rostrata* lectin
DSC	Differential Scanning Calorimetry
EcorL	*Erythrina corallodendron* lectin
FF03	Force field '03
FEP	Free-Energy perturbation
FES	Free-Energy Surface
FSH	Follicle Stimulating hormone
FU-PMPC	Fucosylated PMP-C
GalNac	N-acetylgalactosamine
GalNDns	N-Dansylgalactosamine
GB-SA	Generalized Born-Surface Area
GlcNac	N-acetylglucosamine
GPU	Graphics Processing Unit
GROMOS	Groningen Molecular Simulation
GROMACS	Groningen Machine for Chemical Simulations
GS-IV	*Griffonia simplicifolia* – IV
HCG (hCG)	Human Chorionic Gonadotropin
HeLa cells	A cancerous cell line from Henrietta Lacks
HIV-1	Human Immunodeficiency Virus – 1
HRP	Horseradish Peroxidase
IFN- gamma	Interferon-gamma
IL	Interleukin
kDa	kilo Daltons
LMCI	Locust Trypsin Inhibitor
LectinDB	Lectin Database
LES	Locally Enhanced Sampling
LH	Luteinizing hormone
Man	α-D-Mannose
M2M	Man(α1-2)Man (α1-*O*)Me
M3M6M	Man(α1–3)[Man(α1–6)]Man(α1-*O*)Me

MBP	Mannose Binding Protein
MC	Monte Carlo
MD	Molecular Dynamics
MG63	Human Osteosarcoma Cells
MM	Molecular Mechanics
MMFF	Merck Molecular Force Field
NacLac	N-acetyllactosamine
NAcGlc	N-acetylglucosamine
NF-PMPC	Non-fucosylated PMP-C
NMR	Nuclear Magnetic Resonance
NO	Nitric Oxide
OPLS(-AA)	Optimized Potential for Liquid Simulations (-All Atom)
PBC	Periodic Boundary Conditions
PB-SA	Poisson-Boltzmann-Surface Area
PDB	Protein Data Bank
PES	Potential Energy Surface
PME	Particle Mesh-Ewald
PMF	Potential Mean Force
PMP-C	*pars intercerebralis* major peptide-C
PNA	Peanut agglutinin
PSP-1	Porcine Seminal Plasma Spermadhesin -1
RBC	Red Blood Cell
RESP	Restrained Electrostatic Potential
Rg	Radius of Gyration
RMSD	Root Mean Square Deviation
rRMS	residue-wise RMSD
RNA	Ribonucleic acid
RNase	Ribonuclease (enzyme)
rEcorL	recombinant *Erythrina corallodendron* lectin
TIP(3,4,5)P	Transferable Intermolecular Potential (3,4,5)Point
TNF	Tumor Necrosis Factor
TSH	Thyroid-stimulating hormone
UDA	*Urtica diocia* agglutinin
vdW	van Der Waals (interactions)
WGA	Wheat germ agglutinins
Xyl	α-D-Xylose

Introduction

Prologue

Polypeptides (proteins), polysaccharides (sugars) and polynucleotides (nucleic acids) are three types of molecules that form the main pillars of life especially in biochemical terms. Various biological processes involve interactions between different types of molecules which constitutes a flow of energy essential for life. Different processes occurring in a particular organism depend directly on what constituent molecules are available to participate. It has been very well established now that the information about molecules constituting an organism is stored in polynucleotides, better known as DNA (Deoxyribonucleic acid) and RNA (Ribonucleic acid). Further, given the recent advances in biology, the information stored in DNAs and RNAs could be easily read and decoded. Thus, the focus of scientific research seems to have shifted towards understanding the biomolecular interactions. An enormously large number of interactions are possible in an organism and understanding them seems daunting. Therefore, studying the representative cases is a preferred choice in scientific research. This thesis aims to understand protein-carbohydrate interactions using model proteins from a special class of proteins, called as lectins, and their substrate sugars.

Protein-carbohydrate interactions

Proteins are the effector molecules of a cell participating in various vital processes like enzymes in metabolism, receptors in signal recognition, transporters in translocation of molecules, *etc*. Carbohydrates on the other hand are largely inert molecules. However, they are involved in structural constitution and modification of various biomolecules. For example, glycosylation is a post-translational modification where a carbohydrate entity is covalently attached to a protein or a lipid molecule. Glycosylation has been known to modify structural and functional aspects of many proteins. Protein-carbohydrate interactions are of immense importance because of their involvement in some of the most important cellular processes, like signal recognition and metabolism. The best known model to study protein-carbohydrate interactions are lectins. Therefore, most of the reported literature on such interactions has focused on lectins. The model proteins in this work also involve (legume) lectins namely, Concanavalin A (ConA) and *Erythrina corallodendron* lectin (EcorL) besides a small glycosylated peptide also known as PMP-C.

(Legume) Lectins

Lectins are a class of proteins that specifically bind carbohydrates as their substrates. They have been found to exist almost ubiquitously, including bacteria, fungi, plants and animals. Those derived from leguminous plants like the pulses and beans are called legume lectins. Lectins have also been an important class of proteins in biological research. In view of their property of agglutinating blood cells, lectins have been used in various therapeutical applications like, blood banking and genotype determination. Besides this they have been used to study the suppression of first and second degree immune responses, lymphocyte-stimulation, study of tumor cell surface organization, *etc.*

Lectins have been implicated in a number of biological processes though as of now no enzymatic activity has been assigned to them. However, the lectin-polysaccharide precipitin reaction has been found similar to the antigen-antibody precipitin reaction. In view of their important role as recognition molecules, lectins have also been called as 'decoders of the glycocode'. In order to understand the physiological and immunological properties of the lectins, a large number of studies have focused on their structure and sugar binding ability. As a result, currently a large number of crystal structures of lectins are available in the structural database for proteins, known as PDB (Protein Data Bank). For ConA itself, a large number of crystal structures have been determined. Besides this, a huge amount of data has been reported on lectins from various biochemical and biophysical studies. However, many aspects of lectin structure could not be elucidated even with a large number of crystal structures. For example, no atomically detailed picture could be obtained on conformational changes occurring upon demetalation in the ion binding loop region of ConA.

Molecular Dynamics Simulations

X-ray crystallography is a powerful technique to study biological structures and is capable of providing a high resolution static image of a system. However, techniques which provide time-resolved information of a system are more suitable to observe its live image. One such approach, capable of providing a time-evolved picture of a system, is molecular dynamics (MD) simulations. The biological relevance of the results obtained from MD simulations depends on whether simulations mimic a system properly? This in turn depends on the reliability of various simulation parameters. Simulation parameters have been improved continuously, making the data obtained from MD simulations even more reliable in biological terms. Further, with the accessibility of increased computational power nowadays,

it is now possible to run simulations for biologically relevant time scales. This makes MD simulations a method of choice to analyze biological problems, whenever possible and wherever applicable. Many studies have provided further insights into a system by using MD simulations to complement methods like X-ray crystallography and NMR. Therefore, we have used MD simulations to obtain further insights on topics in lectin biology which have been elusive for long.

Concanavalin A: Conformation of ion binding loop upon demetalation

ConA is the first lectin whose crystal structure was determined and consequently, is a well-studied legume lectin as well. It is a tetrameric lectin with each of its monomers consisting of 237 residues. ConA can bind to both mannose and glucose as its substrates though mannose is about five times stronger a substrate than glucose. The presence of divalent metal ions is absolutely essential for substrate binding to ConA. Mostly, Mn^{2+} and Ca^{2+} have been observed in the metal binding sites of ConA. As a consequence to demetalation, ConA loses its substrate binding ability with no effect on its gross tertiary structure. Using crystallographic analysis, it was observed that in demetalized ConA, the metal ion binding loop and certain portions of substrate binding loops could not be observed due to a higher mobility compared to native ConA. In order to understand the sequential changes in the metal ion binding site upon metalation, several crystal structures of demetalized, partially and fully metalized ConA were determined. However, no concrete information could be obtained on the effects of demetalation/metalation on the conformation of the ion binding and substrate binding loops of ConA. The conformational changes in ion binding loop upon demetalation have remained elusive. Therefore, it is necessary to obtain an atomically detailed picture of demetalation of ConA using a technique other than X-ray crystallography. We have simulated ConA under different compositions i.e. with and without the presence of both metal ions and the substrate trimannoside. The methodology and results have been described in Chapter 2.

Erythrina corallodendron lectin (EcorL): Role of glycosylation

EcorL is a dimeric legume lectin derived from *Erythrina corallodendron* (coral tree). Each monomer is composed of 242 amino acids and has the typical legume lectin fold. It is also the first glycosylated lectin whose crystal structure was solved. The mode of

oligomerization in lectins crystallized before EcorL involved a sideways arrangement of monomers and is known as the "canonical mode" of dimerization. EcorL dimerized differently, with its monomers associating to build an interface like a handshake. Therefore, this mode of oligomerization is known as the "handshake" mode. Using its crystal structure, the site of glycosylation in EcorL was observed to be located at the dimeric interface. This suggested that the oligosaccharide hindered the canonical mode of association by forcing the monomers to associate in an altered mode. However, the crystal structure of recombinant non-glycosylated EcorL (rEcorL) revealed that rEcorL and EcorL had almost identical tertiary structures and even in absence of the covalently attached oligosaccharide, rEcorL dimerized using the handshake mode. This ruled out the role of the oligosaccharide in EcorL dimerization. Therefore, the role of oligosaccharide in EcorL still remained unanswered.

Using biophysical studies it was observed that glycosylated form of EcorL was relatively more stable compared to the non-glycosylated form, rEcorL. Therefore, it appeared that in the absence of any other structural difference among the two forms except glycosylation, the differences in structural stability may originate from the covalently attached oligosaccharide. There have been no reports yet that have resolved the role of the oligosaccharide in enhancing the structural stability of EcorL. Hence, we have studied the role of glycosylation in EcorL using MD simulations. We have simulated both EcorL and rEcorL in order to compare the two forms. The results have been described in Chapter 3 of this thesis.

Role of glycosylation in glycopeptides (A case of PMP-C)

It has been observed that many globular proteins show enhanced structural stability and other physicochemical changes upon glycosylation. One such positive effect imparted by glycosylation is the protection of a polypeptide from proteolytic digestion. For example, therapeutic peptides can be glycosylated to protect them from proteolytic digestion.

Serine protease inhibitors are highly important for normal functioning of cellular processes because they regulate the protease activity in a cell. One such inhibitor is known as *pars intercerebralis* major peptide-C (PMP-C). It is obtained from *pars intercerebralis* region of the brain of common locust *Locusta migratoria*. It is composed of 36 amino acid residues, six cysteines of which form 3 disulfide bonds and provide it a rigid structure. PMP-C exists in two forms, glycosylated and non-glycosylated. The glycosylated form is more stable than the non-glycosylated form, with no other structural difference than glycosylation. Besides the

difference in structural stability, the repertoire of target proteases also differs among the two forms. In order to understand the relationship between structural stabilization and glycosylation and to locate the origin of the biological differences seen among the two forms of PMP-C, we need to perform an atomic level detailed comparison between the two forms. We have performed MD simulations on both forms of PMP-C and compared them. Chapter 4 describes the results obtained from analysis of the MD trajectories.

Chapter 1

Review of Literature

1.1 Lectins

Lectins are proteins that possess at least one non-catalytic domain capable of binding reversibly to specific monosaccharides or oligosaccharides and consequently, have the ability to agglutinate erythrocytes with known carbohydrate specificity (Lam and Ng, 2011). Lectins were first discovered in plants but, later, found to be ubiquitous in nature (Ashwell and Harford, 1982; Rini, 1995b; Sinha *et al.*, 2007). The word lectin comes from Latin *legere* which means to choose or pick out, in order to refer to the selectivity of lectins (Boyd and Shapleigh, 1954). Lectins display a wide range of carbohydrate specificity along with a superb ability to detect even subtle variations in substrate molecules (Sharma and Surolia, 1997).

1.1.1 Lectins in Biology and Medicine

The role of lectins in biology and medicine originates from their carbohydrate binding specificity which leads to interesting biochemical and biophysical properties, most notably their ability to hemagglutinate. This has led to a large number of studies on lectins (Agrawal *et al.*, 2010; Batista *et al.*, 2010; Bies *et al.*, 2004; Cheng *et al.*, 2010; Elgavish and Shaanan, 1997; Fang *et al.*, 2010b; Jiang *et al.*, 2010; Rouge *et al.*, 2010; Sharon and Lis, 2004; Teixeira-Sa *et al.*, 2009). Consequently, a variety of activities have been attributed to lectins including antitumor (Alencar *et al.*, 2010), immuno-modulatory (Araujo-Filho *et al.*, 2010; Lam and Ng, 2010), antifungal (Fang *et al.*, 2010a; Wu *et al.*, 2010), HIV-1 reverse transcriptase inhibitory (Lagarda-Diaz *et al.*, 2009; Singh *et al.*, 2009), anti-insect (Assreuy *et al.*, 2009; Lam and Ng, 2011; Li *et al.*, 2010; Redondo and Alvarez-Pellitero, 2010; Zhang and Monteiro-Riviere, 2010), antibacterial and anti-nematode activities, which may find practical applications (Herman *et al.*, 1985). Animal lectins *viz.* siglecs (sialoadhesins), MBP (mannose binding protein), galectins and tachylactins, *etc.* have been implicated in both intracellular (pre-mRNA splicing, protein folding and transport) and extracellular interactions with juxtaposed cells and matrix *e.g.* host immunity, lymphocyte adhesion, homing and apoptosis (Dagher *et al.*, 1995; Rini and Lobsanov, 1999; Vyakarnam *et al.*, 1997; Yang *et al.*, 1996; Yang and Liu, 2003). In view of the importance of lectins in biology and medicine, several integrated knowledge-based resources like Lectindb (Chandra *et al.*, 2006) and CancerLectinDB (Damodaran *et al.*, 2008) have been developed for comprehensively cataloging various lectins and their sequences, structure, function and biological activities.

1.1.2 Lectin Structure

Lectins are mainly composed of β-sheets as the major secondary structural element, however, some lectins do have additional helical regions (Bouckaert *et al.*, 1999a; Sinha *et al.*, 2007; Wright, 1997). Even though conserved monomeric folds have been observed for diverse plant (legume) as well as animal (galectins, pentraxins and spermadhesins) lectins, the monomeric forms may oligomerize in several different arrangements leading to a variety of quaternary structures (Bouckaert *et al.*, 1999a). The tertiary and quaternary structure of legume lectins has been discussed below.

1.1.3 Tertiary Structure of (Plant) Lectins

Plant lectins mainly have three kinds of β-sheet architectures *viz.* barrel, jelly roll and hevein domain (Sinha *et al.*, 2007).

Barrel Architecture: The barrel architecture (Figure 1.1) has been categorized further as β-prism I, β-prism II and β-trefoil folds. In the β-prism fold, four-stranded β-sheets are arranged on the sides of an equilateral triangle either aligning parallel (β-prism I; atrocarpin, banana and Jacalin lectins) or perpendicular (β-prism II; garlic and snowdrop lectins) to the 3-fold symmetry axis of the triangular plane. Such structures have a stabilizing core region filled with hydrophobic residues (Bennett *et al.*, 1995; Sankaranarayanan *et al.*, 1996). The subunit association or oligomerization in β-prism I occurs through hydrogen bonding and hydrophobic interactions whereas in β-prism II, the strands cross-over to form hybrid β-sheets forming fused oligomers. During oligomerization of both β-prism I and β-prism II folded subunits, the burial of hydrophobic residues takes place to achieve oligomeric stability.

Jelly-Roll Fold: The jelly-roll fold or legume lectin fold (Figure 1.2), first observed in Concanavalin A (ConA), is a β-sandwich containing three β-sheets namely, back (6 stranded, flat), front (7 stranded, curved) and top (5 stranded). This arrangement of β-sheets leads to formation of two hydrophobic cores in the molecule which are known as the primary hydrophobic core (sandwiched between back, front and top β-sheets) and secondary hydrophobic core (sandwiched between front β-sheet and the loops that connect its strands).

Hevein Domain: The hevein domain (Figure 1.3) consists of 43 amino acids and is characteristic to lectins involved in plant defense against microbes and invertebrates (Peumans and Van Damme, 1995). Examples of such lectins, involved in plant defense, are *e.g. Urtica diocia* (UDA) and wheat germ (WGA) agglutinins. The fold name originates from a plant lectin known as "hevein". Hevein domains have a core composed of 3-5 disulfide

bonds formed by conserved cysteines (Rodriguez *et al.*, 1988). For example, WGA has a subunit composed of 171 residues folded into four repetitive hevein domains each having 42-43 residues. Hevien like domains of WGA have an RMS deviation of about 1.15Å with respect to hevein itself.

1.1.4 Quaternary Structure (Oligomerization)

Monomeric subunits of lectins are capable of binding their substrate carbohydrates but with weak interactions. It has also been noticed that almost all plant lectins exist as oligomers *e.g.* all legume lectins are either dimers or tetramers. Not only lectins exist as oligomers, they show a diverse array of different modes of subunit association. Probably, the purpose of having diverse quaternary structures, produced using a variety of subunit organization, is to generate higher order sugar specificities. The multivalency thus generated is a characteristic feature of many but not all lectin types and seems to be a suitable solution to the higher intrinsic flexibility of carbohydrate substrates (Bouckaert *et al.*, 1999a; Rini and Lobsanov, 1999). It ultimately leads to high affinity interactions of the oligomeric lectins with multivalent ligands which otherwise in their monomeric forms are weak interactions.

The oligomerization or quaternary association of monomers occurs mainly through the back β-sheet. Different orientations of the interacting sheets result in seven different types of interfaces observed in various legume lectins (Sinha *et al.*, 2007). Various types of interfaces like "canonical" interface of ConA (Hardman and Ainsworth, 1972), "handshake" interface of EcorL (Shaanan *et al.*, 1991), "X4" of *Griffonia simplicifolia* lectin (Delbaere *et al.*, 1993), "open" interface in peanut agglutinin (PNA) (Banerjee *et al.*, 1996) and several others have been observed in legume lectins. The quaternary structures in various lectins are often found to be similar even in absence of any evolutionary relationship. For example, Lentil lectin, porcine seminal plasma spermadhesins (PSP-1), and human galectin-1 form a canonical legume lectin dimer whereas peanut agglutinin (PNA) and PSP-1-PSP-II heterodimer resemble "open" quaternary structure of GS-IV lectin (Bouckaert *et al.*, 1999a).

1.1.5 Substrate Binding and Specificity

Another major focus in lectin research has been the elucidation of stereo-chemical aspects of substrate recognition which have been studied using various biochemical and structural approaches (Li *et al.*, 2010; Robles *et al.*, 2010). The classical lectin families

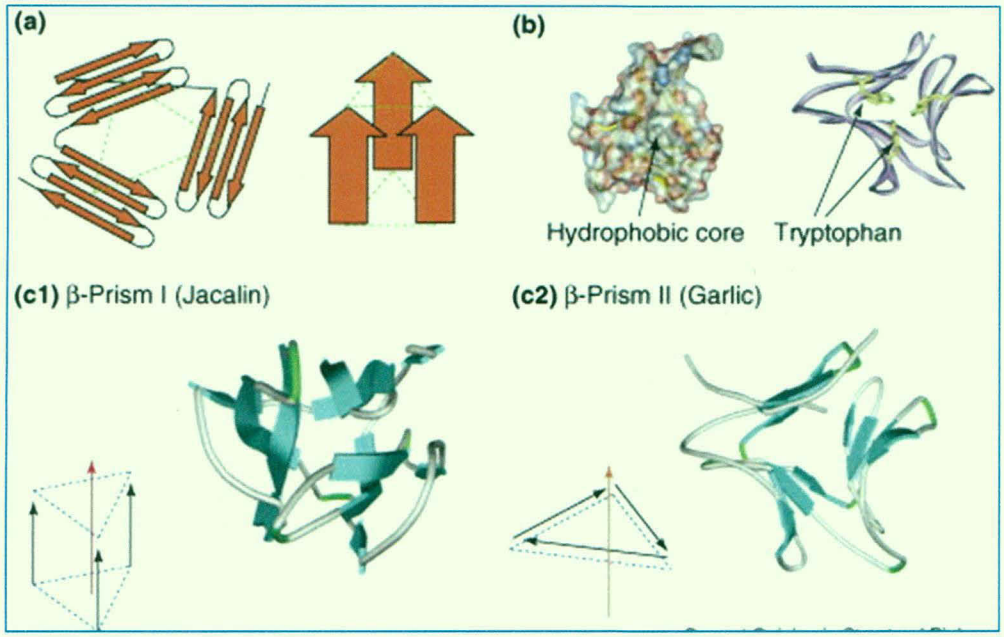


Figure 1.1
Lectins with β-prism structure. **(a)** Schematic representation of 12 stranded β-prism forming three β-sheets in a triangular arrangement. **(b)** Conserved residues in the center of the β-prism form a hydrophobic core with conserved tryptophans. **(c)** β-prism I and II folds. Adapted from (Sinha *et al.*, 2007).

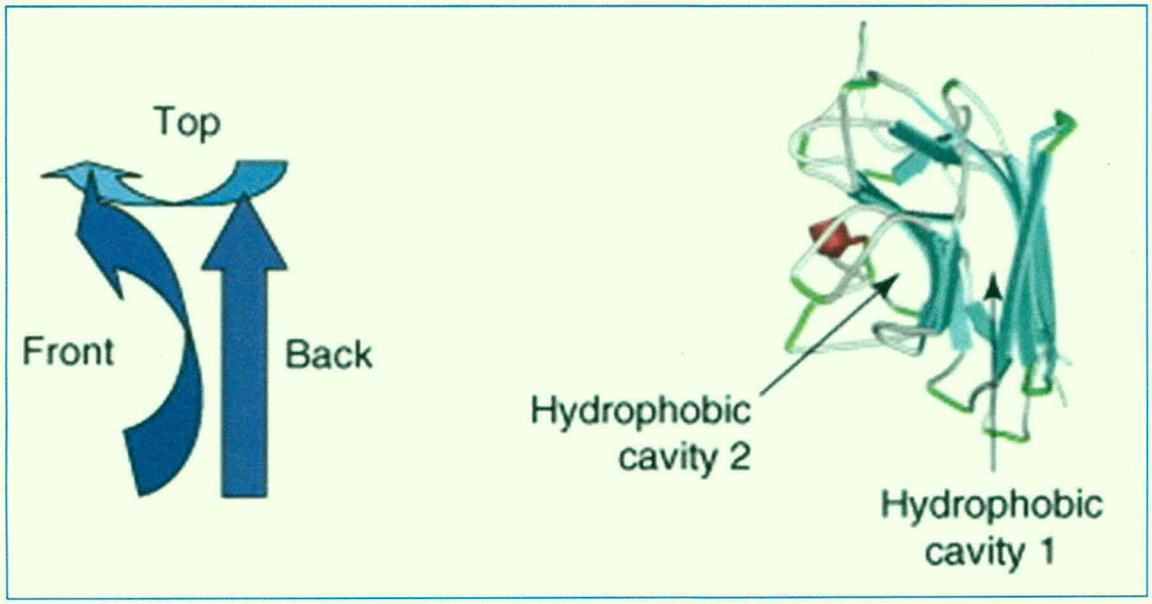


Figure 1.2
Legume lectin or 'jelly-roll' fold. The three anti-parallel β-sheets, 'back', 'front' and 'top' β-sheets, form a β-sandwich. Two hydrophobic cores stabilize this fold. Adapted from (Sinha *et al.*, 2007).

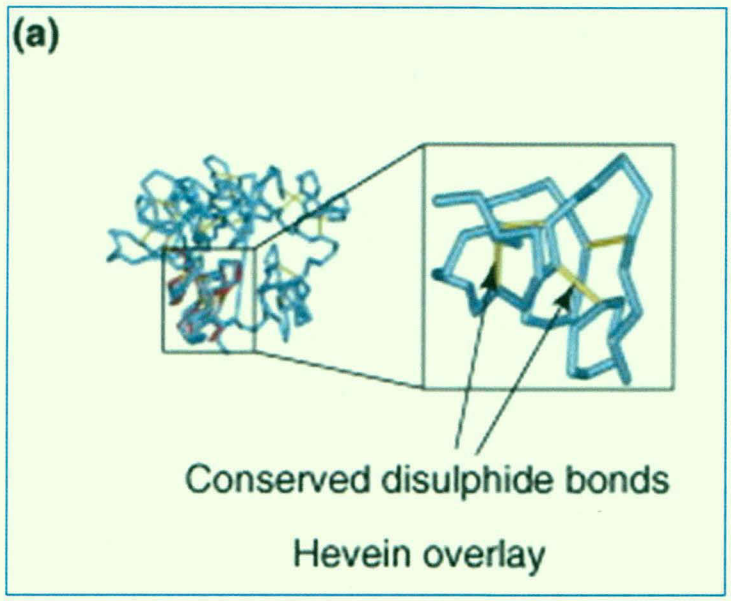


Figure 1.3
Hevein domain. (a) An overlay of single subunit of hevein (red) with one domain of WGA (blue). An expanded form of the domain displaying the conserved disulfide bonds (yellow). Adapted from (Sinha *et al.*, 2007).

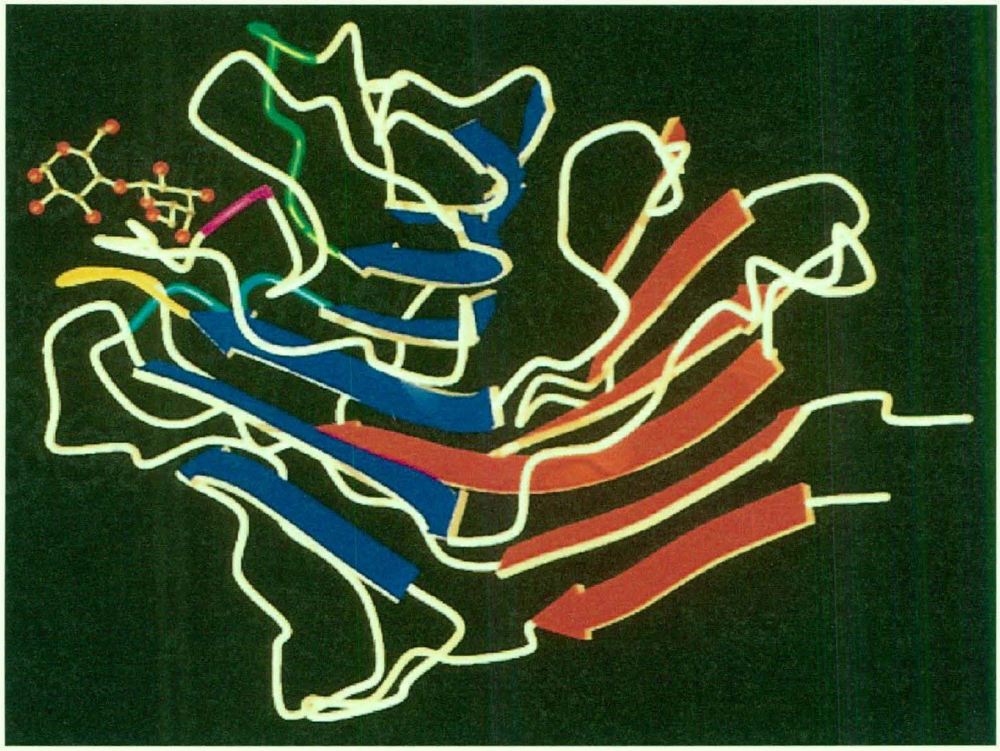


Figure 1.4
3-D structure of peanut agglutinin (PNA) monomer. The back and front β-sheets have been shown as red and blue strands, respectively. Also shown are the saccharide (lactose; ball-and-stick) and four substrate binding loops: A (cyan), B (purple), C (green) and D (yellow). Adapted from (Sinha *et al.*, 2007).

typically have shallow carbohydrate binding sites and in terms of the topology of the binding site they resemble other carbohydrate-binding proteins (Rini, 1995a). For carbohydrate recognition and binding, lectins employ an extensive hydrogen bonding network (due to the abundance of hydroxyl groups on carbohydrate molecules), van der Waals and hydrophobic interactions (with aromatic residues), with each of them playing a significant role (Elgavish and Shaanan, 1997). The key interactions for carbohydrate recognition are common among all lectins, however, different families possess a unique stereochemistry at the combining site for discrimination between ligands (Sharma and Surolia, 1997). Analysis of the crystal structure of substrate bound lectins has revealed certain conserved features of their carbohydrate binding sites. Four key hydrogen bond interactions between the carbohydrate and lectin are provided by a pair of highly conserved Asn and Asp residues, along with a Gly or Arg residue. This setup is further stabilized by stacking interactions with a hydrophobic residue (Phe, Tyr, Trp or Leu). The carbohydrate binding site is formed by four loops *viz.* loop A, B, C and D (Figure 1.4). Loops A, B and C provide the conserved residues namely, Asp, Gly or Arg, Asn and hydrophobic residues. Loop D contours the variable edge of substrate binding site and has a larger role in influencing the substrate specificity than the substrate binding (Sharma and Surolia, 1997). For example, a consensus sequence motif QXDXNXVXY was found to be essential for mannose binding in legume lectins, however, the topology of the backbone in the neighborhood was also found to contribute to mannose binding capability (Ramachandraiah and Chandra, 2000). It has been suggested that size of loop D determines the carbohydrate specificity of legume lectins with the size of loop C being additional contributing factor (Sharma and Surolia, 1997). Therefore, the substrate specificity of a legume lectin could possibly be engineered by suitable alteration of residues in loop D.

1.2 Concanavalin A (ConA)

1.2.1 Discovery of ConA

ConA is legume lectin derived from the seeds of jackbean (*Canavalia ensiformis*). Probably, the earliest reported work on ConA is the isolation of jack bean proteins (Jones and Johns, 1916). Two proteins were isolated from jackbean seeds which differed in their sulfur content and solubility in ammonium sulfate and were called *canavalin* and *concanavalin.* Later, based on their solubility differences the concanavalin content of the jack bean proteins was further divided into concanavalin-A and -B (Sumner, 1919). Sumner was the first to crystallize and characterize ConA (Sumner, 1919). The sugar binding ability of ConA were

first observed when during the amino acid constitution analysis of ConA, a pentosan was reported tightly bound to it (Jones *et al.*, 1924) which was also observed in ConA crystals (Sumner and Graham, 1925). ConA could agglutinate starch, glycogen, yeast, bacteria, RBCs *etc.* but cane sugar prevented the agglutination reaction for both starch and erythrocytes (Sumner *et al.*, 1935; Sumner and Howell, 1936a). The precipitation reaction was similar to the antigen antibody reaction with the precipitate formed being soluble in both acid and alkali (Sumner and Howell, 1936a). ConA could agglutinate erythrocytes in pH range of 5.2-7.5 with its isoelectric point being 7.1 (Agrawal and Goldstein, 1967b).

1.2.2 Interaction with RBCs, microorganisms and eukaryotic cells

The ability of ConA to hemagglutinate RBCs was discovered when Urease was erroneously reported to be the agglutinin from jack bean (Hotchkiss and Tauber, 1931). However, it was later discovered that jackbean agglutinin was in fact ConA (Sumner and Howell, 1935; Sumner *et al.*, 1935). Lectins from different sources were observed to have different "affinity" for different blood groups. Therefore, a large number of lectins were tested for their blood group specificity so that they could be used for medical purposes (Allen and Brilliantine, 1969).

ConA was found to agglutinate rice and corn starch, baker's yeast, mucoproteins, glycogen, certain bacteria and RBCs of certain animals but the mechanism of agglutination was still unclear (Sumner and Howell, 1935; Sumner and Howell, 1936a). However, from ConA's interaction profile it was infered that its interacting partner might be a carbohydrate group either free or attached to a protein or lipid molecule (Sumner and Howell, 1936a). The cell wall of many microorganisms has complex polysaccharides as a major constituent. Consequently, ConA was found to interact with the surface of several microorganisms (Markowitz, 1969; Tkacz *et al.*, 1971). It was also observed that ConA could agglutinate transformed cells (simian virus 40 infected 3T3 cells and polyoma infected rat cells) without any reaction with normal cells and this interaction was reversible in presence of α-methyl-D-mannoside (Inbar and Sachs, 1969).

ConA and other lectins like, favin and phytohemagglutinin were found to be mitogenic as their binding to lymphocytes was associated with the release of a colony stimulating activity factor (Jones, 1972; Ruscetti and Chervenick, 1975; Sela *et al.*, 1975a; Wang *et al.*, 1975). Its interactions with the cell membrane of lymphocytes have also been studied (Yahara and Edelman, 1972). ConA induces liver injury through activation of specific T-lymphocytes

when injected to mice (Trautwein *et al.*, 1998). ConA was also found to be a more potent molecule than antibodies in fighting HIV infection *in vitro* as it was found capable of binding to envelope proteins from different strains of HIV-1 (Pashov *et al.*, 2005; Witvrouw *et al.*, 2005). ConA's ability to bind oligosaccharides has also been utilized in designing specific molecules used for different medical purposes (Yuasa *et al.*, 2007).

Recently, a lectin (AMML) from the roots of *Astragalus mongholicus* was extracted, purified and its effects on cell proliferation, apoptosis and cell cycle of Human cervical carcinoma cell line (HeLa), human osteoblast-like cell line (MG63) and human leukemia cell line (K562) were studied. Along with growth inhibition for HeLa (92%), K562 (84%) and MG63 (48%) cells, the morphological observations showed that AMML-treated HeLa cells displayed outstanding apoptosis characteristics, such as nuclear fragmentation and appearance of membrane-enclosed apoptotic bodies (Yan *et al.*, 2009). Similarly, DrosL, a lectin isolated from *Dioclea rostrata* seeds, was found to induce neutrophil migration, through release of NO and cytokines such as IL-1beta, TNF-alpha and CINC-1 from macrophages (Figueiredo *et al.*, 2009). In vitro immuno-stimulatory effects of Abrus lectins derived peptide fractions (AGP and ABP) were observed in mice bearing Dalton's lymphoma where both these peptides were found to activate splenocytes and induced production of cytokines like IL-2, IFN-gamma and TNF-alpha indicating a Th1 type of immune response (Bhutia *et al.*, 2009).

1.2.3 Biosynthesis of ConA

A circular homology was revealed in ConA protein when its sequence was compared to other lectins (Carrington *et al.*, 1985). The synthesis of ConA involves a circular editing in a precursor glycosylated polypeptide (~34 kDa; proConA) lacking the ability to bind substrate sugars (Bowles *et al.*, 1986; Herman *et al.*, 1985). The precursor polypeptide of ConA is a glycosylated protein which is processed by N-glycanase to give a deglycosylated precursor capable of binding substrate sugars (active lectin) (Sheldon *et al.*, 1996). During editing, this N-glycan is removed (Min *et al.*, 1992; Ramis *et al.*, 2001; Sheldon and Bowles, 1992) and residues 118 and 119 (numbers correspond to ConA sequence) are ligated and a cleavage is performed by an asparaginylendopeptidase through transpeptidation which results in the generation of the terminal residues (Abe *et al.*, 1993; Bowles *et al.*, 1986). Cleaved fragments of the precursor ConA can associate to form a monomer with similar structure and substrate affinities as the native ConA monomer (Abe *et al.*, 1971). The intact monomeric species thus generated, form dimers preferentially compared to the fragmented forms and then the dimers

form tetramers at neutral pH (McKenzie and Sawyer, 1973).

1.2.4 Three-dimensional structure of ConA

ConA is the first lectin and a macromolecular mitogen whose crystal structure was determined. It was found to exist as a tetramer with each of the monomers composed of 237 amino acid residues (Cunningham *et al.*, 1975; Edelman *et al.*, 1972). The molecular weight of ConA was reported to be 96 kDa (~24kDa per monomer) using centrifugation and diffusion experiments (Edelman *et al.*, 1972; Sumner *et al.*, 1938). However, with the availability of the crystal structures the monomeric molecular weight of ConA was confirmed to be ~26kDa (Becker *et al.*, 1975; Greer *et al.*, 1970). Circular dichroism studies indicated ConA to be a predominantly β-sheet containing protein (Kay, 1970; Pflumm *et al.*, 1971; Zand *et al.*, 1971). The crystal structure of ConA (Figure 1.5) revealed further that anti-parallel β-sheets and coil regions constitute the major secondary structural elements (Becker *et al.*, 1975; Bouckaert *et al.*, 2000b). No helical structures were found except for a single helical turn formed by four residues (Reeke *et al.*, 1975). The two β-sheets present in ConA are referred as front and back β-sheets depending on their location in the molecule (Becker *et al.*, 1975). An unusual non-proline *cis* peptide bonds is observed in ConA between Ala207 and Asp208. Apart from ConA, such non-proline *cis* peptide bond is seen only in carboxypeptidase A (Hardman *et al.*, 1982). The *cis*-conformation of this bond is a hallmark of legume lectins and it has been proposed that this unusual *cis* peptide bond is stabilized by the Ca^{2+} ion (Bouckaert *et al.*, 1995). The back β-sheets interact during dimer and tetramer formation (Figure 1.5b) whereas front β-sheet separates the two hydrophobic cores found in ConA (Becker *et al.*, 1975; Derewenda *et al.*, 1989; Reeke *et al.*, 1975). The saccharide binding pocket is formed by Leu9, Tyr12, Asn14, Leu99, Tyr100, Asp208 and Arg228 where Tyr12 can stack with one face of the sugar ring (Derewenda *et al.*, 1989).

Due to sideways association of two monomers during dimerization a large anti-parallel β-sheet composed of 12 β-strands is formed at the back of each dimer and two such β-sheets interact to form the tetramer (Derewenda *et al.*, 1989; Edelman *et al.*, 1972). The intra-dimer contacts are significantly more in numbers than the inter-dimeric contacts. At pH values lower than 5.8, ConA exists as a dimer whereas at pH values higher than 5.8 the tetrameric form is prevalent (Hardman *et al.*, 1971; Kalb and Lustig, 1968; Wang *et al.*, 1971). Availability of crystal structure of ConA at different pH values, ranging from 5.0 to 8.0, has

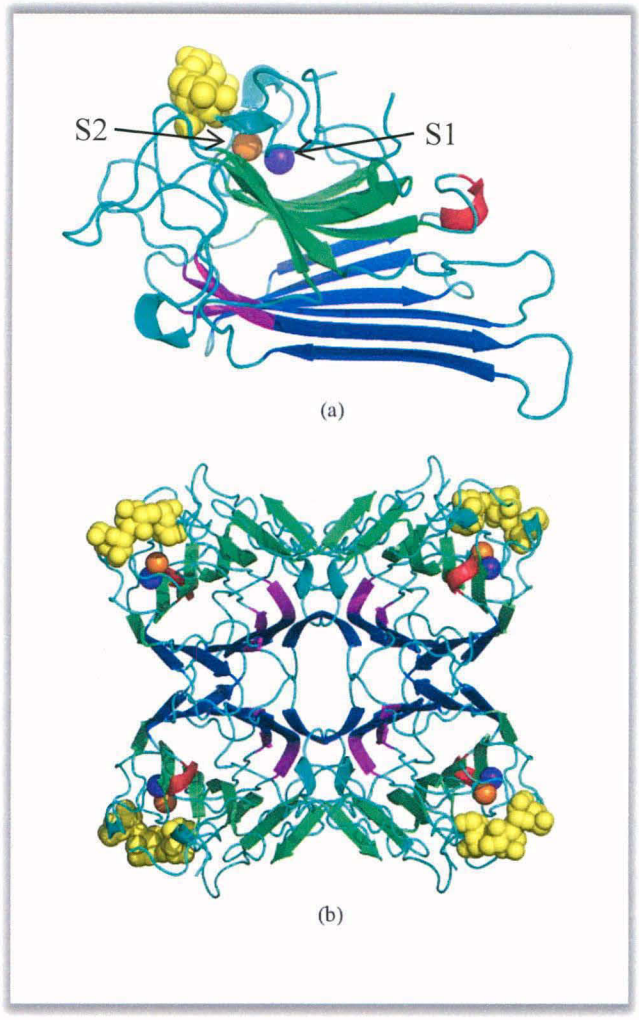


Figure 1.5
Three dimensional structure of ConA. (a) Monomeric ConA with trimannoside (yellow spheres) and metal ions {Mn^{2+} (magenta sphere; site S1) and Ca^{2+} (orange sphere; site S2)}, (b) tetrameric ConA.

helped in understanding of the effects of pH on surface loops, metal binding and molecular packing (Bouckaert *et al.*, 2000b; Lopez-Jaramillo *et al.*, 2004). The tetrameric ConA was approximately tetrahedral with a contact surface area of about 1200Å^2 per monomer (Quiocho *et al.*, 1971) and a water filled cavity at the center of the two concave dimers (Bouckaert *et al.*, 1995). ConA was found to be a tetravalent molecule with each of its monomers possessing a single substrate binding site (Becker *et al.*, 1971). However, no other higher oligomeric structures like trimers or pentamers were observed for ConA (Light-Wahl *et al.*, 1993). ConA has been observed in more than one different crystal forms (Kanellopoulos *et al.*, 1996; Parkin and Craig, 2002). To complete the thermodynamic and modeling knowledge about ConA, recently, a neutron structure of ConA has been solved at 15K to locate the water molecules around the lectin (Blakeley *et al.*, 2004).

1.2.5 Substrate Binding

Goldstein *et al* have studied the protein-carbohydrate interactions extensively using ConA as the model system. These studies indicate that ConA binding polysaccharides contain either α-D-mannose, α-D-glucose or D-fructose and are branched *e.g.* glycogens, amylopectins, dextrans and yeast mannans (Goldstein *et al.*, 1965a; So and Goldstein, 1967). Linear polysaccharides like amylase, isolichenin, nigeran and pullunan consisting of glucose units do not bind to ConA even at higher protein concentrations. However, the binding affinity increases with increased terminal, non-reducing α-linked residues as in branched polysaccharides (Goldstein *et al.*, 1965a; So and Goldstein, 1967). The specificity of the ConA–polysaccharide interactions are comparable to the antigen-antibody interactions with observation of factors like optimum pH, zones of antigen or antibody excess and equivalence zone (Goldstein *et al.*, 1965a; So and Goldstein, 1967; So and Goldstein, 1968b). ConA also binds to glycosylated proteins like IgG, IgM, myeloma IgG, RNase B, soyabean, waxbean lectins *etc.* (Goldstein *et al.*, 1969; Leon, 1967; Person and Markowitz, 1969). The specific binding of ConA to mannose and glucose has been used to isolate ConA with high specificity and yield using gel filtration on Sephadex (Agrawal and Goldstein, 1965; Agrawal and Goldstein, 1972). Affinity column chromatography has also been used to isolate carbohydrate specific immunoglobulins and plant lectins in a single and highly efficient step using affinity chromatography (Biswas *et al.*, 2009; Narahari and Swamy, 2010; Sela *et al.*, 1975a; Sela *et al.*, 1975b; Sultan *et al.*, 2009).

Each monomer of ConA binds a monosaccharide molecule at a single site *i.e.* ConA is a monovalent molecule (Naismith *et al.*, 1994; Yariv *et al.*, 1968). Substrate binding to ConA is accompanied with subtle conformational changes in substrate binding loops only, without affecting the overall tertiary structure of ConA (Becker *et al.*, 1971; Hassing and Goldstein, 1970; Hassing *et al.*, 1971; Naismith *et al.*, 1994; Pflumm *et al.*, 1971). The mannoside binding to ConA (Figure 1.6) involves seven hydrogen bonds in total, four with backbone NH groups, two with Ca^{2+} ion binding residues and one with the surrounding solvent water molecule (Brewer *et al.*, 1973; Derewenda *et al.*, 1989). Hydroxyl groups at C-1 and C-2 atoms of the ring forms of D-glucose and D-mannose are not essential for substrate binding, but those at C-3, C-4 and C-6 are involved in substrate binding to ConA (Goldstein *et al.*, 1965b).

Epimerization and chemical modifications at the essential positions, especially the C-6 hydroxyl group, lead to the loss of ability of sugar to bind to ConA (So and Goldstein, 1967). The crystal structure of ConA-trimannoside complex also revealed that the ConA-

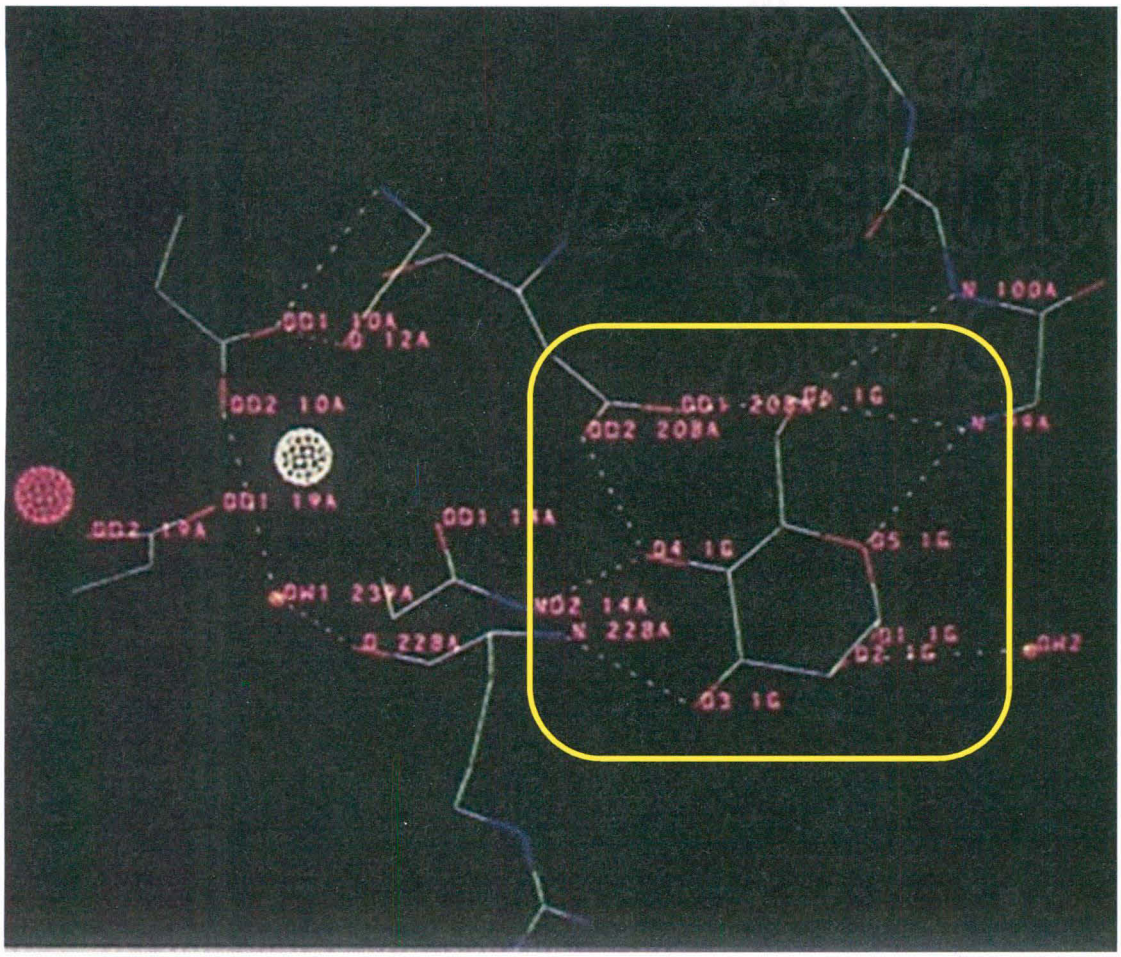


Figure 1.6
Interaction between Me-α-Man and ConA. Hydrogen bonds have been shown using dotted lines. The box is centered on the sugar residue. The seven hydrogen bonds between Me-Man and the binding site residues can be clearly seen. Ca^{2+} and Mn^{2+} ions are depicted as white and pink spheres, respectively; adapted from (Derewenda *et al.*, 1989).

sugar interactions involve hydrogen bonding, hydrophobic interactions and van der Waal contacts (Naismith and Field, 1996). In ConA and other legume lectins, several structurally conserved water molecules also participate in substrate binding by providing the protein-binding surface a complementary shape to that of the sugar (Bouckaert *et al.*, 1999b; Loris *et al.*, 1994; Swaminathan *et al.*, 1998). Based on a computational analysis, Bradbrook *et al* have suggested that unlike glucose, binding of mannose to ConA involves C-2 hydroxyl group forming three additional van der Waals contacts and an interaction between Thr226 and a conserved water molecule (Bradbrook *et al.*, 1998). This has been proposed as the reason for a 3-5 times greater binding of mannose compared to glucose (Derewenda *et al.*, 1989; Goldstein *et al.*, 1965b; Poretz and Goldstein, 1968; So and Goldstein, 1967; So and Goldstein, 1968a; So and Goldstein, 1968b). The trimannoside (α-D-mannopyranosyl-(1-6)-

α-D-mannopyranosyl-(1-3)-α-D-mannopyranoside) can occupy the full substrate binding site and hence, binds to ConA stronger than methyl-α-mannoside (Me-α-D-Man) (Brewer and Bhattacharyya, 1986; Brewer *et al.*, 1985; So and Goldstein, 1968b). The ConA-trimannoside interaction occurs mainly through the 1-6 and 1-3 mannose arms which are placed around the hinge like central reducing mannose (Bouckaert *et al.*, 1999b). It has been proposed that with a binding site overlapping that of the monosaccharide, the 1-6 arm mannose binds stronger than the 1-3 arm mannose mainly due to a favorable entropy (Bouckaert *et al.*, 1999b; Carver *et al.*, 1985; Gupta *et al.*, 1997; Mandal *et al.*, 1994). The thermodynamics experiments also indicated that the binding energy of Me-α-D-mannoside, Me-α-(mannose-α-1-2)-mannoside or M2M, and the trimannoside (M3M6M) to ConA were -5.6, -7.0 and -7.6 kcal/mole, respectively (Dam *et al.*, 2000; Moothoo *et al.*, 1999; So and Goldstein, 1968a). The substrate binding in lectins is enthalpically driven with either a minor or an unfavorable entropic contribution. The ΔH values scale proportionally to the number of substrate residues bound whereas TΔS does not (Dam *et al.*, 2000; Dam and Brewer, 2004; Williams *et al.*, 1992).

1.2.6 Metal binding and Demetalation of ConA

Early studies on ConA had indicated that it can be reversibly deactivated by treating it with mild acid or alkali like any other protein. However, addition of divalent metal ions (Ca^{2+}, Mn^{2+}, Zn^{2+}, Mg^{2+}, Co^{2+} *etc.*) reactivates ConA deactivated by above method, restoring its agglutinating ability. A mixture of Ca^{2+}, Mn^{2+}, and Mg^{2+} is more effective than Ca^{2+} alone (Sumner and Howell, 1936b). Availability of crystal structure helped in revealing the metal binding sites in ConA. ConA has two different metal ion binding sites called as S1 and S2 (Figure 1.5a). Site S1 binds transition metal ions only whereas site S2, binds Ca^{2+} selectively and is occupied only when S1 is occupied by a transitional metal ion like Ni^{2+} or Mn^{2+} (Kalb and Levitzki, 1968). The metal binding sites are placed such that bound metals are 4.25Å apart (Becker *et al.*, 1975; Hardman *et al.*, 1982). Con A exists as a mixture of two conformational states: a "locked" form and an "unlocked" form (Brown *et al.*, 1977). The "unlocked" form binds metal ions and saccharide only weakly and is a predominant conformation of demetalized Con A. The "locked" form possesses full saccharide binding activity and exists in the presence of either both the metal ions or at least Ca^{2+} ion in trace amounts. At equilibrium, about 13% of the demetalized (apo-) ConA was found to exist in locked conformation (Brown *et al.*, 1982). Only a locked conformation of ConA, *i.e.* when at least S2 is occupied, can bind to substrate sugars (Kalb and Levitzki, 1968).

The binding of a transition metal ion to S1 forces the metal binding loop to fold tightly into a rigid structure and concomitantly, forms the calcium ion binding site S2. Sequential binding of metal ions and a cooperativity between the binding of the metal ions is essential in stabilizing S2 into a locked conformation. The *cis*-conformation of Ala207-Asp208 peptide bond is a hallmark of the locked conformation in ConA (Bouckaert *et al.*, 1996; Bouckaert *et al.*, 2000a). Both, metal binding sites, S1 and S2, contain an octahedral coordination shell formed by neighboring residues like Glu8, Asp10, Tyr12, Asn14, Asp19, His24 and two structurally conserved water molecules (Becker *et al.*, 1975; Edelman *et al.*, 1972; Hardman *et al.*, 1982). Asp19 is a common residue binding to both metal ions and lies on the metal binding loop. Thus, the loop needs to fold properly onto the metal ion binding site for Asp19 to interact and stabilize metal ions (Becker *et al.*, 1975; Bouckaert *et al.*, 1995; Reeke *et al.*, 1978).

Demetalized ConA was found to have structural instability, a 7° rotation of the dimers with respect to each other, large structural changes in the residues located in metal binding region (especially Asp19), termini and saccharide binding regions and most notably, the *cis* to *trans*- isomerization of the Ala207-Asp208 peptide bond (Bouckaert *et al.*, 1995; Bouckaert *et al.*, 1996; Jack *et al.*, 1971; Reeke *et al.*, 1978; Shoham *et al.*, 1979). As a consequence to demetalation, the metal ion binding residues shift from their native conformation and ConA assumes the unlocked conformation, with no or weak substrate binding ability (Bouckaert *et al.*, 1995; Koenig *et al.*, 1973). In fact, none of the crystal structures of demetalized ConA have coordinates for the complete metal ion binding loop region. Also, the inter-dimer interactions reduce to about 30% compared to the native metal bound structure. This indicates that metal ions are absolutely essential for substrate binding to and structural stability of ConA (Agrawal and Goldstein, 1968; Sumner and Howell, 1936b; Yariv *et al.*, 1968).

1.3 *Erythrina corallodendron* lectin (EcorL)

EcorL is obtained from seeds of another leguminous plant known as *Erythrina corallodendron* (red coral). It was isolated and characterized much later than ConA, found specific for galactose and N-acetylgalactosamine, and mitogenic towards neuraminidase treated human lymphocytes (Gilboa-Garber and Mizrahi, 1981). Lectins obtained from other species of Erythrina *e.g. Erythrina cristagalli* and *Erythrina indica* were also found to be galactose specific and glycosylated (Horejsi *et al.*, 1980; Iglesias *et al.*, 1982; Lis *et al.*, 1985). EcorL was

TH-20069

found to be composed of 242 amino acid residues showing a high degree of homology to other legume lectins with the site of glycosylation being Asn-17 (Adar *et al.*, 1989). Because of the high homology of EcorL to other legume lectins, many similarities were observed among EcorL and ConA, for example, a similar tertiary fold and similar locations of the metal ion binding and substrate binding sites. However, unlike ConA, EcorL is glycosylated at Asn-17 and the glycan moiety is a heptasaccharide, Manα3(Manα6)(Xylβ2)Manβ4GlcNacβ4-(Fucα3)GlcNac (Ashford *et al.*, 1987). A signal sequence of 26 residues and an additional site for glycosylation due to the presence of the glycosylation motif N-N-X at Asn-113 were observed in EcorL (Arango *et al.*, 1990; Arango *et al.*, 1992).

1.3.1 Structure of EcorL

EcorL was crystallized in its dimeric form and found to have a molecular weight of about 60 kDa (Saper *et al.*, 1987). In the crystal structure of EcorL (Figure 1.7), each monomer is composed of 242 residues, has a molecular weight of ~30 kDa, N-glycosylated at Asn-17, contains Mn^{2+} and Ca^{2+} ions, has a tertiary structure similar and superimposable to other legume lectins like ConA (Shaanan *et al.*, 1991). This was one of the few structures known where the oligosaccharide was observed fully defined and well resolved (Bode *et al.*, 1989; Casasnovas *et al.*, 1997; Stehle *et al.*, 1994; Wyss *et al.*, 1995). With a legume lectin fold EcorL possessed the potential to form canonical dimers as seen in ConA (Becker *et al.*, 1975; Hardman and Ainsworth, 1972), pea lectin (Einspahr *et al.*, 1986), favin (Reeke and Becker, 1986) and Lathyrus ochrus lectin I (Bourne *et al.*, 1990). However, the oligomerization in EcorL occurred in a novel and drastically different mode known as the "handshake mode".

The back β-sheets from monomers docked onto each other to generate dimers and buried ~900Å^2 of surface area at the interface instead of aligning sidewise as in the canonical mode which typically buries ~900-1100Å^2 surface area (Shaanan *et al.*, 1991). Based on the observation of the orientation of the oligosaccharide moiety in the monomer, it was hypothesized that the possible steric clash between the bulky oligosaccharide might prevent the canonical mode of dimerization thus, leading to the novel handshake mode of oligomerization (Shaanan *et al.*, 1991).

1.3.2 Origin of handshake mode of oligomerization

As Asn-17 was located in the region of canonical dimeric interface of legume lectins, it seemed plausible that the oligosaccharide could interfere with the canonical mode of dimer

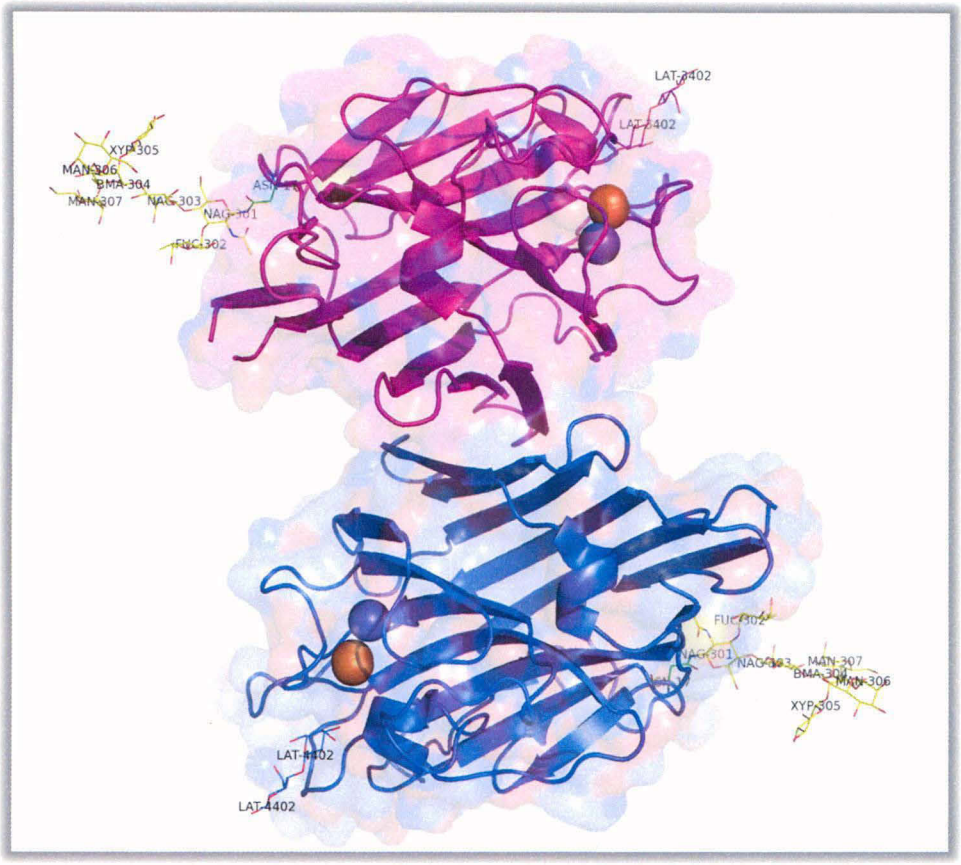


Figure 1.7:
Three dimensional structure of EcorL. The protein backbone has been depicted with cartoons. The oligosaccharide has been shown in yellow lines and its constituting residues have been labeled. The substrate binding site can be located as the binding site for lactose (LAT) molecule.

ization (Shaanan *et al.*, 1991). However, subsequently the recombinant form of EcorL was cloned in *E. coli* and it obviously lacked glycosylation. The recombinant non- glycosylated form of EcorL (rEcorL) was also dimeric and showed similar hemagglutination and carbohydrate specificity as the native glycosylated EcorL (Arango *et al.*, 1992). Surprisingly, the crystal structure of recombinant non-glycosylated form of EcorL (rEcorL) was also found to have similar handshake mode of oligomerization as the native (glycosylated) form despite the absence of the oligosaccharide moiety in the canonical dimer interface (Kulkarni *et al.*, 2004). Other (crystallographic) studies on winged bean-I and -II (Kulkarni *et al.*, 2004; Manoj *et al.*, 2000; Prabu *et al.*, 1998) also revealed that even in absence of the covalently linked oligosaccharide, monomers associated in the 'handshake mode'. This indicated the role of factors other than glycosylation *e.g.* the primary structure, in determination of the mode of quaternary association (Mitra *et al.*, 2003). Thus, crystallographic studies indicated that the tertiary and quaternary structure of EcorL were not dependent on glycosylation. However, computational analysis of factors like buried surface area, shape complementarity

and other interactions indicated a higher stabilization of EcorL over rEcorL (Kulkarni *et al.*, 2004).

1.3.3 Substrate Binding studies on EcorL

EcorL was found to be specific for galactose and N-acetylgalactosamine (Gilboa-Garber and Mizrahi, 1981). Galactose was observed to have a complementary electrostatic potential to the binding site in EcorL and interacts with Phe-131 through aromatic stacking like the stacking observed between mannose and Tyr12 in ConA-mannose complex (Elgavish and Shaanan, 1998). As in other legume lectins, the substrate binding occurs through both hydrophobic and hydrogen bonding interactions (Surolia *et al.*, 1996). Site-directed mutagenesis studies indicated that Asp89, Phe131 and Asn133 were essential for ligand binding by EcorL (Adar and Sharon, 1996). The crystal structure of EcorL also revealed a hydrophobic cavity, surrounded by Tyr108, Pro134 and Trp135, which could accommodate bulky substituents *e.g.* acetamido (NAc) or dansylamido (NDns) at the C-2 of galactose bound to EcorL. In fact, Trp135 in native EcorL closes the metal binding cavity by anchoring the 132-136 loop to rest of the protein through hydrophobic interactions especially with Tyr108 (Adar *et al.*, 1998). The substrate specificity of EcorL for GalNAc and GalNDns could also be altered using single or double mutants in the substrate binding region (Arango *et al.*, 1993). The specific binding of EcorL to N-acetyllactosamine (NacLac) have been used to identify glycosphingolipids with NAcLac as the terminal residues *e.g.* hexaglycosylceramide from bovine buttermilk (Teneberg *et al.*, 1994). Thermodynamics studies indicated that the binding enthalpies for NAcLac, Me-α- or Me-β-galactose were -20 kJ/mol but for galactose it was -14 kJ/mol (Surolia *et al.*, 1996).

1.3.4 Differences between EcorL and rEcorL

The thermal stability differences between the two different modes of dimerization seen in EcorL and ConA have been compared (Surolia *et al.*, 1996). EcorL was found comparatively less stable (T_m = 333 K) and displayed an independent unfolding of the monomeric subunits at higher temperatures. On the other hand, ConA tetramer (T_m = 363 K) unfolded as a single unit. This indicated that EcorL has a weak dimeric interface compared to ConA (Surolia *et al.*, 1996). Unfolding of EcorL was observed to be a two-state process *i.e.*

$$A_2 \Leftrightarrow 2U$$

where A & U are folded and unfolded monomers. However, a non-two-state unfolding was

observed for rEcorL (Mitra *et al.*, 2003). The DSC (differential scanning calorimetry) unfolding profile of rEcorL indicated that during unfolding, the monomers dissociated first (concentration dependent step) and unfolded later (second step) which could be written as the following equation.

$$A_2 \Leftrightarrow 2A \Leftrightarrow 2U$$

The gel filtration profiles of EcorL and rEcorL under denaturing conditions also indicated a greater structural stability for EcorL than rEcorL. As for the role of oligosaccharide in enhancing the stability of EcorL, it was observed that six out of seven residues of the oligosaccharide were involved in water mediated interactions with EcorL suggesting that the deglycosylation might affect the structure and/or folding of this lectin (Mitra *et al.*, 2003).

1.4 Glycoconjugates

Over the past few decades, it has been recognized that covalently attached glycan moieties play a critical role in biology. Glycans have been known to influence protein conformation and stability, cell proliferation and growth, immunological recognition, signal transduction, fertilization, membrane permeability and activity of signaling molecules (Bertozzi and Kiessling, 2001; Corfield, 2004; Kriss *et al.*, 2000; Montreuil, 1980; Varki, 1993). Various glycoproteins are involved in extremely important processes like innate and adaptive immune responses (*e.g.* immunoglobulins, cytokines, T-cell receptors, MHCs) and metabolic regulation (*e.g.* FSH, LH, TSH, and HCG), *etc.* Diseases like rheumatoid arthritis and a number of other autoimmune diseases, and even cancer have been observed to be the consequences of changes in specific glycans or glycopeptides. Therefore, studying glycosylation of polypeptides in relation to their structural and functional aspects could prove rewarding, especially, through development of better protein therapeutics against a number of diseases including cancer (Kriss *et al.*, 2000; Liu *et al.*, 1995; Scanlon *et al.*, 1992). Therefore, apart from glycoproteins, several research groups have studied glycoconjugates involving peptides for understanding the effects of glycosylation on structure and function of proteins and lipids.

Glycans have also been observed to impart additional structural and/or functional properties to the proteins they are covalently attached to. For example, glycosylated polypeptides may display an altered structural stability, protection from proteases and other non-specific protein-protein interactions (Steen *et al.*, 1998). The peptide-carbohydrate hydrogen

bonds may be one of the reasons for such an altered behavior (Schuman *et al.*, 2000). Instead of the glycoproteins, glycopeptides have been widely used to understand the relationship between glycosylation and consequent effects on the polypeptide backbone because they are free from the structural effects of global protein folding (Andreotti and Kahne, 1993; Huang *et al.*, 1996; Kirnarsky *et al.*, 2000). Molecular dynamics and Monte Carlo simulations have also been performed to analyze the differences among the peptides and their respective glycopeptides (Kriss *et al.*, 2000; Schuman *et al.*, 2000).

With advances in chemical synthesis of glycopeptides, studying glycosylated and non-glycosylated forms of various peptides has been made relatively easier (Buskas *et al.*, 2006). And, using these methods various important peptides have been studied in the context of glycosylation. One such peptide, a serine protease inhibitor known as *pars intercerebralis* major peptide-C (PMP-C) represents an interesting example where glycosylation led to an altered structural stability and functional properties (Mer *et al.*, 1996b).

1.4.1 Serine Protease Inhibitors

Serine proteases are the most important players in proteolytic cascades like digestion, blood clotting, metamorphosis, defense mechanisms and immune responses, which once set on are rapid and irreversible. Improper regulation of these proteases results in diseases like arthritis, emphysema, gingivitis and inflammatory infections (Salzet, 2002). Therefore, regulation of proteases is of extreme importance. It is achieved at different levels including gene transcription and translation, zymogen activation and through inhibitors. Protease inhibitors are deployed by most of the organisms involving both vertebrates and invertebrates (Imler and Hoffmann, 2000; Iwanaga, 1993). Some invertebrate protease inhibitors have been found specific towards mammalian proteases too due to which a large number of opportunities have arisen for their medical applications (Hamdaoui *et al.*, 1998; Salzet, 2002).

1.4.2 Mechanism of Serine Protease Inhibition

The peptide bond hydrolysis by a protease is catalyzed through a nucleophilic attack on the target peptide bond by a uniquely reactive serine (Craik *et al.*, 1987; Kraut, 1977). The attack occurs between residues known as P1 and P1' generating an acyl-enzyme complex and a peptide with P1' residue as the new amino terminus (Zakharova *et al.*, 2009). Later, the acyl-enzyme complex is cleaved releasing a peptide with P1 as the C-terminus. The P1 residue of the serine protease inhibitors determines the specificity of the target protease whereas

the P1' residue is either a serine or threonine (Gettins *et al.*, 1993). Most serine protease inhibitors or *serpins* would form 1:1 complexes with the proteases to inhibit them. Possible outcomes are either cleavage of the inhibitor at P1-P1' bond (inhibitor treated as substrate) or partial inhibition and partial cleavage of the inhibitor (Rubin, 1996). To understand the mechanism of the reaction and the dynamics in the active site of serine proteases, MD simulations have also been performed on the acyl-enzyme intermediate of porcine pancreatic elastase and human β-casomorphin-7 peptide complex (Topf *et al.*, 2002). Cleaved inhibitors can't bind and inhibit the protease any further (Gettins *et al.*, 1993). It has been observed that inhibitors with hydrophobic or bulky residues at P1 were chymotrypsin inhibitors whereas basic amino acid residues at P1 indicated specificity for trypsin. Mutating the P1 residue from leucine to arginine makes an inhibitor capable of inhibiting trypsin instead of chymotrypsin *i.e.* specificity swapping (Kellenberger *et al.*, 1995).

1.4.3 Small serine protease inhibitors (Pascifastin family)

Several peptide serine protease inhibitors have been extracted and characterized from locusts, 3 from *Locusta migratoria* (Brehelin *et al.*, 1991; Kellenberger *et al.*, 1995; Nakakura *et al.*, 1992) and 5 from *Schistocerca gregaria* (Hamdaoui *et al.*, 1998). These inhibitors constitute the locust peptide serine protease inhibitor family. These insect serine protease inhibitors have been reported to function in antimicrobial defense, digestion, metamorphosis and development (Hamdaoui *et al.*, 1998). The size of such insect serine protease inhibitors ranges between 29 and 400 amino acid residues. Based on their amino acid sequence and protease inhibition characteristics, insect inhibitors have been classified (Bode and Huber, 1992; Gettins *et al.*, 1993) into

a) *small* (low molecular weight; <10 kDa) serine protease inhibitors, and

b) *Serpins* superfamily (~45-100 kDa).

Small serine protease inhibitors range in size from 29 to 190 amino acid residues and bind to proteases in a substrate-like manner (Laskowski, 1986; Mer *et al.*, 1996a). Small serine protease inhibitors extracted from locusts showed sequence similarity with *pascifastin*, a heterodimeric serine protease inhibitor from signal crayfish, *Pacifastacus leniusculus* (Liang *et al.*, 1997). The sequence comparison also revealed a conserved cysteine rich pattern, **Cys**-X_{9-12}-**Cys**-Asn-X-**Cys**-X-**Cys**-X_{2-3}-Gly-X_{3-6}-**Cys**-Thr-X-**Cys**, in these peptides which leads to formation of three disulfide bonds (Salzet, 2002) as in case of PMP-C (Figure 1.8). Since, pascifastin was the first discovered member of this family, it was named after pascifastin. The

members of *Pascifastin* family are spread widely among the arthropods (Simonet *et al.*, 2002; Simonet *et al.*, 2003; Simonet *et al.*, 2005) being localized mostly, in the neurosecretory cells of *pars intercerebralis* and *corpora cardiaca* region of their brain (Harding *et al.*, 1995; Nakakura *et al.*, 1992). These inhibitors are involved in the regulation of neuropeptide precursor processing, innate immune defense and transition from solitarious to gregarious phase during swarming (Brehelin *et al.*, 1991; Rahman *et al.*, 2002; Salzet, 2002).

1.4.4 pars intercerebralis major peptide-C (PMP-C)

1.4.4.1 Structure of PMP-C

PMP-C is a small protease inhibitor of 36 amino acid residues (Nakakura *et al.*, 1992) and belongs to the *pascifastin* family of serine protease inhibitors. *PMP-C* is the third domain of a precursor polypeptide. The first and second domains of this precursor peptide are a signal sequence peptide and PMP-D2, respectively. PMP-D2 has been observed to display about 40% sequence identity to PMP-C (Kromer *et al.*, 1994). The core region of PMP-C (Figure 1.8) is a β-sheet which consists of three anti-parallel β-strands (β1, residues 9-11; β2, residues 16-19 and β3, residues 26-30). The core β-sheet demarcates a hydrophobic cavity as observed in other peptide serine protease inhibitors (Bode *et al.*, 1986; McPhalen *et al.*, 1985a; McPhalen *et al.*, 1985b; Mer *et al.*, 1996a). The N-terminal region runs almost perpendicular to the β-sheet to which it is tethered by a disulfide bond (Cys 4-Cys19). The protease binding loop, from Thr29 to Ala32 (P2'), lies at the C-terminus being connected to the core by two disulfide bridges, Cys17-Cys28 and Cys14-Cys33 (Mer *et al.*, 1996a). These three disulfide linkages form a "cysteine knot" which stabilizes the peptide inhibitor. The P1 residue in PMP-C was identified to be Leu30 as L30V mutation abrogated α-chymotrypsin inhibition activity of the mutant peptide (Kellenberger *et al.*, 1995).

1.4.4.2 Non-glycosylated vs. glycosylated forms of PMP-C

Four of the eight known locust inhibitors are glycosylated post-translationally on a threonine (Thr) residue after the first Cys of the pascifastin motif (Hamdaoui *et al.*, 1998; Nakakura *et al.*, 1992). The non-glycosylated forms of these peptides have been observed to be structurally identical with the glycosylated forms. The non-glycosylated forms retain their biological activity which poses a question on the need of glycosylation (Kellenberger *et al.*, 1995). In case of PMP-C, a stabilizing role for the fucose moiety has been observed (Mer *et al.*, 1996b). Both fucosylated and non-fucosylated forms of PMPC are capable of inhibiting

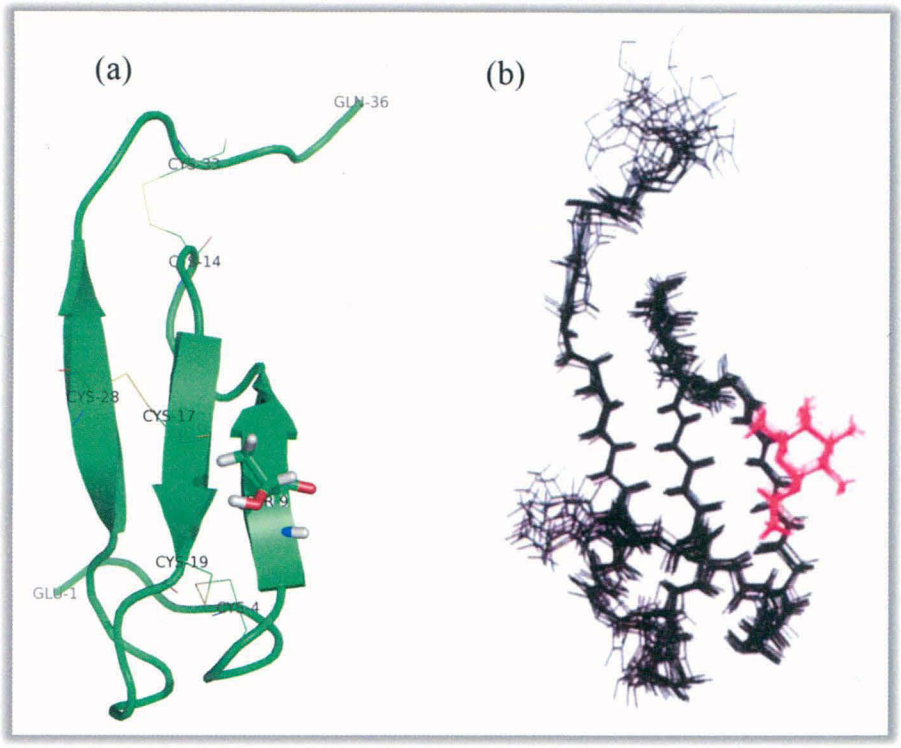


Figure 1.8
Three dimensional structure of *PMP-C*: (a) non-fucosylated form of *PMP-C* showing Thr-9 as the site of fucosylation. The disulfide bonded Cysteines and terminal residues have also been labeled. (b) NMR structure of fucosylated form of *PMP-C* with the red lines depicting the fucose moiety (Mer *et al.*, 1996b).

human leukocyte elastase and α-chymotrypsin but only non-fucosylated form can inhibit high voltage-activated Ca^{2+} currents whereas fucosylated PMPC cannot. Therefore, the effect of the fucose moiety on structural aspects of PMP-C have been investigated (Mer *et al.*, 1996b). It was found to adopt a well-defined orientation with respect to the peptide and interacted with residues like Phe10, Lys11 and Arg18 of PMP-C. A hydrophobic interaction was observed between the methyl groups of Thr16 and the fucose moiety (Mer *et al.*, 1996b). The fucosylated form of PMP-C exhibited lower exchange rates for backbone amide protons of residues involved in the hydrogen bonding network of the β-sheet and of those located in the vicinity of the hydrophobic core (Mer *et al.*, 1996b). As the proton exchange occurs from a global semi-unfolded state, where the hydrogen bonds are broken, it was inferred that fucosylation decreased the frequency of occurrence of such semi-unfolded states by stabilizing the molecule. The extrapolated melting temperatures noted for both forms of PMP-C indicated a higher value for the fucosylated form which emphasized the stabilizing role of fucosylation in case of PMP-C. The enhanced stability was proposed to have been achieved through a de-

crease in the overall dynamics of the peptide (Mer *et al.*, 1996b).

1.5 Molecular Dynamics Simulations

Under native conditions at room temperature, proteins possess considerable internal freedom ranging from local atomic fluctuations and side-chain oscillations to secondary structure displacement and tertiary or quaternary rearrangements. Structural changes proceed through such fluctuations as mentioned above which search for path(s) leading to the structural transition. Especially, the correlated fluctuations caused by perturbations like substrate binding, *etc.* bias the fluctuations in a way that structural transition occurs. One of the widely used methods to obtain protein structures is X-ray crystallography; however, the protein structures obtained from crystallography represent an average structure and are limited to a static image of the molecule. Given the degree of freedom, any momentary structure might differ significantly from this average structure. Therefore, in order to understand and investigate the structure, dynamics and kinetics of proteins at the atomic level we need an alternative approach which would account for the fluctuations occurring at room temperature. Theoretical methods like Molecular Dynamics (MD) and Monte Carlo (MC) simulations were introduced which have been very useful in the analysis of conformational changes in proteins and other biomolecules. Such methods involving computer simulations have been successfully used as explorative tools to facilitate interpretation of biochemical data. Such simulations have also provided access to otherwise inaccessible atomic details (van Gunsteren and Mark, 1992).

MD simulations have become one of the most important theoretical methods to investigate dynamics and kinetics of proteins (Karplus *et al.*, 1980; Wang *et al.*, 2008). Recently, MD simulations have been used extensively to study various processes ranging from biochemical events like substrate binding to biological phenomena like ion transport in membrane located ionic channels. For example, MD simulations have been used in studying the effects of edema factor binding on calmodulin-calcium interactions (Laine *et al.*, 2010), detection and quantification of interior cavities as well as surface pockets of proteins (Bhinge *et al.*, 2004), pathogenesis of human prion proteins (van der Kamp and Daggett, 2010; Zhong, 2010), folding/unfolding events in proteins (Day *et al.*, 2010; Meersman *et al.*, 2010), intra-protein subdomain movements (Bordes *et al.*, 2010), ionic partitioning and transport in ion channels (Faraudo *et al.*, 2010), changes in structure of α-helical peptides upon adsorption onto a carbon nanotube (Balamurugan *et al.*, 2010), modeling ion (Ca^{2+}) binding sites (Yang

et al., 2010), protein dynamics in lipid (membrane) environment (Choutko *et al.*, 2010), RNA stability (Gong and Xiao, 2010), etc. The potentials used in MD simulations have been tested for long and recent studies have also validated their reliability in reproducing the experimental observations (D'Angelo *et al.*, 2010).

1.5.1 The Force field: Potential Energy Function

The molecular mechanics (MM) force-field consists of a suitable empirical potential energy function which can be used to describe interactions between various atoms in a chemical or biological system. For example, the equation given below shows the functional form of a typical MM forcefield used in energy minimization or molecular dynamics simulations (Gelin and Karplus, 1975; Weiner *et al.*, 1984; Weiner *et al.*, 1986).

$$E = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2}\{1 + \cos(n\emptyset - \gamma)\} + \sum_{\substack{i<j \\ non-\\ bonded}} \left\{\frac{A}{r_i^{12}} - \frac{B}{r_i^6} + \frac{q_i q_j}{\varepsilon R_{ij}}\right\} + \sum_{H-bonds} \left[\frac{C_{ij}}{R_{ij}^2} - \frac{D_{ij}}{R_{ij}^{10}}\right]$$

where, K_r and r_{eq} are bond stretching constant and equilibrium bond length whereas K_θ and θ_{eq} are bond-angle bending constant and equilibrium bond angle. V_n is the torsion barrier, n, the periodicity of the barrier and γ the phase. For electrostatic contribution, the Coulombic interaction between charges q_i and q_j on atoms i and j placed at a distance of R_{ij} is calculated. Using Lennard-Jones potential with A and B being repulsive and long-range non-bonded parameters, respectively, the non-bonded interactions are calculated only between atom pairs separated by at least three bonds but within the cut-off distance. Such an empirical energy function contains separable contributions of bond stretching, bond angle bending, dihedral angle twisting terms (Figure 1.9) and non-bonded interactions (hydrogen bonds, van der Waals and electrostatic interactions). Some of the less complex force fields exclude the explicit hydrogen bond term while others include angle dependent terms in hydrogen bonding potential for enforcing co-linearity of hydrogen bonded atoms. Quality of the force field used affects the quality of results obtained from computer simulations. For an improved accuracy the force field parameters may be derived from *ab initio* quantum-mechanical calculations involving electronic degrees of freedom for small molecules. However, it has been suggested that a more complex or elaborate form doesn't necessarily produce a better force field and

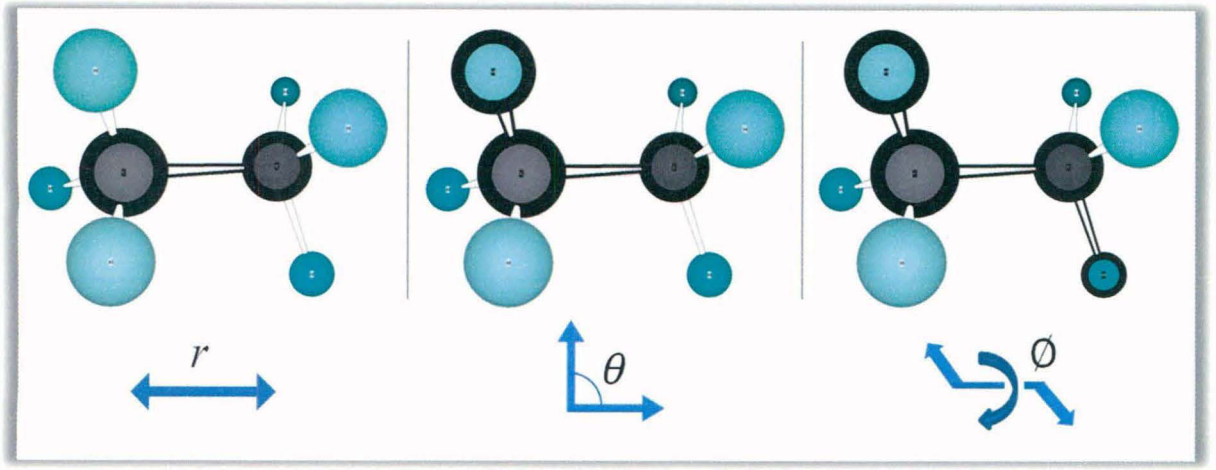


Figure 1.9

Depiction of bond length, bond angle (bending) and torsional angle using a molecule of ethane. The first three summations in potential energy function correspond to these parameters. The grey and blue circles depict carbon and hydrogen atoms, respectively.

better results (van Gunsteren and Mark, 1992). Therefore, most of the force fields for biological systems have been developed by fitting the force field parameters to the experimental data like crystal structure, energy and lattice dynamics, properties of liquids like density and enthalpy of vaporization, free energy of solvation, NMR data, *etc*.

An important factor responsible for a successful force field is the accurate treatment of electrostatic interactions for correctly describing inter- and intramolecular interaction energies. Different methods have been used to handle electrostatic interactions by different force fields. MMFF (Merck Molecular Force Field) uses heuristic algorithms to calculate charges based on electronegativities (Halgren, 1996) while OPLS (Optimized Potential for Liquid Simulations) derives charges based on fitting to experimental properties of liquids like structure, vaporization and sublimation enthalpy (Jorgensen and Tirado-Rives, 1988). AMBER force fields use charges based on quantum-mechanical calculation on fragments to describe electrostatic interactions (Cornell *et al.*, 1995; Weiner *et al.*, 1984). This treatment of charges makes AMBER significantly better than other harmonic force fields (Gundertofte *et al.*, 1996). A further improved version of AMBER force field was developed using restrained electrostatic potential (RESP) approach to derive partial charges which gave the smallest average absolute errors (AAEs) and root mean square deviation (RMSDs), when computed molecular properties were compared with experimental data compared to CHARMm, MM3 and MMFF (Wang *et al.*, 2000). Significant divergence was found on both the potential energy

surface (PES) as well as free-energy surface (FES) of alanine dipeptide calculated by CHARMM27 compared to AMBER force fields (ff03, ff99SB) and OPLS-AA (Vymetal and Vondrasek, 2010).

In a recent benchmarking study, for the experimental NMR verification of MD simulations, 12 different force fields (AMBER, CHARMm, GROMOS, and OPLS-AA) and 5 different water models were tested. These force fields were used to carry out MD simulations on small open chain peptides containing glycosylated glycine and proline residues. The results obtained from these simulations were compared with NMR experiments (Aliev and Courtier-Murias, 2010). AMBER99SB was identified as the most reliable force field for reproducing NMR measured parameters. Three force fields (AMBER99, AMBER99ϕ, and AMBER94) showed the least satisfactory agreement with both the solution NMR and the PDB survey data. For different solvent models considered, a correlation was noted between the number of torsional transitions in GPGG (peptide) and the water self-diffusion coefficient on using TIP3P, TIP4P, and TIP5P models (Aliev and Courtier-Murias, 2010). In another comparison between AMBER03 and GROMOS 53A6 force fields, the analysis of global stability (RMSD, number of base pairs and hydrogen bonds) of Dickerson-Drew dodecamer (a DNA molecule) showed that systems, both with or without specific terminal nucleotide topologies, simulated with AMBER were stable, while the system simulated with the original GROMOS topologies was very unstable after 5 ns (Ricci *et al.*, 2010). Similarly, AMBER (AMBER99/Parmbsc0) and CHARMm (CHARMM27) force fields were used to derive their flexibility descriptors of DNA and RNA duplexes in terms of a consensus conformation where it was observed that whereas for DNA a convergence was seen between both force fields, for RNA the AMBER force fields reproduced the experimental parameters better than CHARMm (Deng and Cieplak, 2010; Faustino *et al.*, 2010). Further, it was observed that with current force fields, simulations beyond hundreds of nanoseconds run an increased risk of undergoing transitions to nonnative conformational states or persisting within states of high free energy for too long. Thus, the obtained population frequencies could be skewed using most of these force fields except for the AMBER99sb force field (Lange *et al.*, 2010; Project *et al.*, 2010). Therefore, there is a need for improved force fields (Guvench *et al.*, 2009; Lindorff-Larsen *et al.*, 2010; Mobley *et al.*, 2009; Vellore *et al.*, 2010). In fact, new force fields are being constantly developed especially, for organic compounds (Raabe and Maginn, 2010; Zhong and Patel, 2010).

1.5.2 Energy Minimization

The potential energy function of any molecule is composed of terms like bond length, angle, dihedrals *etc.* which are directly dependent on coordinates of the system. The variation in energy of a molecule as the coordinates vary can be plotted and is known as the *potential energy surface* or simply, *hypersurface*. The points with lesser energy, compared to neighboring points on the hypersurface are known as the *local minima*. The point with the lowest energy compared to all local minima is known as the *global energy minimum*. Minima correspond to an arrangement of constituent particles (atoms) in a system which is more stable than higher energy arrangements. The process of identifying stable configurations corresponding to local or global minima is known as energy minimization. An ideal minimization algorithm must locate the minima quickly using the least amount of memory and computational time.

Mathematically, if the potential energy function *f(x)* depends on one or more independent variables *x1, x2, x3... etc.* then, at minima, various variables have values such that the derivative of *f(x)* is zero. As given below,

$$\frac{d}{dx}f(x) = 0$$

and

$$\frac{d^2}{dx}f(x) > 0$$

Minimization approaches usually work downhill on the hypersurface and therefore, lack the ability to locate another better minima (or *global* minimum) than the closest. This is due to the fact that once a minimum is achieved the neighboring points have higher energy than the available minimum and these methods cannot move up the energy gradient. This inability of minimization methods can be compensated by starting minimization at multiple points on *hypersurface*.

The first derivative of potential energy (PE) with respect to its coordinates equals to the force on an atom. Similarly, the second derivative of energy can be used to predict where the function would change its direction. The energy minimization involves moving the atoms according to the force acting on them to locations corresponding to a net zero force on the atoms. The derivatives of potential energy *w.r.t.* coordinates are calculated numerically. Most frequently used first-derivative energy minimization methods are Steepest Descent and Conjugate Gradient methods.

1.5.2.1: Steepest Descent (SD) Method: In this method the atoms are moved in the direction of the net force which at step k is represented as a vector s_k which is given as

$$s_k = -\frac{g_k}{|g_k|}$$

where, g_k is the gradient at step k. The newly obtained coordinates are used again to define the direction of the force for the next step. However, a problem with this method is that it takes many small steps to reach the minimum and is non-convergent. The improvement in stereochemistry and energy gets slower after some number of cycles where it cannot make radical changes of conformation any further and the procedure needs to be stopped.

1.5.2.2: Conjugate Gradient (CG) Minimization: In this method, the gradient at each point are orthogonal but directions are conjugate *i.e.* the direction v_k from a point x_k is computed from the gradient at the point (g_k) and the direction vector from the previous step v_{k-1}. Therefore,

$$v_k = -g_k + \gamma_k v_{k-1}$$

where, the scalar constant γ_k is given by

$$\gamma_k = \frac{g_k \cdot g_k}{g_{k-1} \cdot g_{k-1}}$$

1.5.2.3: Newton-Raphson method

Newton-Raphson method of minimization uses the second derivative of potential energy and has the advantage of a quick convergence. The inverse of Hessian matrix of second derivative for a particular system is used assuming that the hypersurface is a quadratic function. For a purely quadratic function, Newton-Raphson method finds minimum in a single step from any point on the hypersurface. However, when the hypersurface is not quadratic or is quadratic only to a first approximation, a number of steps would be required. For each step, Hessian matrix would be required followed by calculation of its inverse. The compute intensive nature of this process limits its application in systems composed of larger number of atoms. Also, far from a minimum the harmonic approximation is not appropriate and minimization may become unstable. Therefore, a more robust method, like SD or CG, must be used initially to bring the system closer to the minimum before using Newton-Raphson method.

1.5.3 Molecular Dynamics simulations: Methodology

Typically, the minimized structure of a given molecule corresponds to the conformation at 0 K. However, at room temperature, the atoms in the molecule have kinetic energy and using this kinetic energy the molecule can cross the barriers in the potential energy sur-

face and jump to lower energy conformations. In principle, given sufficiently long time, starting from any local minimum the system can reach the global energy minimum by crossing barriers in the potential energy surface. MD simulations mimic this situation by assigning velocities to each atom of the system corresponding to the temperature.

For MD calculations, all the particles of a system are given initial positions and velocities. The velocities are assigned so that they follow the Boltzmann distribution corresponding to the temperature of the system and the average kinetic energy of the system equals the temperature. To follow the dynamics, at any instant the force on each constituent particle due to the rest of particles is calculated using a potential function and force field parameters. The particle is then allowed to move for a short time interval according to this force, using Newton's second law. The new position and velocities are calculated from the previous positions and velocities and the net force on the particle. By recording these positions and velocities at successive time steps a trajectory of the system over the desired time period is generated (van Gunsteren and Mark, 1992). Most of the current algorithms for MD simulations are based on the earlier work by Loup Verlet. Verlet studied the dynamics of an isolated system of 864 particles interacting through Lennard-Jones potential and the results shared a good agreement with experimental values known for the time-dependent properties (Jones, 1924; Verlet, 1967).

The equations of motion for an atom in MD simulations is given as

$$v_i(t) = \frac{dr_i(t)}{dt}$$

$$a_i = \frac{dv_i(t)}{dt} = \frac{F_i(t)}{m_i}$$

where $r_i, v_i, and\ m_i$ denote the position, velocity and mass of any atom i which is experiencing a force F_i due to all other atoms of the system. This force is given as the negative gradient of the interaction potential function V as

$$F_i(t) = -\frac{\partial V[r_1(t), r_2(t), \ldots r_n(t)]}{\partial r_i(t)}$$

The atomic motions can be followed using the Velocity Verlet algorithm which has also been implemented in AMBER. For a small time step Δt, the position and velocities at $t + \Delta t$ can be calculated as

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + {}^1/_2\, a_i(t)\Delta t^2$$

and

$$v_i(t + \Delta t) = v_i(t) + {}^1/_2 \{a_i(t) + a_i(t + \Delta t)\}\Delta t^2$$

Since, for computation of $r_i(t + \Delta t)$ and $v_i(t + \Delta t)$ from $r_i(t)$ and $v_i(t)$, higher order terms of Δt are neglected. The value of Δt has to be small enough so that, movements during Δt can be approximated as rectilinear and with a constant acceleration. Therefore, typically time steps of 1-2 fs are used in MD simulations of biomolecules. Thus, movements corresponding to vibrational frequency of covalent bonds 10 fs can be computed in a few steps. A trajectory generated using these calculations can be analyzed later for equilibrium properties of the system. Accuracy of the trajectory and reliability of prediction of properties for a system greatly depends on the length of simulations in relation to the time scale of the process being studied. Longer time periods are required for more complex systems like water-solvated polypeptides as the time required to reach equilibrium for certain properties like potential energy and dielectric relaxation of water is slow. The data obtained from computer simulations is meaningful only when the system has reached its equilibrium state. The complete equilibrium, however, is normally not attainable, thus the system should *at least* be at a metastable equilibrium state before the data becomes meaningful. In order to achieve the equilibrium faster, algorithms like SHAKE have been developed. Since, stretching of covalent bonds corresponds to the fastest motion of the system, SHAKE constrains bond stretching motion and allows use of larger time steps.

1.5.4 Constraint Dynamics: SHAKE algorithm

Early simulations studies on small molecules such as argon (Rahman, 1964), water (Rahman and Stillinger, 1971), nitrogen, and alkanes were not sufficiently reliable for computation of thermodynamic properties mostly due to the shorter time scale accessible through simulations. Therefore, the concept of constraint dynamics was introduced *i.e.* bond lengths, angles and dihedral angles of a macromolecule could be constrained in an MD simulation which resulted in quicker calculations by a factor of 3 (van Gunsteren and Berendsen, 1977). The applications of constraints to MD simulations provided advantages like,

(1) increased time step (thereby reducing computational effort to obtain a given length trajectory),

(2) decreased range of frequencies of motions improved the heat exchange and effective temperature of a system, and

(3) constraining a bond removes the need to treat bond-stretching frequencies quantum mechanically as at room temperature most bond-stretching frequencies are ≥ k_BT/h (k_B is Boltzmann's constant, T is absolute temperature and h is plank's constant)

The SHAKE procedure, which constrains the covalent bonds has been the most popular among distance constraint algorithms (van Gunsteren and Berendsen, 1977). Other improved constraint algorithms are SETTLE (Miyamoto and Kollman, 1992) and another version of SHAKE known as M-SHAKE (Krautler *et al.*, 2001). For small molecules M-SHAKE proved faster compared to the other two, however, in case of macromolecules the M-SHAKE algorithm was found slower (Krautler *et al.*, 2001).

1.5.5 Water Models and explicit solvent simulations

Most biomolecular systems carry out their function either in aqueous environment or in membrane consisting of lipid environment. However, in view of the additional computational cost involved in the simulation of a biomolecule in aqueous or membrane environment, early computer simulations were performed in vacuum. However, it was observed that in absence of solvation the theoretical methods failed to reproduce many properties of a biomolecular system. Therefore, explicit solvent simulation methods which include water around the protein/DNA were developed. Most popular of them has been the TIP3P model, which is a 3 site water model with the *+ve* and *–ve* charges being placed on hydrogen and oxygen atoms, respectively (Jorgensen, 1981). This model was earlier called TIPS and in this model the coulombic and van der Waals (vdW) interactions were calculated for all pairs of water molecules using the following equation,

$$\varepsilon_{mn} = \sum_{i}^{m}\sum_{j}^{n} \frac{q_i q_j e^2}{r_{ij}^2} + \frac{A}{r_{oo}^{12}} - \frac{B}{r_{oo}^{6}}$$

where q_i, q_j and r_{ij} are the charges and distance on interacting atoms i and j, e is the electronic charge, r_{oo} is the distance between two oxygen atoms of water molecules and A and B were chosen such that the results fit for gas phase complexes of or liquid water. This model was re-parameterized to improve the energy and density of liquid water and called TIP3P potential (Jorgensen *et al.*, 1983). Later, even more complex water models were developed where 4 sites (the –ve charges were placed on the bisector of the HOH angle towards the hy-

drogen atoms instead of placing them on the oxygen; (Bernal and Fowler, 1933)) and 5 sites (charges on hydrogen atoms and the lone pair positions both arranged tetrahedrally on oxygen; (Stillinger and Rahman, 1974)) were considered to describe a water molecule. TIP4P potential was found most promising a water model as it gave excellent correlation with diffraction and thermodynamics data on water though its simpler form TIP3P would allow a rapid evaluation but provided lesser structural details (Jorgensen *et al.*, 1983). The proper matching of force field and solvent is critical to obtain correct result in molecular dynamics simulation of biomolecules. This problem has been intensively investigated for proteins (Cerutti *et al.*, 2010) but not for RNA. Recently, the RNA stability under different combinations of amber force fields and solvation models has been tested where it was found that a combination of Amber force field (ff98) and generalized Born implicit solvent model (igb1) were a reasonable choice for RNA simulations (Gong and Xiao, 2010).

1.5.6 Periodic boundary conditions

When simulating a finite size system, one of the factors affecting the quality of simulations is the boundary effects. When vacuum is used as the boundary, mimicking the gas phase of the system, the properties of atoms located at the surface would be distorted and so will be the surface of a non-spherical system. Therefore, simulations *in vacuo* should only be used to obtain qualitative results as the absence of solvent significantly changes the structural and dynamics properties of a protein (van Gunsteren and Mark, 1992). These effects would be reduced by solvating the system with a proper solvent, however, a far better but costlier method to minimize these edge effects is to use periodic boundary conditions. In periodic boundary conditions the atoms of a system are placed in cubic or any periodic space-filling box which is treated as if surrounded by identical translated images of it. The interactions are calculated between the nearest neighbors or images of the atoms, but an atom should not interact with another atoms and a periodic image of that atom simultaneously (nearest image convention).

1.5.7 Particle Mesh-Ewald Summation

All the interactions, especially the charged interactions, depend inversely on the distance between the interacting particles to different orders. Therefore, to reduce the computational effort required to calculate the interaction potential in a biological macromolecule, interactions after a specified distance cutoff were ignored as their contribution to the interaction potential dropped considerably. In simulations involving PBC, typically, a cut off distance of

less than or equal to the size of the simulation box is used. This ensures that any solute molecule and its image in the periodic box do not see (*i.e.* interact with) each other and the system mimics infinite dilution. Such use of cut off distances for truncation typically lead to discontinuity in the energy function and consequent errors in simulations. It was observed that in simulations where Coulombic interactions were truncated after a specified cutoff, the structural elements behaved unrealistically compared to the simulations run without this truncation. Therefore, methods like Ewald pair potential, Winger potentials, particle-mesh techniques and multipole expansions were used for evaluating long range electrostatic forces. In Ewald summation (Ewald, 1921), the summation of interaction energies in real space are replaced with an equivalent summation in Fourier space because of a rapid convergence of the Fourier-space summation compared to its real-space equivalent. As electrostatic energies consist of both short- and long-range interactions, it is maximally efficient to decompose the interaction potential into a short-range component summed in real space and a long-range component summed in Fourier space. However, the most successful method for evaluating electrostatic energies and forces of large periodic systems was inspired by a combination of the Ewald method and the particle-particle particle-mesh method and is known as particle-mesh Ewald (PME) summation (Darden *et al.*, 1993). PME method was later improved further by replacing the Lagrangian interpolation with cardinal B-spline interpolation (Essmann *et al.*, 1995). The PME pair potential and its derivatives are continuous functions of position, thus avoid problems involved with integration of discontinuous functions. PME is fast and has high accuracy with relatively little increase in effort compared to earlier known methods (Darden *et al.*, 1993).

1.5.8 Implicit Solvation: MM/GB-SA and MM/PB-SA methods

MD simulations in explicit solvent environment are computationally expensive. Hence, many fundamental biomolecular processes occurring on a time scale of microseconds cannot be studied using MD simulations even on the high performance machines, like supercomputers or integrated GPUs (graphics processing units) or multicore clusters (Bauer *et al.*, 2010). Therefore, other methods of solvation known as 'implicit or continuum solvation' have been developed by several research groups (Srinivasan *et al.*, 1998; Wang *et al.*, 2008). Such implicit solvent models permit faster calculations without loss of the atomic-detailed description of biomolecules and with no need of pre-equilibration of the solvent environment (Wang *et al.*, 2008). The attractive features of continuum solvent model include

1) it is sufficiently realistic in many cases as the averages over water and counter-ion degrees of freedom are provided implicitly and
2) it is computationally less expensive by removing the need for explicitly including large number of solvent molecules in the calculations (Srinivasan *et al.*, 1998).

The free energy of solvation of a solute can be simplified to a potential mean force (PMF) of a particular configuration which is the reversible work to assemble solute atoms into this configuration in presence of the solvent molecules. The implicit solvation schemes calculate this PMF in two steps:

(1) calculation of the nonpolar contribution to solvation which includes the bond, angle, torsion, vdW interaction terms and non-electrostatic/non-polar solvation energy, and
(2) calculation of reversible work of charging the solute by wrapping both solute coulombic contribution and the solvent-induced contribution.

The nonpolar contribution to solvation is the free energy of solvation of the molecule with all atom-centered partial charges set to zero (Reynolds *et al.*, 1974). The hydrophobic repulsive free energy has been approximated as the unitary free energy of transfer of a non-polar solvent to an aqueous medium, between water and hydrocarbon chains *e.g.* saturated alkanes, branched and cyclic hydrocarbons, was found to be proportional to the surface area of the cavity formed by the solute in water (Hermann, 1972; Reynolds *et al.*, 1974; Sitkoff *et al.*, 1994). Therefore, the non-electrostatic or non-polar solvation (E_{nes}) energy is given as

$$E_{nes} = \gamma SA + b$$

where SA, is surface area of solute, γ and b (the solvation energy for a point solute) are 0.00542 kcal/Å^2 and 0.92 kcal/mol, respectively (Srinivasan *et al.*, 1998; Wang *et al.*, 2008). The electrostatic or polar component of solvation energy is calculated either by Poisson-Boltzmann or generalized Born approach, both of which have been described below.

1.5.9 Poisson Boltzmann approach

The reversible work of charging the solute may be computed by solving Poisson equation or generalized Born method. The Poisson equation is given as

$$\nabla\varepsilon(r)\nabla\emptyset(r) + 4\pi\rho(r) = 0$$

where $\emptyset(r)$, $\varepsilon(r)$, and $\rho(r)$ are the electrostatic potential, position dependent dielectric function and charge density due to solute (Honig and Nicholls, 1995). Poisson equation is mostly solved using the finite difference approach as in Delphi and Grasp programs. A dissolved

electrolyte may be accommodated by using Poisson-Boltzmann (PB) equation instead of Poisson equation

$$\nabla\varepsilon(r)\nabla\emptyset(r) = -\rho(r) - \sum n_i^0 q_i exp[-\beta q_i \emptyset(r)]$$

where $\beta = \frac{1}{kT}$ (k is the Boltzmann constant), n_i^0 and q_i are the number density and charge of the counterions of type i in bulk solution, respectively.

1.5.10 Generalized Born approach

Similarly, another alternative approach based on generalized Born (GB) theory has also been used to calculate the electrostatic contribution to the free energy of solvation of molecules. Unlike PB approach which computes the electrostatic free energy by numerical solution to the PB equation; the GB approach uses an analytical solution or Generalized Born formulae to approximate the electrostatic free energy. In GB approach a molecule in solution is represented as a set of point charges set in spherical cavities embedded in a polarizable dielectric medium (Jayaram *et al.*, 1998). For a successful GB calculation, the GB radii of various atom types must be parameterized which is consistent with the net atomic charges intrinsic to the energy function. Jayaram *et al* (Jayaram *et al.*, 1998) modified the generalized Born model and used the following equation for the calculation of GB energy.

$$\nabla G_{Born}^{GB} = \left(1 - \frac{1}{\varepsilon}\right) \sum_{ij} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 exp\left(-D_{ij}\right)}}$$

where $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$ (α_i being the effective Born radius of atom i) and $D_{ij} = \frac{r_{ij}^2}{2\alpha_{ij}^2}$, the double sum runs over all pairs of atoms including $i=j$ (Srinivasan *et al.*, 1998). In an MD simulation study on DNA and RNA molecules, it was observed that the electrostatic contribution to solvation energies calculated by the computationally expensive Poisson-Boltzmann (PB) method was as good as the relatively less expensive generalized Born (GB) methods of solvation. In this study, the non-electrostatic contribution was estimated using a surface area (SA) dependent term (Srinivasan *et al.*, 1998). Implicit GB models along with ff98 have been shown to be a good choice to simulate RNA molecules (Gong and Xiao, 2010). Several studies have also used GB/SA approach for simulation of protein/peptides.

1.5.11 Advanced Application of MD simulations

1.5.11.1 Free Energy Calculations

Another major challenge for MD and MC simulations have been to determine the free energy change during solvation of molecules, ligand binding to proteins or nucleic acids and environmental effects on enzymatic reactions in solution (Andrew McCammon, 1991; Kollman, 1993). Energy minimization approaches are enthalpy driven and do not include the effects of entropy whereas Free-Energy perturbation (FEP) simulations include the effects of both entropy and enthalpy to binding energies (Kollman, 1993). FEP approach has been successfully used in protein-carbohydrate complexes (Liang *et al.*, 1996; Pathiaseril and Woods, 1999) and ligand-DNA complexes (Treesuwan *et al.*, 2009). Theoretical studies on protein dynamics associated with activated processes, like covalent bond rearrangements and movement of a ligand within a protein through steric bottlenecks, where important energy changes during the transition were associated with local displacement of atoms, have been performed to calculate otherwise experimentally accessible quantities like, rate constants and activation energies (Northrup *et al.*, 1982). For such systems, the simulation of the entire molecule would be rather inefficient as the biological activity is mostly linked to the dynamics of a local region. Therefore, the protein is divided into three regions based on their proximity to the active site *viz.* reaction region (atoms placed within a cutoff distance and directly/indirectly involved in the reaction), reservoir region (consisting of atoms placed beyond a cutoff distance and not participating in the reaction) and the buffer region (atoms lying between reservoir and reaction region) (Brooks *et al.*, 1985). It has been demonstrated that most consistent results were obtained when the reaction zone was treated with MD and the surrounding buffer region was constrained by harmonic forces (Brooks *et al.*, 1985). If free energies can be calculated for different states then the equilibrium constants and rate constants can be calculated. The choice of one of the several states of a system at equilibrium depends on the free energy difference between these states.

The agreement between calculated and experimental free energy has been found to be impressive with a statistical error <1 kcal/mol (Kollman, 1993). The free energy methods have been used to calculate free energy of aqueous and non-aqueous solvation of a molecule, molecular association (ligand binding to proteins) but the most exciting prospect for the free energy calculations has been to predict and help design the macromolecular inhibitors which can be used as drugs and indeed powerful insights into non-covalent interactions of complex molecules in solution have been obtained using these calculations (Kollman, 1993). The binding specificity of rat mannose-binding protein (MBP, a C-type lectin) for mannose and galactose has been studied using free energy simulations both in aqueous solution and in the bind-

ing site of the lectin. The obtained results were found both qualitatively correct and in reasonable agreement with experiment (Liang *et al.*, 1996). Kuhn and Kollman have been able to predict a ligand (a fluorinated analog of biotin) which could bind to avidin even better than biotin using MM-PBSA approach (Kuhn and Kollman, 2000).

1.5.11.2 Locally Enhanced Sampling

The continuous development in the methodologies for efficient conformational sampling of biomolecular systems has led to novel computational methods like multiple histogram methods (Ferrenberg and Swendsen, 1989; Hardin *et al.*, 2000) and locally enhanced sampling (LES) (Verkhivker *et al.*, 1992). Normal MD simulations of nano-second length at room temperature are incapable of overcoming conformational transition barriers and therefore can only sample the neighboring conformations of the starting structure (Golden and Olsen, 2008). Enhanced sampling is achieved by LES MD simulations as a consequence of presence of multiple copies of the protein/ligand and their higher effective temperature. The ligand copies interact with the system in an average way without interaction with each other, being free to move apart, explore different regions of conformational space and consequently, enhancing the statistical sampling. The higher temperatures reduce the barriers and result in more frequent conformational changes (Roitberg and Elber, 1991). The key feature is that the energy function is modified so that energy is identical to the original system when all LES copies have the same coordinates. Another key feature of LES simulations is that global energy minimum occurs when all copies occupy the position of the global minimum in the original system and that it is compatible with MD simulations with explicit solvation and particle-mesh Ewald summation (Simmerling and Elber, 1994). LES MD simulations have been used to study ligand binding pathways in truncated hemoglobins (Golden and Olsen, 2008), activation of the edema factor by calmodulin (Laine *et al.*, 2010), ligand discrimination in H-NOX domains (Zhang *et al.*, 2010), *etc.*

1.5.12 MD simulations on Glycoproteins

With the availability of force field parameters for carbohydrates (Woods *et al.*, 1995) it has been possible to carry out simulation studies on carbohydrates, glycoproteins, glycolipids and protein-carbohydrate complexes. In a recent study, the trans-membrane region of glycophorin A (GpA) was modeled using a coarse grain force field, MARTINI (Marrink *et al.*, 2007). The association of GpA monomers was examined using Monte Carlo simulations, and the results of simulations were found to be in excellent agreement with existing experi-

mental data. This study also provided further insights into the mechanism of protein dimerization (Janosi *et al.*, 2010). The lipid free form of Endothelial protein C receptor (EPCR), a trans-membrane glycoprotein having important regulatory roles in protein C pathway, has been studied by multiple 20ns MD simulations performed using GROMACS4.0 (Hess *et al.*, 2008) and GROMOS96 force field. It was found that, in absence of the phospholipid ligand, the lipid binding groove is narrowed due to structural rearrangement of helices (Chiappori *et al.*, 2010). Based on these observations, it was concluded that the phospholipid ligand was required to maintain the correct conformation of EPCR for protein-protein interactions with Gla domain of Protein C. The specific recognition and binding of carbohydrate substrates to a highly potent anti-HIV lectin, Cyanovirin-N (CVN), and its mutant P51G-m4-CVN has been studied by MD simulations for time scales greater than 50ns using AMBER and GLYCAM06 force fields (Vorontsov and Miyashita, 2009). It was found that O3 and O4 atoms of mannose residues formed a hydrogen bonding network predominantly with the main chain atoms. The binding of PECAM-1 and annexin A2 molecules to heparin fragments of different sizes has been studied using MM/PBSA and MM/GBSA analysis of the MD trajectories obtained from AMBER9, using PARM94 and GLYCAM04 force fields (Gandhi and Mancera, 2009). The simulations revealed that hinge-type conformational change in Ig-domains 2-3 of PECAM-1, makes basic residues accessible on surface and facilitates the binding of a longer heparin with higher affinity. Similarly, from annexin A2 simulations, it was observed that the flexibility of substrate heparin and the loops of annexin A2 allowed the interactions between a surface bound Ca^{2+} ion and the substrate. Computer simulations have also been used to obtain novel and important insights on structural stability of lectins and their interface stability. For example, the thermal unfolding of banana lectin has been studied (Gupta *et al.*, 2009) by MD simulations using AMBER8 package at elevated temperatures of 300, 400, 500 and 600 K. This study revealed the importance of hydrogen bonds in stabilizing the dimer interface. Graph theory based analysis of residue interaction networks helped in elucidating the early events of thermal unfolding of peanut agglutinin (PNA) in several 2ns long MD simulation trajectories (Hansia *et al.*, 2007). This study has been carried out using AMBER7 package and Parm98 force field. Unfolding pathway involved the deoligomerization of the protein wherein the secondary structure was primarily retained but an extensive loss of the tertiary structure was observed (Hansia *et al.*, 2007). Binding of mono- and disaccharides to RicinB has been studied using 3ns long simulations carried out by CHARMm package (Ganguly and Mukhopadhyay, 2006). The discrimination between the substrate galactosides and glucosides

by the binding site-I could be easily observed. Using the thermodynamic integration (TI) method of free energy simulation, galactose was found to be a stronger substrate compared to glucose. Using the linear interaction energy approach, the free energy calculations for binding site-I were also found to be in agreement with the experimental values (Ganguly and Mukhopadhyay, 2006). Binding of phenyl substituted trimannoside to ConA and garlic lectin has been studied by MD simulations (Mazumder and Mukhopadhyay, 2010) using GROMACS package (van der Spoel *et al.*, 2005) and OPLS-AA forcefield (Rizzo and Jorgensen, 1999). Based on the analysis of 10 ns long MD trajectories, it was found that presence of the aromatic phenyl group has no effect on binding of the substituted trimannoside to ConA due to absence of a hydrophobic region near its substrate binding site. However, due to the presence of such a region in garlic lectin a strong binding of the phenyl substituted trimannoside was observed. In another study, the geometrical details of substrate binding in banana lectin have been studied (Sharma and Vijayan, 2010) by MD simulations using GROMACS package and OPLS-AA/GLYCAM06 force fields. It was observed that ligand binding involved a combination of induced fit mechanism and conformational selection. In this thesis we discuss the MD simulations of ConA and EcorL performed using AMBER9 package, GLYCAM-04, GLYCAM -06 and FF03 force fields.

1.5.12.1 MD simulations on ConA

MD simulations have been used to explore ConA structure as well as substrate binding. The differences between binding of mannose and glucose to ConA have been studied in detail using MD simulations (Bradbrook *et al.*, 1998). This simulation study could explain why mannose was a stronger substrate than glucose. The O2 of mannose interacted with Thr226 and Leu99 of ConA either through water mediated interactions or direct hydrogen bonds leading to an increased interaction energy for mannose compared to glucose (Bradbrook *et al.*, 1998). MM-GB/SA approach has also been used to analyze the binding of trimannoside and pentasaccharide to ConA (Bryce *et al.*, 2001). It was observed that as in X-ray studies (i) the 1-3 arm of the pentasaccharide was more flexible than the rest of the oligosaccharide, (ii) the reducing mannose had least interaction energy and (iii) the additional affinity of trimannoside for ConA originated from the 1-3 mannose interaction (Bryce *et al.*, 2001). Further, the effects of water displacement on the thermodynamics of binding of the trimannoside to ConA have also been studied using inhomogeneous fluid solvation theory and MD simulations (Li and Lazaridis, 2004). The importance of involvement of the con-

served water molecules was further emphasized using results obtained from experimental and computational methods involving thermodynamic integration (Kadirvelraj *et al.*, 2008).

1.5.12.2 MD simulations on EcorL

The oligosaccharide moiety of EcorL has the same pentasaccharide core as seen in many other glycoproteins like RNaseB and the proteolytic enzyme from pineapple stem *i.e.* bromelain (Bouwstra *et al.*, 1990; Stubbs *et al.*, 1996; Woods *et al.*, 1994). MD simulations have been used to obtain the solution conformation of the oligosaccharide which was then modeled onto the crystal structure of RNaseB (Woods *et al.*, 1994). Simulations on the oligosaccharide moiety of EcorL have been performed in vacuum (Qasba *et al.*, 1994). MD simulations in explicit water environment have also been carried out for glycosylated EcorL for time scale upto 300ps only (Naidoo *et al.*, 1997). These simulations showed that EcorL oligosaccharide and rest of the crystal structure were relatively stable without any packing constraints and that the presence of solvent and longer simulations were important in molecular mechanical modeling to obtain results which correlate with experimental reports (Naidoo *et al.*, 1997). MD simulations have been used to study the binding of its substrate saccharides (galactose, lactose, and lactosamine) to EcorL. These studies indicated that the α anomer of galactose could form more water mediated interactions with the protein than the β anomer (Bradbrook *et al.*, 2000).

Chapter 2

Role of metal ions in substrate recognition and stability of Concanavalin A: A molecular dynamics study

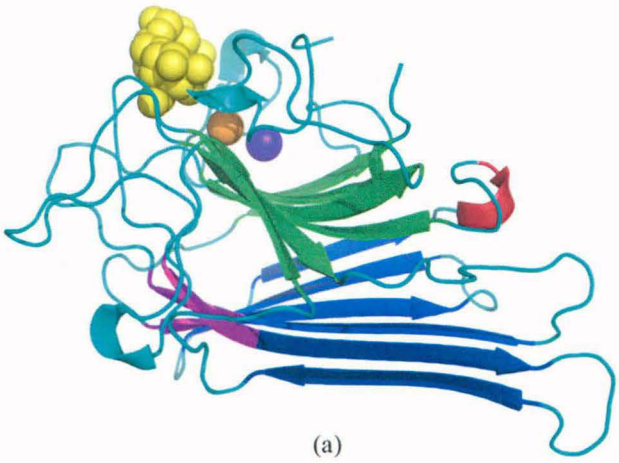
2.1 INTRODUCTION

Lectins are a class of proteins that interact with various glycoproteins and control a variety of processes like defense against pathogens, legume nodulation, mitogenesis of lymphocytes possibly through formation of "micropatches", fertilization, tumor metastasis, and development etc. (Goldstein and Hayes, 1978; Hirsch, 1999; Nangia-Makker *et al.*, 2002; Peumans and Van Damme, 1995; Shur and Hall, 1982). Concanavalin A (ConA; *Canavalia ensiformis* agglutinin) is a prototype member for legume lectins (Ambrosi *et al.*, 2005) and is the first lectin whose crystal structure was solved (Edelman *et al.*, 1972). ConA is a homotetrameric lectin, with each monomer comprising of 237 amino acid residues (Cunningham *et al.*, 1975; Greer *et al.*, 1970). ConA monomers adopt the "jelly-roll" fold, a fold conserved in legume and several other member of lectin family (Bouckaert *et al.*, 1999a). It consists of 3 β-sheets, referred to as 'back', 'front' and 'top' β-sheets (Figure 2.1) with 6, 7, and 5 antiparallel β-strands, respectively (Ambrosi *et al.*, 2005; Edelman *et al.*, 1972; Reeke *et al.*, 1975b).

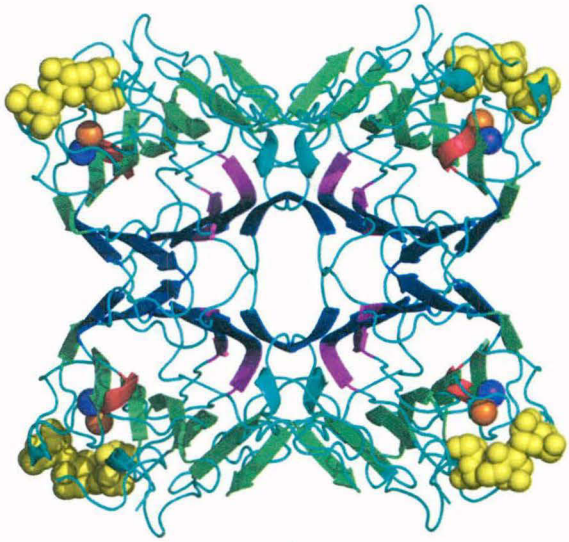
Each monomer of ConA has only one substrate binding site and thus ConA is tetravalent. ConA has its isoelectric point at pH 7.1 (Agrawal and Goldstein, 1967b) and exists as a dimer below pH 6 and as a tetramer near pH 7 (Hardman *et al.*, 1971; McKenzie *et al.*, 1972). It has been proposed recently that the stability and pH-dependence of the tetramer is influenced by two His residues (His 51 & 131) (Del Sol *et al.*, 2007). Legume lectins typically exist as dimers or as tetramers (Ambrosi *et al.*, 2005), with several intra-molecular orientations of the monomers with respect to each other in order to gain increased avidity and specificity for their respective substrates (Drickamer, 1995b; Rini, 1995).

Substrates for ConA include both mannose and glucose though thermodynamic studies indicate that mannose binds strongly to ConA compared to glucose (Gail M. Bradbrook, Research, 1998). The sugar binding pocket in legume lectins is conserved and consists of residues Tyr12, Pro13, Asn14, Thr15, Asp16, Leu99, Asp208, and Arg228 in ConA. The interaction between the lectins and substrate sugars include hydrogen bonds (So and Goldstein, 1967), ring stacking, burial of hydrophobic surface area (Elgavish and Shaanan, 1997) and, a few water mediated interactions (Bradbrook *et al.*, 1998; Pratap *et al.*, 2001).

The importance of metal ion binding in folding of ConA to its native structure has been established in recent biophysical studies on folding and oligomerization (Sinha *et al.*, 2005). Early studies have also suggested that, the removal of the divalent ions from lectin abolishes its sugar binding ability (Yariv *et al.*, 1968) though, the divalent ions don't interact



(a)



(b)

Figure 2.1

The cartoon diagram depicting monomeric (a) and tetrameric (b) ConA (1CVN). The yellow, orange and purple colored spheres depict trimannoside, calcium and manganese ions, respectively.

directly with the saccharide as in calcium dependent animal lectins (Drickamer, 1995a). Despite the importance of metal ions in controlling structural stability and substrate binding ability of ConA, no atomically detailed picture is available for the conformational dynamics of these folding and binding events. Several crystallographic studies have attempted to understand the structural basis of demetalization and consequent abolition of the sugar binding abilities of lectins (Bouckaert *et al.*, 1995; Reeke *et al.*, 1978; Shoham *et al.*, 1978). However, due to the difficulty in interpretation of the electron density for the ion binding loop in demetalized ConA (Reeke *et al.*, 1978) it has not been possible to elucidate the conformational changes associated with demetalization process.

Computational approaches are being increasingly used to explore the structure and interactions of various biological macromolecules. One such computational approach is Molecular Dynamics (MD) which enables us to explore the biological systems at an atomic level and can potentially give valuable insights about dynamic properties of the system. This

makes MD a powerful technique for understanding intricacies of the interactions which govern the folding and binding processes of biological macromolecules. MD simulations have been used in a number of studies *e.g.* in a recent study, it was used to explore the thermal unfolding of PNA and the role of metal ions in its structure and stability (Hansia *et al.*, 2007). In another study on ConA, MD simulations have revealed some extra interactions which were not seen in the static crystal structures. Similarly, MD simulation have been use to study the the protein-ligand interactions in peanut agglutinin (PNA) complexes with T-antigen and lactose (Pratap *et al.*, 2001). These studies demonstrated that MD can be complementary to crystallographic and other experimental methods (Bradbrook *et al.*, 1998; Pratap *et al.*, 2001) and together they provide a greater insight than available while using them individually.

In this work we have carried out extensive explicit solvent MD simulation studies for ConA in presence and absence of metal ions and trimannoside substrate. The objective of this study has been to understand the role of metal ions in substrate binding and structural stability of ConA. The crystal structure of ConA trimannoside complex (1CVN) (Naismith and Field, 1996) has been simulated in aqueous environment for 3 ns with modifications like removal of the ions and/or the trimannoside sugar. Simulations have also been carried out for the monomeric and dimeric forms of ConA for understanding the role of metal ions in the oligomerization process.

2.2 METHODS

The molecular dynamics simulations were performed using parallel version of AMBER9 (Case *et al.*, 2006) and were based on NVT ensemble at a temperature of 300K. Our primary objective was to study the dynamics of the metal ions and substrate binding at physiologically relevant conditions, hence, 300K was the temperature of choice. Moreover, substrate and ion binding does not involve large scale conformational changes in the protein so there was no need to consider volume fluctuations. So, we proceeded with NVT simulations. We believe NPT simulations will also give similar results. The starting coordinates of the ConA complex were from the PDB entry 1CVN (Naismith and Field, 1996), which is a complex of ConA and the N-linked glycan core trimannoside (Manα1-6(Manα1-3)Man) (Naismith and Field, 1996). The simulations were carried out using explicit solvent environment consisting of TIP3P water (Jorgensen *et al.*, 1983) molecules. The water box extended to 7Å beyond the outermost atom of ConA or ConA-sugar complex on either of side of the x, y, and z axes. The simulations were carried out both in the presence and absence of bound

metal ions and trimannoside substrate. Since we have chosen a relatively small solvent box around the protein, no free counter ions were added to the system for charge neutralization. Addition of free counter ions in the solvent might have interfered with protein bound metal ions.

The solvated system was minimized to get rid of any unfavorable contacts and then equilibrated for 20 picoseconds (ps) to bring the temperature from ~0°K to 300K, after which production dynamics was run for 3 nanoseconds (ns). A time-step of 1 femto-second was used in all the simulations. A cut-off radius of 8Å was used for all the simulations. The periodic boundary conditions (PBC) were used to negate the surface effects at the box boundaries. Particle-Mesh Ewald (PME) summation method (Darden *et al.*, 1993) was used for the calculation of electrostatic potential. The bonds containing hydrogen atoms were constrained using SHAKE (Wilfred F. van Gunsteren Review, 1990), and hence concomitantly, bond interactions involving hydrogen atoms were not calculated. The coordinates were saved every pico-second, thus the trajectory from a 3 ns simulation consisted of 3000 frames.

The force fields used for the simulations were ff03 (Duan *et al.*, 2003) for protein parameters and glycam04 (Woods *et al.*, 1995) for sugar parameters. The force field parameters for the ions Mn^{2+} and Ca^{2+} (Bradbrook *et al.*, 1998) were obtained from AMBER contributed parameter database at http://pharmacy.man.ac.uk/amber. The *ptraj* module of AMBER9 was used for the calculation of RMSD, APF (atomic positional fluctuations), B-factors, hydrogen-bonds and radius of gyration (Rg) of the protein. NACCESS v2.1.1 (Hubbard, 1993) was used for surface area calculations.

2.3 RESULTS

The various different environmental conditions and oligomeric states under which simulations were carried out for ConA are summarized in Table 2.1. The first simulation consisted of the native tetrameric crystal structure of ConA in complex with trimannoside sugar and bound Ca^{2+} and Mn^{2+} ions in aqueous environment. In order to mimic the aqueous environment, the water molecules from the crystal were removed and the solvent molecules were introduced from the solvent box. In the subsequent simulations ions and sugar molecule were removed one by one to see the effects of these constituents on the behavior of ConA molecule. This resulted in four environmental conditions viz. presence of both ions and the trimannoside (E1), presence of ions in absence of the sugar (E2), presence of the sugar in the absence of ions (E3), and the absence of both ions and the sugar (E4). Under each of these

Simulation		Concanavalin A (ConA)	Ions (Mn^{2+} & Ca^{2+})	Trimannoside Sugar	Length of Simulation (ns)
E1	Monomer	+	+	+	3
	AB Dimer	+	+	+	3
	Tetramer	+	+	+	3
E2	Monomer	+	+	-	3
	AB Dimer	+	+	-	3
	Tetramer	+	+	-	3
E3	Monomer	+	-	+	7
	AB Dimer	+	-	+	3
	Tetramer	+	-	+	3
E4	Monomer	+	-	-	7
	AB Dimer	+	-	-	3
	Tetramer	+	-	-	3

Table 2.1
The summary of various simulations carried out for monomeric, dimeric and tetrameric structures of ConA. The "+" and "-" indicate presence and absence of the trimannoside substrate or ions in the simulations, respectively.

environmental conditions, simulations were carried out for ConA tetramer for 3 ns each. Monomer and dimer of ConA were also simulated for 3ns each under four similar environmental conditions. Thus, trajectories were collected for a total of 12 simulations. These 12 MD trajectories were analyzed to understand the effect of different environmental conditions on structure of ConA and its substrate recognition.

2.3.1 Effects of metal ions and trimannoside on structure of ConA

2.3.1.1 ConA tetramer in complex with sugar and metal ions (E1)

Figure 2.2 shows the backbone RMSD (Cα, C and N atoms) for the structures sampled during various simulations from their respective starting structures. The 3ns trajectory for the ConA tetramer (bottom panel in Figure 2.2) showed that, in presence of bound metal ions the equilibrated structure deviates by only about 1.5Å from the starting structure. The

major structural changes occur during initial 10-20 ps with a sudden change in the RMSD to ~1Å, and then the structural deviations continue rather slowly for the rest of the simulation to a final value of about 1.5Å. The structural deviations with respect to the original crystal structure were also calculated using ProFit 2.5.3 (Hubbard, 1993), and they were found to vary in the range of 1-1.5Å (Figure 2.3) for the native ConA tetrameric sugar complex in presence of bound metal ions. We also computed the radius of gyration (R_g) as a function of time, as destabilization of the protein core structure would result in large increase of R_g values. The radius of gyration plots (Figure 2.4) for the protein showed only a small increase of about 0.3Å, which would translate into a change of about 5% in the volume.

We also wanted to investigate whether the structural deviation observed above, is confined to few local regions, or over the whole protein. Using the positional fluctuations observed over the MD trajectory, we computed the theoretical B-factors (BF) for the protein and compared with the experimental BF values as seen in the crystal structure 1CVN. Figure 2.5 shows a detailed comparison of the experimental and theoretical B-factor values for chain A of ConA tetramer. The experimental BFs are also depicted in the Cα trace of ConA shown in the inset. As can be seen, the regions having high experimental BFs indeed correspond to the peaks in the computed BF curve and they are mostly confined to the loop regions of ConA. This suggests that there is good agreement between experimental and theoretical B-factor values. In fact the average BF of the tetramer computed from E1 simulation was observed to be 17.85Å^2 (Table 2.2) over the backbone atoms (CA, C, and N). It is interesting to note that the average experimental BF seen in the crystal structure of ConA (Naismith and Field, 1996) is 16.85Å^2, which is very close to the theoretical value observed from our simulation. However, few relevant issues must be kept in mind while comparing the theoretical and experimental BF values. The experimental BFs or "atomic temperature factors" are typically used to quantitate the dynamic disorder in the position of the atoms caused by the temperature-dependent motion. Apart from the thermal motions BFs also have contributions from other sources of experimental errors. Secondly, the MD simulations have been carried out at 300K, while the crystal structure was solved at 150K. Therefore, the comparison between the two can only be qualitative. The apparent agreement between the two is only because of the fact that loop regions have relatively higher mobility compared to tightly packed core of the protein. Figure 2.6a shows the theoretical BFs for all four chains of the tetramer from four different environmental conditions.

If the BF of a region is high it is expected to have high temperature-dependent vibra

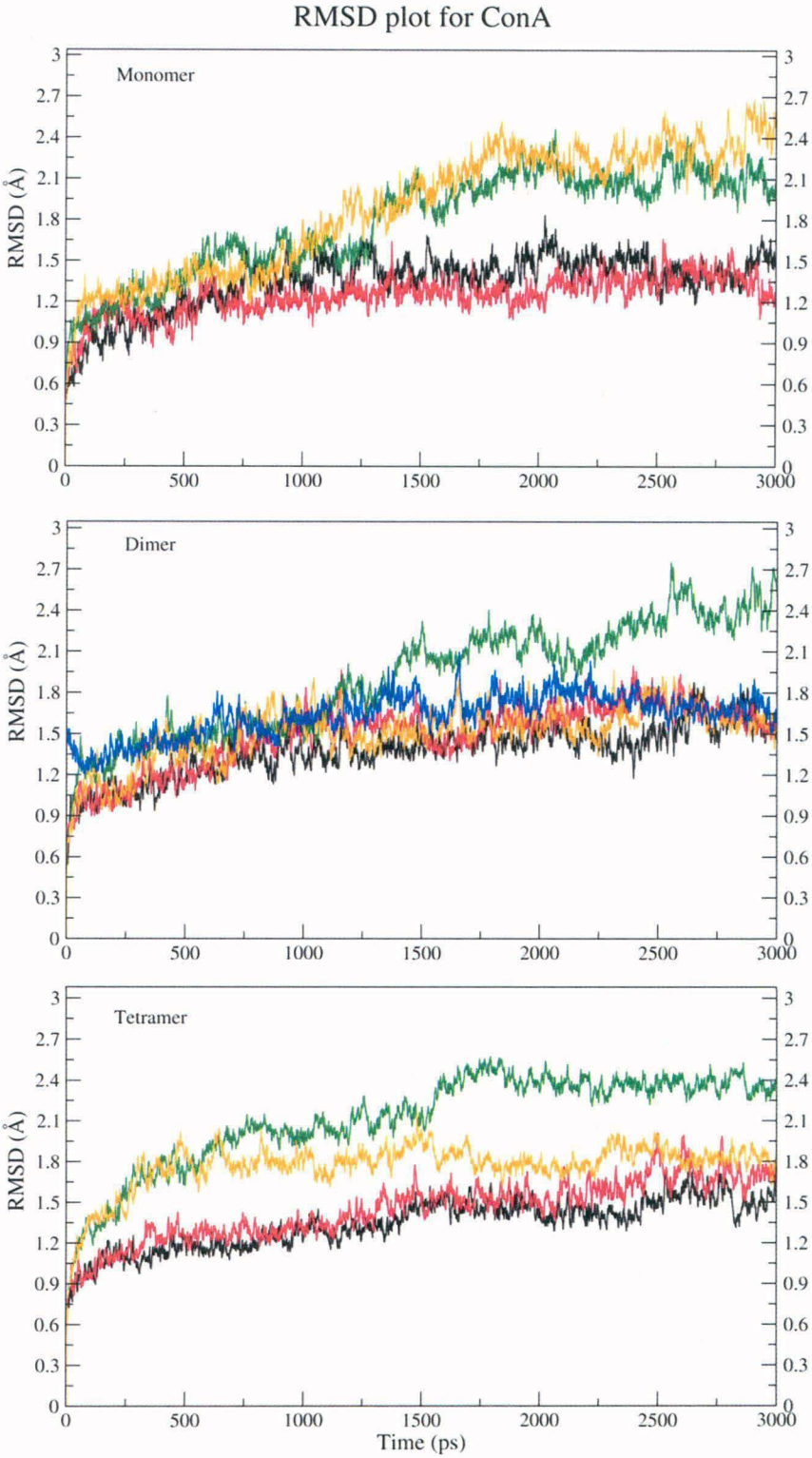


Figure 2.2
RMSD Plot for the structures sampled during simulations with respect to the starting structure. Top, middle and bottom panels show the RMSD plots for monomer, AB dimer, and tetramer, respectively. The results from different simulations namely E1, E2, E3 and E4 have been shown using black, red, green and orange colored lines, respectively.

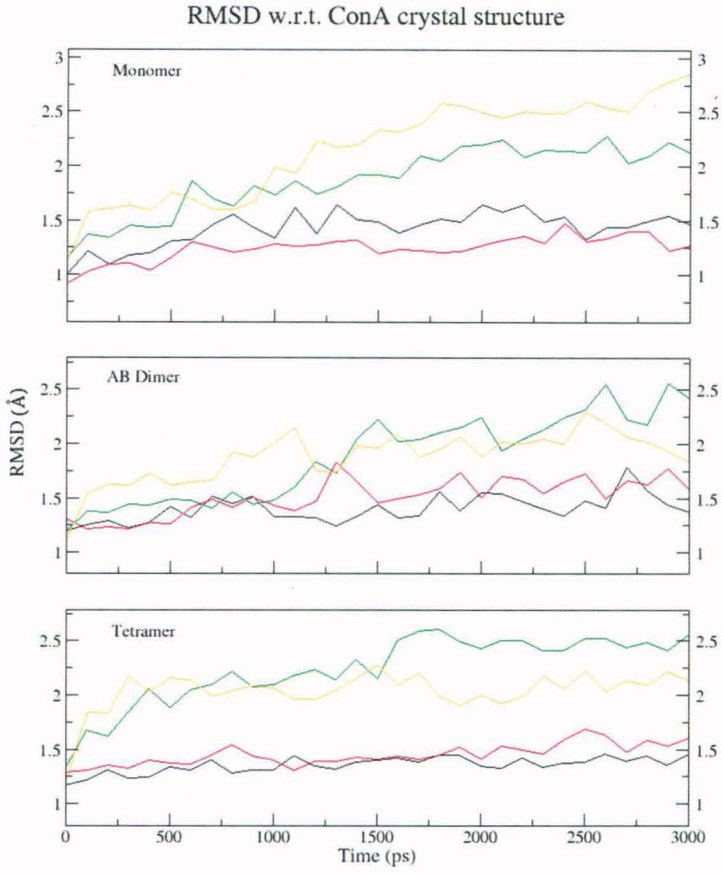


Figure 2.3
The plot showing RMSD between the crystal structure (1CVN) of ConA tetramer and the structures extracted from the trajectories of all the four simulations (E1 to E4) of the monomer (top), dimer (middle) and tetramer (bottom).

tions and consequently, it can undergo structural changes easily and frequently. So we explored the regions with high experimental BF values. Only one region, which is the loop 160-164, showed a consistent high BF value (>40Å^2) in all of the chains of the tetramer. Besides this, the regions 183-187 and 201-205 showed a high BF value but not uniformly in all the chains i.e. region 183-187 was more flexible in chain C only, while 201-205 region has high BF only in chain B, in the rest of the chains these regions had normal (approximately $\leq$ 40Å^2) BF values. All these regions mentioned above correspond to loop regions of the molecule or amino acid stretches involved in ligand binding.

Thus, the analysis of structurally flexible regions in ConA suggests that, the 1.5Å RMSD between the starting structure and the final structure from dynamics run may indeed arise from structural deviations in the loop regions only, while the core of the ConA tetramer remains intact and relatively rigid during the dynamics run. This inference is also supported by the observation that there is only about 2% change in the solvent accessible non-polar sur-

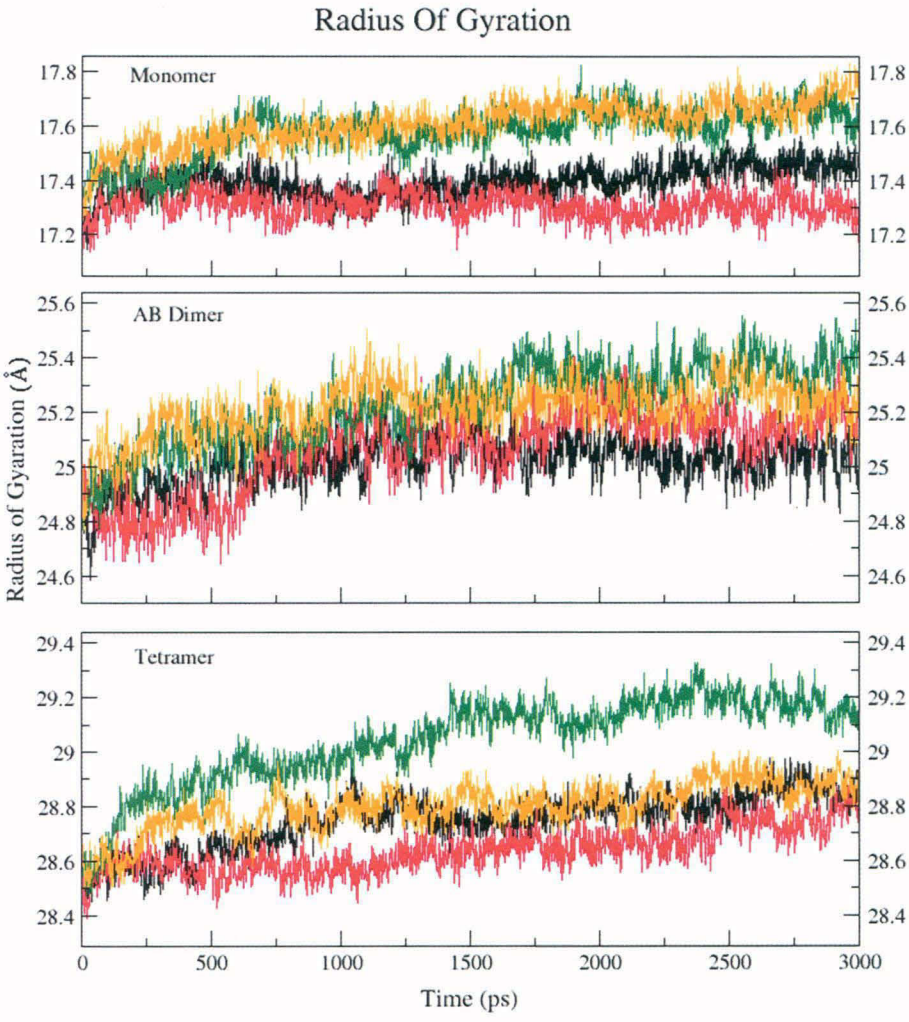


Figure 2.4
Plot depicting the variation of the radius of gyration for the monomer (top), AB dimer (middle), and tetramer (bottom) of ConA in various MD trajectories. Black, red, green and orange colored lines depict the results for E1, E2, E3 & E4 simulations, respectively.

face area (ASA) for the trimannoside bound tetramer (Figure 2.7). This means that the ConA-trimannoside complex is already in its best folded and stable state so that only a slight alteration is observed in its packing after 3ns simulation (as decrease in exposed non-polar surface area). Next we examined whether there were any changes at the secondary structure level during the simulation. For this we identified the hydrogen bonds present in the structures sampled during the dynamics run and analyzed separately the hydrogen bonds involving main-chain to main-chain (M-M) and those involving side chains (S-M/S). The M-M hydrogen bonds, which can be thought of as a measure of the secondary structure indicated a slow drop in numbers, ~8% over the 3ns (Figure 2.8a). In contrast to M-M hydrogen bonds, the hydrogen bonds involving side-chains showed a comparatively fast decline in number (about 20%) over 3ns (Figure 2.8b). This slow decline in the number of M-M hydrogen bonds and a

Oligomeric status	Simulation	RMSD (Å)	Radius of Gyration (Å)	Non-polar ASA (%)	Average BF (Å^2)
Monomer	E1	1.41 ± 0.09	17.57 ± 0.03	54.15 ± 0.70	14 ± 18
	E2	1.38 ± 0.09	17.27 ± 0.77	55.30 ± 0.56	12 ± 15
	E3	2.13 ± 0.10	17.74 ± 0.04	53.01 ± 0.50	24 ± 47
	E4	2.38 ± 0.13	17.65 ± 0.79	54.64 ± 0.65	22 ± 49
Dimer	E1	1.61 ± 0.11	25.22 ± 0.06	53.85 ± 0.30	17 ± 16
	E2	1.65 ± 0.08	25.11 ± 1.13	54.31 ± 0.29	20 ± 21
	E3	2.43 ± 0.12	25.61 ± 0.08	54.41 ± 0.33	29 ± 91
	E4	1.65 ± 0.09	25.23 ± 1.13	54.65 ± 0.21	22 ± 29
Tetramer	E1	1.54 ± 0.08	29.07 ± 0.03	52.17 ± 0.25	18 ± 17
	E2	1.72 ± 0.09	28.69 ± 1.28	52.59 ± 0.26	19 ± 20
	E3	2.37 ± 0.05	29.48 ± 0.04	52.79 ± 0.28	31 ± 75
	E4	1.83 ± 0.08	28.78 ± 1.29	53.68 ± 0.38	21 ± 30

Table 2.2
The converged values for various structural parameters (RMSD from the starting structure, Radius of gyration, Percentage of non-polar ASA) extracted from the last 500 ps of the 3ns simulations. Standard deviations (SD) are given in the brackets. The table also shows the computed B-factor values averaged over all the residues of the protein. The large SD values for computed average B-factors arise from the presence of specific loop regions with very high B-factor values.

relatively faster decline in the number of S-M/S hydrogen bonds suggest that, there is only a minor change in the secondary structure elements, but there has been significant rearrangement of the conformations of the side chains during the simulation. Based on these results, it can be concluded that during the dynamics the core of the ConA tetramer remains relatively rigid and dynamic structural changes in the core essentially correspond to relaxation of side chain orientations. These observations reiterate the fact that the conditions needed for a simulation of native ConA were met sufficiently enough so that no anomalous behavior was observed during the 3ns simulation.

2.3.1.2 ConA tetramer in absence of trimannoside (E2)

The trajectory from the simulation in absence of bound trimannoside (E2) was analyzed to understand the effects of substrate binding on the dynamic fluctuations in the ConA

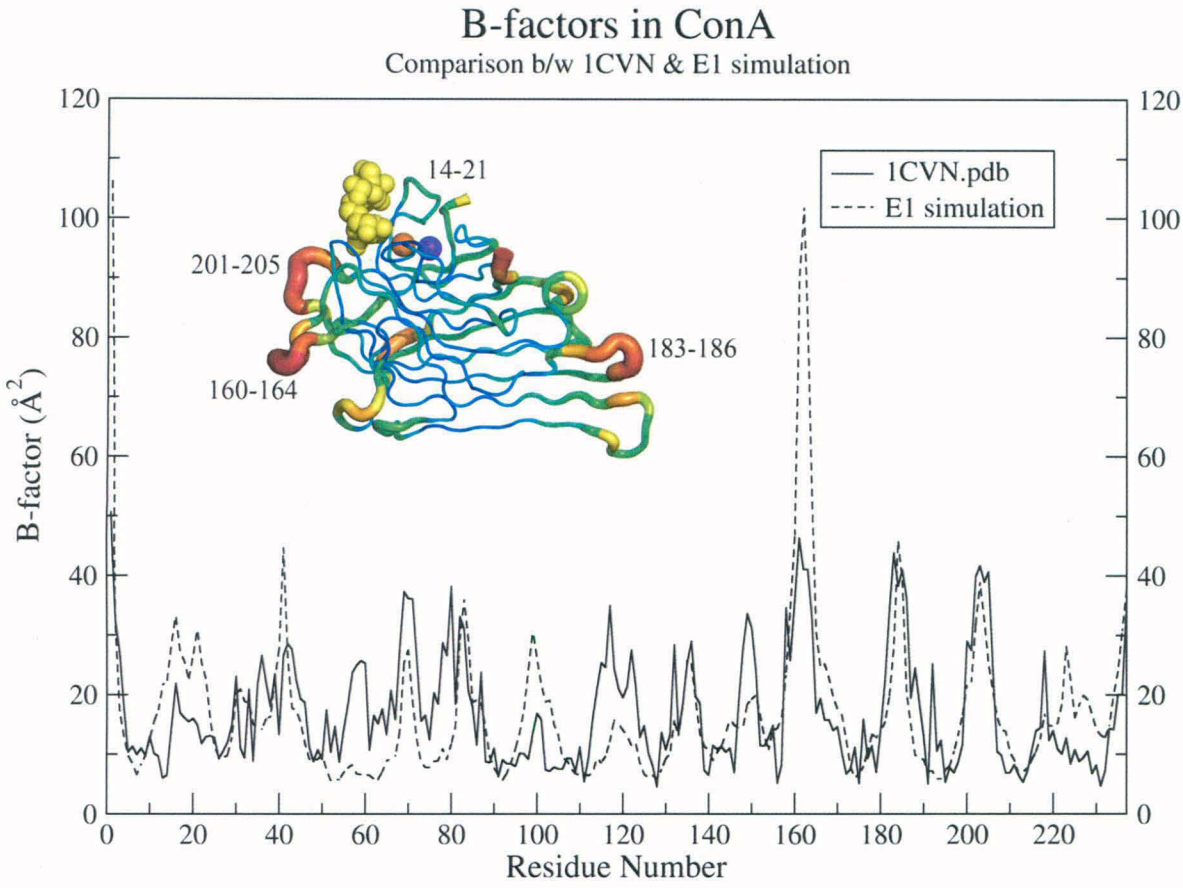


Figure 2.5
Comparison between the theoretical and experimental BFs. The plot shows the computed (dashed line) as well as experimental (solid line) BF for chain A of the ConA tetramer. The theoretical BFs have been computed using the trajectory of the simulations for substrate bound tetramer in presence of metal ions (E1). The inset shows the C-alpha trace for the crystal structure of the substrate bound ConA monomer with thickness and color of the chain reflecting the experimental BF values of the constituent residues.

tetramer. However, the structural deviations were negligibly different when compared to the simulation involving substrate bound ConA. The final RMS deviation with respect to the starting structure of the simulation was ~1.5Å at the end of 3ns simulation (Figure 2.2). The RMSD with respect to the crystal structure was ~1.5Å as well (Figure 2.3). The change in radius of gyration (Rg) was also similar as in E1 i.e. a change in Rg of ~0.4Å (Figure 2.4).

Similarly, no appreciable changes were observed in the non-polar ASA (Figure 2.7) or M-M and S-M/S hydrogen bonds (Figure 2.8) when compared to the E1 simulation. However, significant differences were observed in the thermal flexibility of the amino acid stretch 201-205 and also the region 160-164 (Fig. 2.6a). The 201-205 loop region is involved in substrate binding and had increased BF values (>60Å^2) in all the chains except chain A. The dynamic fluctuations of the loop corresponding to 160-164 appear to be correlated to the mo-

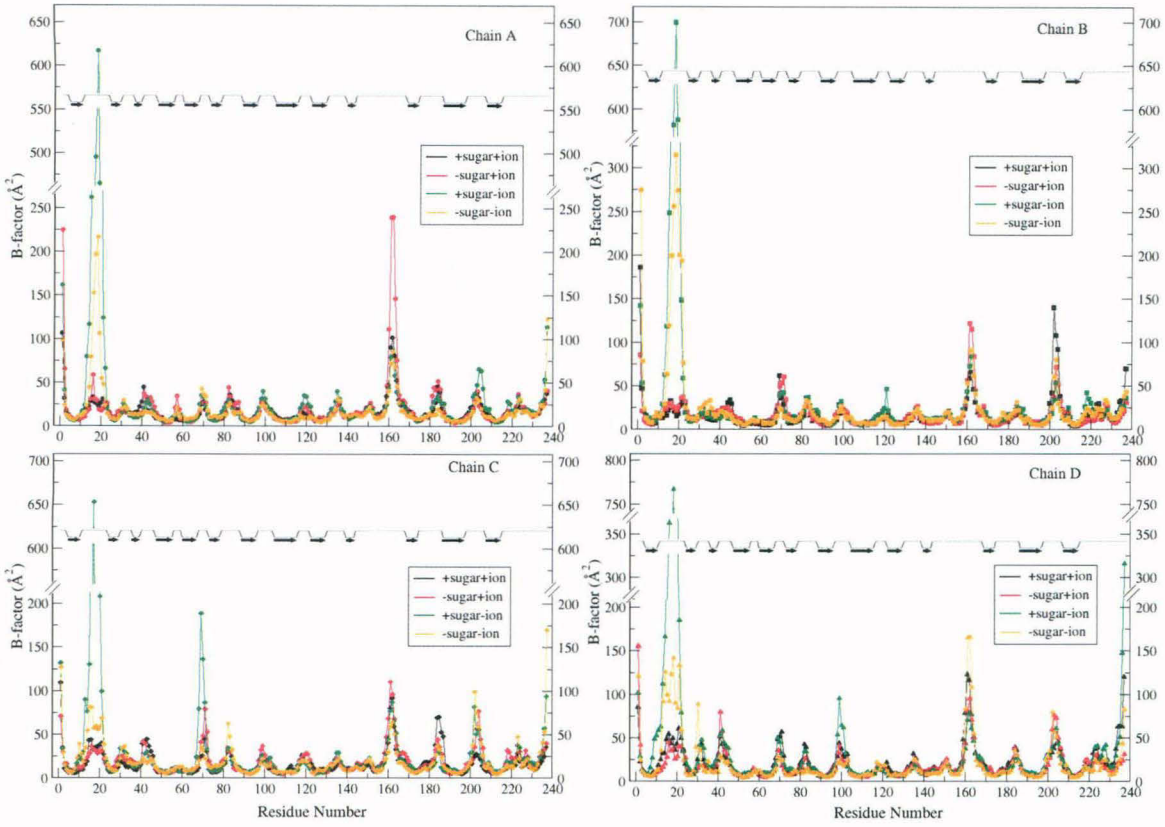


Figure 2.6(a)
The plot showing the theoretical residue-wise BFs computed from the MD trajectories for the four chains of ConA tetramer during the four different simulation conditions, E1, E2, E3 and E4. The results from the four different simulations (E1, E2, E3, and E4) have been shown using black, red, green, and orange lines, respectively.

vements of the 201-205 region, even though this stretch is not directly involved in substrate binding.

2.3.1.3 ConA tetramer in absence of metal ions (Mn^{2+} and Ca^{2+}) but in complex with sugar (E3)

The E3 simulation was run for ConA in the absence of the metal ions but in complex with the trimannoside sugar. As can be seen from Figure 2.2, in this simulation the RMSD has increased continuously to more than 2Å at the end of 1.5 ns and at about 1.8ns the RMSD had reached the maximum value of ~2.5Å. The radius of gyration for ConA increased by about 0.7Å (Figure 2.4) with a corresponding ~8% increase in the volume. Analysis of the local structural fluctuations indicated that, the large RMSD of the overall structure resulted from high local mobility of certain specific amino acid stretches of ConA protein. In the absence of ions the segment which showed highest thermal motion was the region corresponding to amino acids 14-21 (Figure 2.6a). The minimum computed BF for this region was more

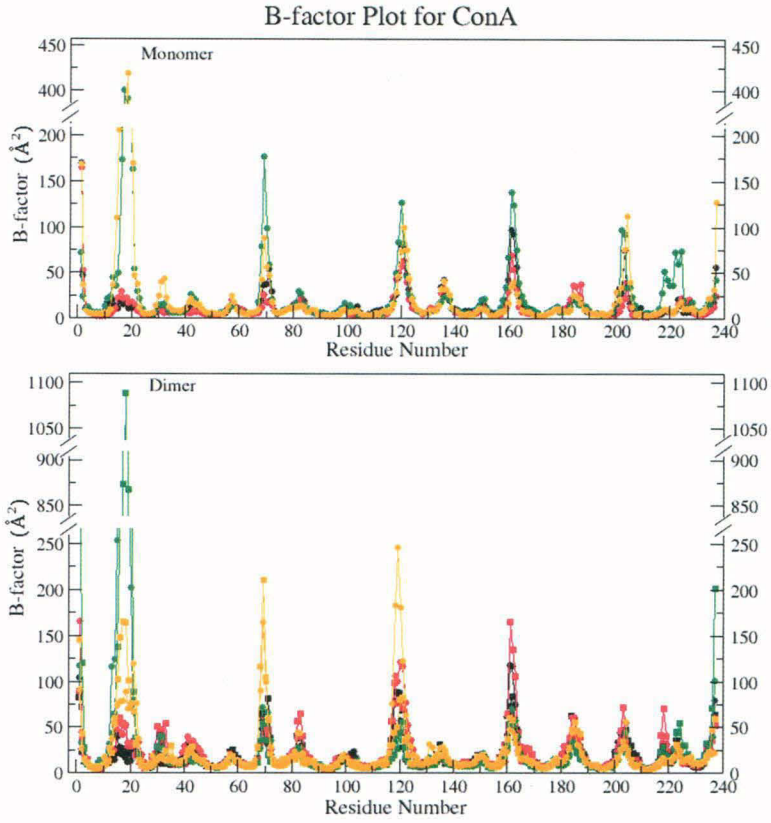


Figure 2.6(b)
The plot showing the theoretical residue-wise BFs computed from the MD trajectories for ConA monomer and AB dimer during the four different simulation conditions, E1, E2, E3 and E4. The results from the four different simulations (E1, E2, E3, and E4) have been shown using black, red, green, and orange lines, respectively.

than 600Å^2 for chain A (Figure 2.6a), while the other chains displayed even higher BF values for this stretch. The 68-71 region has highest BF value in chain C (BF >180Å^2) while in the rest of the chains this region was comparatively less mobile (BF <50Å^2). Similarly, 98-101 residue region had a high BF value (~100 Å^2) for chain D compared to other chains (BF <40Å^2). The 160-164 residue region had BF values ~80Å^2 for all the chains. The 201-205 residue region had BF values about 60Å^2.

The non-polar ASA of the tetramer remained around 52.5% throughout the simulation, except for transient time periods where there was about 1% increase observed (Fig. 2.7). The M-M hydrogen bonds dropped by about 7.6% within 400ps of simulation (Fig. 2.8a) but then there was no further drop. And, similar to the previous simulations, there was about 20% drop in S-M/S hydrogen bonds (Figure 2.8b). These results further confirm that, absence of bound metal ions has only a minor effect on the overall core of the tetrameric structure of ConA, while the metal ions seem to predominantly alter the mobility of the specific loop

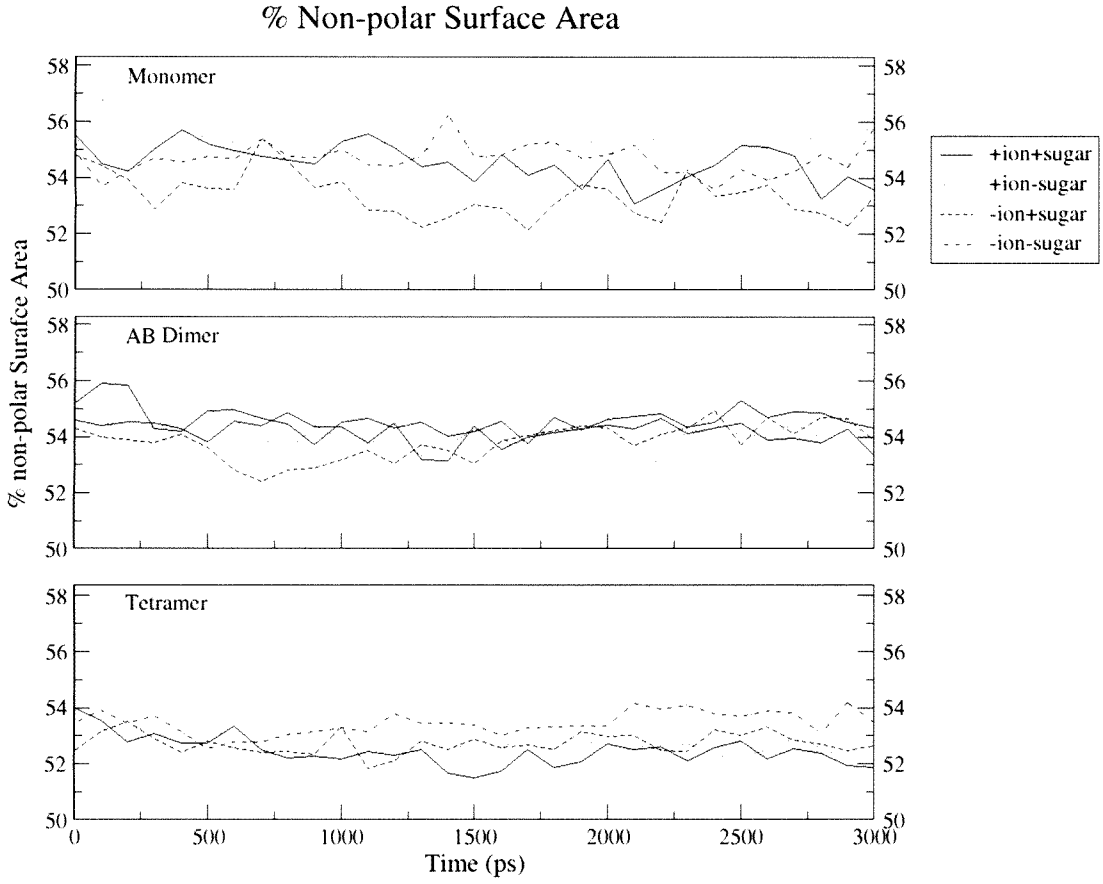


Figure 2.7
The variation of the non-polar solvent accessible surface area for ConA during the 3ns simulations have been shown for monomer (top), AB dimer (middle), and tetramer (bottom), respectively..

regions. Figure 2.9 shows the residue-wise RMSD (rRMS) with respect to the crystal structure (1CVN) for the final structures obtained from the E1 and E3 simulations. As can be seen, the results essentially agree with the results obtained from the analysis of computed BFs *i.e.* the structural deviations are mostly restricted to the loop regions and maximum deviations in absence of the ions are confined to the ion binding loop region only. Some of the residues in the ion binding loop region moved as far as 17Å with respect to their position in the starting structure (Figure 2.9). To further confirm the observation that the absence of metal ions does not lead to any major change in the core of the protein, we also analyzed the RMSD of the core (β-sheet) and loop regions separately over the entire 3ns trajectory (Figure 2.10). The β-sheet core region showed RMSD of only 0.75Å (Figure 2.10) over the entire 3ns trajectory both in presence and absence of bound metal ions, thus suggesting that the core region is highly stable even upon demetalization. In contrast to the core of the tetramer, the RMSD values for the loop region were different in E1 and E3 simulations.

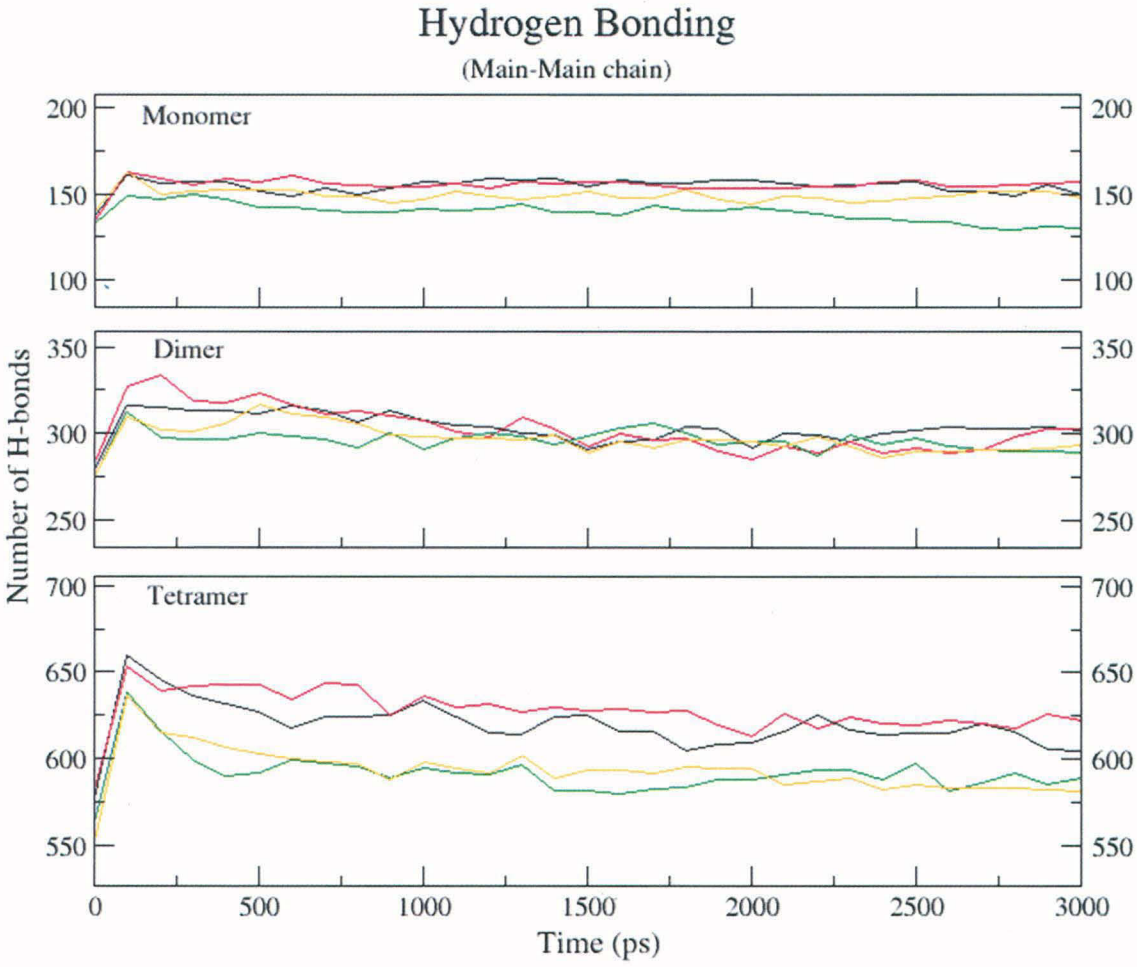


Figure 2.8(a)
The total number of hydrogen bonds involving main chain atoms in structures sampled at an interval of 100 ps during the 3 ns simulations of monomer, AB dimer and tetramer of ConA. The panels have used the same depiction scheme for different simulations as in previous figures.

In presence of bound metal ions, the loop regions of all four chains in the tetramer had RMSD between 1.0 to 1.5 Å during the 3ns simulation, while in absence of metal ions the average RMSD of the loop regions increased to as high as 3.5Å (Figure 2.10). Thus the loop region was found to have maximum structural deviations from its native conformation upon demetalization in all the four chains of ConA. We also analyzed the hydrogen bonds within the protein and plotted them according to their time of occupancy (Figure 2.11). The total number of hydrogen bonds within the protein and those with occupancy higher than 95% were found to be comparable in E1 and E3 simulation of the tetramer (Figure 2.11). This also reinforced the conclusion that the structure of the core of ConA remains unaffected even in the absence of the metal ions and demetalization only changes the conformation of the loop regions.

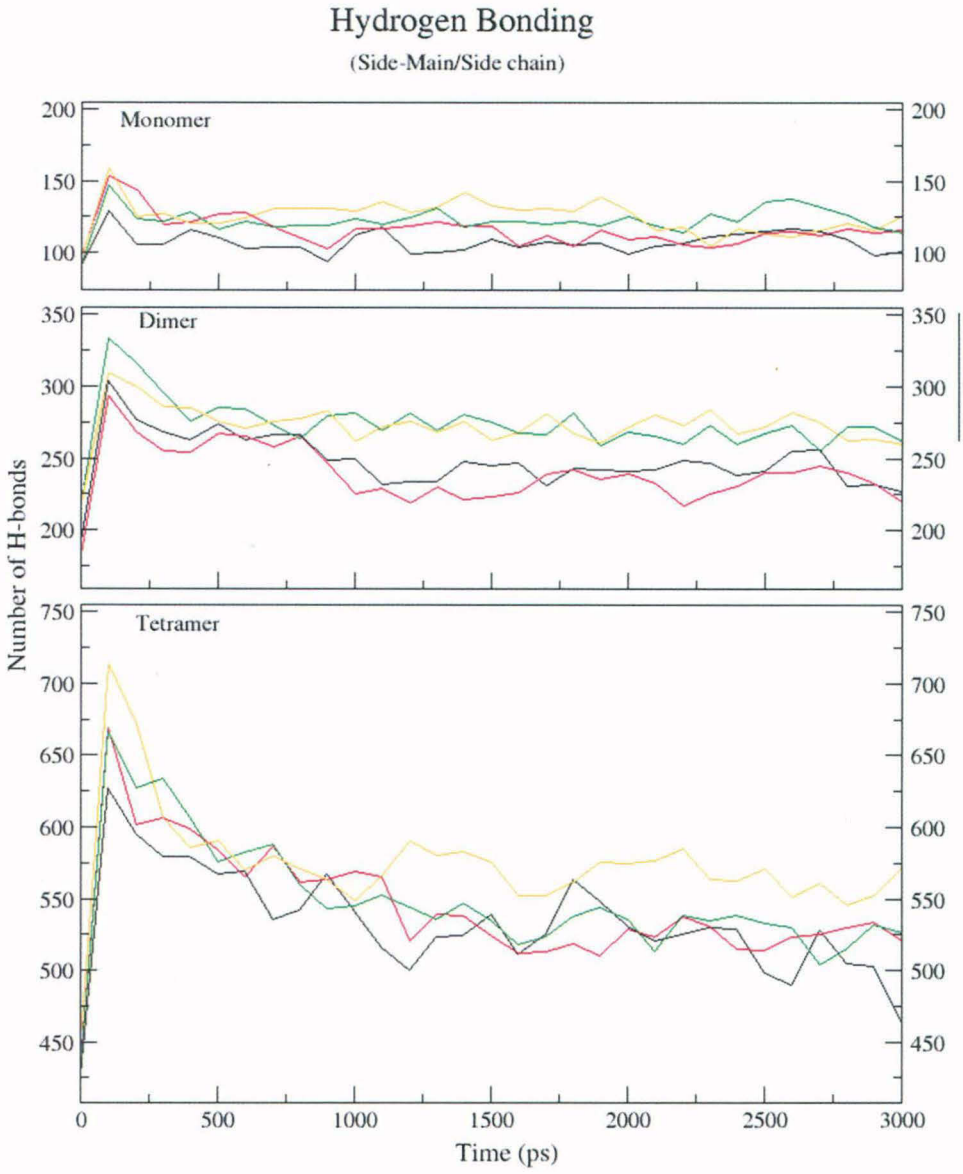


Figure 2.8(b)
The total number of hydrogen bonds involving side chains atoms in structures sampled at an interval of 100 ps during the 3 ns simulations of monomer, AB dimer and tetramer of ConA. The panels have used the same depiction scheme for different simulations as in previous figures.

2.3.1.4 ConA tetramer in absence of trimannoside and metal ions (E4)

In the E4 simulation, where neither the ions nor the sugar were included, the RMSD increased to about 1.8Å as early as 500 ps and remained stable at that value during rest of the simulation (Figure 2.2) whereas the R_g increased by only 0.4 Å (Figure 2.4). The tetrameric core of the protein remained essentially unperturbed as seen from analysis of non-polar ASA (Fig. 2.7) as well as M-M and S-M/S hydrogen bonds (Figure 2.8). The BF values were highest for 14-21 residue region (Figure 2.6) similar to the E3 simulation. The other regions hav-

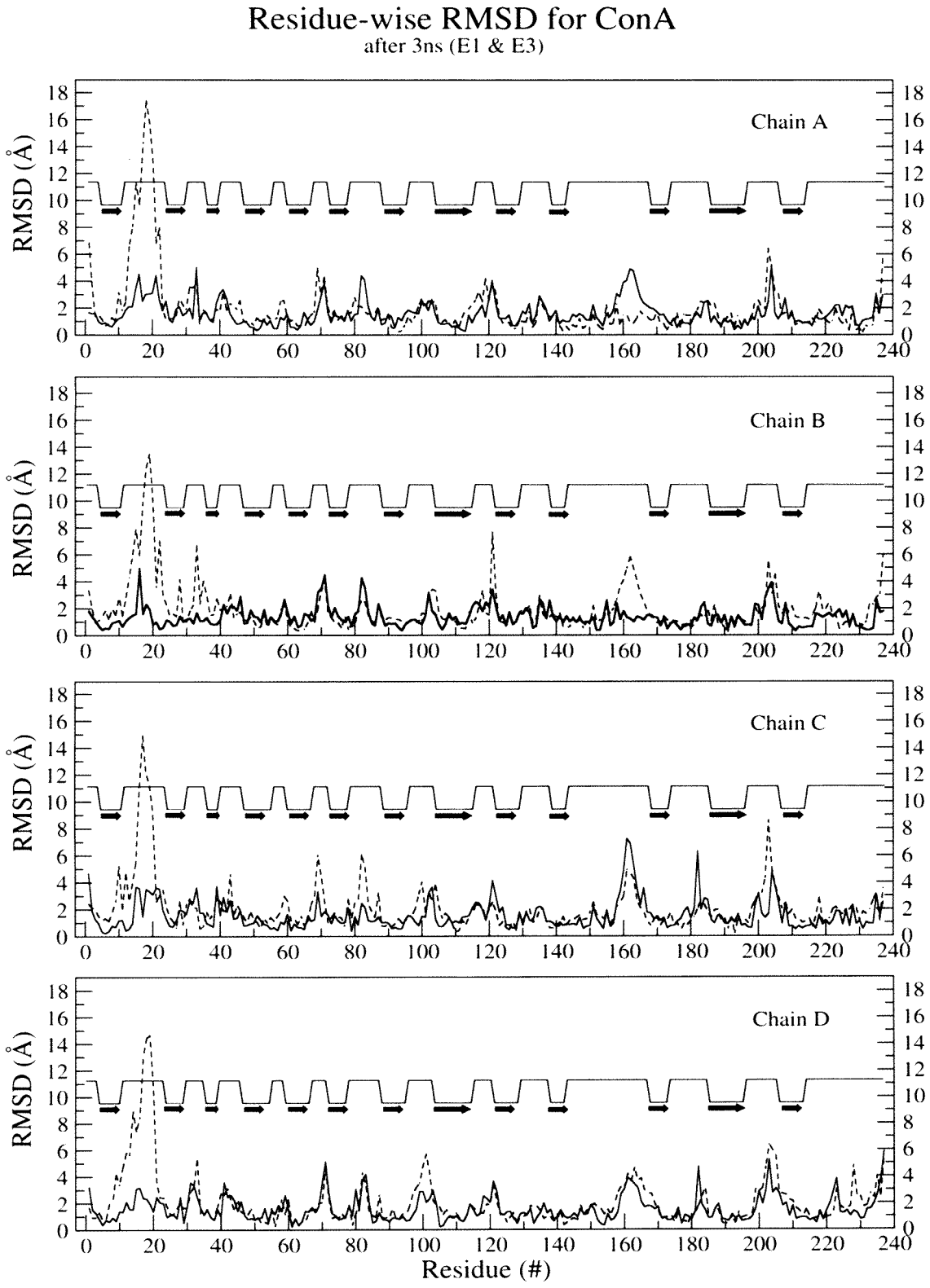


Figure 2.9
Residue-wise RMSD (rRMS) plot for the final structures obtained from E1 (solid line) & E3 (dashed line) simulations w.r.t the crystal structure. The respective chain name has been mentioned in the panels and the secondary structural states for various residues are annotated (the arrows indicate β sheets and lines indicate loop regions).

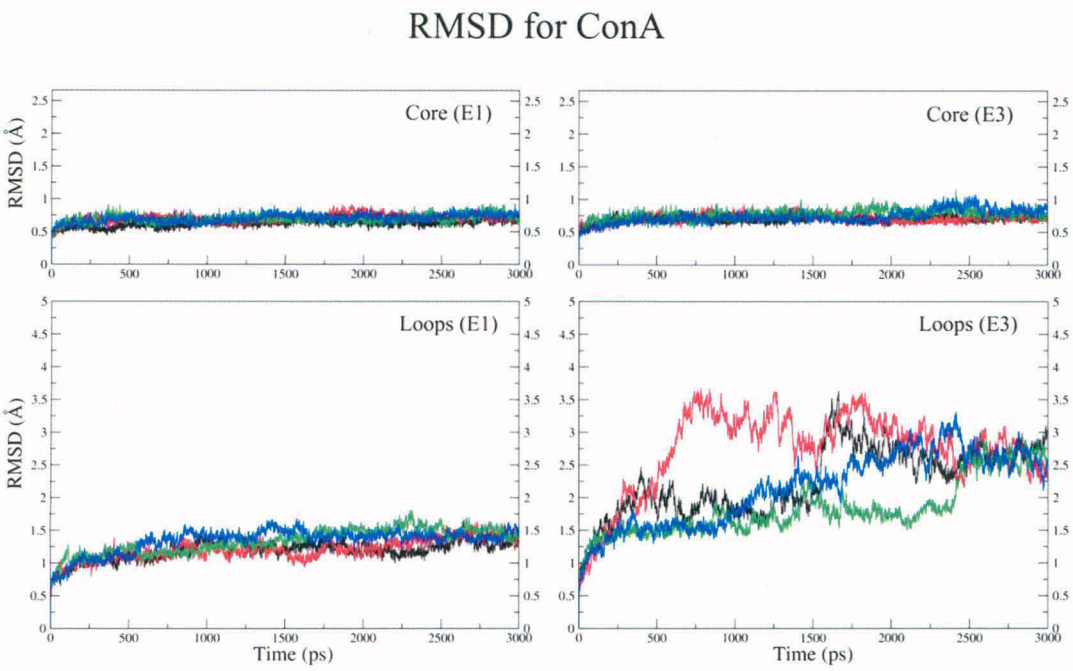


Figure 2.10
RMSD values for the core and the loop regions of the four chains of tetrameric ConA for the E1 and E3 trajectories.

ing high BF in all the chains is 160-164 and 201-205 with all chains showing BF values above 60Å^2. In 160-164 region chain D showed BF >160Å^2 while in 201-205 region chain A showed BF less than 40Å^2. Thus the behavior of the protein during E4 simulation was very similar to the E3 simulation.

2.3.2 Effects of metal ions on the binding of trimannoside

The trajectories obtained from simulations E1 and E3 were analyzed to investigate the effects of bound metal ions (Mn^{2+} and Ca^{2+}) on substrate recognition by ConA. Figure 2.12 shows the orientation of the trimannoside molecule with respect to the ConA during the simulation in presence (E1) and absence (E3) of metal ions. The coordinates of substrate bound ConA have been extracted at an interval of 100ps from both E1 & E3 simulations, and only the final protein structure is shown for the reasons of clarity. During the simulation in presence of metal ions the trimannoside remained bound to ConA throughout the 3ns trajectory, however, different degrees of the trimannoside mobility within the binding site were seen among different chains. The different degrees of mobility are also evident within the trimannoside molecule, as seen by the atomic positional fluctuation (APF) values of the trimannoside bound to chain A (henceforth, triA). The α(1-6)-mannose (MAN 240 in 1CVN)

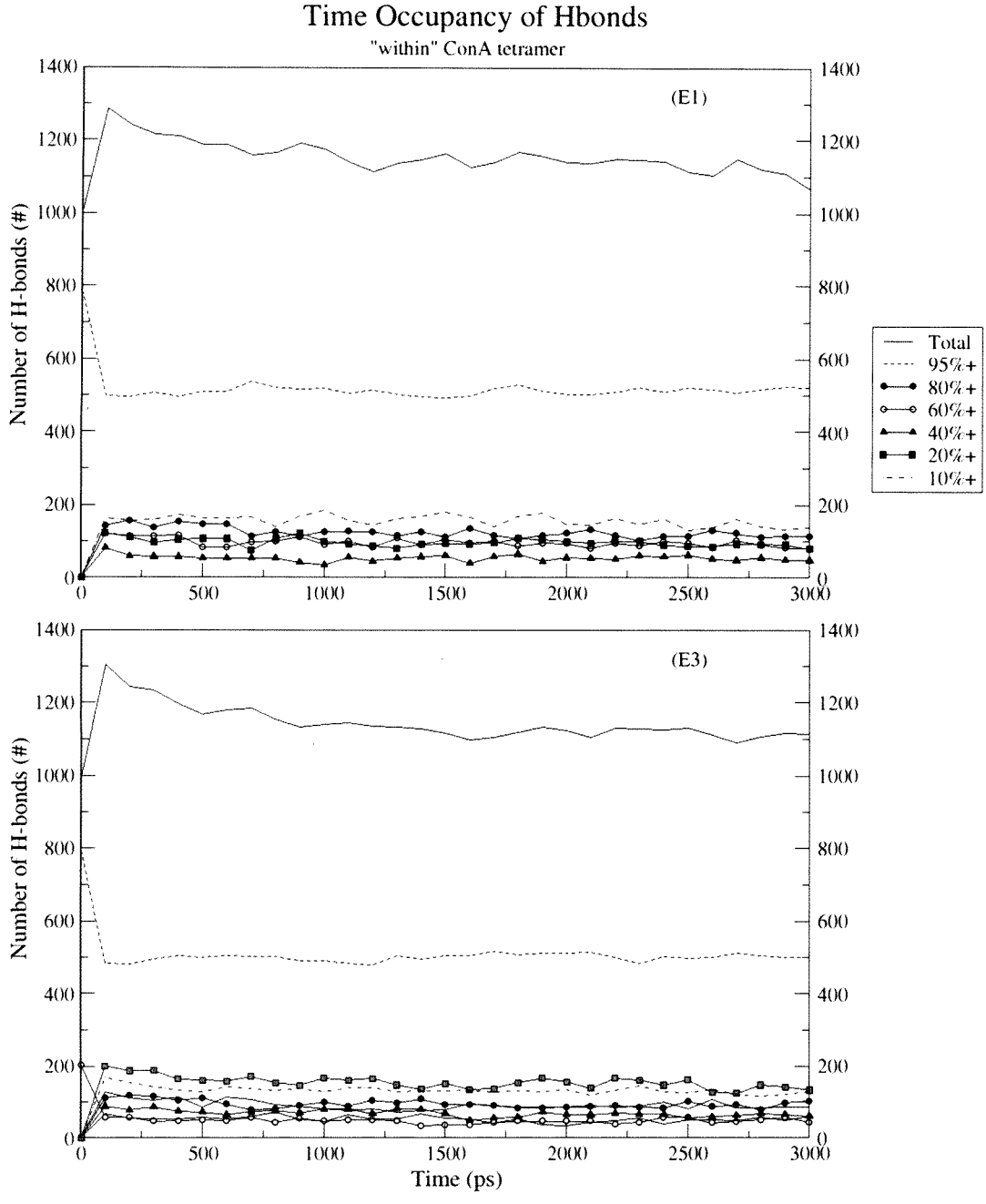


Figure 2.11
The percentage occupancy of various hydrogen bonds involving main chain and side chains in structures sampled at an interval of 100 ps during the 3 ns simulations of substrate bound ConA tetramer (top panel) in presence and (bottom panel) absence of the ions.

had APF of about 1.25Å whereas α(1-3)-mannose (MAN 242) had an APF value >2.5Å. The overall APF of triA was ~1.9Å. The trimannoside bound to chain B (henceforth, triB) in E1 was found to be the least mobile compared to others, with an overall APF value of only ~1.3Å. The trimannoside bound to chain C (henceforth, triC) and to chain D (henceforth,

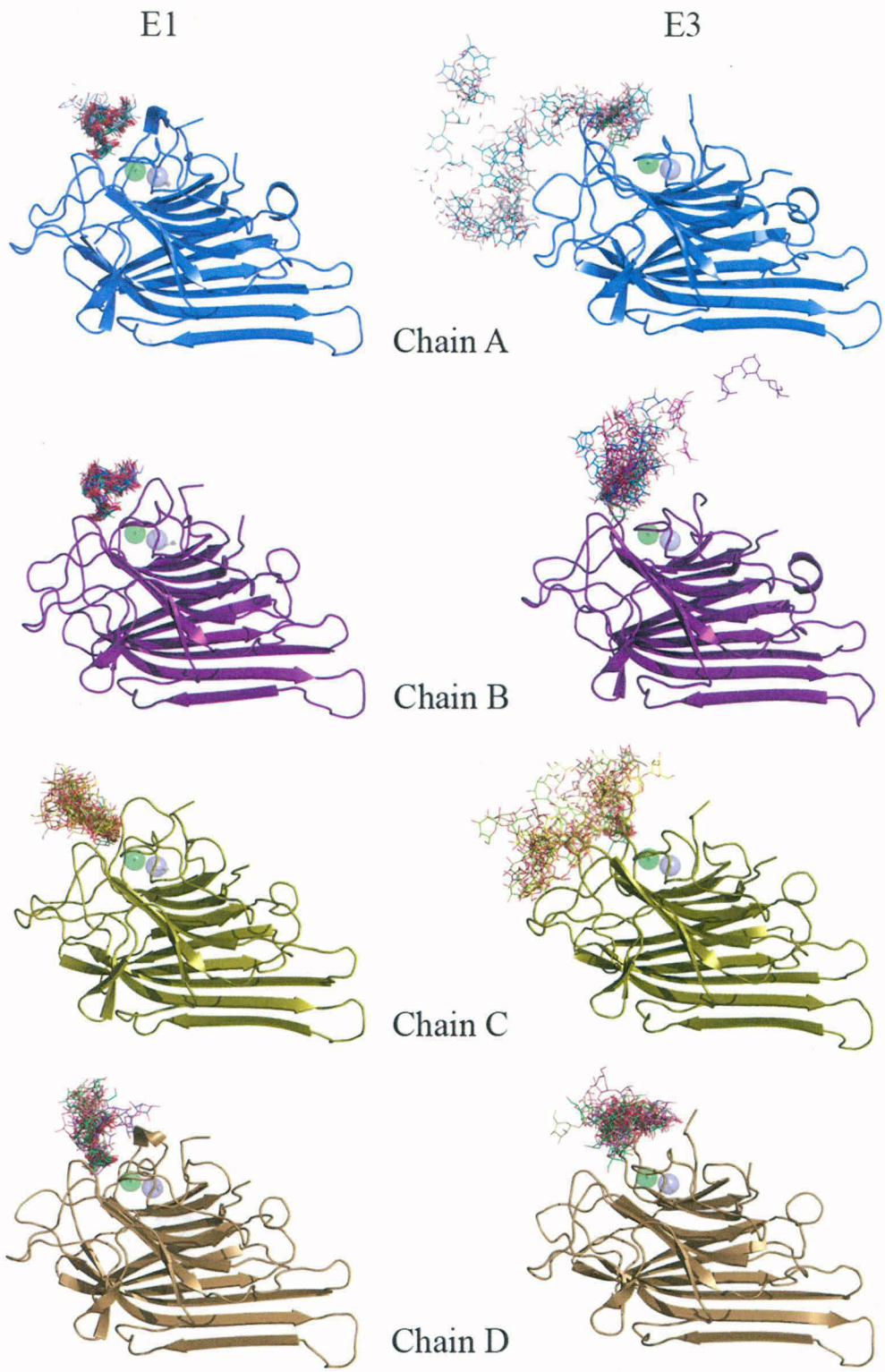


Figure 2.12:
The orientations of trimannoside in the substrate binding site during the simulations for the substrate bound ConA tetramer in presence (E1; left panel) and absence (E3; right panel) of the metal ions. The frames extracted from E1 and E3 trajectories at an interval of 100ps were superposed on the starting structure. It is clearly visible from the figure that the trimannoside remains bound to ConA to a large extent in presence of metal ions (E1), while it drifts away in their absence (E3).

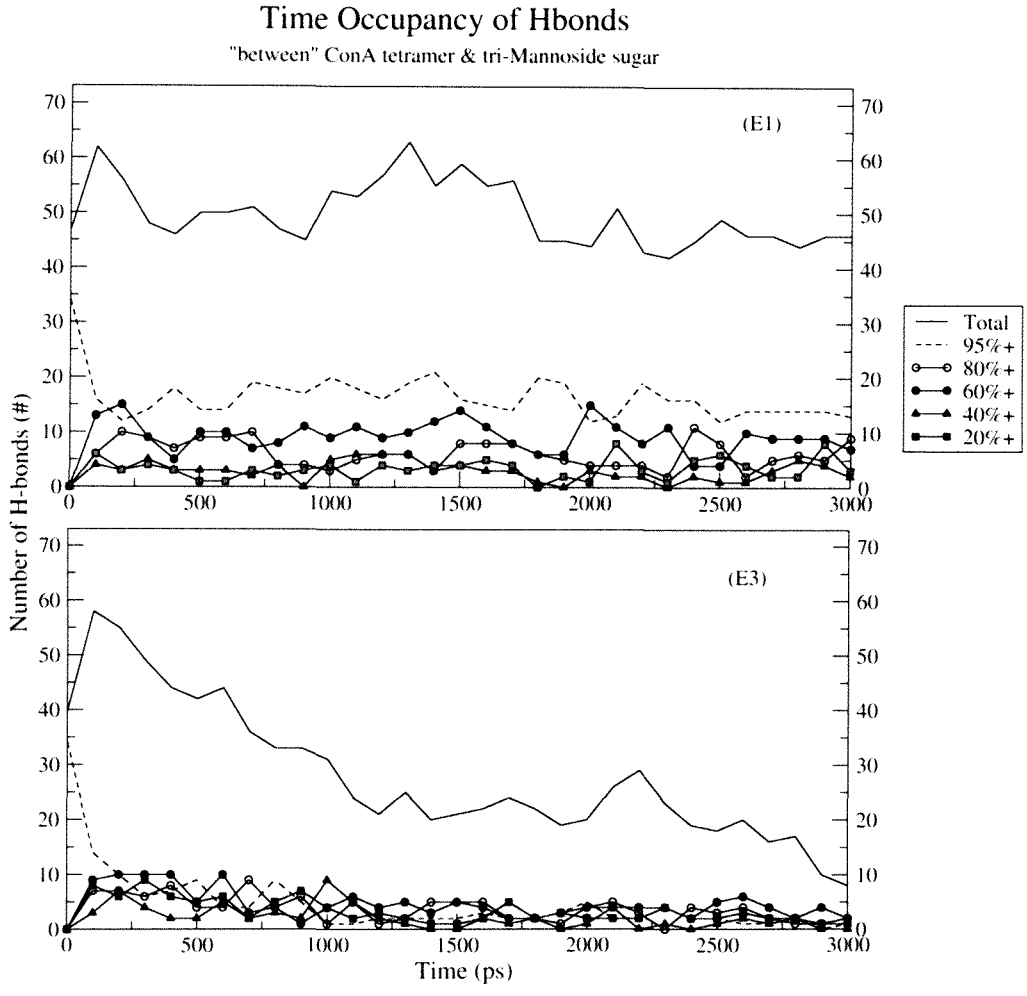


Figure 2.13
The percentage occupancy of various hydrogen bonds between trimannoside and ConA in structures sampled at an interval of 100 ps during the 3 ns simulations of substrate bound ConA tetramer (top panel) in presence and (bottom panel) absence of the ions.

triD) were found to be about two times more mobile compared to triA & triB, with their overall APF values being ~3.5Å and ~3.7Å, respectively.

During the E3 simulation the sugar molecule drifted away from its binding site (Figure 2.12). The dissociation of the trimannoside from its binding site occurred at different times for different chains. TriA was found to be the most mobile molecule after demetalization, so much so that it quickly moves out of its binding pocket. The APF value for it in the E3 simulation was ~13.6Å (compared to ~1.9Å during E1). TriB and triC were almost equally mobile with overall APF values of ~6.8Å and 9Å, respectively. So, as could be expected with such high mobility, the triA, triB and triC molecules move out into the solution by the end of the 3ns simulation. Strangely, triD had APF value of ~4.3Å and remained almost bound to chain D. Figure 2.13 shows the occupancy of hydrogen bonds between ConA and

Oligomeric status	Simulation name	Chain A	Chain B	Chain C	Chain D
Tetramer	E1	11	15	5	15
	E3	0	0	3	5

Table 2.3:
The total number of hydrogen bonds between each trimannoside and various monomers of ConA obtained from last 100 ps of the 3ns MD simulations in presence and absence of bound metal ions.

trimannoside during E1 and E3 simulations. The total number of hydrogen bonds between ConA and trimannoside remained almost constant in the simulation performed in the presence of the ions. However, in the absence of divalent ions the number of hydrogen bonds reduced dramatically. On an average, both simulations showed that number of probable hydrogen bonds between the lectin and the sugar were around 15 per monomer/chain at the start of the simulation which gradually stabilized at ~11 for the E1 simulation, but continuously fell from ~15 to ~2 for the E3 simulation due to the drift of the ligand from the binding site.
We have also calculated the residency time of all the possible hydrogen bonds between the protein and the trimannoside during the 3 ns simulations for E1 & E3. During E1 simulation for the tetramer there were a total of 130 different hydrogen bonds between ConA and trimannoside that were present at various points on the simulation time scale. However, 65% of those were transient hydrogen bonds present only for 10-25% of the simulation time. Only a set of the 29 hydrogen bonds persisted more than 50% of the simulation time, thus they can be called dynamically stable hydrogen bonds. Interestingly, the demetalization of the lectin led to an increase in the total number of different hydrogen bonds that existed during the 3ns simulation to 179, but only 5 of those hydrogen bonds were persistent for a period longer than 50% of the simulation time. Therefore, the number of dynamically stable hydrogen bonds between the trimannoside and ConA reduced dramatically upon demetalization.

Table 2.3 lists the total number of hydrogen bonds between ConA and the sugar during last 100 ps of the 3 ns simulation. Thus, the results of our simulations are in agreement with the experimental observation that metal ions are essential for substrate recognition by ConA. However, as discussed earlier, in absence of metal ions the overall structure of the ConA tetramer essentially remained intact and structural deviations are essentially confined to the loop regions. Therefore, we analyzed the conformations of the substrate and ion bind-

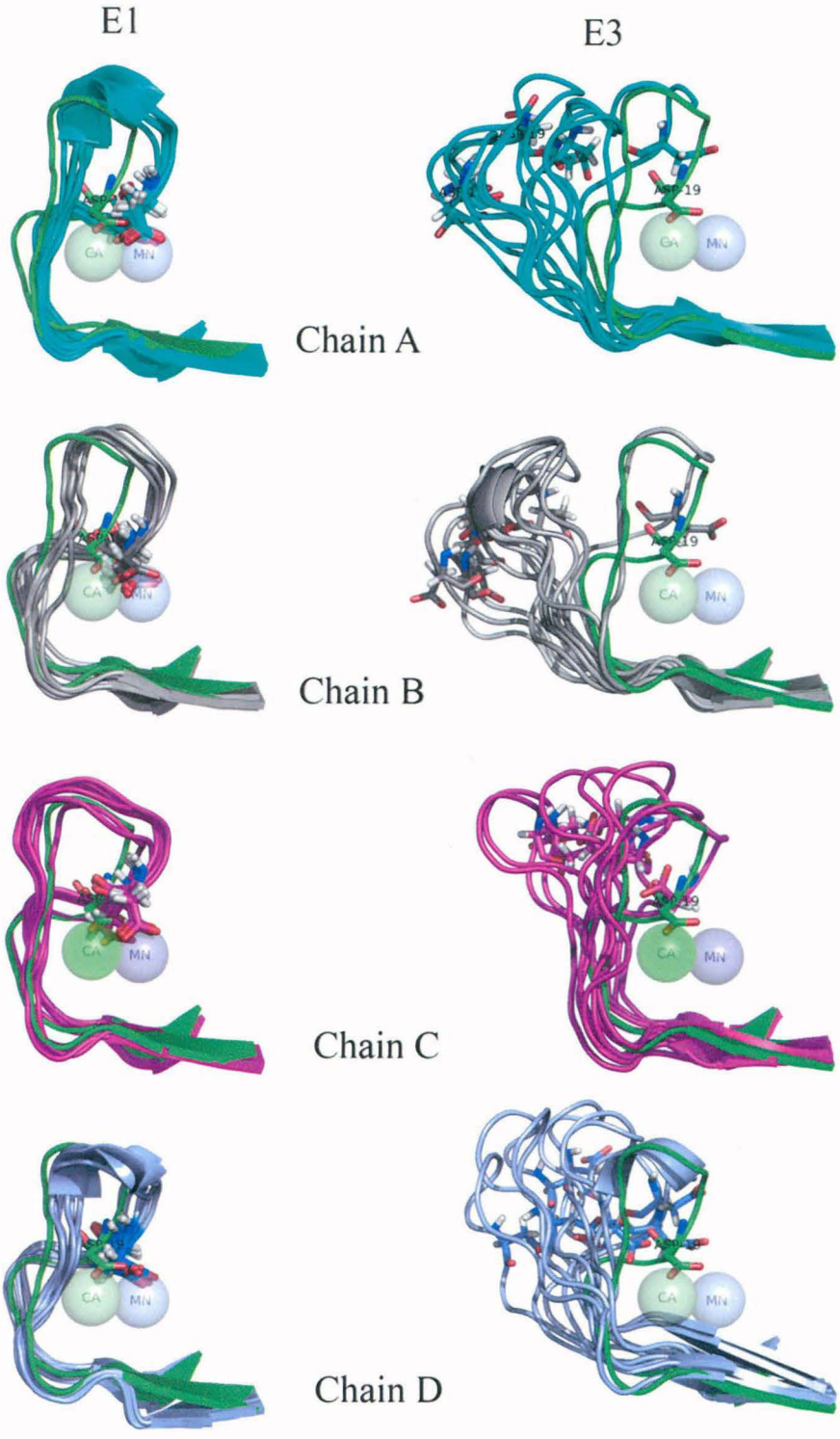


Figure 2.14
The cartoon diagram depicting the movement of the ion binding loop (7-25 region) for all four chains of the substrate bound ConA tetramer in presence (E1) and absence (E3) of the metal ions. The frames extracted from E1 (left) and E3 (right) trajectories at an interval of 500 ps were superposed on the crystal structure (green). The large structural rearrangements of this region are clearly visible. Asp19 has been shown in sticks to depict the disruption of the Mn^{2+} binding site in absence of metal ions. The transparent spheres represent the Ca^{2+} and Mn^{2+} ions.

ing loops in detail to understand the structural basis of the interplay between the metal ion binding and substrate recognition.

2.3.3 Effect of metal ions on conformation of the ion-binding loop

In order to understand the effect of bound metal ions on the conformation of the ion-binding loop, the frames were extracted from the E1 and E3 simulations at an interval of 500 ps. Figure 2.14 shows the structural deviations in the ion-binding loop upon optimal superposition of the protein core. The conformation of the ion binding loop of the four chains during the E1 simulation *i.e.* in the presence of the ions and the sugar has been depicted in the left panel of Figure 2.14, while the right panel depicts the conformation of this loop during the E3 simulation. The loop region was found to be highly ordered during the simulation in presence of the metal ions and hence this region from different frames aligned very well (Figure 2.14). The ion binding loop region obtained from the final structure of (chain A in) the E1 simulation also superposed very well with the corresponding region (of chain A) in the crystal structure (1CVN) (Figure 2.15a). This indicates that in presence of metal ions the ion-binding loop retains its native conformation throughout the 3 ns simulation. On the contrary, in the absence of the ions (E3 simulation) the ion binding loop region appeared highly disordered for all the four chains (Figure 2.14). The superposition of the final structure of the E3 simulation on the crystal structure of metal ion bound ConA (Figure 2.15b) indicate that, Asn19 (of chain A) has moved as far as ~15Å with respect to its native position in crystal structure, and Gly18 has moved even further to a distance of ~17 Å. In the native conformation the metal ions are tightly coordinated by a cluster of negatively charged residues. In absence of metal ions, strong electrostatic repulsion makes such a conformation for this loop region energetically unfavorable, leading to a large conformational change in the ion-binding loop. This large movement of the ion-binding loop obstructs the substrate binding site resulting in drifting of trimannoside from binding site. Thus, our molecular dynamics simulations provided an atomically detailed picture of the conformational changes associated with demetalization process and consequent destruction of the saccharide binding site on ConA. Table 2.2 summarizes the converged values for various structural parameters extracted from the last 500 ps of the 3ns simulations.

2.3.4 The monomeric and dimeric ConA simulations

Detailed analysis of the results from the simulations of the monomer and dimer of ConA indicate that, the monomer and the dimer don't behave dramatically different comp-

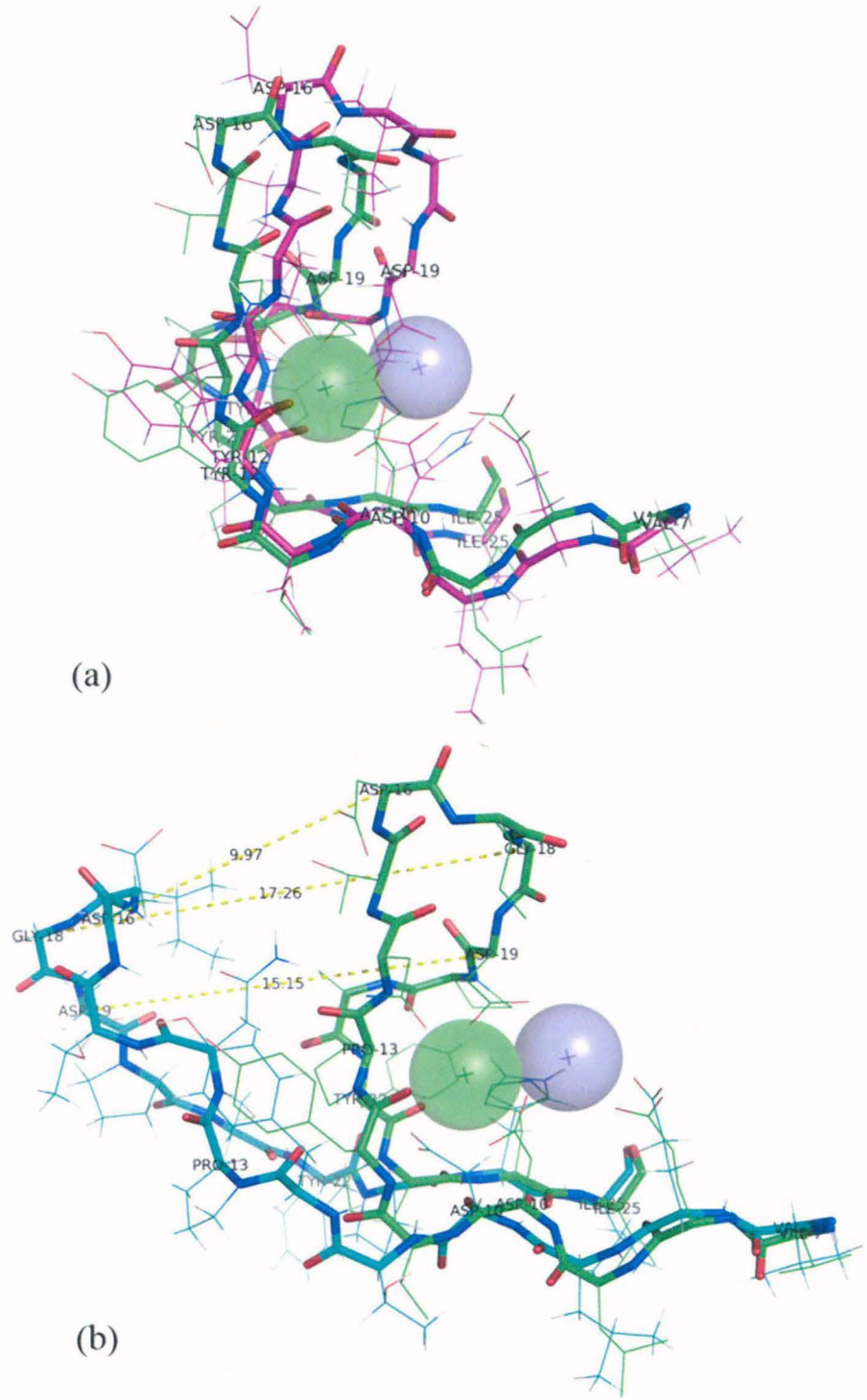


Figure 2.15
The conformation of the fragment containing ion binding loop (7-25) in the final structure obtained from the 3 ns simulations for substrate bound ConA tetramer in **(a)** presence (E1) and **(b)** absence of the ions (E3). The final structures (C-alphas shown in cyan) obtained from E1 and E3 have been superposed on the crystal structure (C-alpha shown in green). The dashed lines depict the distances between equivalent C-alpha atoms of the superposed structures. The light-green and purple colored transparent spheres are the Ca^{2+} and Mn^{2+} ions, respectively.

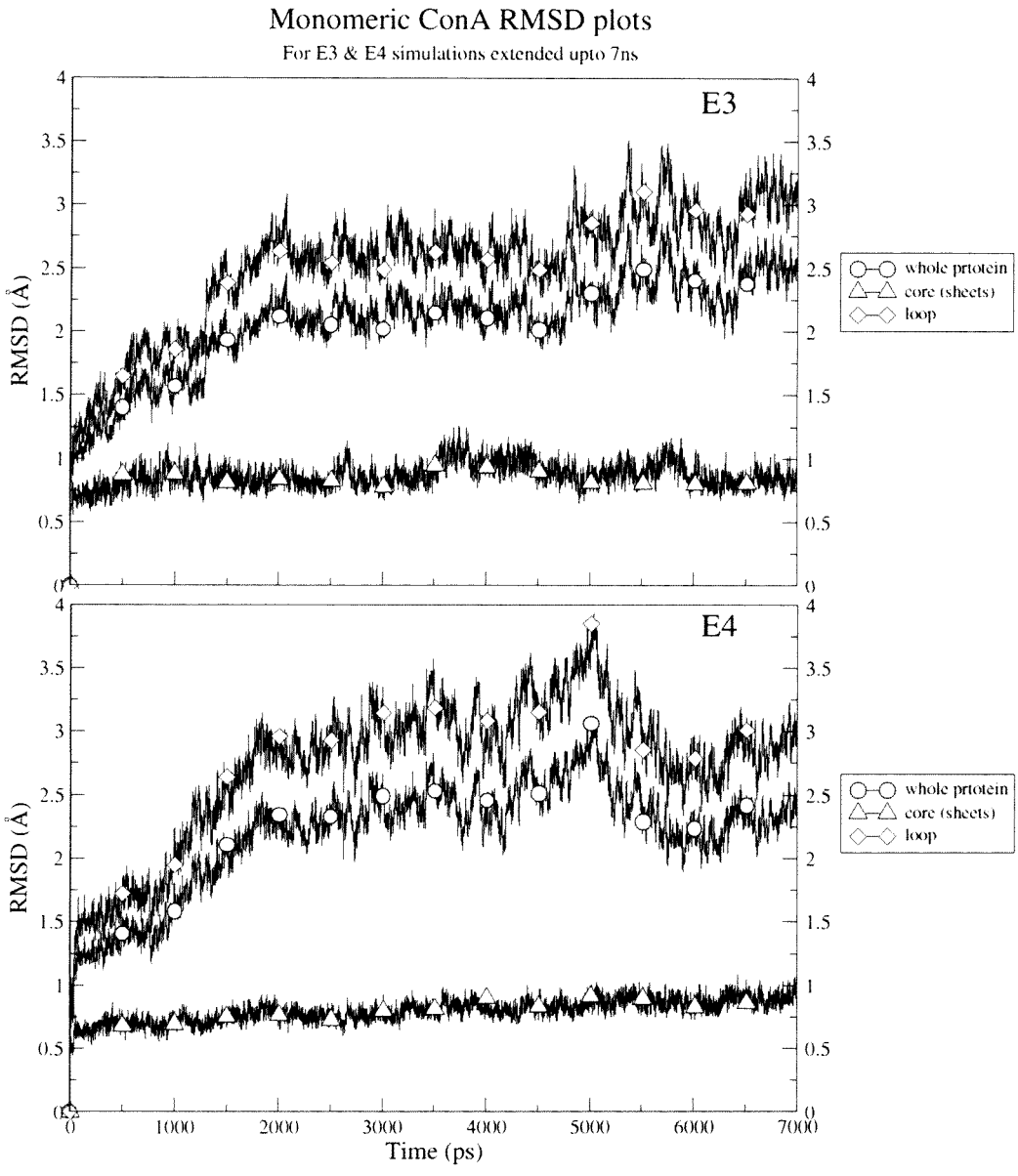


Figure 2.16
RMSD plot of the monomeric ConA for the 7ns simulations in absence of metal ions. The RMSD values for the loops, β-sheet core and whole protein have been plotted separately.

ared to the tetramer. The various structures sampled by the monomer and the dimer show a maximum RMSD of ~2.5 Å which is similar to the maximum RMSD of the tetrameric state of ConA. The only simulation where monomer differs from dimer and the tetramer simultaneously is E4 where in the absence of both ions and the trimannoside sugar, the monomer showed RMSD value of ~2.5Å, while the dimer and tetramer systems exhibited an RMSD value of about 1.5-2Å (Figure 2.2). It is possible that the structural destabilization of the monomer in absence of metal ions as reflected in the RMSD of 2.5Å (Figure 2.2) could have been a part of the unfolding event of ConA due to demetalization. Therefore, it was desirable

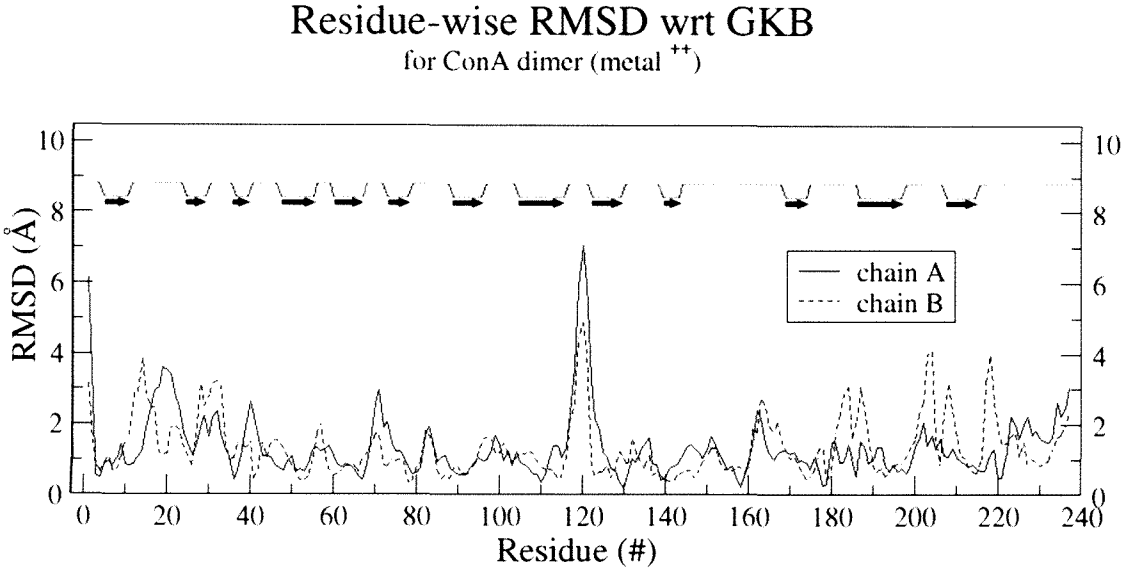


Figure 2.17
The residue-wise RMSD plot for ConA dimer (final structure from E2 simulation) with respect to 1GKB. The rRMSDs for the two chains of ConA dimer have been plotted separately. The secondary structural states for various residues are annotated (the arrows indicate β sheets and lines indicate loop regions).

to further extend the simulations and see if the core unfolds as well in longer simulations. Since the monomer and the tetramer showed similar conformational features during the 3ns simulations and simulations on the monomer required less computational resource, only the demetalized monomer simulations were extended up to 7ns. It was found that the monomer does not show any further structural destabilization. There was only a small increase in backbone RMSD for the monomer (Figure 2.16). The core and loop regions also didn't show any dramatic change in their RMSD values (Figure 2.16). However, the fluctuations in RMSD values are relatively larger compared to the simulations in the presence of the metal ions, thus indicating reduced structural stability. The RMSD data from 7ns of simulations of the demetalized ConA monomer also suggests that demetalization only affects the loop conformations, while the core of the protein remains intact, in agreement with crystal structure data which show that the core structure of the demetalized ConA is very similar to the metal ions bound form, but the loop regions have large conformational disorder. However, in the view of the orders of the magnitude difference in experimental time scale of the folding process and the time scale of our simulations our comparisons with experimental observations are only qualitative.

These results also indicated that, even though oligomerization results in increased stability of ConA tetramer, the monomer and dimer could also be stable in isolation. The stabil-

ity of the monomer and dimeric form of ConA can be rationalized by the observation that, change in non-polar ASA for the monomer and dimer is comparable to that of the tetramer (Figure 2.7). Thus, dissociation of the ConA tetramer does not lead to significant increase in non-polar ASA. The monomer and dimer simulations also showed higher mobility in the loop region (Figure 2.6b). However, the dimer showed exceptionally raised BF for the ion binding loop in absence of the metal ions. The BF for the ion binding loop of the dimer was found to be >1000 $Å^2$ which is about double the value for the BF value (~600-700 $Å^2$) of this region in the tetramer, and more than double of the BF value for the monomer (~400$Å^2$) in E3 simulation. The dimer structure used in our simulations comprised of the A & B chain of ConA tetramer. However, the crystal structure of an uncomplexed dimeric ConA (1GKB) with both metal ions bound was also available in PDB (Kantardjieff *et al.*, 2002).

Hence, the results from AB-dimer simulations were compared with 1GKB. It was found that in terms of RMSD values the structure sampled during the simulation show similar deviations from 1CVN as well as 1GKB. This suggests that the choice of AB-dimer from 1CVN is almost as good as simulating 1GKB. Figure 2.17 shows the residue-wise RMSD between 1GKB and the final structure obtained from the E2 simulation of the ConA dimer. Most of the regions in both chains show an RMSD of about ≤1.5Å, except for some regions notably 116-121 region, which is a loop region at the inter-dimer interface.

2.4 DISCUSSION

The MD simulations reported in this work have given us further insights into the process of demetalization and its consequences on substrate binding ability of a legume lectin. To the best of our knowledge, these are the longest explicit water MD simulation on this protein. Simulations have also been carried out for the monomeric and dimeric form of ConA under similar environmental conditions. It is encouraging to note that for the substrate and metal ion bound native ConA tetramer the computed BF values obtained from our MD simulations are in close agreement with the experimentally determined BF values. Comparative analysis of the total of twelve 3ns MD trajectories has given novel insights into the dynamic aspects of the structural stability and substrate recognition of ConA. The results from our simulations indicate that the core of the ConA tetramer has a tightly packed structure and it remained intact even upon demetalization. The effect of the demetalization however, appeared to be concentrated in the loop regions. This observation was based on the analysis of RMSD, rRMS, secondary structure content and non-polar ASA for the structures sampled

during various simulations. However, metal ions play a major role in stabilizing the conformation of the key loop regions of this protein which are important for binding of metal ions as well as the trimannoside substrate. The native conformation of this loop is stabilized by the coordination of the two metal ions with a cluster of negatively charged amino acids. However, in the absence of metal ions, probably due to strong electrostatic repulsion, this conformation of the ion-binding loop becomes energetically unfavorable. Therefore, upon demetalization this loop is characterized by high mobility and in fact some of the amino acids in this loop moved as far as 17 Å with respect to its position in the native crystal structure. This large movement of the ion binding loop obstructs the binding of the substrate resulting in drifting of the trimannoside from ConA during simulations in the absence of metal ions. The results of these simulations thus, support the long known experimental fact that ion binding is essential for the sugar binding (Kalb and Levitzki, 1968; Yariv *et al.*, 1968).

Our studies have provided additional insights into the dynamics of the interplay between metal and sugar binding to ConA. The metal ions interact with the side chains on the ion-binding loop and these interactions keep the loop in a fixed conformation. This in turn provides a scaffold necessary for the sugar molecule to bind to lectin both by decreasing the thermal motion of this region and thus, possibly, by properly orienting the side chains for a better and optimal presentation to the sugar atoms. In contrast to the role of metal ions, substrate binding does not seem to affect the structure of the loop regions appreciably. Similarly, the simulations carried for monomer and the dimer suggest that monomeric and dimeric form of ConA also have stable structures and substrate binding ability. The comparison of dimeric simulations with 1GKB showed that the choice of 1CVN, a tetrameric crystal, for obtaining coordinates for the dimeric form, was as good as using a dimeric crystal structure 1GKB.

2.5 CONCLUSIONS

We have carried out explicit solvent molecular dynamics simulations for ConA under different environmental conditions to understand the effect of bound metal ions on its structure and substrate binding ability. These simulations have revealed dynamic conformational changes associated with demetalization of ConA. We find that demetalization leads to large conformational changes in the ion binding loop, while the β-sheet core of the protein remains relatively unperturbed. The increased flexibility and structural destabilization of the ion binding loop in absence of the bound metal ions results in drifting of the substrate sugars. These results are in agreement with experimental observations about the abolition of the sugar bind-

ing ability upon demetalization. We also find that the amino acid stretches of ConA having high B-factor values in the crystal structure show relatively higher mobility in the simulations. These results on dynamic aspects of the metal ion binding to legume lectins and their mode of substrate recognition add to our appreciation of the design of their combining sites (Sharma and Surolia, 1997).

Chapter 3

Role of glycosylation in structure and stability of *Erythrina corallodendron* lectin (EcorL): A molecular dynamics study

3.1 INTRODUCTION

Proteins are known for the diversity of functions they display within and around the cells. They are the drivers, effectors as well as modulators of processes that drive the cellular machinery. The versatility of proteins stems from their ability to generate additional diversity by co- and post-translational modifications. Of these, glycosylation is by far the most complex and profuse form of modification. It confers structural complexity, variety as well as differential physicochemical properties to proteins. The sugar moiety decorating the protein may range from a small monosaccharide to large branched polysaccharides. Broadly speaking, glycosylation can be classified as *N*-linked (nitrogen of asparagine or arginine) or *O*-linked (hydroxyl oxygen of serine, threonine, tyrosine, hydroxylysine or hydroxyproline). Of these, the asparagine linked N-glycosylation is the most recurrent form observed in eukaryotes (Nalivaeva and Turner, 2001; Wyss and Wagner, 1996). It was discovered recently that glycosylation occurs even at prokaryotic level and shares several similarities with the eukaryotic glycosylation (Upreti *et al.*, 2003; Weerapana and Imperiali, 2006).

Glycosylated proteins are involved in various phenomena like host-pathogen interactions, symbiotic associations, cell adhesion, cell-cell recognition, embryonic development and differentiation (Mitra *et al.*, 2006; Varki, 1993). The specificity of signaling and receptor binding by these proteins can be attributed to their glycosylation pattern. Additionally, glycosylation can also potentially alleviate several physical instabilities of proteins that are a result of exposure to elements such as extreme temperatures, pH, amphipathic interfaces or hydrophobic surfaces and chemical denaturants (Sola and Griebenow, 2009). Based on studies over several years on a number of glycoproteins, it has been proposed that glycosylation leads to a decrease in the overall structural dynamics of the protein and sometimes the effects can even be observed on regions far from the site of glycosylation (Gervais *et al.*, 1997; Joao *et al.*, 1992; Mer *et al.*, 1996b; Wormald and Dwek, 1999).

Recently, studies on antibodies have been carried out to understand the functional consequences of variation in glycan moieties attached to the protein (Burton and Dwek, 2006; Kaneko *et al.*, 2006). The work by Kaneko *et al* suggests that antibody function depends greatly on the specific sequence of oligosaccharides attached to the Fc chain (Kaneko *et al.*, 2006). It has been observed that one class of Fc-FcγR interactions generate pro-inflammatory effects while another class has anti-inflammatory effects. The distinction arises due to differential sialylation of the Fc core glycan, which provides a switch from innate anti-

inflammatory activity to a pro-inflammatory one upon antigenic challenge. However, in certain cases the glycans may have no function at all and might be completely replaceable (Wang *et al.*, 1996; Wyss and Wagner, 1996) . There also have been contradictory reports e.g. in one study the carbohydrate moiety was found to enhance the folding and stability of the covalently attached protein (Chu *et al.*, 1978) whereas in another study on the same protein no increase in thermostability could be seen (Schulke and Schmid, 1988). The secondary or tertiary structures of the glycoproteins e.g. RNase (Joao *et al.*, 1992; Williams *et al.*, 1987) and *pars intercerebralis* major peptide-C (PMP-C) (Mer *et al.*, 1996b) were found unaffected by glycosylation though an enhanced thermostability was observed for both. The deglycosylation of several natural glycoproteins did not lead to substantial changes in their conformations though the glycoproteins experienced decreased thermostability (Wang *et al.*, 1996). In the case of soybean agglutinin, it was observed that glycosylation not only affects the overall stability of the protein but also brings about a change in its unfolding kinetics (Sinha and Surolia, 2007). In case of asialoglycoprotein receptor, deglycosylation was found to have no significant effect on cell-surface targeting, function or turnover (Breitfeld *et al.*, 1984). The vesicular stomatitis virus producing non-glycosylated viral glycoprotein under the presence or effect of tunicamycin was found to have comparable specific infectivity (Gibson *et al.*, 1978). Deglycosylation of hCG had no effect on the structure and binding to its receptor, but its function was found to be compromised (Kalyan and Bahl, 1983). For horseradish peroxidase isoenzyme C (HRP) which has 8 N-linked glycans, it was found that glycosylation enhanced the kinetic stability significantly but not the thermodynamic stability (Tams and Welinder, 1998). The importance of glycosylation is further highlighted by the existence of a quality control system, the calreticulin-calnexin cycle, which prevents the release of misfolded proteins. Glycans are crucial for correct folding of proteins, though a few proteins do fold in their absence. It prompts local conformational changes in the protein structure at the site of glycosylation which in turn influences not just the stability, but also the oligomerization status of proteins in many cases (Mitra *et al.*, 2006). These studies suggest that, the effect of glycosylation on structure, stability and function of the protein depends on several factors like sequence of the protein, the type of glycan motif attached and the site of attachment etc. Therefore, it would be interesting to understand the effect of glycosylation on structure, stability and function at molecular level.

One such glycoprotein whose crystal structure is available both in glycosylated and non-glycosylated form is a lectin from *Erythrina corallodendron* (EcorL). Figure 3.1a shows

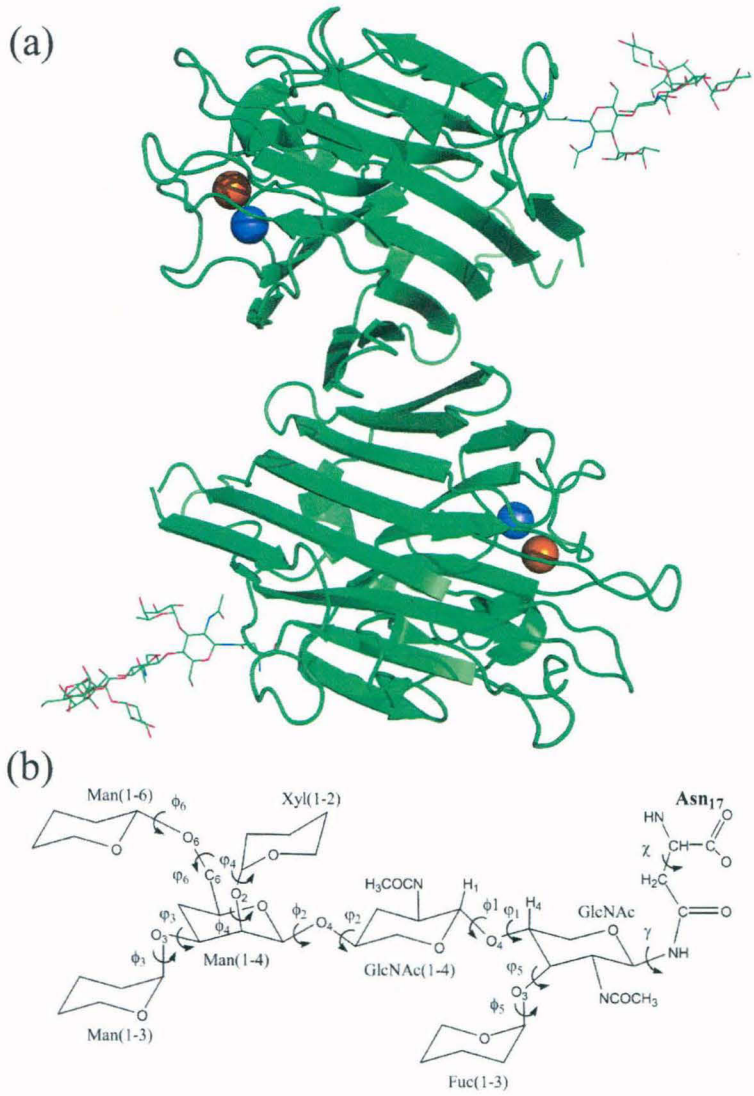


Figure 3.1: EcorL structure

(a) The cartoon depiction of the crystal structure of EcorL (PDB entry 1AX0). The orange and purple spheres represent the Ca^{2+} and Mn^{2+} metal ions, respectively. The covalently attached heptasaccharide is depicted using lines.

(b) The chemical structure of the heptasaccharide depicting various dihedral angles (both φ and ψ) along with the glycosidic dihedral angle (γ) and the chi (χ) angle of the Asn-17 side chain. The monosaccharides have been labeled along with all the glycosidic dihedral angles.

the crystal structure of the dimeric form of EcorL (Shaanan *et al.*, 1991). The glycan moiety was found to be a heptasaccharide (Figure 3.1b) Manα3(Manα6)(Xylβ2)-Manβ4-GlcNAcβ4-(Fucα3)GlcNAc (Ashford *et al.*, 1987) predominantly, with Asn-17 being the site of glycosylation (Adar *et al.*, 1989). Even though another glycosylation site has been proposed at Asn-113 (Young *et al.*, 1995), none of the EcorL crystal structures have shown the presence of an oligosaccharide on this location. One of the surprising aspects of the structure of EcorL was that in spite of having a typical legume-lectin fold (Chandra *et al.*, 2001;

Shaanan *et al.*, 1991), it showed a completely novel mode of quaternary association not seen before in other legume lectins. Initially, this new mode of quaternary association was attributed to the presence of the oligosaccharide (Shaanan *et al.*, 1991). However, subsequent elucidation of the crystal structure of the recombinant form of EcorL (rEcorL) (Arango *et al.*, 1992; Kulkarni *et al.*, 2004) which lacked glycosylation revealed that it had identical quaternary association as the native glycosylated form. This indicated that glycosylation *per se* could not have been the cause for a new mode of quaternary association. Instead, it was found that non-glycosylated EcorL had decreased thermal stability and a drastically changed refolding pathway compared to the native glycosylated form of EcorL (Mitra *et al.*, 2003) indicating the role of glycosylation in the folding and thermostability.

During the last decade, major progress has been made in development of accurate forcefield parameters for the simulations of sugar molecules (Guvench *et al.*, 2008; Woods *et al.*, 1995) and easy availability of high performance computing resources have permitted long nano second scale simulations for large oligomeric glycoproteins. Recently, we were able to successfully simulate the effects of demetalation on saccharide binding and the conformational changes in the ion-binding loop in Concanavalin A (Kaushik *et al.*, 2009). Similar studies involving MD simulations on ConA, jacalin, galectin and other plant lectins have also provided novel insights into their structure and substrate specificity (Gupta *et al.*, 2009; Hansia *et al.*, 2007; Meynier *et al.*, 2009; Pratap *et al.*, 2001; Sharma *et al.*, 2009). This encouraged us to explore, in the current study, the effects of glycosylation on the structure and stability of EcorL by comparing the glycosylated and non-glycosylated forms of EcorL using molecular dynamics (MD) simulations. Molecular dynamics simulations have been used earlier to study the structure of the EcorL oligosaccharide successfully (Naidoo *et al.*, 1997). However, due to lack of adequate computational resources the protein was kept rigid and only the dynamics of the oligosaccharide was studied. MD simulations have also been used to study the retardation of unfolding of bovine pancreatic RNaseA due to glycosylation (Choi *et al.*, 2008). In this work, we report a series of long MD simulations in explicit solvent environment carried on glycosylated as well as non-glycosylated dimers of EcorL for understanding effect of glycosylation on the structure and stability of this important glycoprotein.

3.2 METHODS

The coordinates for the dimeric non-glycosylated form of EcorL were obtained from C and D chains of the PDB entry 1SFY. The coordinates for the dimeric glycosylated EcorL

were obtained by superposing the monomer coordinates from the PDB entry 1AX0 on the non-glycosylated EcorL dimer. The glycosylation has been retained at Asn-17 only, as none of crystal structures of EcorL have coordinates for the glycan at Asn-113. The recombinant and glycosylated forms of EcorL have almost identical structure and all atoms RMSD between monomeric chains of 1SFY and 1AX0 was only 0.8Å. Therefore, in order to generate a glycosylated structure with protein coordinates exactly identical to the non-glycosylated counterpart, a glycosylated dimer was generated by using protein coordinates from 1SFY and oligosaccharide coordinates from 1AX0. Metal ions were included for both the dimers for the simulations but ligand coordinates were removed as it has been known that the ligand binding induces conformational changes in the protein though they may be minor. Thus, we obtained the metal-bound and ligand-free coordinates of both glycosylated and non-glycosylated EcorL dimers. MD simulations were carried out using AMBER9 package (Case *et al.*, 2006). The ff03 (Duan *et al.*, 2003) and glycam06 (Woods *et al.*, 1995) force-fields were used for the polypeptide and heptasaccharide segments respectively of the glycoprotein. The oligomeric EcorL was solvated inside a TIP3P (Jorgensen *et al.*, 1983) water box extending 10Å from the protein surface along all three axes. Energy minimization was carried out for the solvated protein to remove steric clashes if any between the protein and the solvent molecules. To remove bad contacts in an experimental structure of a protein, an energy minimization of the protein alone might be warranted before solvating it. However, significant proportion of EcorL structure is composed of loops and minimization in vacuum might have altered the conformation of these loops. In addition, water molecules have been known to participate in structuring certain residues and the EcorL structures do not have steric clashes and therefore, we have minimized the solvated protein. The temperature of the system was raised to 300K over a 20 picosecond (ps) constant volume (NVT) simulation and then equilibrated at 300K for 200ps by carrying out constant pressure (NPT) simulations at 1 atmospheric pressure. Subsequently, the production dynamics was run for a period of 5 nanoseconds (ns) under NPT conditions. A time step of 1 femtosecond (fs) and a non-bonded cutoff radius of 8Å were used in all the simulations. The periodic boundary conditions were used to mimic the infinite solvent system and the electrostatic potential was calculated using particle-mesh Ewald summation method (Darden *et al.*, 1993). The coordinates were stored every 1ps during the dynamics so that the 5ns trajectory consisted of a set of 5000 structures. The non-glycosylated form of EcorL was also simulated for 5ns under identical conditions at 300K and 1 atm pressure. Each of the simulations for glycosylated and non-glycosylated

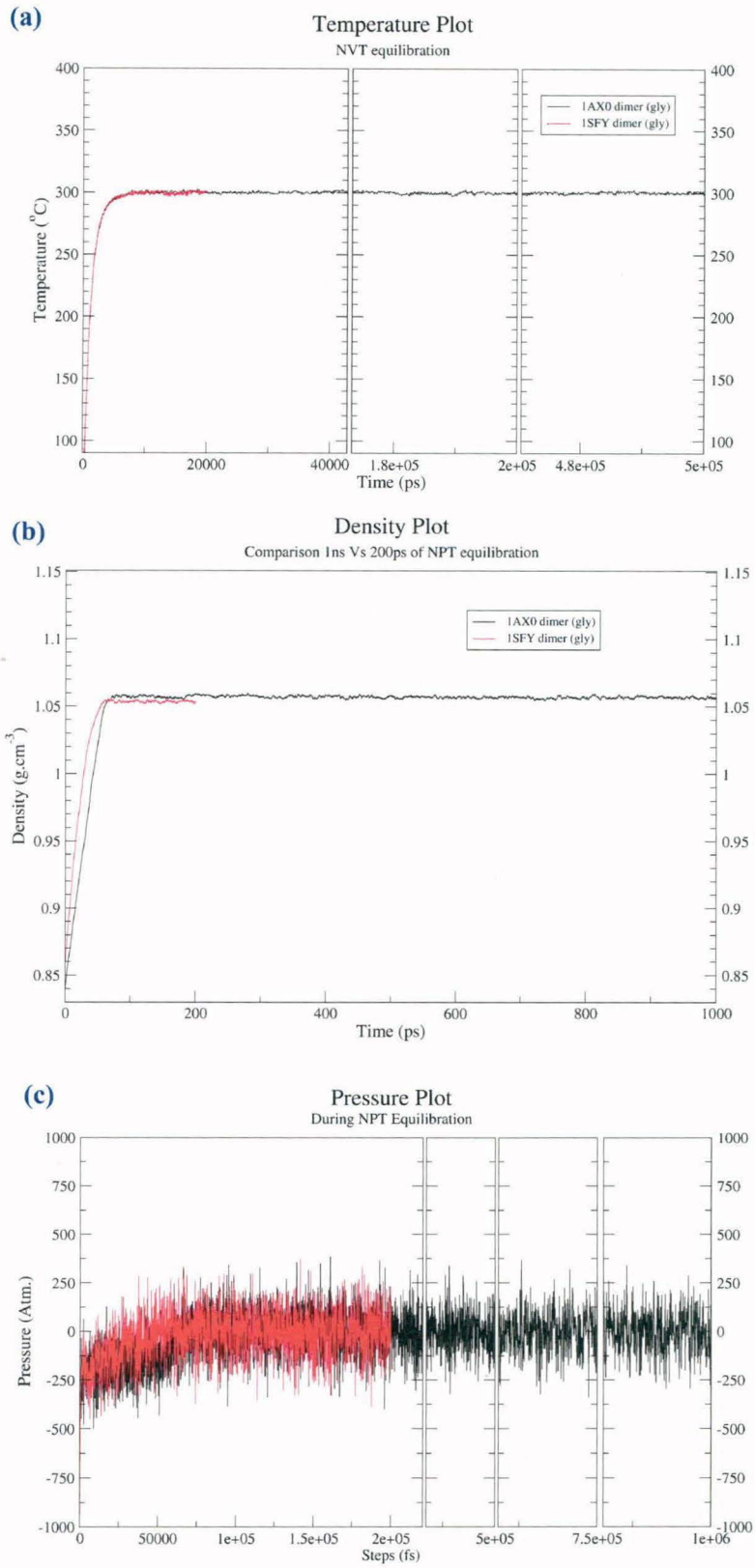


Figure 3.2: Effect of equilibration length on temperature, density and pressure
Comparison of (a) temperature, (b) density and (c) pressure plots of the system after shorter and longer periods of equilibration. Temperature plots after 20ps (red) and 500ps (black) of NVT equilibration have been compared. Density and pressure plots for 200ps (red) have been compared with 1ns long (black) NPT equilibration.

EcorL were repeated four times using different random number seeds for assigning different starting velocities. These simulations have been called S1, S2, S3 and S4, respectively. MD simulations for both glycosylated and non-glycosylated EcorL were also carried out at elevated temperatures of 400K, 500K and 600K. For these high temperature simulations at 300K and 400K, the temperature of the system was gradually raised to the desired value and the system was equilibrated for 200ps at constant pressure. During the constant pressure equilibration at higher temperatures, it was seen that the density of the system dropped from ~1.05 g/cm^3 at 300K to ~0.96 and 0.83 g/cm^3 at 400K and 500K, respectively. At 600K the density of the system kept dropping to lower values. Hence, the production dynamics was initiated after a 400ps of equilibration at 600K. At 400K, 500K and 600K, only a single MD run was carried out for glycosylated and non-glycosylated EcorL. Thus, the high temperature simulations gave a total of six trajectories for glycosylated and non-glycosylated EcorL.

As mentioned earlier, the glycosylated EcorL simulations were carried out on a hybrid structure using protein coordinates from non-glycosylated form and sugar coordinates from the glycosylated EcorL crystal structure. In view of the very high structural similarity (backbone atoms RMSD < 0.4 Å) between the glycosylated (1AX0) and non-glycosylated (1SFY) EcorL structures, simulations of 1AX0 dimer or 1SFY-oligosaccharide hybrid dimer would give very similar results for glycosylated EcorL. However, for the purpose of comparison, we also carried out a simulation on the glycosylated EcorL crystal structure i.e. 1AX0. The 1AX0 dimer was simulated for 10ns in a single run. To examine whether increasing the equilibration for longer periods would affect the system properties *viz.* temperature, pressure or density any further, we ran this simulation with thermalization period increased to 500ps and NPT equilibration period extended to 1ns. The temperature, pressure and density values from this simulation were very similar to the corresponding values obtained from simulations with 20ps of thermalization and 200ps of equilibrations (Figure 3.2).

The trajectories were analyzed using the *ptraj* module of AMBER9 (Case *et al.*, 2006) package. Various structural and dynamic parameters were calculated using ptraj from the trajectories like root mean square deviation (RMSD) with respect to the starting structure, radius of gyration (Rg), B-factors, and hydrogen bonds. RMSD and B-factors were calculated for the backbone atoms only whereas Rg was calculated only for protein residues. Hydrogen bonds were calculated with distance, angle and percent occupancy cut-offs as 3.5Å, 120°, and 10%, respectively, in a window of 100ps of each trajectory. The hydrogen bonds were div-

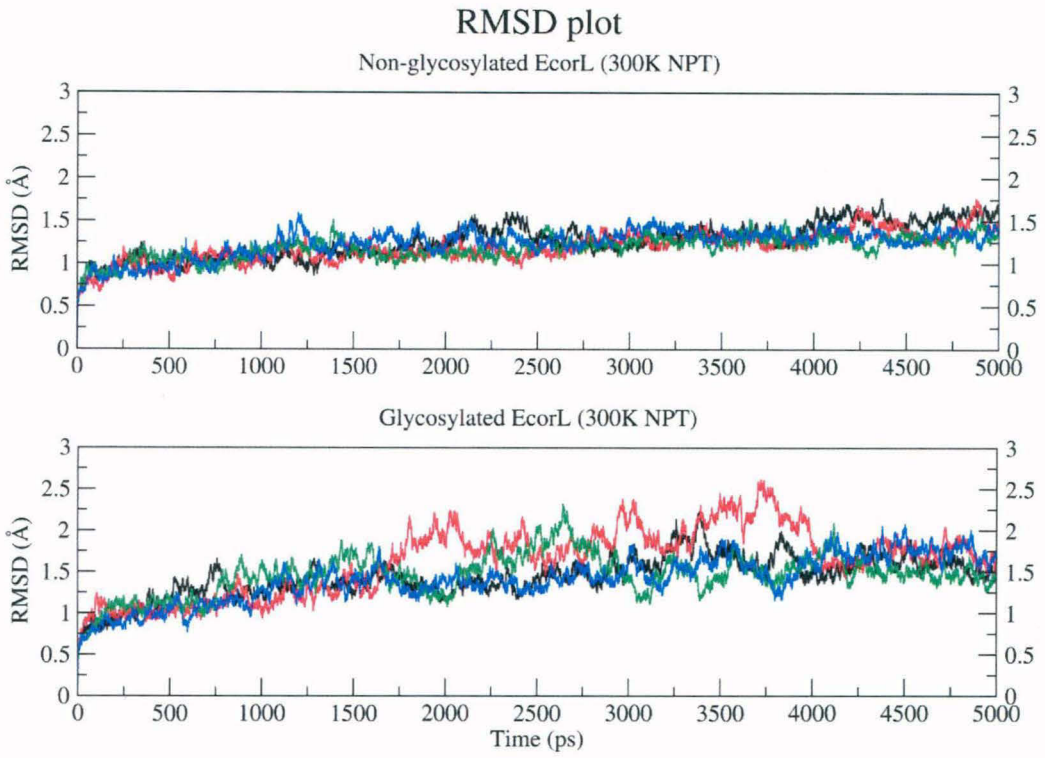


Figure 3.3: RMSD Plot.
The RMSD plot for rEcorL and EcorL at 300K (NPT) simulations. The RMSD values from the four independent simulations have been shown in different colors.

ided into two categories, the backbone interactions (M-M) which exist between main chain atoms (M) only and the side-chain interaction (M/S-S) which exist between a side chain atom (S) and either a backbone atom (M) or another side chain atom (S). NACCESS (Hubbard, 1993) was used to calculate the solvent accessible surface area (ASA). ProFit (Martin, 2008) was used for the calculation of residue-wise RMSDs (rRMS). PyMol (DeLano, 2002) was used for depiction of structural models.

3.3 RESULTS AND DISCUSSION

3.3.1: Structural features and dynamic behavior of glycosylated and non-glycosylated EcorL: Similar or different?

The structural changes in glycosylated and non-glycosylated EcorL were analyzed by RMSD plots. Figure 3.3 shows the RMSD with respect to the starting structures for each of the four simulations for glycosylated and non-glycosylated EcorL as a function of time. As can be seen, both glycosylated and non-glycosylated EcorL remain close to their respective starting structures throughout the simulations with a final RMSD close to 1.5Å. However, non-glycosylated EcorL had a peak RMSD of ~1.7Å, whereas glycosylated EcorL showed slightly higher RMSD values intermittently with a peak RMSD value of 2.6 Å. The 1AX0

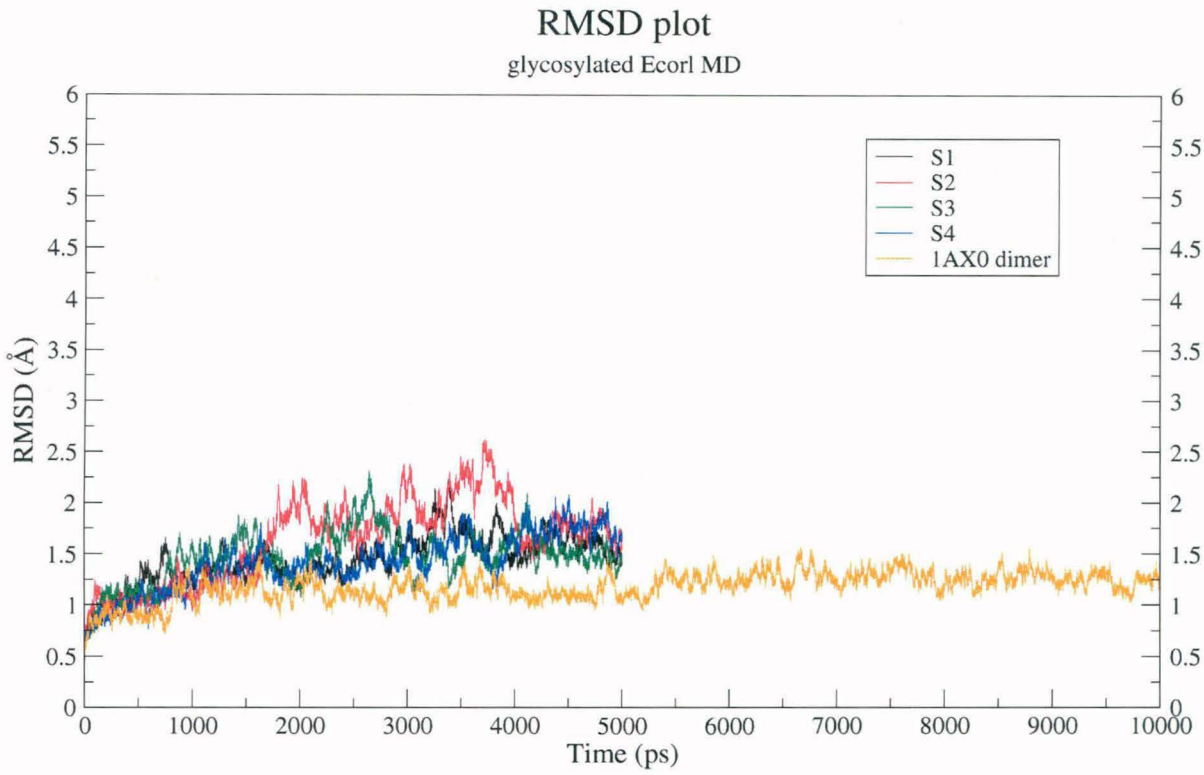


Figure 3.4: Comparison between 1SFY hybrid and 1AX0 glycosylated EcorL.
The RMSD plot for all the simulations of glycosylated EcorL including both 1SFY generated EcorL structures (S1-4; 5ns) and 1AX0 structure (1AX0 dimer; 10ns). It can be seen that both structures of glycosylated EcorL show similar RMSDs.

dimer was simulated for 10ns in a single trajectory and its structural deviation has been compared to other 5ns glycosylated and non-glycosylated EcorL simulations in Figure 3.4 where it is clear that the RMSD for 1AX0 dimer over 10ns is similar to other EcorL simulations. Therefore, we used the 5ns trajectories for further analysis as each of them had been repeated four times with different random number seeds. Figure 3.5 shows the residue-wise breakup of RMSD (rRMS) between the starting structure and the final structures at the end of the 5ns simulations for both glycosylated and non-glycosylated EcorL. It was also found that there were no regions with large differences in rRMS values (Figure 3.5) between the starting and the final structures *i.e.* rRMS was found to be similar across all the residues. This indicated that there were no significant structural differences between the final conformations of glycosylated and non-glycosylated EcorL neither in any local region nor in terms of the overall structure. Analysis of Rg values of the sampled conformations in all eight room temperature trajectories indicated that it was close to 24.3Å for both glycosylated and non-glycosylated EcorL across all simulations (Figure 3.6). We also computed the theoretical B-factor values from the fluctuations for various residues over the entire 5ns trajectories.

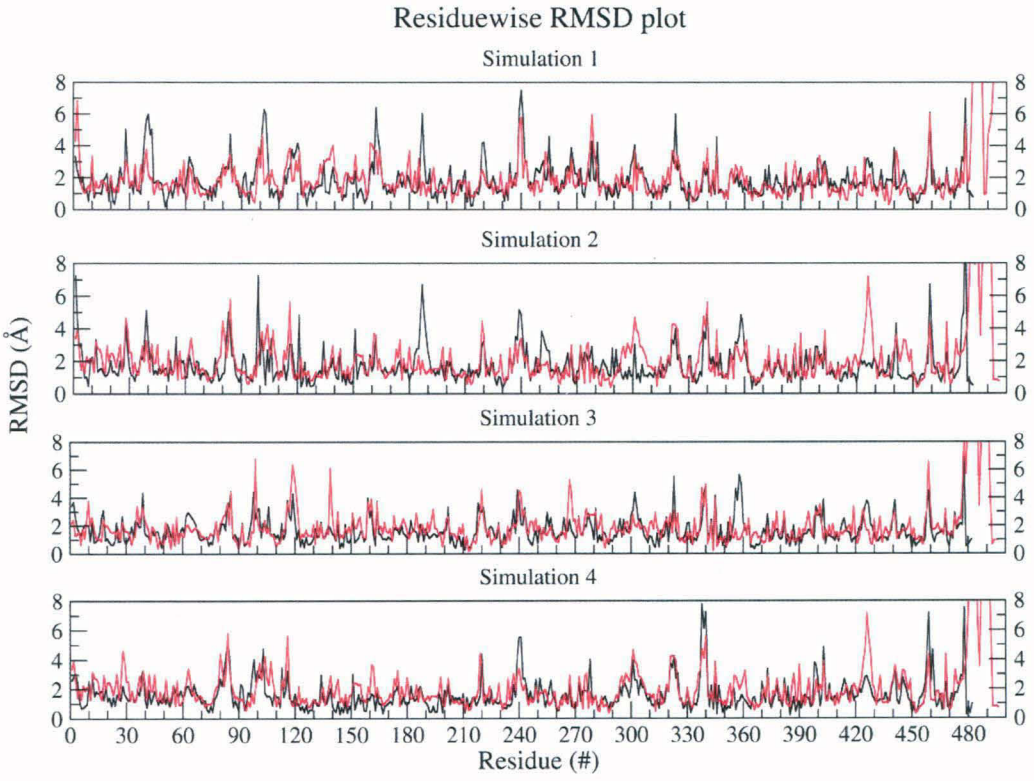


Figure 3.5: Residue-wise RMSD Plot
The residue-wise RMSD (rRMS) values for all the residues of the final structure of rEcorL (black) and EcorL (red) simulations have been plotted.

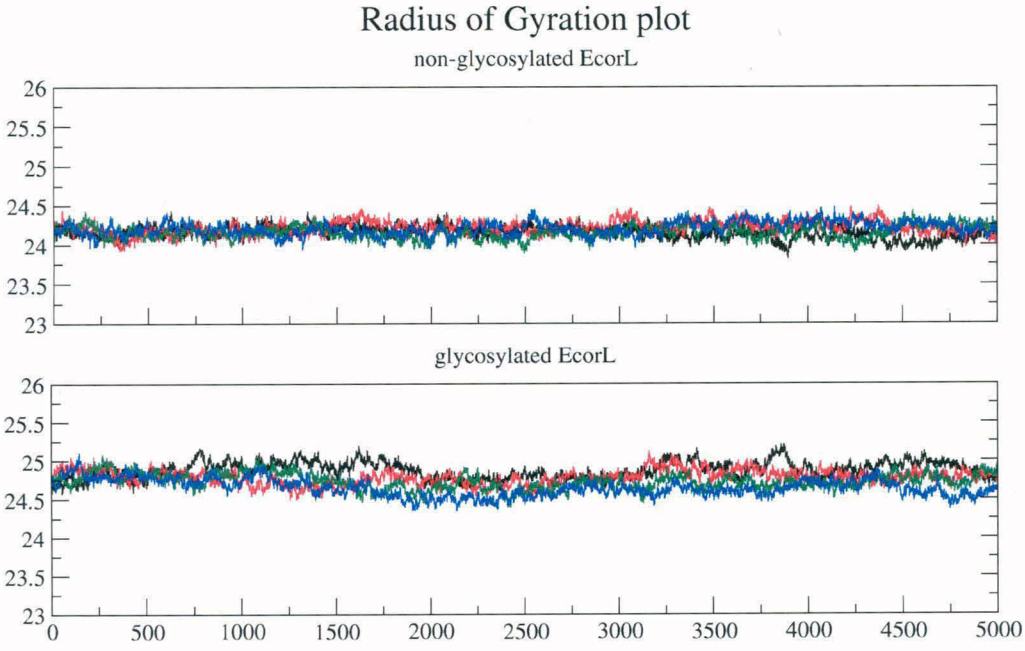


Figure 3.6: Radius of Gyration Plot
The radius of gyration values for rEcorL and EcorL has been plotted. The R_g values from all the four simulations have been plotted in different colors.

The comparative analysis of B-factors (Figure 3.7) for glycosylated and non-glycosylated EcorL revealed that during the simulations all the regions of both the proteins behaved in a very similar manner except for the loop regions. Most non-loop regions had almost identical B-factor values and showed minor variation if at all. The loop regions had sharp peaks in the B-factor plot. The average value of B-factors for non-glycosylated EcorL at 300K, over all the residues including the chain ends, were found to be 19, 18, 19, 23 $Å^2$ for the 4 different simulations whereas for glycosylated EcorL, the corresponding values were 22, 23, 29, 21 $Å^2$. Thus, glycosylated EcorL had slightly higher average B-factor values compared to non-glycosylated EcorL. However, there was no region in the protein which showed significantly different mobility due to the presence of the glycosyl moiety.

We next proceeded to analyze differences in the MD trajectories of glycosylated and non-glycosylated EcorL in terms of accessibility of side chains and also hydrogen bonds involving backbone as well as side chains. Figure 3.8 shows the variation of percentage of nonpolar solvent accessible surface area (ASA) over the different MD trajectories for both glycosylated and non-glycosylated EcorL. Table 3.1 shows the mean and standard deviation values of non-polar ASA for various simulations. The non-polar ASA for both forms of EcorL remained around 52-54% indicating that both the forms of EcorL have stable structures at 300K. As can be seen from Figure 3.8 and Table 3.1, glycosylated EcorL has lower non-polar ASA compared to non-glycosylated EcorL in three out of the four trajectories and this is most prominent in third simulation (S3). This suggests marginally higher stability of glycosylated EcorL over its non-glycosylated counterpart. We have also tried to identify if some specific region of non-glycosylated EcorL was contributing to the higher non-polar ASA. We first calculated the average non-polar ASA (ASA_{mean}) from all four simulations for each residue and then subtracted the average non-polar ASA observed in non-glycosylated EcorL from the average non-polar ASA found in glycosylated EcorL. From simulation 3 (S3; which showed maximum non-polar ASA difference) we identified residues which showed a difference of more than 10% in ASA_{mean} and they are 16-17, 25, 55, 64, 97-98, 161-163, 187 and 237 (Figure 3.9). Visualization of these residues on the structure indicated that only some of them were present adjacent to residues which interact with sugar moiety while many were present far from sugar moiety.

During the molecular dynamics simulations, many of the non-covalent interactions like hydrogen bonds keep breaking and forming. Interactions which have very short resident times or occupancy are termed as "transient interactions", while interactions having relatively

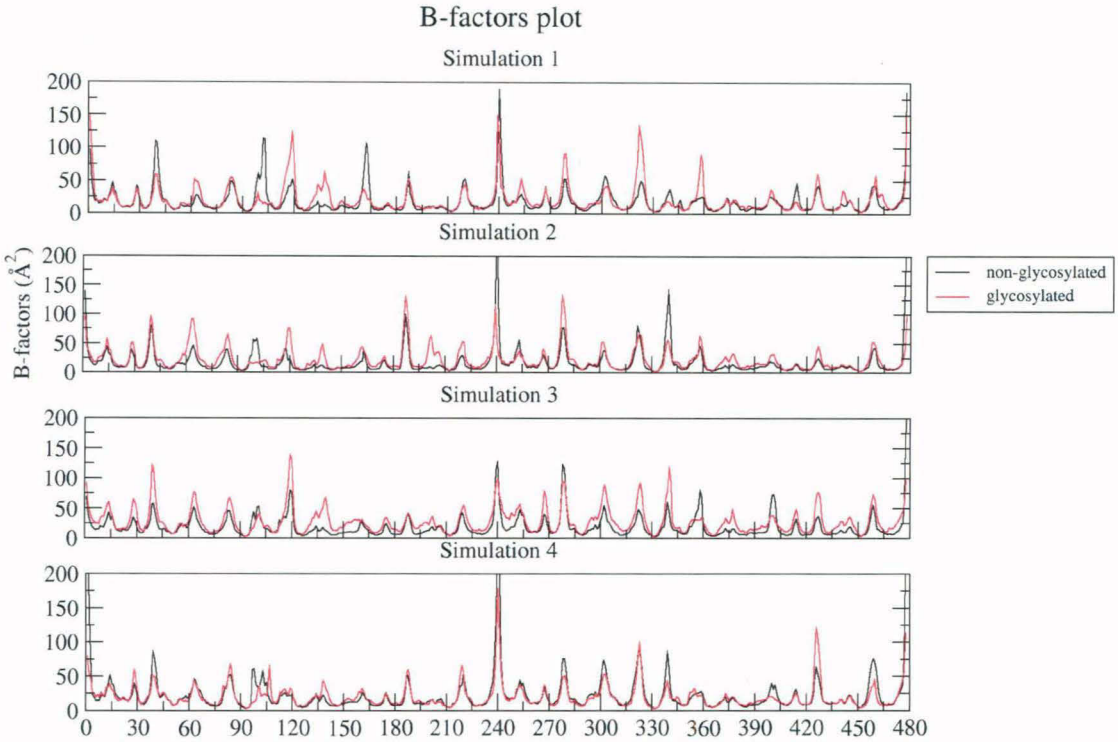


Figure 3.7: B-factor Plot
The B-factor values for rEcorL (black) and EcorL (red) has been plotted.

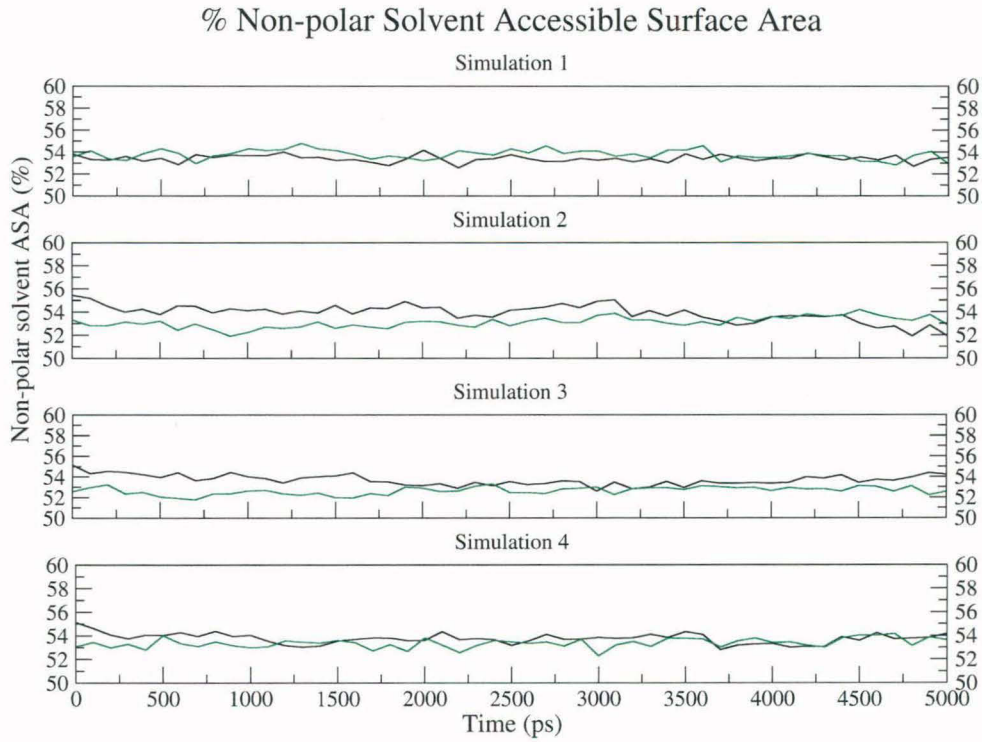


Figure 3.8: Non-polar solvent accessible surface area plot
The non-polar solvent accessible surface area has been plotted for both rEcorL (black) and EcorL (green). The ASA values from all the four simulations have been plotted in different panels.

Simulation	% non-polar ASA	
	Non-glycosylated EcorL	EcorL
S1	53.4±0.3	53.9±0.4
S2	53.9±0.7	53.1±0.5
S3	53.7±0.5	52.6±0.4
S4	53.7±0.4	53.4±0.4

Table 3.1: Non-polar ASA for non-glycosylated and glycosylated EcorL.
The mean and standard deviation value of non-polar ASA for non-glycosylated and glycosylated EcorL from all four simulations.

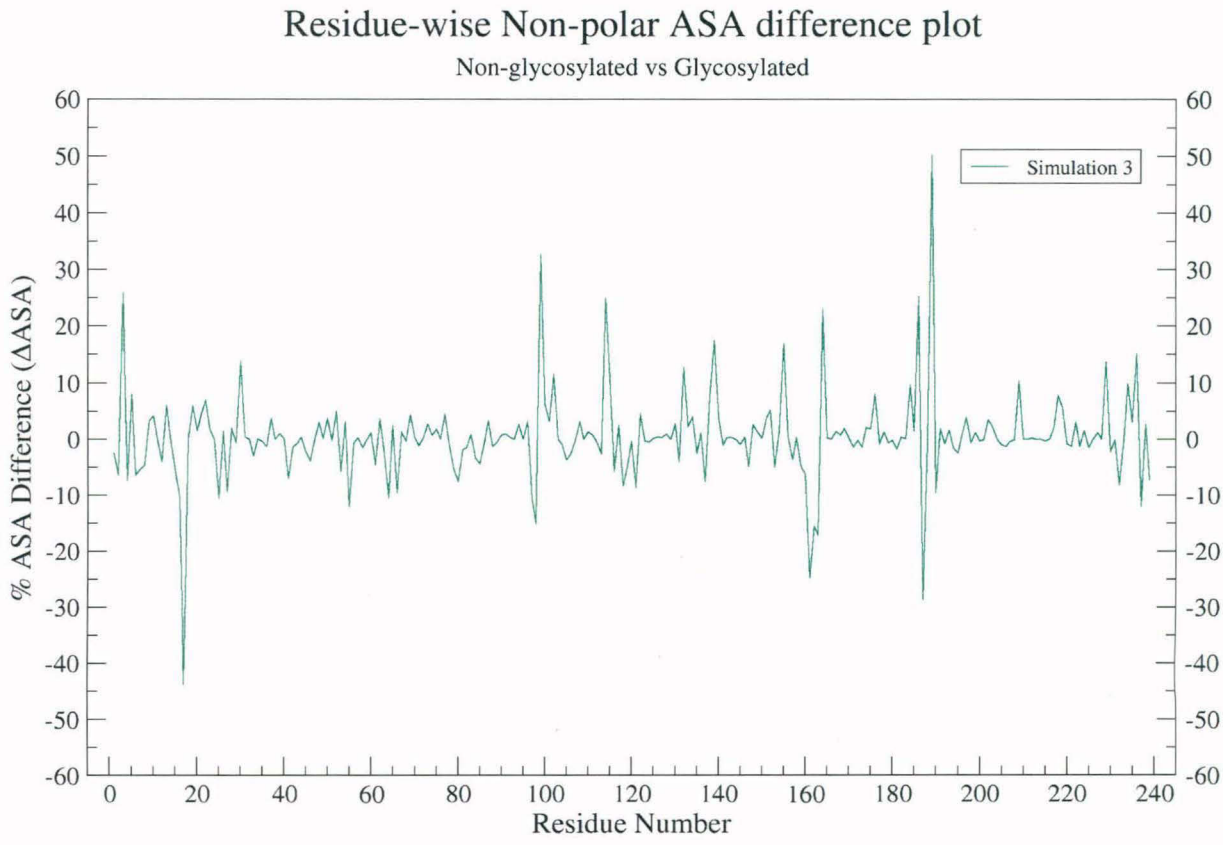
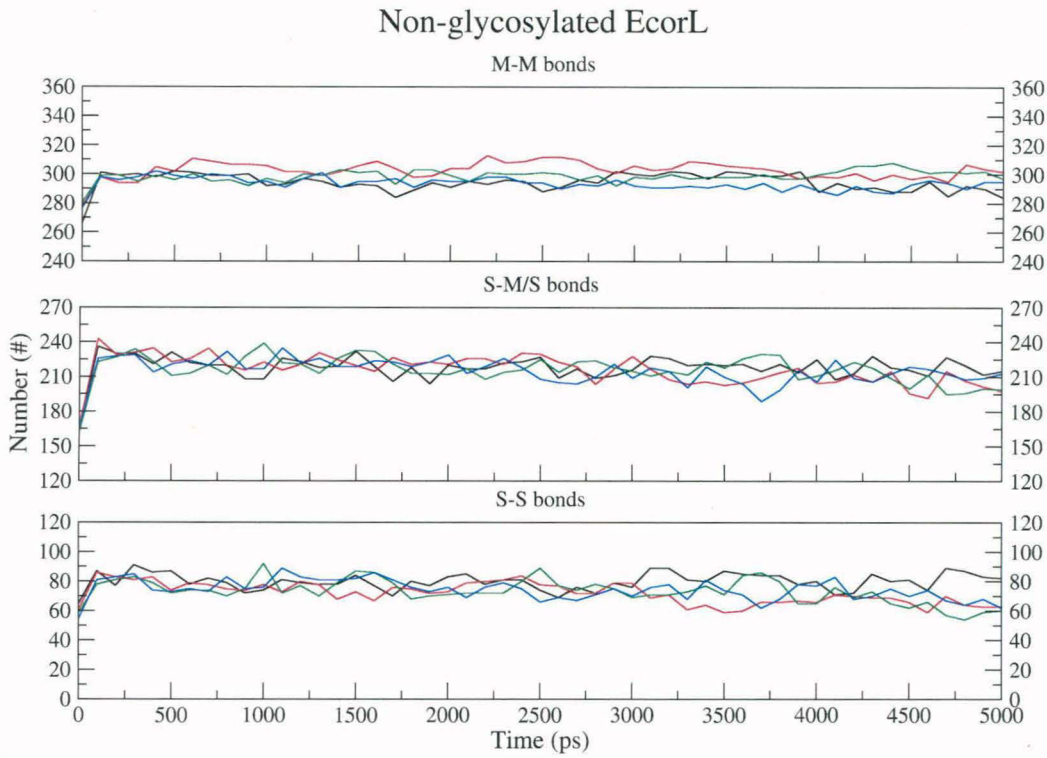


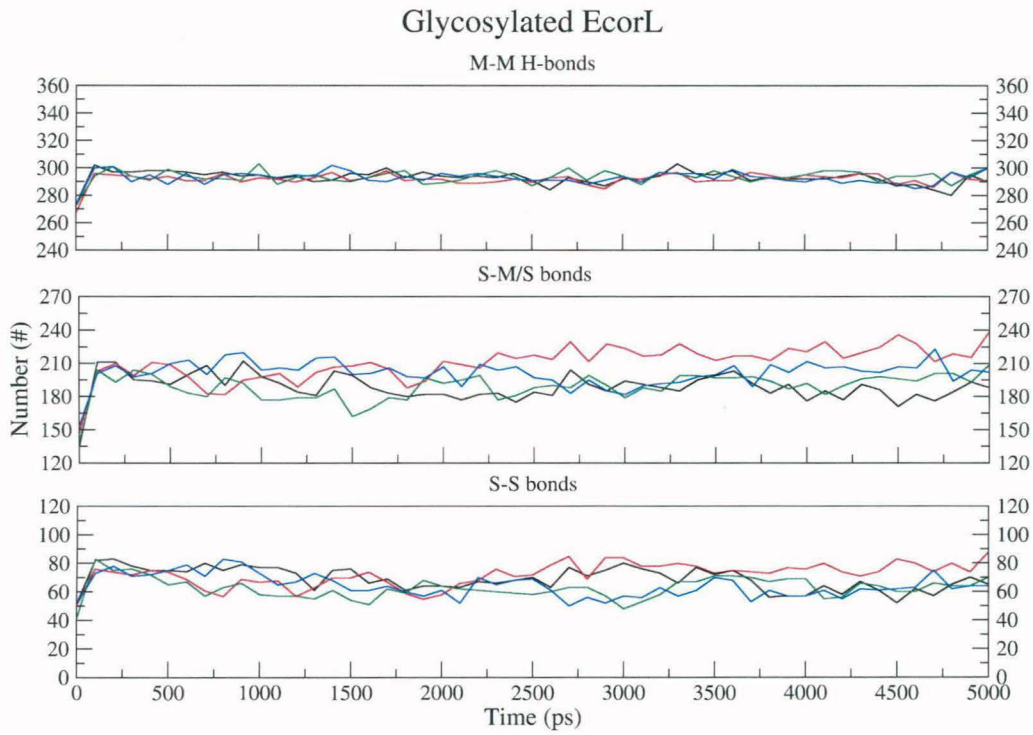
Figure 3.9: Residues specifically responsible for a higher ASA in rEcorL
The residue-wise difference between mean ASA values among the glycosylated and non-glycosylated EcorL. The change in ASA was calculated by subtracting the mean ASA of a residue in non-glycosylated EcorL from that of the glycosylated form, as described in the text.

longer resident times are termed "dynamically stable interactions". We have analyzed hydrogen bonds in each window of 100ps in the MD trajectory and hydrogen bonds which are persistent for 10% of the simulation time have been termed 'dynamically stable interactions'. Figure 3.10 shows the number of dynamically stable hydrogen bonds involving main chain - main chain (M-M), main chain - side chain (M-S) and side chain – side chain (S –S) for the non-glycosylated EcorL (Figure 3.10a) and glycosylated EcorL (Figure 3.10b) in the various 5ns trajectories sampled at 300K. As can be seen, the M-M count is similar in both forms of EcorL indicating that the intra-backbone interactions are similar among the two forms, though glycosylated EcorL seemed to have a more consistent and stable count. This stability might be considered as a signature of the structural stability at the fold level. Figure 3.10 also shows the number of hydrogen bonds involving side chains (the second panel in both Figure 3.10a and 3.10b shows total hydrogen bonds formed by side-chain atoms), which are likely to reflect the packing interactions present in the two forms. The total side-chain hydrogen bonds in non-glycosylated EcorL were observed to be more than 225 (Figure 3.10a; middle panel) whereas in glycosylated EcorL such bonds were observed to be about 200 (Figure 3.10b; middle panel). Although, at the end of 5ns, both of the forms had a similar average count for hydrogen bonds involving side chain interactions, glycosylated EcorL showed a relatively larger variation in the count of side chain interactions among different simulations. It may be noted that even though for individual trajectories there was a difference in the number of dynamically stable S-M/S and S-S hydrogen bonds between glycosylated and non-glycosylated EcorL, the differences in the numbers were of the same order as the differences which arise in multiple MD runs for the same structure. This suggests that the overall packing interactions in glycosylated and non-glycosylated EcorL are also very similar.

Similar final RMSD and almost identical number of backbone interactions strongly suggested that at the fold level there is no difference among the two forms. These observations `supported the previous studies that the tertiary structure of a glycoprotein was unaffected by glycosylation (Joao *et al.*, 1992; Mer *et al.*, 1996b). However, some differences were still observed among the two forms *e.g.* intermittent high RMSD value, slightly increased Rg and B-factors of the loops in glycosylated EcorL as compared to non-glycosylated EcorL which suggested an increased structural flexibility of the loops, possibly due to glycosylation. A decreased and varying count of side chain interactions also appeared as a constant feature in all four simulations. The decreased side chain interactions might have



(a)



(b)

Figure 3.10: Intra-molecular hydrogen bonds analysis
The intramolecular hydrogen bonds have been plotted for both rEcorL **(a)** and EcorL **(b)**. The hydrogen bonds have been classified as main chain-main chain (M-M) hydrogen bonds, total side chain (M/S-S) bonds and side chain-side chain (S-S) interactions. The values from all the four simulations have been plotted in different colors.

		Φ = O5-C1-Ox-Cx					Ψ = C1-Ox-Cx-Cx-1				
		X-ray	S1	S2	S3	S4		X-ray	S1	S2	S3	S4
GlcNAc-β(1-4)-GlcNAc (x=4)	**φ1**	**-73.7±8.4**	-74±9	-74±10	-73±9	-68±9	**Ψ1**	**116.8±15.6**	129±7	130±8	131±8	128±7
Man-α(1-4)-GlcNAc (x=4)	**φ2**	**-88±10.8**	-90±12	-84±15	-86±19	-80±14	**Ψ2**	**107.9±20.3**	103±19	104±17	110±16	115±22
Man-α(1-3)-Man (x=3)	**φ3**	**72.5±11.0**	68±9	69±10	69±10	70±11	**Ψ3**	**-112.3±22.5**	-121±16	-96±9	-98±9	-91±11
Xyl-β(1-2)-Man (x=2)	**φ4**	**-91.5±6.6**	-89±7	-85±17	-88±16	-90±16	**Ψ4**	**-105.8±3.9**	-117±17	-110±18	-114±17	-112±24
Fuc-β(1-3)-GlcNAc (x=3)	**φ5**	**-68.2±9.6**	-74±11	-67±13	-74±12	-67±12	**Ψ5**	**-101.7±8.1**	-99±9	-96±9	-98±9	-91±11
Man-α(1-6)-Man (x=6)	**φ6**	**65.4±9.0**	70±13	71±16	67±14	70±18	**Ψ6**	**182.6±5.1**	177±32	180±22	157±45	160±45

Table 3.2: Φ & Ψ angles of the individual disaccharides from the EcorL heptasaccharide
The dihedral angles for individual disaccharides are given as phi (φ) and psi (ψ) values which refer to the O_5-C_1-O_x-C_x and C_1-O_x-C_x-C_{x-1} angles, respectively. The columns S1, S2, S3 and S4 correspond to dihedral angles obtained from the four independent simulations carried out with different initial velocities.

		Non-glycosylated EcorL				Glycosylated EcorL			
	Chain	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
γ (Cγ-N-C1-O5)	A	-	-	-	-	74±9	73±10	73±9	72±9
	B	-	-	-	-	73±10	73±13	75±10	73±9
χ (N-Cα-Cβ-Cγ)	A	-96±49	-97±49	-103±62	-124±57	-68±12	-68±13	-63±11	-70±11
	B	-102±52	-76±32	-107±52	-111±65	-59±8	-47±8	-60±11	-70±11

Table 3.3: Glycosidic dihedral angle (γ) and chi (χ) angle for Asn17 of EcorL
The glycosidic dihedral angle (γ) measured as Cγ-N-C1-O5 (between Cγ-N of Asn-17 and C1-O5 of GlcNAc) and the chi (χ) angle measured as N-Cα-Cβ-Cγ for the side chain of Asn-17 have been calculated for both chains of the EcorL dimer. The columns S1, S2, S3 and S4 correspond to dihedral angles obtained from the four independent simulations carried out with different initial velocities.

led to increased Rg values and B-factors of some residues or vice versa without affecting the stability of the molecule. Instead, it was found that glycosylated EcorL had lesser exposure of non-polar regions to solvent compared to non-glycosylated EcorL indicating an advantage of extra stability for glycosylated EcorL over non-glycosylated EcorL. These observations from our MD simulations indicate that effects of glycosylation on EcorL manifests mainly in the rearrangement of the side chains so as to decrease the ASA. This could be the reason for the greater thermal stability of glycosylated EcorL over non-glycosylated EcorL. Thus, the results of our long MD simulation studies are in agreement with the reported biophysical studies regarding stability differences between glycosylated and non-glycosylated EcorL. Our simulation results are also in agreement with observations from crystallographic studies (Kulkarni *et al.*, 2004) that, even though glycosylation does not lead to a change in the overall structure of EcorL, it affects the side chain packing.

3.3.2: Conformation of the oligosaccharide moiety at 300K

We also analyzed the conformational changes occurring in the oligosaccharide moiety of glycosylated EcorL during the 5ns of simulations. The flexibility of the heptasaccharide has already been investigated in earlier studies (Naidoo *et al.*, 1997) using MD simulations. However, in those studies the majority of the EcorL protein was restrained and the simulations were for short durations. Since, our current study involved multiple numbers of long simulations without any restraints, it might give a more realistic picture of the conformational dynamics of the oligosaccharide moiety. Table 3.2 lists the average values of the various dihedral angles in the oligosaccharide along with the standard deviations over the 5 ns simulations for each of the 4 trajectories for glycosylated and non-glycosylated EcorL at 300K. For the purpose of comparison, we have also listed in Table 3.2, the average and standard deviation values of the corresponding dihedral angles in the N- and O-linkage conformations in various crystal structures. These values have been reported in literature by Petrescu *et al* (Petrescu *et al.*, 1999) based on statistical analysis of crystallographic data. The most dramatic effect of glycosylation was seen in the χ angle of Asn-17 residue (Table 3.3). For non-glycosylated EcorL, the average value of the χ dihedral angle of the Asn-17 side chain was found to vary between -76° and -124° in four different simulations and the standard deviations were as high as 65°. However, in glycosylated EcorL, the average value of χ ranged between -47° and -70° with a maximum standard deviation of only 13°. Thus, glycosylation seems to have restricted the rotation of the Asn-17 side chain to a relatively

smaller range. Similarly, the glycosidic dihedral angle (γ) connecting the oligosaccharide to the protein was found to be rigid with values close to 73° (Table 3.3) in both the chains of glycosylated EcorL with a standard deviation of only about 10°. The φ and ψ dihedral angles (Figure 3.1b) for each of the six disaccharide linkages were not only compared with their statistical averages in the crystallographic data (Petrescu *et al.*, 1999), but also with the values reported in the earlier work by Naidoo *et al* (Naidoo *et al.*, 1997) which involved short simulation of the oligosaccharide. It is interesting to note that, most of the φ and ψ values seen in our simulations (Table 3.2) were in good agreement with the values obtained from statistical analysis of crystal structures (Petrescu *et al.*, 1999). For example, the average values for φ1 (GlcNAc(1-4)GlcNAc linkage) in our simulations ranged between -68° and -74°, whereas φ1 value of -73.7±8.4° has been reported by Petrescu *et al* (Petrescu *et al.*, 1999). Similarly, the ψ1 value was found to vary between 128° and 131° in our simulations and was 116.8±15.6° in the work of Petrescu *et al* (Petrescu *et al.*, 1999). Thus, most φ and ψ values obtained from our simulations were in close agreement with the values reported by Petrescu *et al*. However, comparison of the corresponding φ and ψ values reported by Naidoo *et al* with the φ and ψ values reported by Petrescu *et al* indicate that only φ values were in agreement. The ψ values showed no direct correlation except for ψ1 and ψ5. The reason for such a disagreement could be that, the length of simulations carried out by Naidoo *et al* (Naidoo *et al.*, 1997) was very short, limiting the conformational sampling for the oligosaccharide. On the other hand, longer simulations carried out in the current work permitted efficient sampling of the conformational space, thus leading to better agreement with experimental values.

Even though most dihedral angles of the oligosaccharide moiety showed only limited flexibility, due to the combinatorial effect of the six disaccharide linkages, the oligosaccharide was highly mobile and sampled a large volume of space around the location of its tethering (Figure 3.11a) during the 5ns simulation time. So, we analyzed if during its motion the oligosaccharide made contacts with the amino acid residues in protein. We searched for any oligosaccharide atoms that were closer than 5Å to the protein atoms. It was found that most of the oligosaccharide residues, except the ones which are sequential neighbors due to covalent connectivity, were interacting more with the solvent than the protein (Figure 3.11a). The interactions, whenever occurred, were of transient nature and no stable non-covalent contact could be located between the protein and oligosaccharide atoms. Figure 3.11b shows for one trajectory, the residues that were found uniformly closer (≤5Å) to

the oligosaccharide, the oligosaccharide contacting residues for all the four simulations are listed in Table 3.4. As can be seen from Figure 3.11c all these contacting residues 4-8, 15-17, 53 and 55 were in spatial proximity of the Asn 17 which is covalently connected to the sugar moiety. Hence, our analysis suggests that the oligosaccharide shows high degree of flexibility and it interacts predominantly with the solvent. It may be noted that Naidoo *et al* (Naidoo *et al.*, 1997) had also made the similar observations about the oligosaccharide's preferential location being the solvent.

3.3.3 Interactions of Lys55 and Tyr53 with oligosaccharide

Our main objective has been to understand how the oligosaccharide moiety can help in the folding process. Analysis of the glycosylated EcorL indicated that a total of 13 residues showed interactions with the oligosaccharide. However, out of these 13 residues 11 residues are proximal in sequence to the site of glycosylation i.e. Asn17. Proximity in one dimensional sequence will automatically result in proximity in terms of spatial distance. Thus, these 11 residues are likely to be proximal to the oligosaccharide both in folded as well as unfolded state, irrespective of whether they preferentially interact with oligosaccharide or not. Therefore, interaction of the oligosaccharide with these 11 residues is unlikely to help much in the folding process. However, Lys55 and Tyr53 are sequentially far from Asn17, the site of glycosylation, and preferential interaction between oligosaccharide and these two residues would result in the formation of tertiary native contacts. Thus, it might function as a folding nucleus. Therefore, we carried out detailed analysis of interactions between glycan and these two residues.

Earlier studies have reported hydrogen bonding interactions between oligosaccharide and Tyr53 as well as Lys55, based on the analysis of crystal structure of glycosylated EcorL (Mitra *et al.*, 2003). In view of the known flexibility of the oligosaccharide group, it was not known whether these interactions seen in the static crystal structure will be stable under dynamic conditions, which are of significance in solution. We analyzed the dynamic behavior of these oligosaccharide-protein interactions in various MD trajectories. One of the novel observation from our multiple trajectory MD simulation studies is that, hydrogen bonds seen in the crystal structure do not persist during the dynamic simulations, they constantly form and break during the simulation. Analysis of hydrogen bonds involving sugar residues which persist for at least 10ps in each 100ps window indicated that, apart from the proximal oligosaccharide residues like GlcNac 301 and Fuc 302 which form hydrogen bonds with

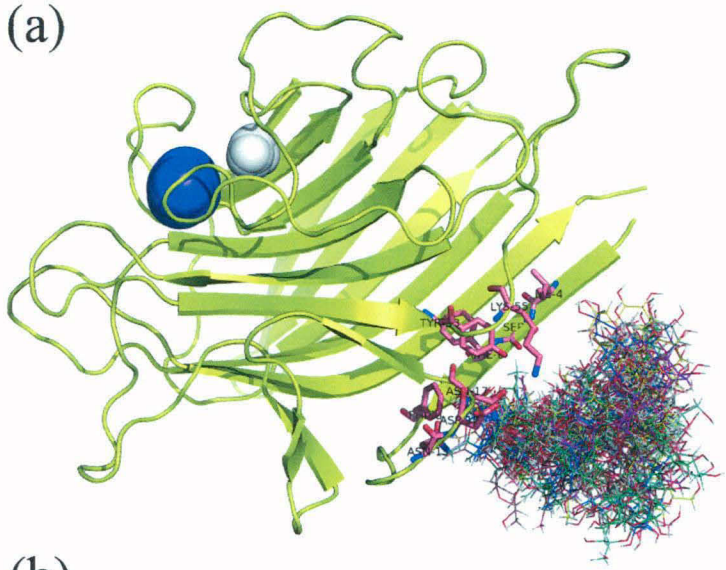
Figure 3.11: Oligosaccharide solution conformation and interacting residues

(a) A cartoon depiction of EcorL and the various conformations of the oligosaccharide sampled during the 5ns simulation. The structures were extracted at an interval of 100ps and aligned with each other using the protein co-ordinates. However, only one conformation of the protein has been shown for the purpose of clarity. The Mn^{2+} and Ca^{2+} ions are depicted as purple and white spheres, respectively. The residues at a distance closer than 5Å to oligosaccharide have been shown as sticks.

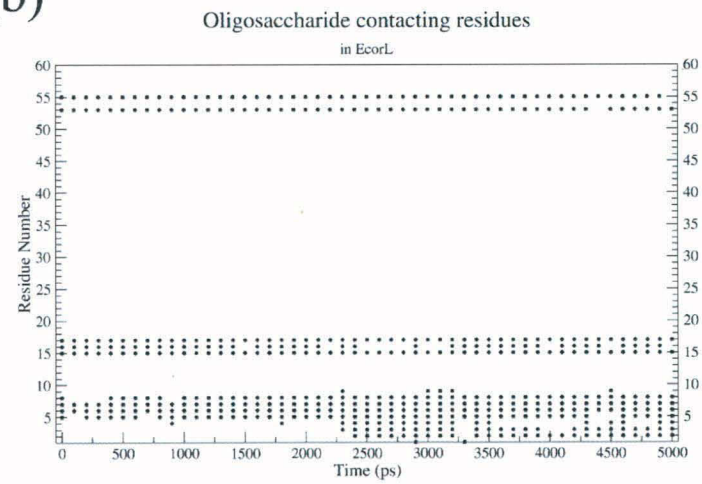
(b) A graph showing residues which had contact (closer than 5Å) with the oligosaccharide. Residues, e.g. 5, 6, 7, 15, 17, 53 and 55, were seen to be in 5Å distance of oligosaccharide throughout the simulations.

(c) The residues found to be in contact with the oligosaccharide were also located in the crystal structure (1AX0). The green transparent cartoon depicts the protein and the residues contacting oligosaccharide are shown in sticks. The oligosaccharide itself is shown in lines.

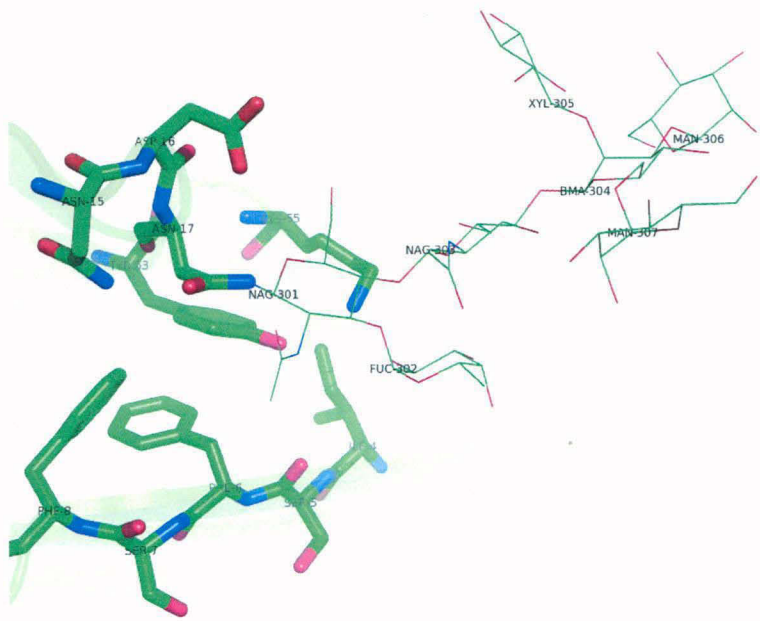
(a)



(b)


Oligosaccharide contacting residues
in EcorL
Residue Number
Time (ps)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5
10
15
20
25
30
35
40
45
50
55
60

(c)


XYL-305
MAN-306
BMA-304
MAN-307
NAG-302
NAG-301
FUC-303
ASN-16
ASN-15
ASN-17

S1		S2		S3		S4	
Chain A	**Chain B**	**Chain A**	**Chain B**	**Chain A**	**Chain B**	**Chain A**	**Chain B**
1		1				1	1
2		2		2	2	2	2
3		3		3	3	3	3
4	4	4		4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8		8	8	8	8	8	8
9		9		9	9	9	
		10			10		
		11					
	12	12	12	12	12	12	
			13		13		
		14	14	14	14		
15	15	15	15	15	15	15	15
16	16	16	16	16		16	16
17	17	17	17	17		17	17
		18					
53	53	53	53	53	53	53	53
	54					54	54
55	55	55	55	55	55	55	55
	56			56			
	57						
	58						
	59						
	64						
	96						
	117						
	207			207		207	
	208						
	209						
	210						
231							
				234			

Table 3.4: Residues within 5Å to the oligosaccharide in EcorL.
A list of various residues that were found closer than 5Å to the oligosaccharide and calculated from structures obtained from the trajectories at an interval of 100ps for all the four simulations (S1, S2, S3 and S4) performed at 300K. Residue differences among the chains can also be noted.

Lys55 in the crystal structure, even distal sugar residues like GlcNac 303, Xyl 305 and Man 307 can also form hydrogen bonds with Lys55. Figure 3.12 shows interactions of these oligosaccharide residues with Lys55 from a single trajectory. For the sake of clarity, results have been shown from a single trajectory only instead of all 4 trajectories. The top 5 panels of Figure 3.12 actually correspond to interactions of Lys55 with five different sugar residues, namely Fuc 302, Man 307, Xyl 305, GlcNac 303 and GlcNac 301. Each of these panels depicts the occupancy of various hydrogen bonds formed between a particular oligosaccharide residue and Lys55. For example, the topmost panel shows the occupancy of hydrogen bonds formed between Lys55 and Fucose residue (Fuc 302) of the oligosaccharide. It can be seen clearly that hydrogen bonds between O2 of Fucose and Lys55 form during the first nanosecond only. Similarly, the second panel shows that hydrogen bonds between O6 of Man 307 and Lys55 form only after 2 ns during the simulation. These results indicate that different hydrogen bonds form at different time points in the trajectory and persist for different lengths of time. The other three trajectories also show similar results but differ in the type of hydrogen bonds, their occupancy and the time line of their formation and disappearance on the trajectory (data not shown). Even though, the trajectory from simulation 4 (top 5 panels in Figure 3.12) shows lack of hydrogen bonds between oligosaccharide and Lys55 for considerable period of time, averaging of results from all four simulations indicated presence of two hydrogen bonds between oligosaccharide and Lys55 during most of the 5ns simulation (Figure 3.12, bottom most panel). It is also clear from the bottom panel of Figure 3.12 that on average two hydrogen bonds existed during most of the 5ns simulation, thus indicating that the bonds between the oligosaccharide residues and Lys55 are not transient. This dynamic view of interactions between protein residues and oligosaccharide is entirely different from the static picture seen in the crystal structure (Mitra *et al.*, 2003). The hydroxyl oxygen of Y53 was observed to interact, almost exclusively, with the carboxyl oxygen of acetyl group of the first *NAcGlc* (GlcNAc-301) unit of the oligosaccharide.

Figure 3.13a shows that Tyr53 of non-glycosylated EcorL (glycan-free enzyme) interacts with only two other residues *viz.* Asn17 & Lys55 and that the average number of such hydrogen bonds over the four trajectories was around 0.25 i.e. no hydrogen bond was present during 75% of the simulation time. On the contrary, in Figure 3.12, on average more than 1.5 hydrogen bonds exist between K55 and the oligosaccharide during the simulations *i.e.* at least two hydrogen bonds are present 75% of the simulation time. Despite similar spatial proximity the average number of hydrogen bonds between Tyr53 and Asn17 in the

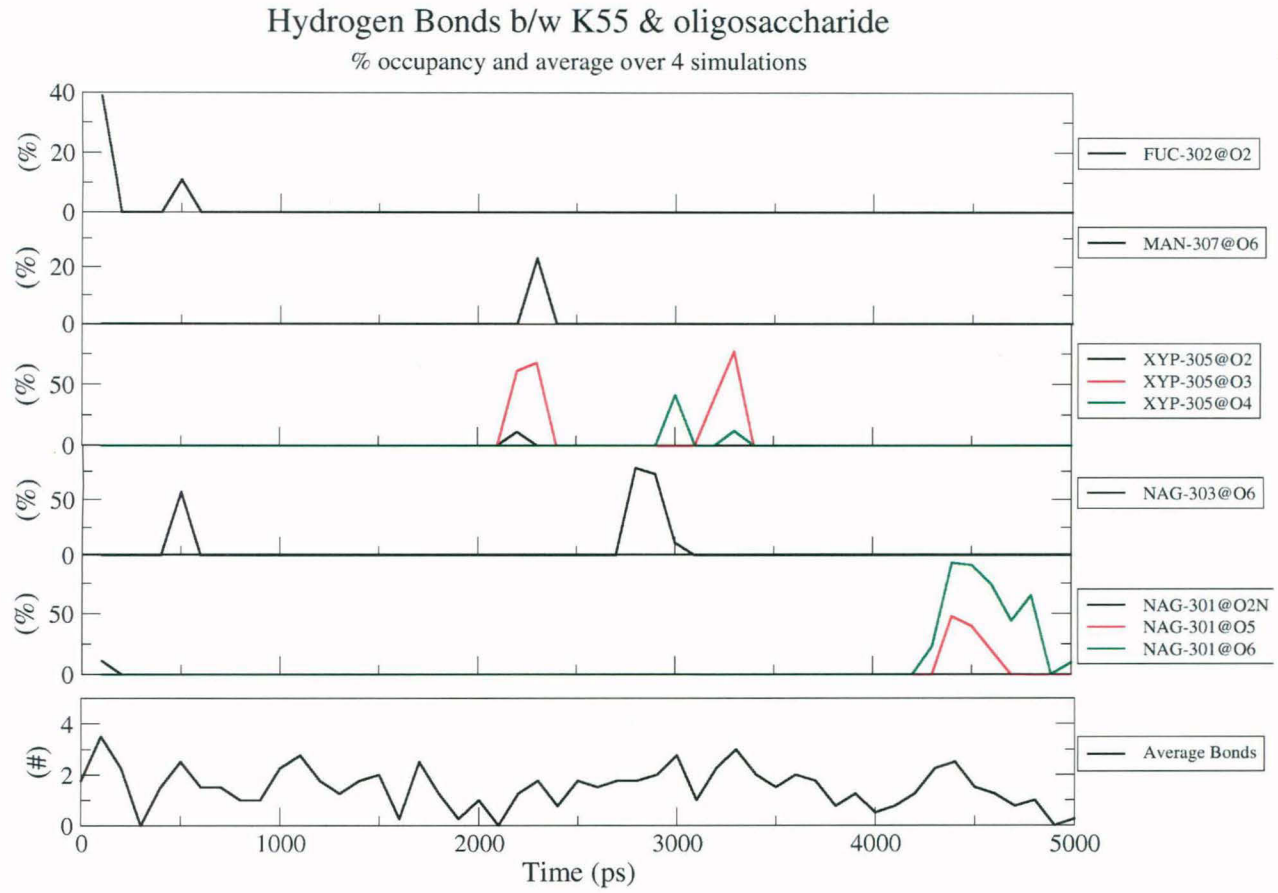


Figure 3.12: Interactions between Lys55 and the oligosaccharide
The time occupancy plot and average number of hydrogen bonds observed between K55 and the oligosaccharide. Different panels show bonds b/w K55 and a particular residue of the oligosaccharide from a single trajectory. It also depicts the discontinuity of the hydrogen bond formation over one single trajectory. The last panel, however, shows the average number of such hydrogen bonds b/w K55 and the oligosaccharide residues occurring over all four trajectories.

non-glycosylated structure is only 0.25 (Figure 3.13a). Comparison of Figure 3.13a with Figure 3.12 indicates that there is a preferential interaction between oligosaccharide and Lys55 which does not arise entirely from spatial proximity. This also suggests that even though Y53 and K55 could interact intermittently with N17 in non-glycosylated EcorL, glycosylation led to formation of stronger and dynamically stable interactions of Y53 and Lys55 with the oligosaccharide in glycosylated EcorL. The higher stability and number of these interactions clearly show that the hydrogen bonds between K55, Y53 and oligosaccharide residues in glycosylated EcorL are not a mere replacement of weak intra-protein hydrogen bonds in non-glycosylated EcorL. This is further supported by results from simulations on glycosylated EcorL mutants, as discussed later.

The interactions with oligosaccharide led to no conspicuous effects on the orientation of Y53 side-chain with respect to the backbone. However, K55 side chain seemed to have

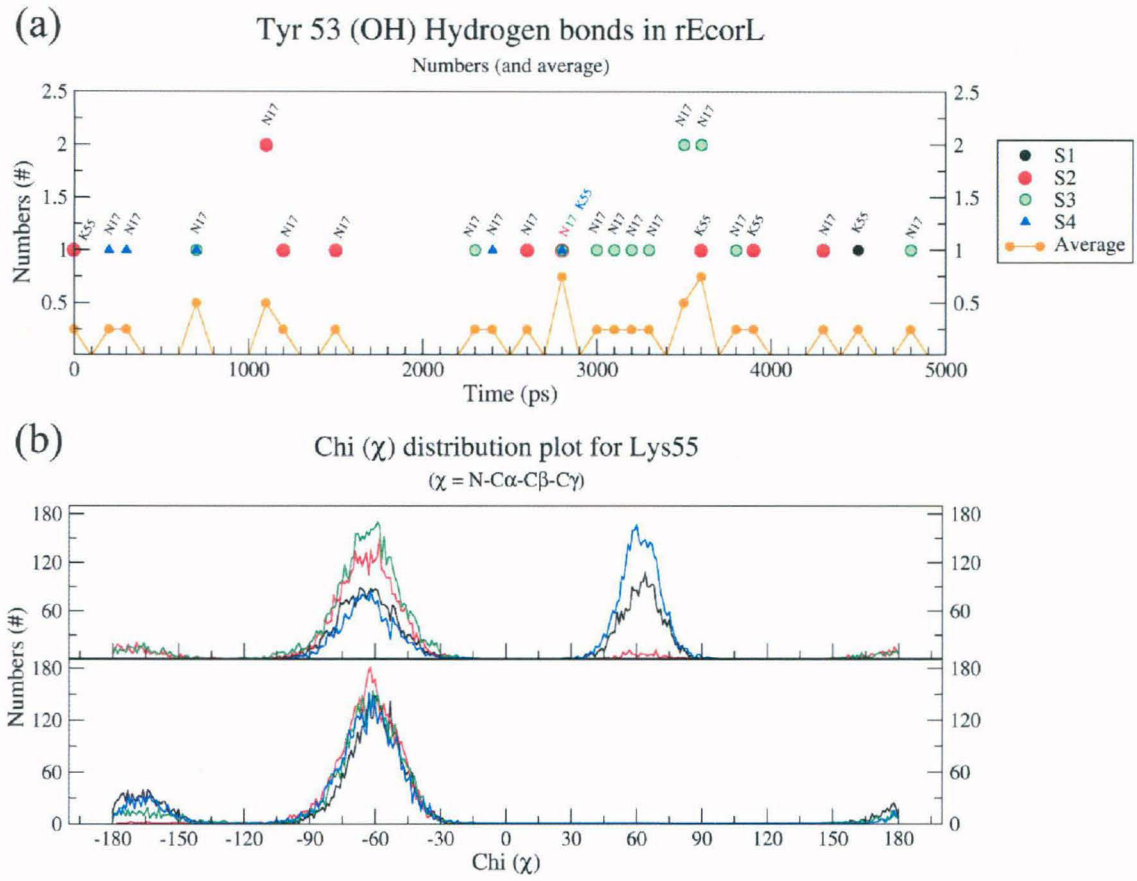


Figure 3.13: Effects glycosylation on orientations of Lys53 and Tyr55

(a) Hydrogen bonding interactions of Y53 in rEcorL have been plotted from all the four simulations and indicated with different colors. Residues with which the hydrogen bonds were observed have been labeled. Also shown is the average number of hydrogen bonds involving Y53 from all four rEcorL simulations at 300K.

(b) Dihedral angle value distribution for χ of K55 of rEcorL and EcorL has been plotted. The χ values from all the four simulations have been plotted in different colors.

been constrained as could be seen from the range of values sampled by the χ angle. In non-glycosylated EcorL, the χ of K55 was found to exist in two rotameric states, one corresponding to χ value 65° and the other one with a χ value close to -65° (Figure 3.13b). In glycosylated EcorL, only the peak corresponding to the χ value of -65° was observed. These results suggested that due to interactions with the oligosaccharide atoms the K55 side chain gets locked in a specific conformation in contrast to the two rotameric states seen in the non-glycosylated EcorL.

3.3.4 Effect of Y53A and K55A mutation on oligosaccharide interactions

Since, the analysis of the simulations on glycosylated EcorL indicated preferential interactions between oligosaccharide moiety and the residues Y53 and K55, we wanted to analyze the effect of mutations at these sites on the interactions between protein and

Oligosaccharide residues	Wild Type EcorL				Mutant EcorL			
	Tyr 53		Lys 55		Ala 53		Ala 55	
	Ch A	Ch B	Ch A	Ch B	Ch A	Ch B	Ch A	Ch B
NAG 301	10.5±0.4	9.4±0.3	10.1±0.8	8.4±0.6	10.7±0.6	10.4±0.3	9.8±1.1	9.3±0.7
FUC 302	12.9±0.9	13.0±0.4	12.3±1.5	12.0±0.8	13.5±1.0	13.0±0.7	11.3±1.7	10.8±1.2
NAG 303	14.8±0.8	13.3±0.4	12.5±1.3	10.0±0.7	15.8±0.8	15.2±0.7	13.6±1.3	12.2±1.2
MAN 304	18.7±1.2	17.4±0.5	15.3±1.9	13.4±0.9	20.4±1.2	19.3±1.1	17.2±2.0	14.9±1.6
XYL 305	17.2±2.4	**14.6±0.7**	**12.9±2.9**	10.3±1.1	19.8±2.1	**18.4±1.7**	**16.9±3.5**	13.7±2.1
MAN 306	18.4±1.8	17.6±0.9	14.5±2.3	14.7±1.1	20.6±2.2	19.1±1.7	17.1±3.4	13.8±2.1
MAN 307	21.6±1.3	**18.5±1.3**	18.0±2.0	**13.9±1.6**	23.1±1.5	**23.6±1.1**	20.4±2.0	**19.3±1.6**

Table 3.5: Comparison of the distance from COM of various oligosaccharide residues to Cβ of residue 53 and 55 in wild type and mutant EcorL. Distance (Å) between the COM of oligosaccharide residues and the Cβ atom of residues 53 and 55 in wild type and double mutant have been measured and tabulated. It is clear that Y53A,K55A double mutant EcorL shows weaker interaction compared to the wild type EcorL.

Oligosaccharide residues	Wild EcorL		Mutant EcorL	
	Ch A	Ch B	Ch A	Ch B
NAG 301	79	27	166	41
FUC 302	229	37	552	106
NAG 303	244	49	351	116
MAN 304	637	93	1037	270
XYL 305	910	83	1597	390
MAN 306	997	132	2490	584
MAN 307	1141	206	944	499

Table 3.6
A comparison of B-factors ($Å^2$) of the oligosaccharide residues for both the chains (chain A & B) in wild type glycosylated EcorL (from 1AX0) versus the double mutant EcorL (Y53A, K55A; refer manuscript text for further details). It can be seen clearly that the B-factors in the mutant are higher than the wild type EcorL.

FIGURE 3.14: Variation in the distance from COM of various oligosaccharide residues to Cβ of residue 53 and 55 in wild type and mutant EcorL.
The distance between the center of mass of each oligosaccharide residue and Cβ atom (CB) of residue 53 and 55 has been calculated for both wild type EcorL and double mutant EcorL. The difference between mutant and wild type in terms of average distances mentioned above has been plotted for both residues 53 and 55 for both chains.

oligosaccharide. Simulations were carried out on a double-mutant of glycosylated EcorL (DM-EcorL; Y53A, K55A) and conformations and orientations of the oligosaccharide were analyzed. This hypothetical double mutant of native glycosylated EcorL (1AX0 dimer) was created by replacing both Y53 and K55 side chains with Ala side chain. DM-EcorL and native glycosylated EcorL were simulated for 5 and 10ns, respectively. The side chain of Ala lacks hydrogen bond donor or acceptor atoms therefore hydrogen bonds cannot be used to analyze these interactions. The interactions between residues 53 (Tyr or Ala) and 55 (Lys or Ala) with oligosaccharide residues were analyzed by measuring the distance between center of mass of each oligosaccharide residue from C_β atom (CB) of residue 53 and 55. As can be seen from Table 3.5 and Figure 3.14, compared to the wild type EcorL the distance between oligosaccharide residues and residues 53, 55 in DM-EcorL was higher by at least 5Å. Additionally, the B-factors of oligosaccharide residues were found to be higher compared to the corresponding values in the wild type EcorL (Table 3.6). These results further demonstrate that the observed interactions between oligosaccharide and residues at position 53 and 55 arise from specific hydrogen bonding interactions and not from their spatial proximity. Thus, our simulations indicated that glycosylation at N17 leads to formation of dynamically stable interactions between sites which are separated in sequence and in locking

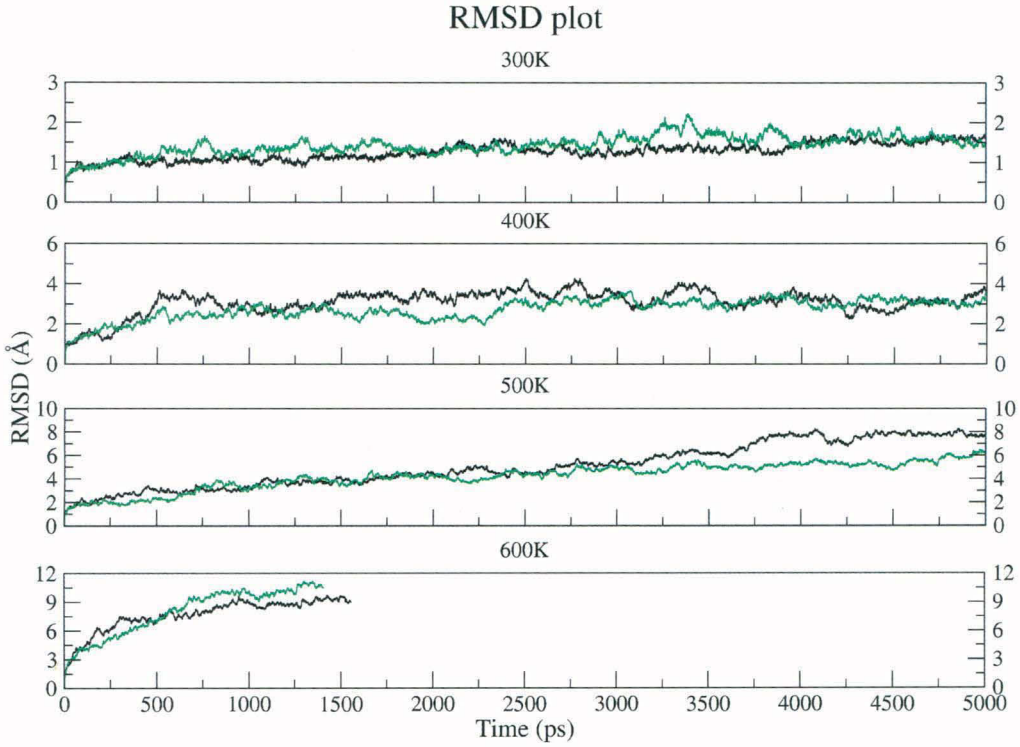


Figure 3.15: RMSD plot for conformations sampled during high temperature simulations
RMSD plot for non-glycosylated EcorL (black) and glycosylated EcorL (green) at 300K and higher temperatures (400K, 500K, and 600K) from simulations.

long flexible side chains in specific conformations. The stable interactions between the oligosaccharide and the residue 53 and 55 could act as a folding nucleus and help in enhanced folding as seen in earlier experimental studies (Mitra *et al.*, 2003). The folding nucleus proposed here involved hydrogen bonding interactions but generally folding nucleus involves hydrophobic residues. The observations about hydrophobic interactions driving the process of protein folding are mostly based on folding studies on non-glycosylated proteins. In case of glycoproteins hydrogen bonding interactions might initiate the process of folding. However, a caveat may be noted before we infer the formation of a folding nucleus from our simulation results. Since, we have not carried out folding simulations; our results do not unambiguously demonstrate that these interactions are indeed a folding nucleus. This is a hypothesis based on the results of our simulations.

3.3.5 Thermal unfolding of the glycosylated and non-glycosylated EcorL

As discussed earlier, the analysis of the results of various simulations at room temperature indicated that both glycosylated and non-glycosylated EcorL had similar structural features, except for a decreased nonpolar ASA and altered side chain-backbone interactions in glycosylated EcorL. Some additional destabilizing factors e.g. low pH

(Esposito and Daggett, 2005), might have improved the chances of observing the differences in stability among the two forms of EcorL. However, to avoid generating complications in the results we have not tried to test the effects of other parameters on the structure of EcorL. The lower nonpolar ASA would suggest higher stability for glycosylated EcorL compared to non-glycosylated EcorL. Earlier experimental studies (Mitra *et al.*, 2003) have also shown that deglycosylation leads to lower stability of non-glycosylated EcorL compared to glycosylated EcorL. We wanted to further investigate the differences in the stability of glycosylated and non-glycosylated EcorL using *in silico* thermal unfolding studies by carrying out MD simulations at elevated temperatures. MD simulations were carried out for glycosylated and non-glycosylated EcorL at 400K, 500K and 600K. Unfolding of the structures at various elevated temperatures was monitored by RMSD plots over the trajectories (Figure 3.15) and compared with the RMSD plot from 300K simulation. As can be seen, at 300K the RMSD increased to ~1Å within first 500ps and for the rest of the 4.5 ns simulation it only increased another 0.5Å. However, at 400K RMSD for both glycosylated and non-glycosylated EcorL had increased to ~3Å at about 1ns and for initial 3.5ns, the RMSD of glycosylated EcorL was less compared to non-glycosylated EcorL. At 500K, the RMSD for both the forms rose to ~5Å in the initial 3ns of the simulation but in the subsequent 2ns run the final RMSD for glycosylated EcorL increased to ~6Å whereas non-glycosylated EcorL had RMSD value of ~7.5Å rising at the end of 5ns MD run. The lower RMSD of glycosylated EcorL at 400K for a major period and also a lower final RMSD at 500K suggest differences in the stability of glycosylated and non-glycosylated EcorL. The relatively lower deviation from the native structure in the glycosylated EcorL compared to the non-glycosylated form was also seen in regions which specifically interacted with the oligosaccharide. Figure 3.16 shows the conformation of the β strand containing the amino acids Tyr 53 and Lys 55 which preferentially interact with the oligosaccharide and could possibly form the folding nucleus. As can be seen, at elevated temperatures this β strand has relatively lower deviation from its native structure in the presence of glycosylation. This suggests that the dynamically stable contacts between the oligosaccharide and the residues Tyr 53 and Lys 55, cannot only act as potential folding nucleus, they can also impart additional stability to the glycosylated form. Additionally, the analysis of non-polar ASA of the structures of glycosylated and non-glycosylated EcorL sampled at elevated temperature simulations showed that glycosylated EcorL had relatively lower non-polar ASA compared to non-glycosylated EcorL even at higher temperatures which was prominent in 500K and 600K simulations (Figure 3.17).

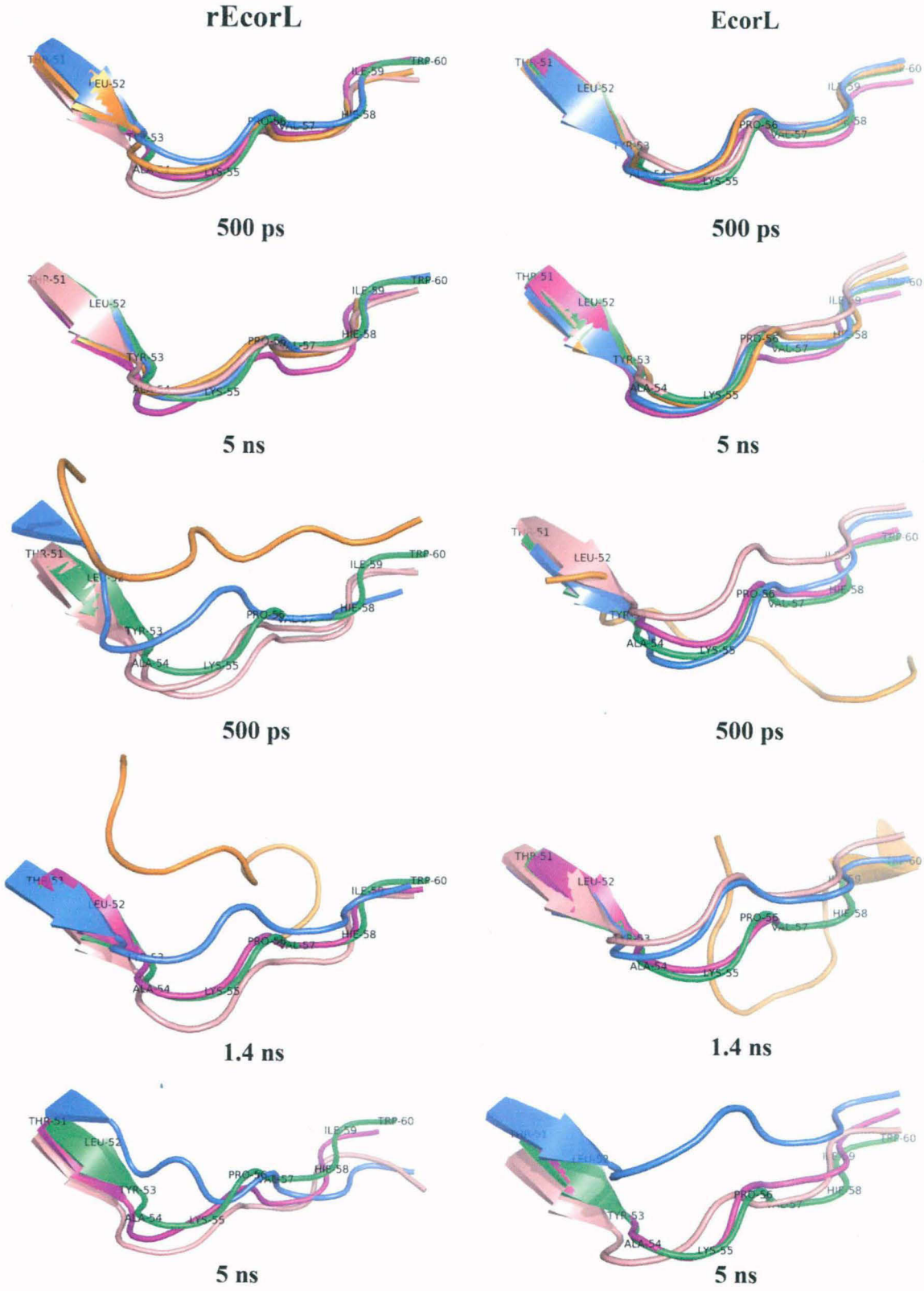


FIGURE 3.16
The conformation of the strand containing amino acids K55 and Y53 in 300K and higher temperature simulations has been shown. The initial time point chosen was 500ps as at least several hundred ps would be required before we can see any structure destabilizing effect on the protein. The final conformation of the strand was obtained from the last conformation at the end of the trajectories (5ns).

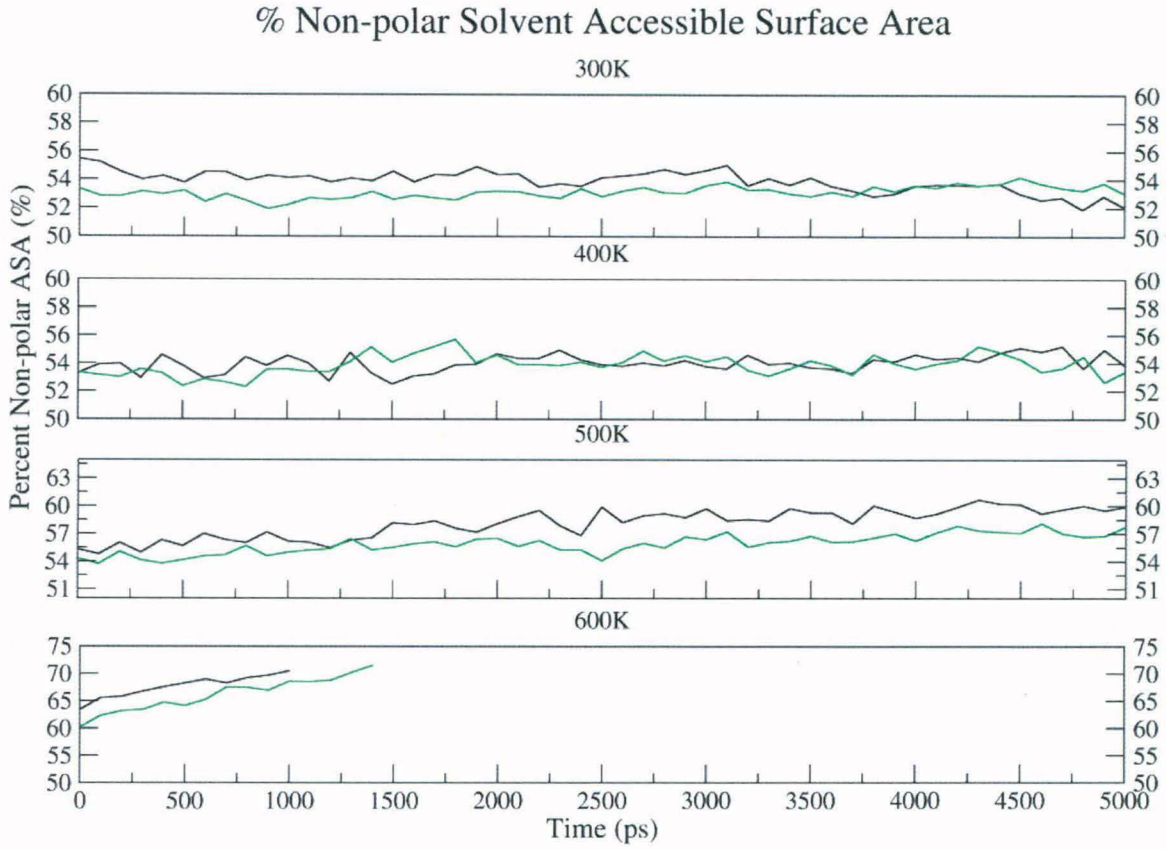


Figure 3.17: ASA plot for conformations sampled during high temperature simulations
The ASA plot for rEcorL (black) and EcorL (green) at 300K and higher temperatures (400K, 500K, and 600K) from simulations. ASA was calculated every 100ps from the trajectory.

Thus, the results from these elevated temperature simulations were also in agreement with results from simulations at 300K in terms of stability differences between glycosylated and non-glycosylated EcorL. At 600K, the simulations could not be continued beyond 1.5ns and within this period the RMSD for non-glycosylated and glycosylated EcorL had steeply increased to ~9Å and >10.5Å, respectively. This indicated that both the forms now had a hugely unfolded structure compared to the native structure at 300K. At such high temperatures, the contribution to stability due to glycosylation was not adequate to compete with temperature induced unfolding. Thus, the high temperature simulations were also in qualitative agreement with the earlier experimental studies which suggested lower stability of non-glycosylated EcorL compared to glycosylated EcorL. We also carried out detailed analysis of the rRMS to understand the mechanistic details of temperature induced unfolding of glycosylated and non-glycosylated EcorL. It was found that, in all the high temperature simulations only the loop regions of the protein showed higher RMSD from native structures while the primary hydrophobic core of the proteins consisting of the β-sheet including the

dimer interface remained unperturbed even at 600K.

3.4 CONCLUSION

A series of long MD simulations (5ns each) in explicit solvent environment for glycosylated as well as non-glycosylated dimers of EcorL were carried out at various temperatures for understanding the effect of glycosylation on the structure and stability of this glycoprotein. Detailed analysis of these multiple MD trajectories indicated that, glycosylation led to neither local nor overall change in EcorL structure. These results have interesting implications for efforts towards expression of glycoproteins in heterologous systems for therapeutic purposes. Even though the structure of the glycosylated EcorL was found to be very similar to its non-glycosylated counterpart, the percent non-polar solvent accessible surface area (ASA) for glycosylated EcorL was found to be less compared to non-glycosylated EcorL for most of the simulation time. This indicates a higher stability of glycosylated EcorL compared to non-glycosylated EcorL. In order to understand the effect of glycosylation on thermal unfolding of EcorL, a series of MD simulations were also carried out on glycosylated as well as non-glycosylated EcorL at elevated temperatures of 400K, 500K and 600K. The unfolding of the structures at elevated temperatures was measured by RMSD from native structure and non-polar accessible surface areas. Interestingly, in many of the high temperature simulations the glycosylated EcorL showed lower RMSD and non-polar ASA compared to non-glycosylated EcorL. Thus, these high temperature simulations are also in qualitative agreement with earlier experimental studies which suggested slightly higher stability for glycosylated EcorL compared to the non-glycosylated form.

Analysis of the various conformations sampled by the oligosaccharide moiety indicated that, during the simulations the oligosaccharide moiety was highly flexible. However, the average values of various dihedral angles was in close agreement with average values of the corresponding dihedrals reported in literature based on analysis of various crystal structures. In various trajectories, the oligosaccharide moiety preferentially interacted with Lys 55 and Tyr 53 which are separated in sequence from the site of glycosylation Asn 17. It may be noted that, hydrogen bonds between oligosaccharide and these residues as seen in the crystal structure do not persist during the dynamics, they constantly form and break during the simulation. Apart from the proximal oligosaccharide residues like GlcNac 301 and Fuc 302 that form hydrogen bonds with Lys55 in the crystal structure, even distal sugar residues like GlcNac 303, Xyl 305 and Man 307 can also form hydrogen bonds with Lys55.

This dynamic view of interactions between protein residues and oligosaccharide is entirely different from the static picture seen in the crystal structure. We believe that it constitutes a novel observation from the current studies. In non-glycosylated EcorL the Lys 55 side chain sampled two rotameric states, whereas in glycosylated EcorL Lys 55 remained predominantly in one rotameric state due to interactions with the sugar groups. It is possible that glycosylation helps in forming long range contacts between amino acids which are separated in sequence and thus provides a folding nucleus. Interestingly, the β strand containing Tyr 53 and Lys 55 which might work as folding nucleus also shows relatively lower deviations from native structure in glycosylated EcorL compared to the non-glycosylated counterpart. Thus, our simulations are not only in agreement with experimental observations about higher stability of glycosylated EcorL compared to non-glycosylated EcorL but also provide novel insights into possible molecular mechanisms by which glycosylation can help in enhanced folding of the glycoprotein by formation of folding nucleus involving specific contacts with the oligosaccharide moiety. However, it must be noted that unambiguous demonstration of involvement of oligosaccharide moiety in folding nucleus would require folding simulations which are currently impossible for such large proteins. Nevertheless, our simulations provide valuable clues which can help in design of experiments for better understanding the folding and stability of glycoproteins.

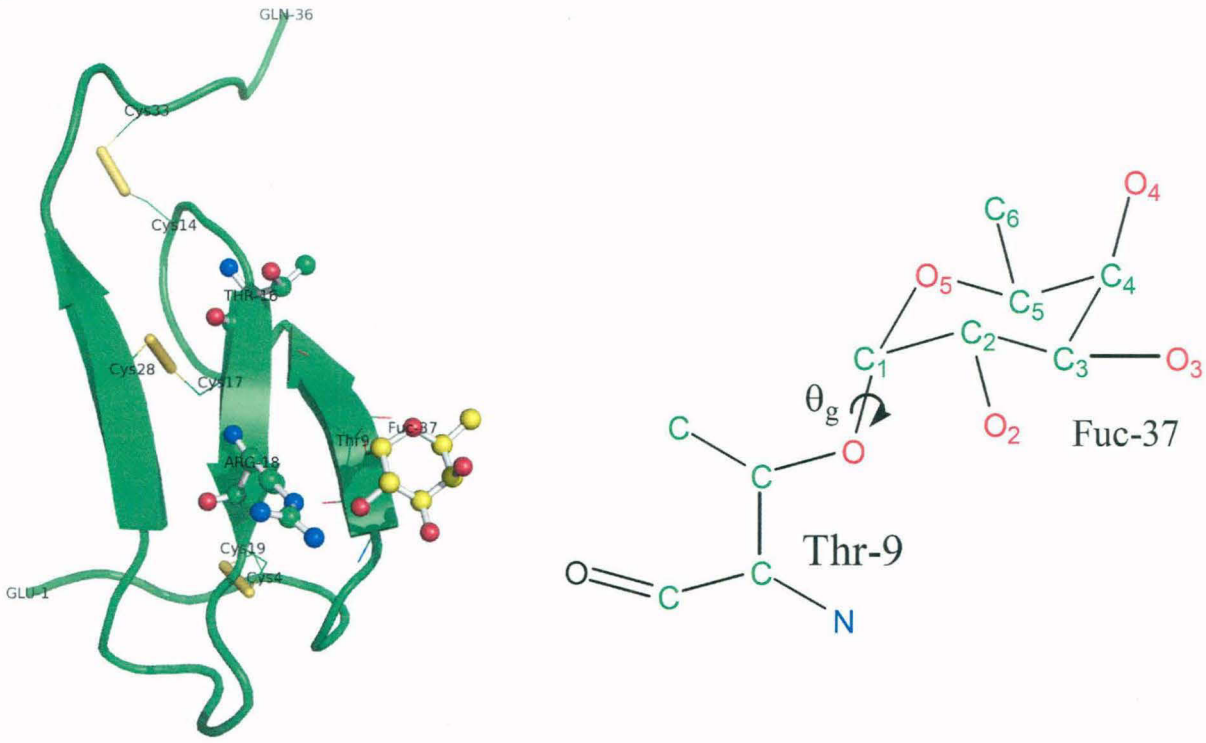
Chapter 4

Role of glycosylation and disulfide bonds in structure and function of *pars intercerebralis* major peptide-C (PMP-C): A molecular dynamics study

4.1 INTRODUCTION

Pars intercerebralis major peptide-C (PMP-C) is a 'small' serine protease inhibitor composed of 36 amino acid residues. PMP-C is extracted from the brain and fat bodies of the migratory locust, *Locusta migratoria* (Nakakura *et al.*, 1992). PMP-C belongs to a large group of regulatory proteins, known as "serine protease inhibitors" or *serpins* which act as substrates for serine proteases (Plotnick *et al.*, 1996). Serpins regulate the activity of various proteases for which they act as a substrates. Serine proteases are a ubiquitous and well explored class of enzymes constituting over one third of all known proteolytic enzymes (Page and Di Cera, 2008; Zakharova *et al.*, 2009). They catalyze peptide bond hydrolysis through a nucleophilic attack on the target peptide bond using a uniquely reactive serine (Craik *et al.*, 1987; Kraut, 1977). The attack occurs between residues known as P1 and P1' generating an acyl-enzyme complex and a peptide with P1' residue as the new amino terminus (Zakharova *et al.*, 2009). Later, the acyl-enzyme complex is cleaved releasing a peptide with P1 as the C-terminus. They are involved in primary processes like digestion, protein maturation and turnover, immune responses *etc.* with their direct or indirect malfunctioning leading to several clinical syndromes (Bots and Medema, 2006; Megyeri *et al.*, 1995; Menasche *et al.*, 2005; Page and Di Cera, 2008; Trapani, 2001). An imbalance between proteases and their inhibitors could affect the survival of the organism (Rubin, 1996). Several hereditary diseases like emphysema, cirrhosis and angioedema arise due to deficiency of such inhibitors (Rubin, 1996). Viruses, tumor cells or virus-infected cells survive against cytotoxic responses of the immune system using several intracellular inhibitors of granzymes to escape cell death (Bots *et al.*, 2005; Medema *et al.*, 2001). Therefore, serine protease inhibitors are being used as antiviral agents (Reiser and Timm, 2009). They have been found to be protective against tissue damage occurring through free radical production or apoptosis by inhibiting processes like adhesion and migration (Rubin, 1996). In view of the biological importance of the serine protease inhibitors, several studies have attempted to understand structural features of serine protease inhibitors and intricacies of their interaction with proteases (Gettins *et al.*, 1993; Lawrence *et al.*, 1995; Matheson *et al.*, 1991; Plotnick *et al.*, 1996; Roussel *et al.*, 2001; Rubin, 1996; Wei *et al.*, 1994).

PMP-C belongs to the pascifastin family of serine protease inhibitors found in arthropods. A unique feature of pascifastin family is the presence of conserved cysteines which are involved in disulfide bonds. PMP-C has been found to exist in both fucosylated (FU-PMPC) and non-fucosylated (NF-PMPC) forms. Fucosylation occurs on a threonine (Thr-9) residue, making it an O-linked glycopeptide. Both, FU-PMPC and NF-PMPC are capable of inhibiting human leukocyte



(a) (b)

Figure 4.1

a) A cartoon depiction of PMP-C with all three disulfide bonds intact has been shown. The starting, disulfide linked and terminal residues have been labeled. The site of glycosylation (Thr-9), Thr-16, Arg-18 and Fucose residues are shown in ball and sticks representation.
b) A schematic representation of fucosylation at Thr-9 of PMP-C. The glycosidic dihedral angle (θ_g) has been depicted.

elastase and α-chymotrypsin but NF-PMPC can inhibit high voltage-activated Ca^{2+} currents whereas FU-PMPC cannot. Glycosylation of PMP-C has been observed to enhance its structural stability compared to its non-glycosylated form (Mer *et al.*, 1996b). The three dimensional structure (Figure 4.1a) of PMP-C consists of a β-sheet formed by 3 anti-parallel β-strands, two loop regions connecting the β strands and small flexible regions at N- and C-terminus (Mer *et al.*, 1996b; Roussel *et al.*, 2001). The three β-strands are connected by 3 disulfide bonds between three pairs of cysteines namely, Cys 4-Cys 19, Cys 14-Cys 33 and Cys 17-Cys 28. A hydrophobic core exists on one side of the β-sheet in a manner similar to other serpins (Bode *et al.*, 1986; Fujinaga *et al.*, 1982; McPhalen *et al.*, 1985a; McPhalen *et al.*, 1985b). NMR studies on fucosylated as well as non-fucosylated forms of PMP-C indicate that glycosylation does not alter the structure of PMP-C significantly and structural differences arising from glycosylation are localized to the site of glycosylation (Mer *et al.*, 1996b). However, FU-PMPC exhibited an enhanced structural stability

compared to NF-PMPC. The proton exchange rate for several hydrophobic core residues was found to be decreased in FU-PMPC which is a sign of a compact hydrophobic core relative to NF-PMPC. Based on these observations it has been proposed that there is an overall decrease in the dynamic fluctuations in fucosylated PMP-C and this might be the reason for its enhanced stability. Apart from enhancing structural stability fucosylation is also known to enhance the functional properties of PMP-C. The fucosylated PMP-C is known to be a more potent protease inhibitor compared to its non-fucosylated counterpart. Therefore, it is necessary to understand the molecular basis of the effect of glycosylation on the structural stability and conformational flexibility of this functionally important peptide.

It is well known that glycosylation affects structural and dynamic features of many proteins and peptides, and consequently, has significant effect on their functional properties. However, lack of atomically detailed crystal/NMR structures of glycoproteins and glycopeptides along with their glycan moieties have often been a major bottleneck in understanding the effect of glycosylation on structure and dynamics. Even though crystallographic and NMR studies (Mer *et al.*, 1996a; Mer *et al.*, 1996b; Roussel *et al.*, 2001) have given novel insights on the structure of PMP-C, effects of glycosylation on structure and stability of PMP-C and its interaction with proteases, the crystal and NMR structures of PMP-C lack coordinates of the glycan moiety . Therefore, it is necessary to complement results from experimental studies with computer simulations for comprehensive understanding of the effect of glycosylation on structure and function. In recent years, computer simulations of glycans and glycoproteins are being increasingly used to address biological problems, because of easy availability of high performance computing resources and enormous progress in development of reliable forcefield parameters. In our earlier work, we have used MD simulations to understand dynamic interplay between ion binding and substrate recognition by Concanavalin A (ConA) (Kaushik *et al.*, 2009), and effect of glycosylation on structure and stability of the Erythrina corallodendron lectin (EcorL) (Kaushik *et al.*, 2011). Similarly, MD simulations have been used by several other groups to study various aspects of glycoproteins and obtain insights on elusive matters (Chiappori *et al.*, 2010; Gandhi and Mancera, 2009; Ganguly and Mukhopadhyay, 2006; Gupta *et al.*, 2009; Hansia *et al.*, 2007; Hess *et al.*, 2008; Janosi *et al.*, 2010; Marrink *et al.*, 2007; Vorontsov and Miyashita, 2009). MD simulations have been used to analyze serine proteases as well (Topf *et al.*, 2002). These studies prompted us to use molecular dynamics simulations for understanding the role of fucosylation on structural stability and conformational flexibility of PMP-C. There has been one report on unfolding MD simulations on PMP-C (Choi *et al.*, 2010). However, the structural model of PMP-C used in this simulation study did not appear to

have incorporated disulfide bonds between the conserved cysteines which impart stability to both fucosylated and non-fucosylated PMP-C.

In this work, we have carried out multiple explicit solvent MD simulations on fucosylated and non-fucosylated forms of native PMP-C with disulfide bonds and also on fucosylated and non-fucosylated forms of non-native PMP-C lacking disulfide bonds to understand relative contributions of disulfide bonds and fucosylation towards structural stability and conformational flexibility of this novel protease inhibitor. The main objective of the study has been to explore the conformations sampled by the fucose moiety and identify the residues of PMP-C with which fucose moiety is involved in stable interactions over long time scale. We have also tried to understand, whether fucosylation indeed lowers the flexibility of the PMP-C as seen in NMR studies. Since, residues 28-32 of PMP-C are known to interact with the protease inhibitor, we have also attempted to understand how exactly fucosylation at Thr 9 which is away from the protease interacting segment affects its function as an inhibitor.

4.2 METHODS

4.2.1: Starting structures for MD simulations on non-fucosylated and fucosylated PMP-C

The coordinates for the non-fucosylated form of PMP-C were obtained from PDB. The PDB entry 1PMC (Mer *et al.*, 1996a), had the NMR derived coordinates for 36 models of PMP-C. Since, the 36 structural models were very similar to each other and had maximum backbone (N, Cα and C) RMSD of 2.26 Å, one of them was chosen as the starting structure for MD simulations on non-fucosylated PMP-C. The structure of the fucosylated form of PMP-C is not available in PDB. However, NMR studies on fucosylated PMP-C have indicated that structure of fucosylated PMP-C is similar to its non-fucosylated counterpart and these studies have also determined the dihedral angles which define the orientation of fucosyl moiety with respect to the Thr 9 side chain. Based on these dihedral values the fucosylated-Thr 9 was modeled in the starting structure for fucosylated PMP-C using XLEAP module of AMBER 9 (Case *et al.*, 2006), while all other amino acids had coordinates identical to those in non-fucosylated PMP-C. In order to build the fucosylated Thr, fucose residue was added as the last residue (Fuc 37) and relevant bonds, angles and dihedrals were defined between Thr 9 and Fuc 37. Figure 4.1b shows the glycosidic dihedral angle (θ_g) which defines the orientation of the fucosyl moiety of Fuc 37 with respect to the Thr 9. The ff03 (Duan *et al.*, 2003) and Glycam06 (Kirschner *et al.*, 2008) force fields were used for assigning atom types, charges, equilibrium bond lengths, bond angles, dihedral and van der Waals parameters for the

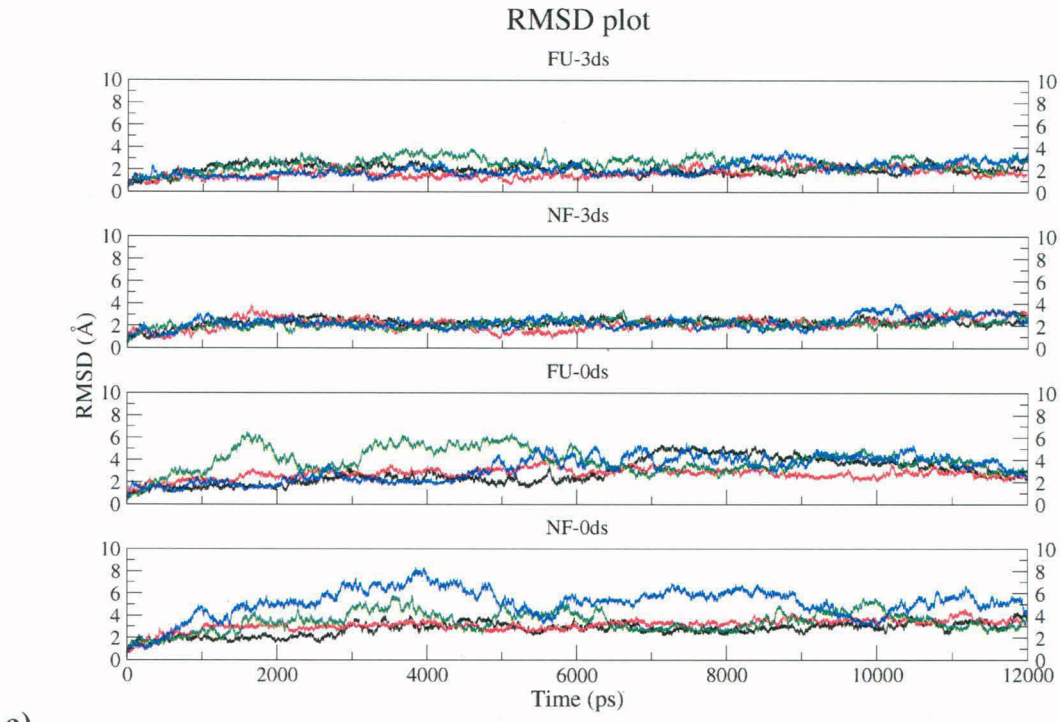
simulations. Both the fucosylated and non-fucosylated PMP-C structures were solvated using TIP3P (Jorgensen *et al.*, 1983) water box which extended 10Å from the outer most atom of the peptide along X, Y and Z axes. The periodic boundary conditions (PBC) were used to negate the surface effects at the box boundaries. Particle-Mesh Ewald (PME) summation method (Darden *et al.*, 1993) was used for the calculation of electrostatic potential. The native form of PMP-C contains 3 disulfide bonds between six cysteine residues. Apart from disulfide bonded PMP-C, simulations were also carried out on reduced form of fucosylated and non-fucosylated PMP-C which lacked the disulfide bonds.

4.2.2: Molecular Dynamics

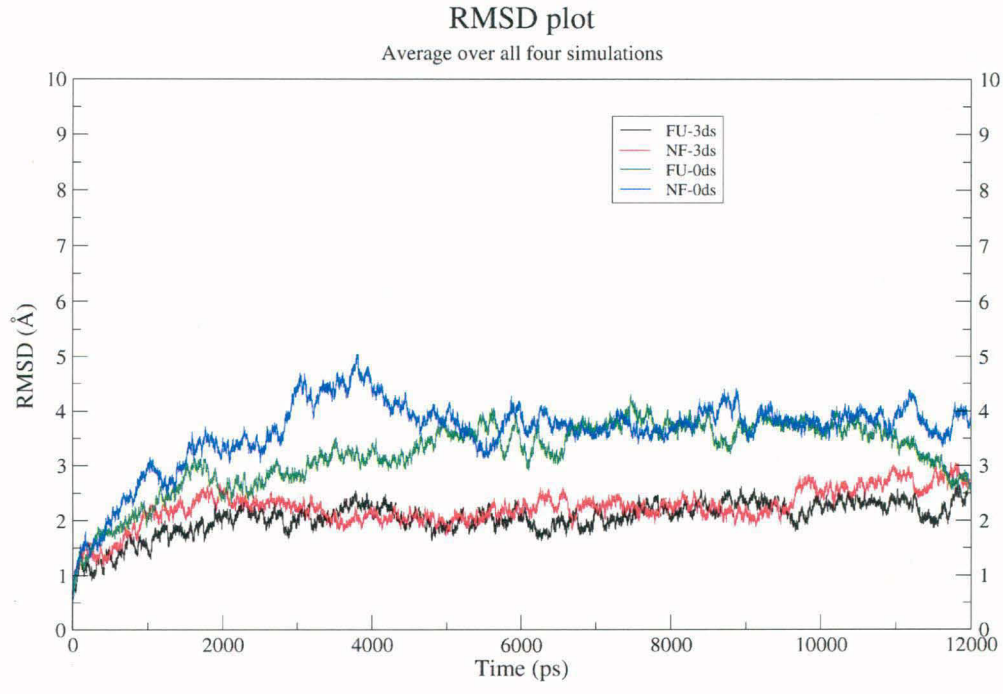
Both the solvated PMP-C structures were minimized using the SANDER module of AMBER9 (Case *et al.*, 2006) till the RMS gradient of potential energy reached a value of 0.001 kcal/mole/Å. Minimization was followed by equilibration dynamics which was carried out in two steps. The first step involved a 20ps constant volume (NVT) equilibration to bring the system to the desired temperature of 300K, followed by the second step involving a 200ps constant pressure (NPT) equilibration to bring the system to the desired pressure *i.e.* 1 atm. After equilibrating the system to the desired temperature and pressure, molecular dynamics simulation was continued under NPT conditions for 12 ns. A step size of 1 fs was used during the entire equilibration and production phases of the simulations. Since the conformations sampled during a particular MD run is often dependent on the initial velocities assigned to the atoms in the molecule, for efficient sampling of the conformational space it is necessary to carry out multiple MD runs with different starting velocity assignments. Calculations of dynamic averages of various structural and thermodynamic parameters from multiple trajectories are statistically more meaningful than parameters computed from single simulations. Therefore, for each of the two structures, four different simulations were run using different seeds for the random number generators for initial velocity assignments. The coordinates were stored at an interval of 1000 steps of MD integration, thus, generating 12000 structures for each MD trajectory. Thus, the four trajectories provided us with a total of 48000 conformations for both the fucosylated as well as non-fucosylated PMP-C.

4.2.3: Analysis of MD trajectories

PyMol (DeLano, 2002) software was used for depiction analysis of three dimensional structures. Ptraj module of AMBER9 (Case *et al.*, 2006) and in-house PERL scripts were used to analyze various conformational features of the structures sampled during the MD simulations. For



a)



b)

Figure 4.2
The RMSD plot has been shown for native (3ds) and non-native (0ds) forms of both fucosylated and non-fucosylated PMP-C.

a) Individual RMSD plots for various simulations with different trajectories being colored differently.
b) Average RMSD values have been plotted for the four types of simulations.

each of the structures sampled during the simulations, root mean square deviation (RMSD) of the coordinates with respect to the starting structure was computed to analyze conformational changes during the MD simulation. Apart from RMSD, radius of gyrations (R_g) of the whole molecule and theoretical B-factors (BF) for each residue of the fucosylated and non-fucosylated PMP-C were also computed. Radius of gyration (R_g) indicated the change in overall shape of the molecule, while theoretical B-factor (BF) values were a measure of the mobility of the various residues. The hydrogen bonds were calculated using a cut off value of 3.5Å for the donor-acceptor distance and 120^0 for the donor-hydrogen-acceptor angle. The percentage occupancy of hydrogen bonds was calculated using a window size of 100ps and only those hydrogen bonds which persisted for at least 10% of the time in each window were considered as stable during dynamics. The hydrogen bonds were divided into two groups, namely hydrogen bonds involving main chain atoms only (M-M) and hydrogen bonds involving side chains (S-M/S). During analysis of inter residue contacts, the distances between methyl groups of Fuc 37 (Fuc^{37}-CH_3), Thr 16 (T^{16}-CH_3) and δ-methylene group of Arg 18 (R^{18}-CH_2) were calculated using the centers of masses of the respective groups. The *mean* values for various parameters were calculated by averaging over all values obtained from a particular trajectory. Thus, there were four mean values obtained for a particular parameter each representing four different trajectories. The combined mean values were also calculated considering all the values of the parameter from all four trajectories.

4.3 RESULTS AND DISCUSSION

4.3.1: Analysis of the structural changes during dynamics simulation

Molecular dynamics simulations were carried out on both fucosylated and non-fucosylated forms of PMP-C. Both forms contained three disulfide bonds between six cysteine residues; therefore, we refer to the simulations of fucosylated and non-fucosylated forms as FU-3ds and NF-3ds, respectively. However, presence of these covalent disulfide bonds between β strands is likely to significantly restrict the main chain conformational space accessible to these molecules. This structural rigidity might overshadow the structural stability arising from glycosylation and thus, deciphering the effect of glycosylation on structure of PMP-C from FU-3ds and NF-3ds would be a difficult task. Therefore, we have performed another set of simulations as well on fucosylated and non-fucosylated PMP-C which lacked disulfide bonds. We refer to these simulations as FU-0ds and NF-0ds. Simulations on each of these 4 forms of PMP-C were repeated 4 times using different random number seeds for initial velocity assignments. This resulted in a total of 16 MD trajectories

		Simulation 1	Simulation 2	Simulation 3	Simulation 4	Combined
RMSD	FU-3ds	3.2 (2.0± 0.4)	3.4 (1.7± 0.5)	4.0 (2.4± 0.6)	3.8 (2.0± 0.6)	4.0 (2.0± 0.6)
	NF-3ds	3.3 (2.3±0.4)	4.0 (2.2±0.6)	3.6 (2.2±0.4)	4.0 (2.3±0.6)	4.0 (2.2±0.5)
	FU-0ds	5.4 (2.9±1.1)	4.2 (2.8±0.5)	6.5 (3.9±1.1)	5.5 (3.2±1.1)	6.5 (3.2±1.1)
	NF-0ds	4.3 (2.8±0.6)	4.7(3.2±0.5)	5.8 (3.4±0.9)	8.3 (5.1±1.3)	8.3 (3.6±1.3)
Rg	FU-3ds	11.5 (10.7±0.2)	11.4 (10.7±0.2)	11.9 (10.9±0.3)	12.0 (11.0±0.4)	12.0 (10.8±0.3)
	NF-3ds	11.9 (11.0±0.3)	11.7 (10.7±0.4)	11.7 (10.9±0.2)	11.3 (10.6±0.2)	11.9 (10.8±0.3)
	FU-0ds	11.5 (10.5±0.3)	11.8 (10.8±0.4)	12.3 (11.1±0.4)	12.6 (10.8±0.5)	12.6 (10.8±0.5)
	NF-0ds	11.1 (9.7±0.4)	12.0 (10.5±0.6)	12.5 (10.7±0.5)	12.6 (11.2±0.5)	12.6 (10.5±0.7)

Table 4.1
The RMSD and Rg values from different simulations representing different conditions of simulations have been tabulated. The values contain three entries *i.e.* peak, mean and standard deviation for each of the four simulations [format: peak (mean ± standard deviation)]. The last column of the table *(combined mean)* tabulates the averages calculated using all the values from all four simulations.

of 12 ns length each. Figure 4.2a shows the variation of backbone RMSD (for N, Cα, and C atoms) for the four different forms of PMP-C with respect to their respective starting structures and for each form results from four different trajectories obtained by different random number seeds are shown. Figure 4.2b shows similar RMSD plots for four different forms of PMP-C which have been obtained by calculating the averages from the four simulations with different random number seeds. Table 4.1 shows the peak RMSD for each trajectory and also mean and standard deviation values for RMSDs from each trajectory. As can be seen from Figure 4.2 and Table 4.1, the mean RMSD for native fucosylated PMP-C with disulfide bonds varies between 1.7-2.4Å and the peak RMSD ranged between 3.2-4.0Å. On the other hand, the mean RMSD for native non-fucosylated PMP-C (NF-3ds) varied between 2.2-2.3Å while the peak values ranged between 3.3–4.0Å. The mean RMSD values obtained from averaging of four trajectories for both FU-3ds and NF-3ds were found to be 2.0 and 2.2 Å, respectively (Table 4.1). These results indicate that, native structures of both fucosylated and non-fucosylated PMP-C have a rigid backbone conformation because of the presence of disulfide bonds. These results also suggest that, backbone conformations of native fucosylated PMP-C and its non-fucosylated counterpart were very similar. In contrast to native PMP-C with disulfide bonds, non-native PMP-C which lacked disulfide bonds showed significant differences between the fucosylated and non-fucosylated forms. As can be seen from Figure 4.2a

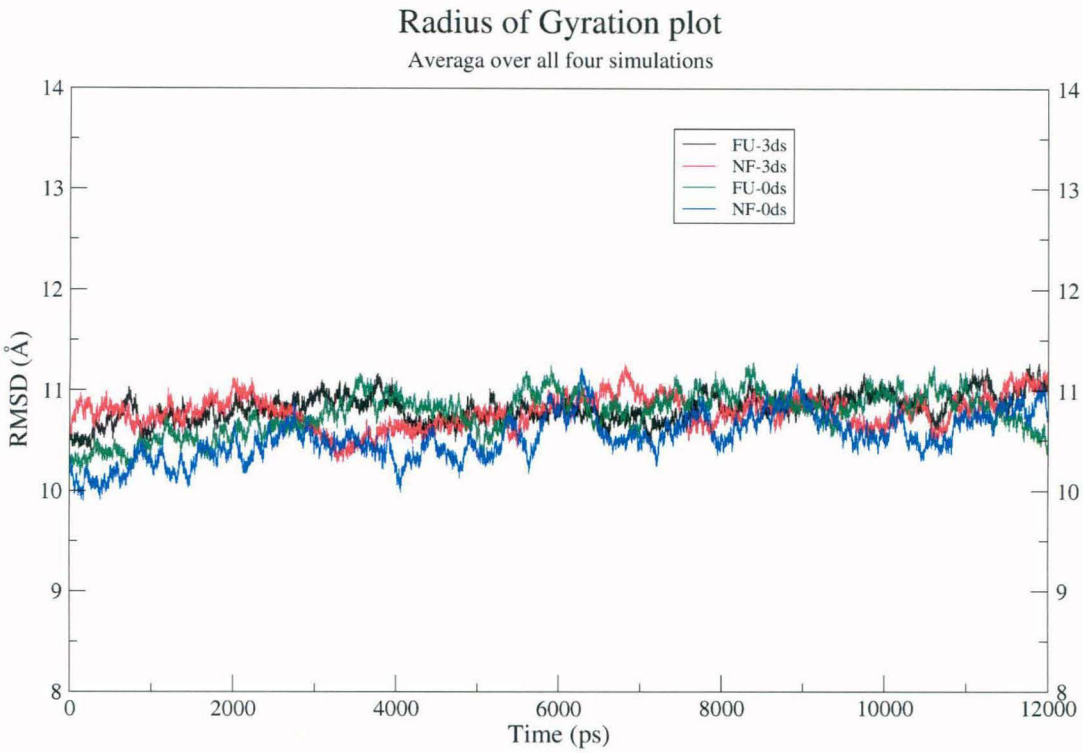


Figure 4.3
The average Rg values have been plotted for 3ds and 0ds forms of both fucosylated and non-fucosylated PMP-C. Different trajectories have been colored differently.

and Table 4.1, for the four trajectories obtained for FU-0ds the mean RMSD varied between 2.8 and 3.9Å whereas peak RMSD values ranged between 4.2 and 6.5Å. On the other hand, mean RMSD for NF-0ds varied between 2.8–5.1Å while the peak values ranged between 4.3–8.3Å. These results indicate that in absence of the disulfide bonds the conformational flexibilities are higher and hence, different simulations with different random number generators sample different regions of the available conformational space. Therefore, even at room temperature these structures unfold, deviating as high as 8Å from the native structure, but they again refold to structures which are within 3 to 4Å of the native state. Interestingly, the degree of unfolding seen in the non-fucosylated forms lacking disulfide bonds is higher than their fucosylated counterparts. This suggests that glycosylation helps in increasing the stability of the fucosylated form. It may be noted that, because of the presence of the disulfide bonds in FU-3ds and NF-3ds simulations the stability induced byglycosylation was not prominent because of the rigid structure adopted by the polypeptide backbone. The effect of glycosylation on the stability of non-native PMP-C lacking disulfide bonds is more prominent in Figure 4.2b, which shows the mean RMSD values obtained from averaging over multiple trajectories. As can be seen from Figure 4.2b for the disulfide bonds containing

Figure 4.4
A cartoon presentation of superposition of native PMP-C (green) with the most deviated structure (cyan) of the non-native PMP-C (NF-0ds) extracted from the simulations. The disulfide bonds in the native (NF-3ds) PMP-C structure have been shown as sticks. The terminal residues have been labeled to indicate their relative spatial positions.

native forms FU-3ds and NF-3ds, the two trajectories almost superimpose indicating that glycosylation does not significantly alter the backbone structure or the stability of the native PMP-C. However, in case of non-native PMP-C the trajectories for FU-0ds and NF-0ds converge to an RMSD of 4Å, while the corresponding RMSD values for native PMP-C simulation is 2Å. This indicates that loss of disulfide bonds makes PMP-C more flexible. Comparison of the FU-0ds and NF-0ds trajectories indicate that, while FU-0ds trajectory show an RMSD lower than 4Å during most part of the 12ns simulation, the non-fucosylated PMP-C attains a peak RMSD of 5Å. These results clearly indicate that, in absence of disulfide bonds the glycosylation imparts additional stability to the PMP-C structure. Figure 4.3 shows the radius of gyration plots obtained from averaging over multiple simulations. As can be seen from Figure 4.3, there are no appreciable differences between the R_g values for the four different forms of PMP-C except that higher fluctuations are seen in the R_g values for PMP-C structures lacking disulfide bonds.

Since, RMSD plots indicated that maximum deviations from the native PMP-C NMR structure was observed in absence of fucosylation and disulfide bonds, we extracted the relevant structures from the NF-0ds trajectory and superposed it on the NMR structure of native PMP-C. As can be seen from Figure 4.4 most striking changes occurred in the terminal regions of the peptide whereas the core β-sheet regions were found to be similar in both structures. This indicates that presence of fucose moiety and disulfide bonds help in reducing the flexibility of the terminal region

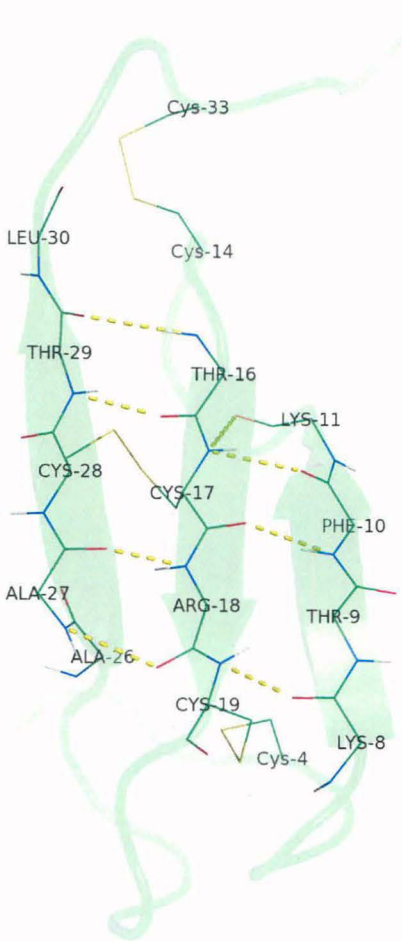


Figure 4.5
The intra-sheet (inter-strand) backbone hydrogen bonds, calculated using a distance cut off of 3.5Å for the backbone carbonyl oxygen (O) and nitrogen atoms (N), have been depicted for PMP-C. Seven of the hydrogen bonds are clearly visible whereas the hydrogen bond between Lys 11 (O) and Cys 17 (NH) appears hidden behind the cartoon depiction of the PMP-C backbone.

Table 4.2

Sr No	Donor residue (O)	Acceptor residues (@NH)	NMR structure Distance	Average duration (ns)			
				FU-3ds	NF-3ds	FU-0ds	NF-0ds
1	8	19	2.7	11	11.8	12	11.5
2	10	17	2.9	10.7	9.2	11.3	11.9
3	11	17	3.5	0.0	0.0	0.1	0.0
4	16	29	2.7	0.0	0.0	7.1	7.2
5	17	10	3.1	11.9	11.2	11.9	11.9
6	18	27	2.9	8.5	8.2	8.8	10.8
7	27	18	2.5	8.5	6.2	9.3	10.3
8	29	16	3.5	0.0	0.0	0.0	1.6

Table 4.2
The average occurrence of the hydrogen bonds, calculated in the β-sheet of the NMR structure of PMP-C using a distance cut off of 3.5Å (ref. Fig. 4.5), during the four types of simulations viz. fu-3ds, nf-3ds, fu-0ds and nf-0ds.

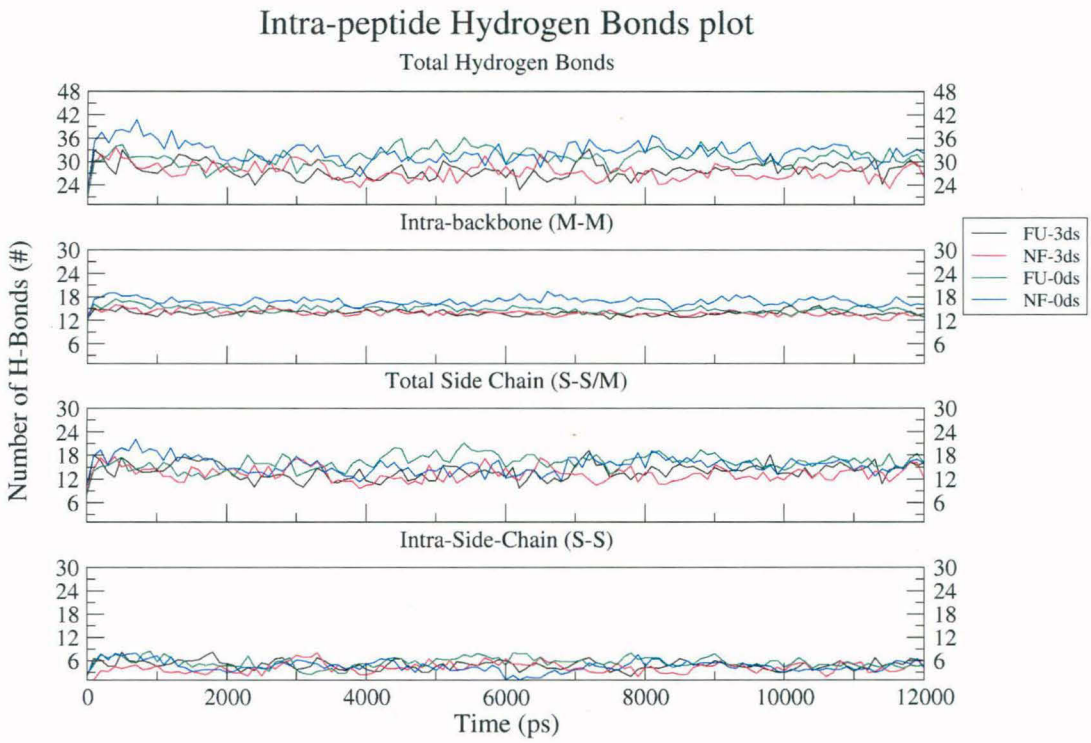


Figure 4.6
The average number of total intra-peptide hydrogen bonds, obtained from the four simulations, has been plotted for each of the fucosylated and non-fucosylated forms of PMP-C. Different trajectories have been colored differently. The average values for total intra-peptide (top panel), total intra backbone (M-M; second from top), total side-chain (S-S/M; third from top) and intra side-chain (S-S; bottom panel) hydrogen bonds for various simulations have been plotted.

and make it more ordered, while the β sheet core of PMP-C maintains the similar structure because of the presence of inter strand hydrogen bonds. Therefore, we proceeded to analyze in details the intra molecular hydrogen bonds in PMP-C.

4.3.2 Analysis of hydrogen bonds

The NMR structure of PMP-C which was the starting conformation for our MD simulations, was found to have eight β-sheet carbonyl oxygen and nitrogen atom pairs within a distance cut off of 3.5 Å and therefore, capable of forming inter strand hydrogen bonds. We monitored the breaking and formation of these eight hydrogen bonds in all the MD trajectories. Table 4.2 shows the average period of occurrence for these eight hydrogen bonds calculated from each of the four trajectories.As can be seen only five out of the eight hydrogen bonds persisted for higher than 7ns in the disulfide bonded PMP-C structures, while six out of those eight hydrogen bonds persisted in non-native PMP-C structure lacking disulfide bonds. The extra hydrogen bond present in non-native form of PMP-C involved oxygen of Thr 16 and NH group of Thr 29. Interestingly, this hydrogen bond was

present for longer than 7ns in all eight of simulation of PMP-C lacking disulfide bonds. Since, the disulfide bond between Cys 17 and Cys 28 is immediately adjacent to this hydrogen bond (Figure 4.5), it is possible that presence of this disulfide bond alters the backbone conformation of the adjacent residues and this leads to the disappearance of this hydrogen bond. The two hydrogen bonds which were missing in both the native as well as non-native PMP-C involved residues Lys 8, Lys 11 and Thr 29 which are present towards terminal regions of the β strands and flexibility of the terminal regions could be the reason for absence of these two hydrogen bonds during the dynamics. Thus, our inter-strand hydrogen bond analysis indicates that, neither glycosylation nor presence of disulfide bonds affect the inter strand hydrogen bonds in PMP-C structure in a significant way. All the four different forms of PMP-C have similar inter strand hydrogen bonds, except that presence of Cys 17-Cys 28 disulfide bond leads to disappearance of hydrogen bond between NH of Thr 29 and O of Thr 16. Figure 4.6 shows the variation of all the different types of intra molecular hydrogen bonds involving main chain and side chain atoms in the combined MD trajectories of various forms of PMP-C. The peak, mean and standard deviations in the numbers of different types of hydrogen bonds for individual trajectories are listed in Table 4.3. The detailed analysis of all these intra molecular hydrogen bonds in various forms of PMP-C also indicated that all four forms of PMP-C had very similar number of intra molecular hydrogen bonds. Here also, it was observed that upon loss of the disulfide bonds there was an increase in both hydrogen bonds involving main chain (~2 H-bonds) and hydrogen bonds involving side chains (~1 hydrogen bond) leading to a difference of ~3 hydrogen bonds in total intra molecular hydrogen bonds. Thus, our analysis suggests that both fucosylated as well as non-fucosylated forms of non-native PMP-C lacking disulfide bonds have higher number of hydrogen bonds compared to their native counter parts having disulfide bonds.

4.3.3 Glycosidic dihedral angle and fucose-peptide interactions

We also analyzed the variation of the glycosidic dihedral angle (θ_g) in various simulations for fucosylated native (FU-3ds) PMP-C and fucosylated non-native (FU-0ds) PMP-C which lack disulfide bonds. Figure 4.7a shows the variations of θ_g over the various trajectories for FU-3ds and Fu-0ds, while distributions of the corresponding glycosidic dihedral angles are shown in Figure 4.7b. For FU-3ds, the mean value of θ_g ranged between -74.5° and -73.5° in various simulations giving rise to a combined mean value of -74.0±12.1°. For FU-0ds the mean of θ_g ranged between -78.9° to -86.4° with a combined mean value of -83±14° (Figure 4.7b). The plot for θ_g values and their distribution (Figure 4.7) clearly showed that fucose moiety in both native and non-native

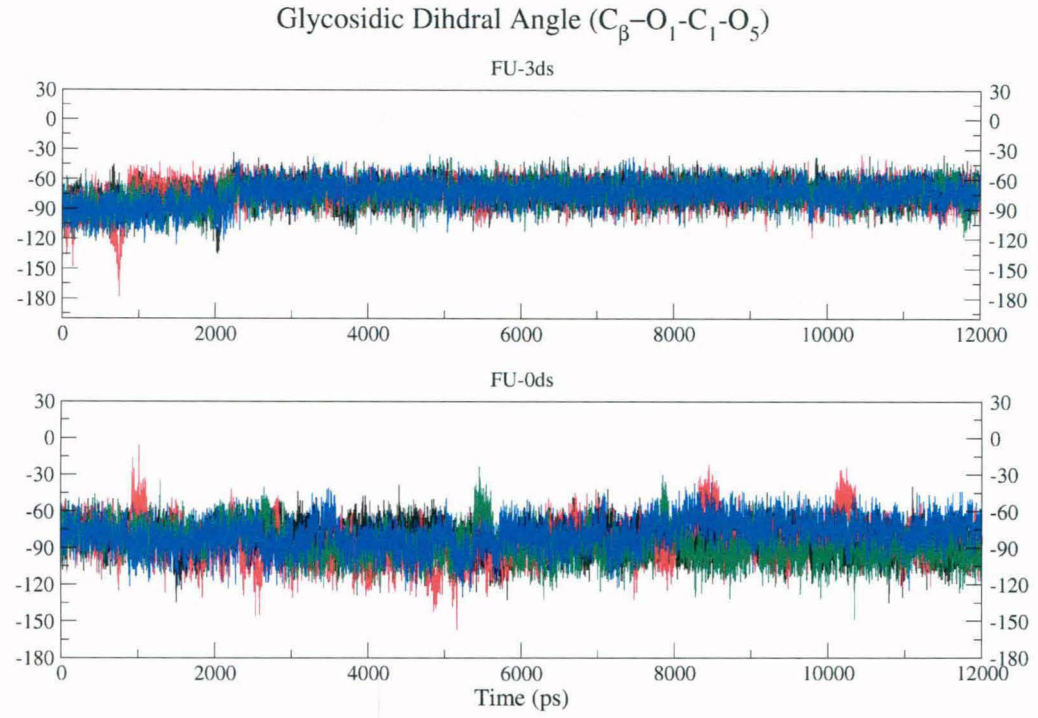
		Simulation 1	Simulation 2	Simulation 3	Simulation 4	Combined
TOT	FU-3ds	37.0 (26.2±4.3)	38.0 (28.0±4.0)	38.0 (27.2±3.4)	40.0 (28.2±4.5)	40.0 (27.4±4.2)
	NF-3ds	35.0 (26.0±3.2)	38.0 (28.3±4.0)	45.0 (27.6±5.2)	38.0 (27.0±3.9)	45.0 (27.2±4.2)
	FU-0ds	38.0 (26.8±3.9)	42.0 (33.6±4.3)	39.0 (28.9±4.2)	39.0 (28.0±3.9)	42.0 (29.3±4.8)
	NF-0ds	44.0 (32.9±4.0)	49.0 (36.0±4.4)	38.0 (29.5±3.4)	42.0 (30.3±4.4)	49.0 (32.2±4.8)
M-M	FU-3ds	17.0 (11.4±1.8)	18.0 (14.5±1.6)	18.0 (15.1±1.1)	18.0 (14.7±1.2)	18.0 (13.9±2.1)
	NF-3ds	16.0 (13.0±1.3)	18.0 (14.3±1.2)	18.0 (13.8±1.4)	18.0 (15.1±1.6)	18.0 (14.1±1.6)
	FU-0ds	20.0 (14.7±2.3)	20.0 (15.7±1.7)	19.0 (13.6±1.8)	19.0 (15.4±1.5)	20.0 (14.9±2.0)
	NF-0ds	20.0 (16.1±1.4)	24.0 (20.3±2.1)	22.0 (16.2±1.6)	19.0 (15.0±1.9)	24.0 (16.9±2.7)
M/S-S	FU-3ds	24.0 (14.8±4.1)	23.0 (13.5±3.6)	22.0 (12.1±3.3)	24.0 (13.1±4.0)	24.0 (13.4±3.9)
	NF-3ds	22.0 (13.0±3.0)	25.0 (13.9±3.7)	29.0 (13.8±4.6)	21.0 (11.9±3.3)	29.0 (13.2±3.8)
	FU-0ds	21.0 (12.1±3.1)	25.0 (17.8±3.6)	24.0 (15.2±4.0)	21.0 (12.6±3.4)	25.0 (14.4±4.2)
	NF-0ds	28.0 (16.8±3.8)	28.0 (15.8±4.0)	22.0 (13.3±2.8)	25.0 (15.2±3.6)	28.0 (15.3±3.8)
S-S	FU-3ds	14.0 (6.4±2.7)	11.0 (4.4±2.5)	12.0 (4.5±2.4)	12.0 (3.2±2.4)	14.0 (4.6±2.8)
	NF-3ds	12.0 (3.7±2.5)	11.0 (4.7±2.4)	13.0 (4.1±2.4)	10.0 (4.2±2.3)	13.0 (4.2±2.4)
	FU-0ds	8.0 (3.1±2.0)	13.0 (5.8±2.5)	10.0 (4.5±2.8)	10.0 (3.2±1.8)	13.0 (4.2±2.5)
	NF-0ds	18.0 (5.0±3.1)	12.0 (5.0±2.3)	10.0 (3.7±2.2)	13.0 (4.0±2.9)	18.0 (4.4±2.7)

Table 4.3
The hydrogen bonds values for the total intra-peptide, M-M, S-M/S and M-sol hydrogen bonds from different simulations denoting different conditions have been tabulated. The values contain three entries viz. peak, mean and standard deviation for each of the four simulations [format: peak (mean ± standard deviation)].. The last column shows values for hydrogen bonds where the three entries have been calculated after combining all the values from all four simulations for that particular condition.

PMP-C had similar orientations and the mean values obtained from simulations were close to the experimental value of -57.9±0.6° (Mer *et al.*, 1996b).

Since, the glycosidic dihedral angle in fucosylated PMP-C was restricted to a value around -80°, we wanted to investigate if interactions between the fucose moiety and specific residues of the peptide locked the fucose moiety in an essentially fixed orientation. The starting NMR structure of PMP-C used in our simulation had hydrophobic interactions between methyl groups of Thr-16 and fucose ring (Mer *et al.*, 1996b). Therefore, we have monitored the distance between the center of masses (COM) of the methyl groups of Thr-16 (T^{16}-CH_3) and fucose (Fuc^{37}-CH_3) in various conformations sampled in FU-3ds trajectories. As can be seen from Figure 4.8, the initial distance

a)



b)

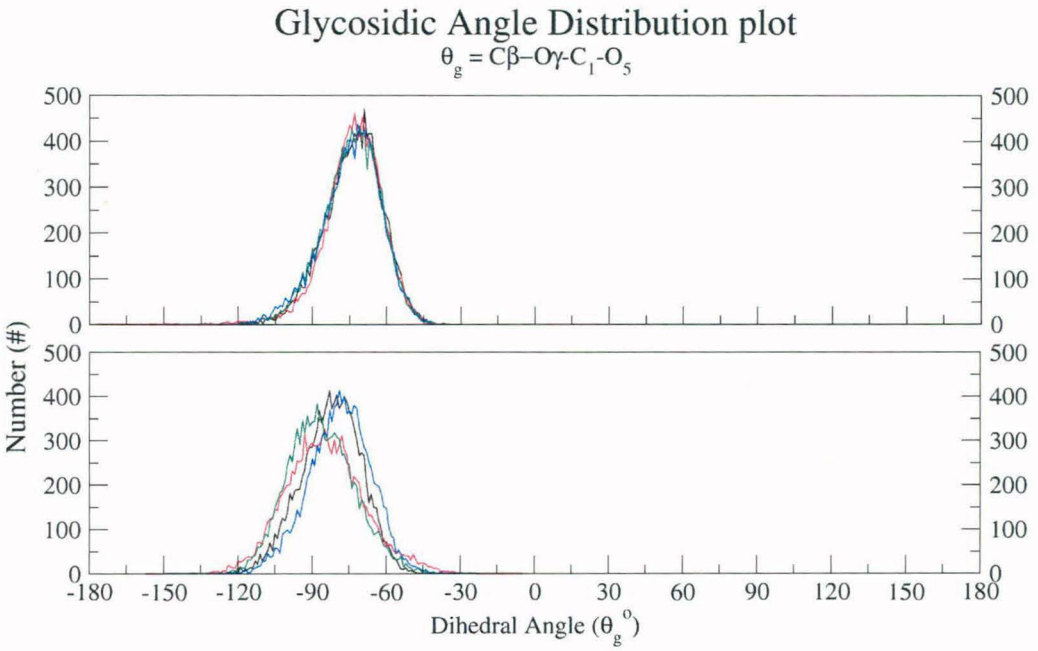


Figure 4.7

a) The glycosidic dihedral angle (C_β-O_1-C_1-O_5) for 3ds and 0ds fucosylated PMP-C has been plotted. The dihedral angles from different trajectories have been colored differently.

b) The glycosidic dihedral angle distribution plot showing the population of each dihedral value. It can be seen that in fu-3ds simulations, the values from the four simulations are highly similar. This is not the case in fu-0ds simulations of PMP-C.

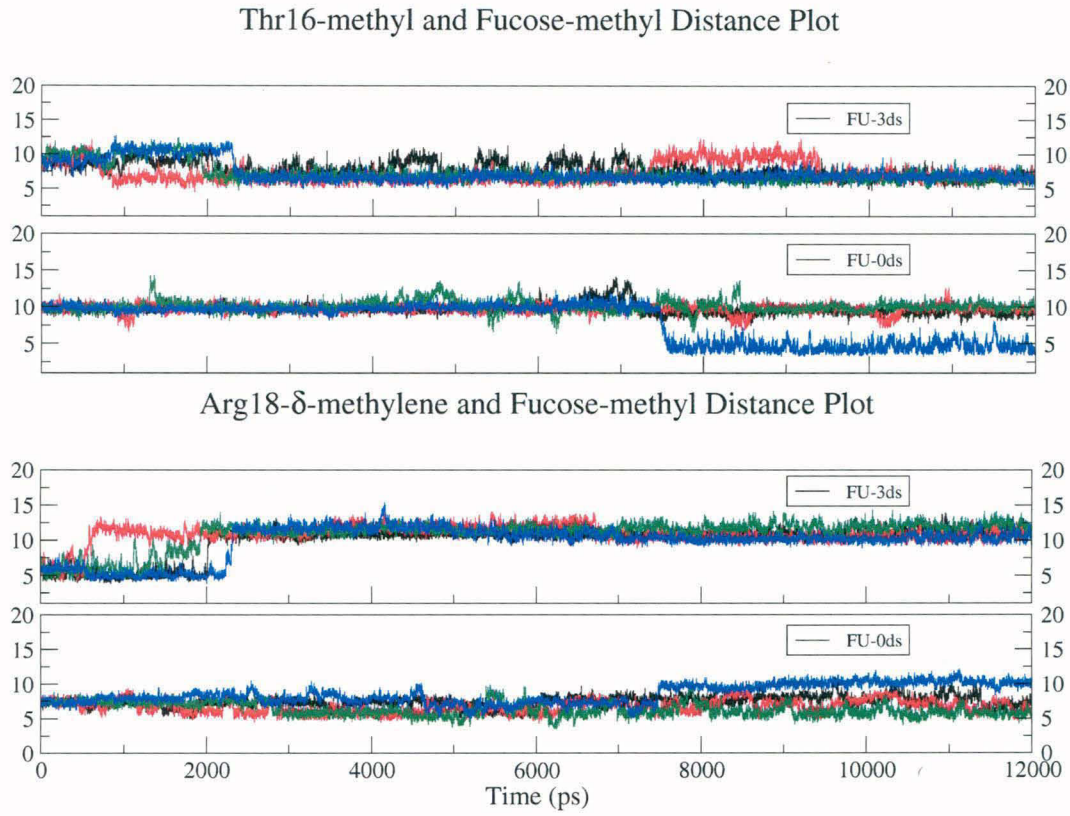


Figure 4.8
The distance between the center of masses (COM) of methyl groups of fucose and Thr16 (top panel) and δ-methylene group of Arg-18 (bottom panel) have been plotted for the FU-3ds and FU-0ds trajectories. The distance values from different trajectories have been colored differently.

		Simulation 1	Simulation 2	Simulation 3	Simulation 4	Combined
T16-Fuc methyl distance	FU-3ds	11.7 (7.7±1.1)	12.4 (7.3±1.4)	12.4 (7.3±1.4)	12.8 (7.4±1.5)	12.8 (7.4±1.4)
	FU-0ds	14.1 (9.8±0.7)	12.8 (9.8±0.6)	14.3 (10.1±0.8)	12.3 (8.0±2.6)	14.3 (9.4±1.7)
R18-Fuc methyl distance	FU-3ds	13.9 (10.0±2.2)	14.1 (10.9±1.4)	14.4 (10.9±2.0)	15.4 (9.8±2.3)	15.4 (10.4±2.1)
	FU-0ds	10.9 (7.4±0.8)	9.5 (6.8±0.9)	10.1 (6.3±1.0)	12.1 (8.6±1.4)	12.1 (7.3±1.3)
Glycosidic Dihedral angle	FU-3ds	-134.5 (-73.5±11.9)	-177.3 (-73.9±12.7)	-120.3 (-74.3±11.7)	-125.3 (-74.5±12.2)	-177.3 (-74.0±12.1)
	FU-0ds	-134.8 (-82.4±12.2)	-156.6 (-84.8±16.3)	-148.5 (-86.4±13.7)	-130.1 (-78.9±12.6)	-156.6 (-83.1±14.1)

Table 4.4
The distance between the COM of methyl groups of Thr16 and Arg18 with respect to the COM of methyl group of fucose from different simulations have been tabulated. The values contain three entries *viz.* peak, mean and standard deviation for each of the four simulations [format: peak (mean ± standard deviation)]. The last column tabulates the combined average of the respective COM distances.

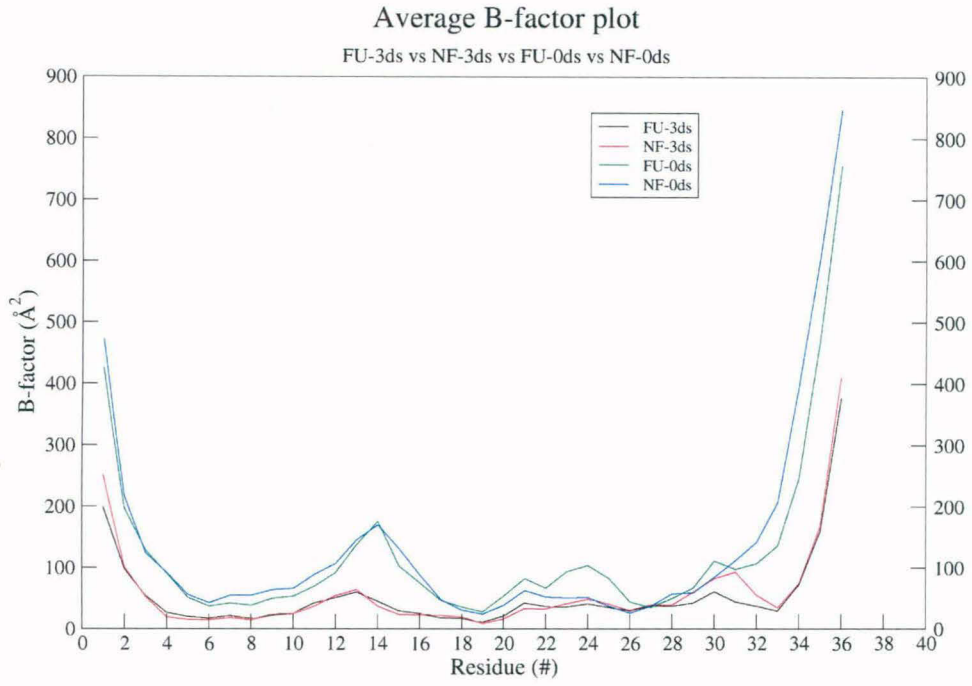


Figure 4.9
The average B-factors have been plotted for both fucosylated (3ds, black; 0ds, green) and non-fucosylated (3ds, red; 0ds, blue) forms of PMP-C. It can be clearly seen that the native (3ds) forms of PMP-C differ in their average B-factor values in the 29-32 region.

between the methyl groups was more than 8Å, but after 2 ns of simulation the two methyl groups came as close as 6Å and remained in contact during the most part of the remaining simulation period. This suggests that the hydrophobic interaction between these methyl groups was stable during our simulations. In FU-0ds simulations, the distance between these two methyl groups was about 10Å throughout the simulation period in three of the four simulations. However, in one of the simulations, this distance remained 10Å for initial 7 ns but dropped to ~4.5Å for the later part. We also analyzed the distance between the COM of R^{18}-$C_\delta H_2$ and Fuc^{37}-CH_3 **(**Figure 4.8**).** It was observed that, in FU-3ds simulations starting with a distance of ~6Å, the two groups moved further apart as the simulations progressed. Within 2 ns of simulation, they moved ~12Å apart and stayed there for the rest of the simulation period. The mean distance ranged between 10 and 11Å with a peak of 15Å. This indicates that in fucosylated PMP-C the fucose moiety prefers to interact with Thr 16 rather than Arg 18. Table 4.4 lists peak, mean and standard deviation values for the variation of the distance between fucose methyl group and Thr 16 as well as Arg 18. As can be seen from Figure 4.8 and Table 4.4, in some of the FU-0ds simulations a reverse trend is observed i.e. Arg 18 to fucose methyl distance was lower compared to Thr 16 to fucose distance (Figure 4.8).

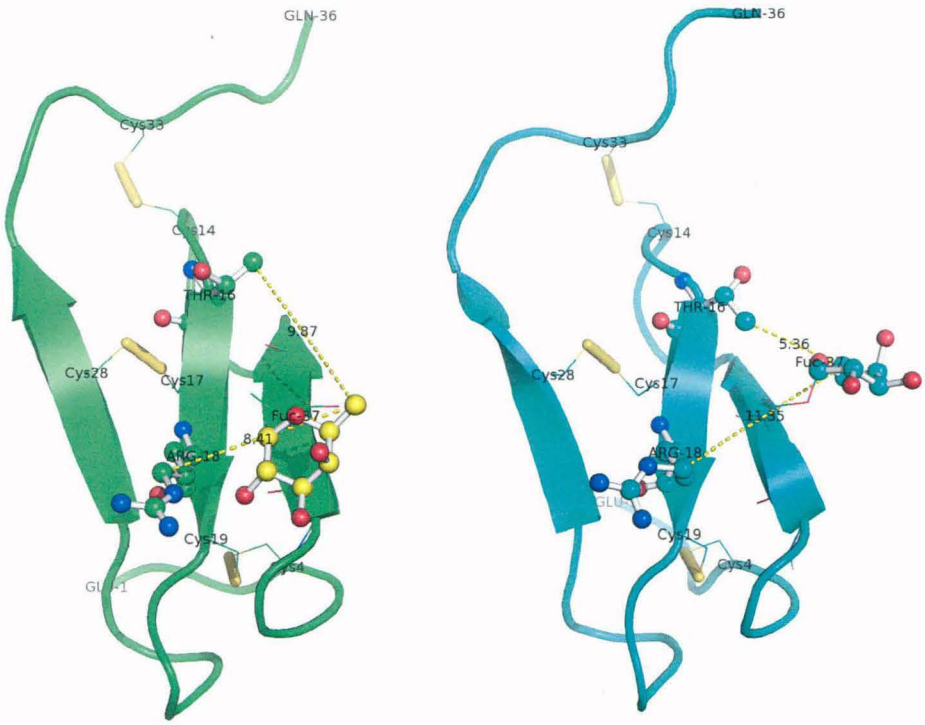


Figure 4.10
Cartoon depiction of the NMR structure of PMP-C and the structure obtained after 12ns of simulations for the fucosylated native (fu-3ds) form of PMP-C. Thr-16, Thr-29, and Fucose-37 residues have been shown in ball and stick representation using PyMol.

4.3.4 Effect of glycosylation on flexibility of PMP-C

In order to identify regions of PMP-C which show higher mobility, we computed theoretical B-factors for each residue in various forms of PMP-C from different molecular dynamics trajectories. Figure 4.9 shows B-factor plots for four different forms of PMP-C obtained by averaging over multiple simulations for the same structures. As can be seen from Figure 4.9, terminal residues 1-2 and 35-36 have very high B-factors exceeding 100Å^2 in all four forms of PMP-C. For all the residues of PMP-C, the non-native forms lacking disulfide bonds have higher B-factors compared to native forms which contain three disulfide bonds between the beta strands (Figure 4.9). This result is consistent with the earlier observation from RMSD and R_g plots which suggested higher flexibility for non-native forms of PMP-C. Comparison of the B-factor plots for the fucosylated and non-fucosylated forms of PMP-C shows certain interesting features. As can be seen from Figure 4.9, both the native forms, FU-3ds and NF-3ds, have very similar B-factors for all the residues except for the amino acid stretch 29-32. The fucosylated form has lower B-factors for the residue stretch 29-32 compared to the non-fucosylated form. Interestingly, this region interacts with the active site of the protease when PMP-C functions as a protease inhibitor (Kellenberger *et al.*, 1995). According to Kellenberger *et al* (Kellenberger *et al.*, 1995), Leu30 most probably

interacts with the P1 pocket of the protease and L30V mutation abrogated α-chymotrypsin inhibition activity of the mutant peptide. Thus, the lower mobility of the residue stretch 29-32 in the fucosylated PMP-C compared to the non-fucosylated counterpart correlates well with the functional differences between fucosylated and non-fucosylated PMP-C. Hence, our simulations provide a structural rationale for the stability and functional differences between fucosylated and non-fucosylated PMP-C.

We also analyzed how exactly the glycosylation at Thr 9 results in a mobility difference in the amino acid stretch 29-32 which is far apart from the site of glycosylation both in sequence and also in three dimensional structure. As can be seen from Figure 4.10 the disulfide bonds Cys17-Cys28 and Cys14-Cys33 covalently connect the residue stretch 29-32 to the central β sheet of PMP-C and side chains of Thr-16 and Arg-18 of the central β sheet are in contact with the fucose moiety. Thus, using a novel network of side chain interactions and disulfide bonds, fucosylation at Thr 9 alters the flexibility of the stretch 29-32 located at a distal site. It may be noted that FU-0ds and NF-0ds simulations do not show any significant differences in the mobility of the stretch 29-32 (Figure 9). This further establishes the role of disulfide bonds in the network of interactions which connect the site of glycosylation to the residue stretch 29-32 which interacts with the P1 site of the protease when PMP-C functions as a protease inhibitor. Thus, our molecular dynamics simulation studies have provided for the first time mechanistic details of the effect of glycosylation on the function of PMP-C.

4.4 CONCLUSIONS

We have carried out explicit solvent molecular dynamics simulations for 12 ns on fucosylated and non-fucosylated forms of native PMP-C having disulfide bonds and also on fucosylated and non-fucosylated forms of non-native PMP-C lacking disulfide bonds. Each simulation has been repeated four times to ensure consistency of the results and also to enhance the conformational sampling. The various conformational parameters for different forms of PMP-C have been calculated from individual MD trajectories as well as by averaging over multiple trajectories for the same molecule. Thus, effective simulation time for each molecule is 48 ns. Analysis of the various MD trajectories indicated that the fucosylated and non-fucosylated forms of disulfide bonded native PMP-C had very similar conformations. The differences between fucosylated and non-fucosylated forms of the peptide became more pronounced when the non-native forms of PMP-C, lacking disulfide bonds, were analyzed. The non-fucosylated form of PMP-C without disulfide bonds showed larger deviations from the starting structure than the fucosylated

form. However, analysis of the structural changes seen in the trajectories for non-fucosylated PMP-C lacking disulfide bonds revealed that structural deviations were restricted to the terminal regions while core β sheet retained its hydrogen bonded structure even in the absence of disulfide bonds as well as fucosylation. Fucosylation helps in lowering the flexibility of the terminal regions and hence fucosylated form of non-native PMP-C does not show large deviations from the structure of the disulfide bonded native form. Since, disulfide bonds can form only after the formation of β sheet core, based on the results of our simulations it is tempting to speculate that fucosylation helps in folding and once the molecule is folded the disulfide bonds lock it in a rigid conformation which is required for its function. Our detailed analysis of the persistence of inter β strand hydrogen bonds during the simulation time period indicated that out of the eight hydrogen bonds between the carbonyl and amino groups of the β sheet core, six hydrogen bonds were present during most part of the 12 ns trajectories in non-native forms of PMP-C both in presence and absence of fucose moiety. Only the two terminal hydrogen bonds broke during the simulations. On the other hand, only five out of these eight hydrogen bonds were present in the native forms of PMP-C. This result was surprising as one would *a priori* expect the disulfide bonded β sheet core to have more number of hydrogen bond than the non-native form lacking disulfide bonds. Detailed analysis of the trajectories indicated that presence of a disulfide bond between Cys 17 and Cys 28 alters the backbone conformation of the adjacent residues and this leads to the disappearance of the inter strand main chain hydrogen bond between Thr 16 and Thr 29. Analysis of theoretical B-factor or mobility of different regions of PMP-C in various forms showed certain features with interesting implications for function of PMP-C. The fucosylated form of disulfide bonded native PMP-C had lower B-factor compared to its non-fucosylated counterpart for the residue stretch 29-32, which interacts with the active site of the protease when PMP-C functions as a protease inhibitor. On the other hand, fucosylation did not affect mobility of this stretch in the non-native PMP-C lacking disulfide bonds. Our analysis revealed that disulfide bonds Cys17-Cys28 and Cys14-Cys33 covalently connect the residue stretch 29-32 to the central β sheet of PMP-C and using a novel network of side chain interactions and disulfide bonds fucosylation at Thr 9 alters the flexibility of the stretch 29-32 located at a distal site. Thus, our simulations have provided for the first time a molecular basis for the role of fucosylation in structure and function of PMP-C as a protease inhibitor and explain how presence of disulfide bonds between conserved cysteines and fucosylation help in enhancing its function as a protease inhibitor.

Chapter 5

Summary and Conclusions

Lectins are a class of proteins which bind sugar molecules and function in crucial events like host-pathogen interactions, protein folding pathways, *etc.* The main objective of this thesis has been to analyze the structural stability and protein-sugar interactions in various legume lectins using molecular dynamics simulations. Specifically, we have analyzed the role of metal ions and glycosylation in the structural stability of legume lectins *viz.* ConA and EcorL. Both these lectins are globular proteins and the sugar moiety is known to play a crucial role in the structural stability, conformational flexibility or substrate binding of these two proteins. Therefore, we have used all atom MD simulations in explicit solvent environment to understand how interactions with the glycan moiety affect the structure and dynamics of the polypeptide chain. As mentioned before, MD simulations have been instrumental in obtaining insights beyond the realms of experimental reach. For example, the atomic details of ConA demetalation have eluded X-ray crystallography despite numerous attempts. Similarly, the role of glycosylation in enhancing the structural stability of EcorL had not been deciphered. We have delved on these objectives and found several interesting facts which seem to answer the above mentioned queries. Another interesting question that we analyzed in this thesis was the structural and functional differences among the glycosylated and non-glycosylated forms of PMP-C, a serine protease inhibitor. The results obtained from the computer simulations help establish a direct link between glycosylation of this serine protease inhibitor and the effects of glycosylation on its function. A summary of the results obtained from the molecular dynamics simulations performed on the legume lectins (ConA and EcorL) and the serine protease inhibitor PMP-C are described below.

Role of Metal Ions in Substrate Recognition and Stability of ConA

Despite the importance of metal ions in controlling structural stability and substrate binding of ConA, no atomically detailed picture was available for the conformational dynamics of the folding of the ion binding loop and the substrate binding events. Crystallographic studies did not provide adequate information for understanding the structural basis of demetalization and consequent abolition of sugar binding abilities of lectins due to difficulty in interpretation of electron density for ion binding loop in demetalized ConA. We have performed 3ns explicit solvent MD simulations using crystal structure of ConA trimannoside complex. To understand the role of metal ions in substrate binding and structural stability of ConA simulations were carried out in the presence and absence of metal ions and trimannoside substrate. These simulations have also been carried

out for the monomeric and dimeric forms of ConA to understand whether oligomerization affects the behavior of ConA under otherwise identical conditions of simulations like demetalization.

The observation that computed B-factors from our simulations of native ConA tetramer were in close agreement with experimentally determined B-factors, validated our simulations. Comparative analysis of the total of twelve 3ns MD trajectories has given novel insights into the dynamic aspects of structural stability and substrate recognition of ConA. The core of ConA tetramer, comprising of β-sheets, was found to be a tightly packed structure as it remained intact even upon demetalization. The metal ions were found to play a major role in stabilizing the conformation of ion binding loop of ConA and in binding of the trimannoside. In contrast to metal ion binding, substrate binding does not seem to affect the structure of the loop regions appreciably. The simulations carried for monomeric and dimeric ConA suggested both these forms of ConA have stable structures and a similar substrate binding ability. This indicated that oligomerization has no effect on the metal ion binding phenomenon in legume lectins or at least in ConA.

Native conformation of the ion binding loop is stabilized by coordination of two metal ions with a cluster of negatively charged amino acid side chains. In absence of metal ions, due to a strong electrostatic repulsion, the folded conformation of the ion-binding loop becomes energetically unfavorable. Hence, the ion-binding loop exhibited a very high mobility upon demetalation. In fact, it moved away from the protein with certain residues were observed to be shifted as far as 17Å with respect to their spatial position in the crystal structure. Consequently, the substrate binding site is disrupted and the trimannoside could be observed drifting from substrate binding site. The interactions between metal ions and certain side chains on the ion-binding loop help in decreasing the thermal motion of the ion binding loop. This forces the loop into a conformation requisite of sugar binding. Possibly, during this process, the orientation of side chains of residues on the loop is also altered for optimal presentation to the sugar atoms. Not only the results from these simulations supported the long known experimental fact that ion binding is essential for the sugar binding, our studies have provided novel insights into the dynamics of interplay between metal and sugar binding to ConA for the first time.

Role of glycosylation in folding and structural stability of the EcorL

EcorL is another member of legume lectin family having the typical legume lectin fold as ConA. However, it differs from ConA in being glycosylated and having an altered

mode of quaternary association, also known as the 'hand shake' mode of dimerization. This mode of dimerization was not observed in legume lectins before crystal structure of EcorL was determined. The covalently bound oligosaccharide was proposed as the reason for the altered mode of oligomerization. However, as the crystal structures for other leguminous lectins and recombinant non-glycosylated EcorL (rEcorL) were solved, it became clear that the oligosaccharide *per se* could not have been responsible for the hand shake mode of dimerization. Further, compared to rEcorL, EcorL was observed to be structurally more stable and display an altered unfolding pathway. In view of these facts, it was pertinent to ask what role glycosylation played in the structural stability and folding of EcorL. Therefore, we have performed four MD simulations each of 5ns for both EcorL and rEcorL. Simulations were carried out in explicit solvent environment at room temperature and atmospheric pressure.

Both glycosylated and non-glycosylated forms of EcorL were observed to have similar structural properties except for a lesser non-polar solvent accessible surface area (ASA) of EcorL compared to rEcorL. Analysis of the results obtained from simulations indicated that the dynamic view of interactions between protein and oligosaccharide residues is entirely different from the static picture observed in the crystal structure. The oligosaccharide moiety had dynamically stable interactions with Lys 55 and Tyr 53, both of which are separated in sequence from the site of glycosylation, Asn 17. It is possible that glycosylation helps in forming long range contacts between amino acids which are separated in sequence and thus provides a folding nucleus around these interactions. The results from our simulations supported previous experimental observations on the stability of EcorL and oligosaccharide. Our simulations not only revealed the conformations sampled by the oligosaccharide, but also provided novel insights into possible molecular mechanisms by which glycosylation can help in folding of the glycoprotein by formation of folding nucleus involving specific contacts with the oligosaccharide moiety.

Exploring effects of glycosylation on small peptides

Serine proteases are an important family of proteins and their inhibitors can be of immense importance as drugs. The imbalance between a protease and its inhibitor is critical for the survival of the organism. Some inhibitors, like *pars intercerebralis* major peptide-C (PMP-C), exist in glycosylated and non-glycosylated forms. They display similar structural but non-identical functional properties. Fucosylated PMP-C was observed to show enhanced structural stability compared to the non-fucosylated form. Therefore, to understand the role of

glycosylation and disulfide bonds in structure and function of small peptides, we have performed 12 ns long multiple MD simulations for both fucosylated and non-fucosylated forms of PMP-C both in presence and absence of disulfide bonds.

The results from our simulations on the native forms of PMP-C, having all the three disulfide bonds intact, showed that there was no structural change induced in PMP-C as a consequence to fucosylation. Both fucosylated and non-fucosylated forms of PMP-C were observed to have a rigid structure. The results from simulations on the non-native forms of PMP-C, lacking disulfide bonds, clearly showed that fucosylated PMP-C had less structural deviation from its starting structure compared to its non-fucosylated counterpart. This proved that fucosylation indeed enhanced the structural stability of the peptide. The core β sheet of the peptide contains several inter-β-strand hydrogen bonds. Almost all of them remained intact during simulations of both native and non-native forms of PMP-C, depicting that the β sheet core is very stable structure both in presence or absence of the disulfide bonds. Interestingly, our simulations also revealed that fucosylation of PMP-C led to a decreased thermodynamic flexibility in the residue stretch 29-32 of PMP-C which is known to interact with the active site of the target proteases. On the other hand, fucosylation did not affect mobility of this stretch in the non-native PMP-C lacking disulfide bonds. Our analysis revealed that disulfide bonds Cys17-Cys28 and Cys14-Cys33 covalently connect the residue stretch 29-32 to the central β sheet of PMP-C and using a novel network of side chain interactions and disulfide bonds fucosylation at Thr 9 alters the flexibility of the stretch 29-32 located at a distal site. Thus, our simulations have provided for the first time a molecular basis for the role of fucosylation in structure and function of PMP-C as a protease inhibitor and explain how presence of disulfide bonds between conserved cysteines and fucosylation help in enhancing its function as a protease inhibitor.

Conclusions

Using the results obtained from the MD simulations of ConA, EcorL and PMP-C, following observations/conclusions could be made:

1. Metal ions stabilize the ion binding loop in lectins in a conformation which is conducive to the substrate binding. Upon demetalation the ion binding loop becomes thermally unstable and moves far from its native conformation disrupting the substrate binding site. This leads to the loss of ability of the lectin to bind substrate sugars.

2. MD simulations carried out for understanding the effect of glycosylation on structure and stability of EcorL revealed that the observed differences in the thermodynamic stability between glycosylated and non-glycosylated EcorL arise from lower non-polar solvent accessible surface area in glycosylated EcorL. Our simulations on EcorL have also provided novel insights into possible molecular mechanisms by which glycosylation can help in folding of EcorL by formation of folding nucleus involving specific contacts with the oligosaccharide moiety.
3. Both forms of PMPC are very stable and structurally similar. Fucosylation stabilizes the region of PMPC interacting with the active site of proteases, possibly leading to functional differences. A network of disulfide bridges and the intra-peptide interactions might be responsible for propagating the effects of fucosylation from the site of glycosylation to the distal regions.

Publications

The work reported in this thesis has resulted in the following publications:

1. Kaushik, S., Mohanty, D. and Surolia, A. (2009) **The role of metal ions in substrate recognition and stability of concanavalin A: A molecular dynamics study.** *Biophys J*, **96**, 21-34.

2. Kaushik, S., Mohanty, D. and Surolia, A. (2011) **Role of glycosylation in structure and stability of Erythrina corallodendron lectin (EcorL): A molecular dynamics study.** *Protein Sci*, **20**, 465-481.

BIBLIOGRAPHY

Abe, Y., Iwabuchi, M. and Ishii, S. I. (1971) Multiple forms in the subunit structure of concanavalin A. *Biochem Biophys Res Commun*, **45**, 1271-1278.

Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J Biol Chem*, **268**, 3525-3529.

Adar, R., Richardson, M., Lis, H. and Sharon, N. (1989) The amino acid sequence of Erythrina corallodendron lectin and its homology with other legume lectins. *FEBS Lett*, **257**, 81-85.

Adar, R. and Sharon, N. (1996) Mutational studies of the amino acid residues in the combining site of Erythrina corallodendron lectin. *Eur J Biochem*, **239**, 668-674.

Adar, R., Moreno, E., Streicher, H., Karlsson, K. A., Angstrom, J. and Sharon, N. (1998) Structural features of the combining site region of Erythrina corallodendron lectin: role of tryptophan 135. *Protein Sci*, **7**, 52-63.

Agrawal, B. B. and Goldstein, I. J. (1965) Specific binding of concanavalin A to cross-linked dextran gels. *Biochem J*, **96**, 23contd-25c.

Agrawal, B. B. and Goldstein, I. J. (1967a) Physical and chemical characterization of concanavalin A, the hemagglutinin from jack bean (Canavalia ensiformis). *Biochim Biophys Acta*, **133**, 376-379.

Agrawal, B. B. and Goldstein, I. J. (1967b) Protein-carbohydrate interaction. VI. Isolation of concanavalin A by specific adsorption on cross-linked dextran gels. *Biochim Biophys Acta*, **147**, 262-271.

Agrawal, B. B. and Goldstein, I. J. (1968) Protein-carbohydrate interactin. XV. The role of bivalent cations in concanavalin A-polysaccharide interaction. *Can J Biochem*, **46**, 1147-1150.

Agrawal, B. B. L. and Goldstein, I. J. (1972) [34] Concanavalin A, the Jack Bean (Canavalia ensiformis) phytohemagglutinin
Methods in Enzymology. In: *Complex Carbohydrates Part B* (Ginsburg, V., Ed.) Academic Press, pp. 313-318.

Agrawal, P., Kumar, S. and Das, H. R. (2010) Mass spectrometric characterization of isoform variants of peanut (Arachis hypogaea) stem lectin (SL-I). *J Proteomics*, **73**, 1573-1586.

Alencar, N. M., Oliveira, R. S., Figueiredo, J. G., Cavalcante, I. J., Matos, M. P., Cunha, F. Q., Nunes, J. V., Bomfim, L. R. and Ramos, M. V. (2010) An anti-inflammatory lectin from Luetzelburgia auriculata seeds inhibits adhesion and rolling of leukocytes and modulates histamine and PGE2 action in acute inflammation models. *Inflamm Res*, **59**, 245-254.

Aliev, A. E. and Courtier-Murias, D. (2010) Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *J Phys Chem B*, **114**, 12358-12375.

Allen, N. K. and Brilliantine, L. (1969) A survey of hemagglutinins in various seeds. *J Immunol*, **102**, 1295-1299.

Ambrosi, M., Cameron, N. R. and Davis, B. G. (2005) Lectins: tools for the molecular understanding of the glycocode. *Org Biomol Chem*, **3**, 1593-1608.

Andreotti, A. H. and Kahne, D. (1993) The effects of glycosylation on peptide backbone conformation. *Journal of the American Chemical Society*, **115**, 3352-3353.

Andrew McCammon, J. (1991) Free energy from simulations : Current Opinion in Structural Biology 1991, 1 : 196-200. *Current Opinion in Structural Biology*, **1**, 196-200.

Arango, R., Rozenblatt, S. and Sharon, N. (1990) Cloning and sequence analysis of the Erythrina corallodendron lectin cDNA. *FEBS Lett*, **264**, 109-111.

Arango, R., Adar, R., Rozenblatt, S. and Sharon, N. (1992) Expression of Erythrina corallodendron lectin in Escherichia coli. *Eur J Biochem*, **205**, 575-581.

Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontiens, F. G., Rozenblatt, S. and Sharon, N. (1993) Modification by site-directed mutagenesis of the specificity of Erythrina corallodendron lectin for galactose derivatives with bulky substituents at C-2. *FEBS Lett*, **330**, 133-136.

Araujo-Filho, J. H., Vasconcelos, I. M., Martins-Miranda, A. S., Gondim, D. M. and Oliveira, J. T. (2010) A ConA-like lectin from Dioclea guianensis Benth. has antifungal activity against Colletotrichum gloeosporioides, unlike its homologues, ConM and ConA. *J Agric Food Chem*, **58**, 4090-4096.

Ashford, D., Dwek, R. A., Welply, J. K., Amatayakul, S., Homans, S. W., Lis, H., Taylor, G. N., Sharon, N. and Rademacher, T. W. (1987) The beta 1----2-D-xylose and alpha 1----3-L-fucose substituted N-linked oligosaccharides from Erythrina cristagalli lectin. Isolation, characterisation and comparison with other legume lectins. *Eur J Biochem*, **166**, 311-320.

Ashwell, G. and Harford, J. (1982) Carbohydrate-specific receptors of the liver. *Annu Rev Biochem*, **51**, 531-554.

Assreuy, A. M., Fontenele, S. R., Pires Ade, F., Fernandes, D. C., Rodrigues, N. V., Bezerra, E. H., Moura, T. R., do Nascimento, K. S. and Cavada, B. S. (2009) Vasodilator effects of Diocleinae lectins from the Canavalia genus. *Naunyn Schmiedebergs Arch Pharmacol*, **380**, 509-521.

Balamurugan, K., Gopalakrishnan, R., Raman, S. S. and Subramanian, V. (2010) Exploring the Changes in the Structure of alpha-Helical Peptides Adsorbed onto a Single Walled Carbon Nanotube Using Classical Molecular Dynamics Simulation. *J Phys Chem B*.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. and Vijayan, M. (1996) Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. *J Mol Biol*, **259**, 281-296.

Batista, F. A., Goto, L. S., Garcia, W., de Moraes, D. I., de Oliveira Neto, M., Polikarpov, I., Cominetti, M. R., Selistre-de-Araujo, H. S., Beltramini, L. M. and Araujo, A. P. (2010) Camptosemin, a tetrameric lectin of Camptosema ellipticum: structural and functional analysis. *Eur Biophys J*, **39**, 1193-1205.

Bauer, B. A., Davis, J. E., Taufer, M. and Patel, S. (2010) Molecular dynamics simulations of aqueous ions at the liquid-vapor interface accelerated using graphics processors. *J Comput Chem.*

Becker, J. W., Reeke, G. N., Jr. and Edelman, G. M. (1971) Location of the saccharide binding site of concanavalin A. *J Biol Chem*, **246**, 6123-6125.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. and Edelman, G. M. (1975) The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides.
. *Journal of Biological Chemistry*, **250**, 1513-1524.

Bennett, M. J., Schlunegger, M. P. and Eisenberg, D. (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci*, **4**, 2455-2468.

Bernal, J. D. and Fowler, R. H. (1933) A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *The Journal of Chemical Physics*, **1**, 515-548.

Bertozzi, C. R. and Kiessling, L. L. (2001) Chemical glycobiology. *Science*, **291**, 2357-2364.

Bhinge, A., Chakrabarti, P., Uthanumallian, K., Bajaj, K., Chakraborty, K. and Varadarajan, R. (2004) Accurate detection of protein:ligand binding sites using molecular dynamics simulations. *Structure*, **12**, 1989-1999.

Bhutia, S. K., Mallick, S. K. and Maiti, T. K. (2009) In vitro immunostimulatory properties of Abrus lectins derived peptides in tumor bearing mice. *Phytomedicine*, **16**, 776-782.

Bies, C., Lehr, C. M. and Woodley, J. F. (2004) Lectin-mediated drug targeting: history and applications. *Adv Drug Deliv Rev*, **56**, 425-435.

Biswas, S., Agrawal, P., Saroha, A. and Das, H. R. (2009) Purification and mass spectrometric characterization of Sesbania aculeata (Dhaincha) stem lectin. *Protein J*, **28**, 391-399.

Blakeley, M. P., Kalb, A. J., Helliwell, J. R. and Myles, D. A. (2004) The 15-K neutron structure of saccharide-free concanavalin A. *Proc Natl Acad Sci U S A*, **101**, 16405-16410.

Bode, W., Papamokos, E., Musil, D., Seemueller, U. and Fritz, H. (1986) Refined 1.2 A crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. *Embo J*, **5**, 813-818.

Bode, W., Meyer, E., Jr. and Powers, J. C. (1989) Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry*, **28**, 1951-1963.

Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem*, **204**, 433-451.

Bordes, F., Barbe, S., Escalier, P., Mourey, L., Andre, I., Marty, A. and Tranier, S. (2010) Exploring the Conformational States and Rearrangements of Yarrowia lipolytica Lipase. *Biophys J*, **99**, 2225-2234.

Bots, M., Kolfschoten, I. G., Bres, S. A., Rademaker, M. T., de Roo, G. M., Kruse, M., Franken, K. L., Hahne, M., Froelich, C. J., Melief, C. J. *et al.* (2005) SPI-CI and SPI-6 cooperate in the protection from effector cell-mediated cytotoxicity. *Blood*, **105**, 1153-1161.

Bots, M. and Medema, J. P. (2006) Granzymes at a glance. *J Cell Sci*, **119**, 5011-5014.

Bouckaert, J., Loris, R., Poortmans, F. and Wyns, L. (1995) Crystallographic structure of metal-free concanavalin A at 2.5 A resolution. *Proteins*, **23**, 510-524.

Bouckaert, J., Poortmans, F., Wyns, L. and Loris, R. (1996) Sequential structural changes upon zinc and calcium binding to metal-free concanavalin A. *J Biol Chem*, **271**, 16144-16150.

Bouckaert, J., Hamelryck, T., Wyns, L. and Loris, R. (1999a) Novel structures of plant lectins and their complexes with carbohydrates. *Curr Opin Struct Biol*, **9**, 572-577.

Bouckaert, J., Hamelryck, T. W., Wyns, L. and Loris, R. (1999b) The crystal structures of Man(alpha1-3)Man(alpha1-O)Me and Man(alpha1-6)Man(alpha1-O)Me in complex with concanavalin A. *J Biol Chem*, **274**, 29188-29195.

Bouckaert, J., Dewallef, Y., Poortmans, F., Wyns, L. and Loris, R. (2000a) The structural features of concanavalin A governing non-proline peptide isomerization. *J Biol Chem*, **275**, 19778-19787.

Bouckaert, J., Loris, R. and Wyns, L. (2000b) Zinc/calcium- and cadmium/cadmium-substituted concanavalin A: interplay of metal binding, pH and molecular packing doi:10.1107/S0907444900013342. *Acta Crystallographica Section D*, **56**, 1569-1576.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P. and Fontecilla-Camps, J. C. (1990) X-ray crystal structure determination and refinement at 1.9 A resolution of isolectin I from the seeds of Lathyrus ochrus. *J Mol Biol*, **214**, 571-584.

Bouwstra, J. B., Spoelstra, E. C., De Waard, P., Leeflang, B. R., Kamerling, J. P. and Vliegenthart, J. F. (1990) Conformational studies on the N-linked carbohydrate chain of bromelain. *Eur J Biochem*, **190**, 113-122.

Bowles, D. J., Marcus, S. E., Pappin, D. J., Findlay, J. B., Eliopoulos, E., Maycox, P. R. and Burgess, J. (1986) Posttranslational processing of concanavalin A precursors in jackbean cotyledons. *J Cell Biol*, **102**, 1284-1297.

Boyd, W. C. and Shapleigh, E. (1954) Specific Precipitating Activity of Plant Agglutinins (Lectins)
10.1126/science.119.3091.419. *Science*, **119**, 419-.

Bradbrook, G. M., Thomas Gleichmann, Stephen J. Harrop, Jarjis Habash, James Raftery, Joseph Kalb (Gilboa), Joseph Yariv, and, I. H. H. and Helliwell, J. R. (1998) X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *Journal of the Chemical Society, Faraday Transactions*, **94**, 1603-1611.

Bradbrook, G. M., Forshaw, J. R. and Perez, S. (2000) Structure/thermodynamics relationships of lectin-saccharide complexes: the Erythrina corallodendron case. *Eur J Biochem*, **267**, 4545-4555.

Brehelin, M., Boigegrain, R. A., Drif, L. and Coletti-Previero, M. A. (1991) Purification of a protease inhibitor which controls prophenoloxidase activation in hemolymph of Locusta migratoria (insecta). *Biochem Biophys Res Commun*, **179**, 841-846.

Breitfeld, P. P., Rup, D. and Schwartz, A. L. (1984) Influence of the N-linked oligosaccharides on the biosynthesis, intracellular routing, and function of the human asialoglycoprotein receptor. *J Biol Chem*, **259**, 10414-10421.

Brewer, C. F., Sternlicht, H., Marcus, D. M. and Grollman, A. P. (1973) Interactions of saccharides with concanavalin A. Mechanism of binding of alpha- and beta-methyl D-glucopyranoside to concanavalin A as determined by 13C nuclear magnetic resonance. *Biochemistry*, **12**, 4448-4457.

Brewer, C. F. and Bhattacharyya, L. (1986) Specificity of concanavalin A binding to asparagine-linked glycopeptides. A nuclear magnetic relaxation dispersion study. *J Biol Chem*, **261**, 7306-7310.

Brewer, F., Bhattacharyya, L., Brown, R. D., 3rd and Koenig, S. H. (1985) Interactions of concanavalin A with a trimannosyl oligosaccharide fragment of complex and high mannose type glycopeptides. *Biochem Biophys Res Commun*, **127**, 1066-1071.

Brooks, C. L., 3rd, Brunger, A. and Karplus, M. (1985) Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers*, **24**, 843-865.

Brown, R. D., Brewer, C. F. and Koenig, S. H. (1977) Conformation states of concanavalin A: Kinetics of transitions induced by interaction with Mn2+ and Ca2+. *Biochemistry*, **16**, 3883-3896.

Brown, R. D., 3rd, Koenig, S. H. and Brewer, C. F. (1982) Conformational equilibrium of demetalized concanavalin A. *Biochemistry*, **21**, 465-469.

Bryce, R. A., Hillier, I. H. and Naismith, J. H. (2001) Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys J*, **81**, 1373-1388.

Burton, D. R. and Dwek, R. A. (2006) Immunology. Sugar determines antibody activity. *Science*, **313**, 627-628.

Buskas, T., Ingale, S. and Boons, G. J. (2006) Glycopeptides as versatile tools for glycobiology. *Glycobiology*, **16**, 113R-136R.

Carrington, D. M., Auffret, A. and Hanke, D. E. (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature*, **313**, 64-67.

Carver, J. P., MacKenzie, A. E. and Hardman, K. D. (1985) Molecular model for the complex between Concanavalin A and a biantennary-complex class glycopeptide. *Biopolymers*, **24**, 49-63.

Casasnovas, J. M., Springer, T. A., Liu, J. H., Harrison, S. C. and Wang, J. H. (1997) Crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature*, **387**, 312-315.

Case, D. A., T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman *et al.* (2006) AMBER 9, University of California, San Francisco.

Cerutti, D. S., Freddolino, P. L., Duke, R. E. and Case, D. A. (2010) Simulations of a Protein Crystal with a High Resolution X-ray Structure: Evaluation of Force Fields and Water Models. *J Phys Chem B*, **114**, 12811-12824.

Chandra, N. R., Prabu, M. M., Suguna, K. and Vijayan, M. (2001) Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng*, **14**, 857-866.

Chandra, N. R., Kumar, N., Jeyakani, J., Singh, D. D., Gowda, S. B. and Prathima, M. N. (2006) Lectindb: a plant lectin database. *Glycobiology*, **16**, 938-946.

Cheng, J., Lu, T. H., Liu, C. L. and Lin, J. Y. (2010) A biophysical elucidation for less toxicity of agglutinin than abrin-a from the seeds of Abrus precatorius in consequence of crystal structure. *J Biomed Sci*, **17**, 34.

Chiappori, F., Merelli, I., Milanesi, L. and Rovida, E. (2010) Exploring the role of the phospholipid ligand in endothelial protein C receptor: a molecular dynamics study. *Proteins*, **78**, 2679-2690.

Choi, Y., Lee, J. H., Hwang, S., Kim, J. K., Jeong, K. and Jung, S. (2008) Retardation of the unfolding process by single N-glycosylation of ribonuclease A based on molecular dynamics simulations. *Biopolymers*, **89**, 114-123.

Choi, Y., Kim, H., Lee, J. H., Park, S., Jeong, K. and Jung, S. (2010) Structural stabilization of a rigid beta-sheet cluster of fucosylated proteinase inhibitor PMPC (Pars intercerebralis major peptide C) against thermal denaturation: An unfolding molecular dynamics simulation study. *J Mol Graph Model*, **28**, 487-494.

Choutko, A., Glattli, A., Fernandez, C., Hilty, C., Wuthrich, K. and van Gunsteren, W. F. (2010) Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle. *Eur Biophys J.*

Chu, F. K., Trimble, R. B. and Maley, F. (1978) The effect of carbohydrate depletion on the properties of yeast external invertase. *J Biol Chem*, **253**, 8691-8693.

Corfield, A. P. (2004) *Structure/function of O-glycans*. John Wiley & Sons, Ltd.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, **117**, 5179-5197.

Craik, C. S., Roczniak, S., Largman, C. and Rutter, W. J. (1987) The catalytic role of the active site aspartic acid in serine proteases. *Science*, **237**, 909-913.

Cunningham, B. A., Wang, J. L., Waxdal, M. J. and Edelman, G. M. (1975a) The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F3.
. *Journal of Biological Chemistry*, **250**, 1503-1512.

Cunningham, B. A., Wang, J. L., Waxdal, M. J. and Edelman, G. M. (1975b) The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F3. *J Biol Chem*, **250**, 1503-1512.

D'Angelo, P., Della Longa, S., Arcovito, A., Anselmi, M., Di Nola, A. and Chillemi, G. (2010) Dynamic Investigation of Protein Metal Active Sites: Interplay of XANES and Molecular Dynamics Simulations. *J Am Chem Soc*.

Dagher, S. F., Wang, J. L. and Patterson, R. J. (1995) Identification of galectin-3 as a factor in pre-mRNA splicing. *Proc Natl Acad Sci U S A*, **92**, 1213-1217.

Dam, T. K., Roy, R., Das, S. K., Oscarson, S. and Brewer, C. F. (2000) Binding of multivalent carbohydrates to concanavalin A and Dioclea grandiflora lectin. Thermodynamic analysis of the "multivalency effect". *J Biol Chem*, **275**, 14223-14230.

Dam, T. K. and Brewer, C. F. (2004) Multivalent protein-carbohydrate interactions: isothermal titration microcalorimetry studies. *Methods Enzymol*, **379**, 107-128.

Damodaran, D., Jeyakani, J., Chauhan, A., Kumar, N., Chandra, N. R. and Surolia, A. (2008) CancerLectinDB: a database of lectins relevant to cancer. *Glycoconj J*, **25**, 191-198.

Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, **98**, 10089-10092.

Day, R., Paschek, D. and Garcia, A. E. (2010) Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins*, **78**, 1889-1899.

Del Sol, F. G., Cavada, B. S. and Calvete, J. J. (2007) Crystal structures of *Cratylia floribunda* seed lectin at acidic and basic pHs. Insights into the structural basis of the pH-dependent dimer-tetramer transition. *J Struct Biol*, **158**, 1-9.

DeLano, W. L. (2002) The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA.

Delbaere, L. T., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. and Dauter, Z. (1993) Structures of the lectin IV of Griffonia simplicifolia and its complex with the Lewis b human blood group determinant at 2.0 A resolution. *J Mol Biol*, **230**, 950-965.

Deng, N. J. and Cieplak, P. (2010) Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophys J*, **98**, 627-636.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T. and Campbell, J. (1989) The structure of the saccharide-binding site of concanavalin A. *Embo J*, **8**, 2189-2193.

Drickamer, K. (1995a) Increasing diversity of animal lectin structures. *Curr Opin Struct Biol*, **5**, 612-616.

Drickamer, K. (1995b) Multiplicity of lectin-carbohydrate interactions. *Nat Struct Biol*, **2**, 437-439.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem*, **24**, 1999-2012.

Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J. and Wang, J. L. (1972) The covalent and three-dimensional structure of concanavalin A. *Proc Natl Acad Sci U S A*, **69**, 2580-2584.

Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. and Suddath, F. L. (1986) The crystal structure of pea lectin at 3.0-A resolution. *J Biol Chem*, **261**, 16518-16527.

Elgavish, S. and Shaanan, B. (1997) Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem Sci*, **22**, 462-467.

Elgavish, S. and Shaanan, B. (1998) Structures of the Erythrina corallodendron lectin and of its complexes with mono- and disaccharides. *J Mol Biol*, **277**, 917-932.

Esposito, L. and Daggett, V. (2005) Insight into ribonuclease A domain swapping by molecular dynamics unfolding simulations. *Biochemistry*, **44**, 3358-3368.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, **103**, 8577-8593.

Ewald, P. P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, **369**, 253–287.

Fang, E. F., Lin, P., Wong, J. H., Tsao, S. W. and Ng, T. B. (2010a) A lectin with anti-HIV-1 reverse transcriptase, antitumor, and nitric oxide inducing activities from seeds of Phaseolus vulgaris cv. extralong autumn purple bean. *J Agric Food Chem*, **58**, 2221-2229.

Fang, E. F., Wong, J. H., Lin, P. and Ng, T. B. (2010b) Biochemical and functional properties of a lectin purified from korean large black soybeans--a cultivar of glycine max. *Protein Pept Lett*, **17**, 690-698.

Faraudo, J., Calero, C. and Aguilella-Arzo, M. (2010) Ionic Partition and Transport in Multi-Ionic Channels: A Molecular Dynamics Simulation Study of the OmpF Bacterial Porin. *Biophys J*, **99**, 2107-2115.

Faustino, I., Perez, A. and Orozco, M. (2010) Toward a consensus view of duplex RNA flexibility. *Biophys J*, **99**, 1876-1885.

Ferrenberg, A. M. and Swendsen, R. H. (1989) Optimized Monte Carlo data analysis. *Physical Review Letters*, **63**, 1195 LP - 1198.

Figueiredo, J. G., Bitencourt, F. S., Mota, M. R., Silvestre, P. P., Aguiar, C. N., Benevides, R. G., Nascimento, K. S., de Moura, T. R., Dal-Secco, D., Assreuy, A. M. *et al.* (2009) Pharmacological analysis of the neutrophil migration induced by D. rostrata lectin: involvement of cytokines and nitric oxide. *Toxicon*, **54**, 736-744.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr. and James, M. N. (1982) Refined crystal structure of the molecular complex of Streptomyces griseus protease B, a serine protease, with the third domain of the ovomucoid inhibitor from turkey. *Proc Natl Acad Sci U S A*, **79**, 4868-4872.

Gail M. Bradbrook, T. G., Stephen J. Harrop, Jarjis Habash, James Raftery, Joseph Kalb (Gilboa), Joseph Yariv, Ian H. Hillier and John R. Helliwell (1998) X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *Journal of the Chemical Society, Faraday Transactions*, **94**, 1603-1611.

Gandhi, N. S. and Mancera, R. L. (2009) Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology*, **19**, 1103-1115.

Ganguly, D. and Mukhopadhyay, C. (2006) Binding diversity of the two binding sites of ricin B lectin. *Biopolymers*, **83**, 83-94.

Gelin, B. R. and Karplus, M. (1975) Sidechain torsional potentials and motion of amino acids in porteins: bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A*, **72**, 2002-2006.

Gervais, V., Zerial, A. and Oschkinat, H. (1997) NMR investigations of the role of the sugar moiety in glycosylated recombinant human granulocyte-colony-stimulating factor. *Eur J Biochem*, **247**, 386-395.

Gettins, P., Patston, P. A. and Schapira, M. (1993) The role of conformational change in serpin structure and function. *Bioessays*, **15**, 461-467.

Gibson, R., Leavitt, R., Kornfeld, S. and Schlesinger, S. (1978) Synthesis and infectivity of vesicular stomatitis virus containing nonglycosylated G protein. *Cell*, **13**, 671-679.

Gilboa-Garber, N. and Mizrahi, L. (1981) A new mitogenic D-galactosephilic lectin isolated from seeds of the coral-tree Erythrina corallodendron. Comparison with Glycine max (soybean) and Pseudomonas aeruginosa lectins. *Can J Biochem*, **59**, 315-320.

Golden, S. D. and Olsen, K. W. (2008) Identification of ligand-binding pathways in truncated hemoglobins using locally enhanced sampling molecular dynamics. *Methods Enzymol*, **437**, 459-475.

Goldstein, I. J., Hollerman, C. E. and Merrick, J. M. (1965a) Protein-Carbohydrate Interaction. I. The Interaction of Polysaccharides with Concanavalin A. *Biochim Biophys Acta*, **97**, 68-76.

Goldstein, I. J., Hollerman, C. E. and Smith, E. E. (1965b) Protein-Carbohydrate Interaction. Ii. Inhibition Studies on the Interaction of Concanavalin a with Polysaccharides. *Biochemistry*, **4**, 876-883.

Goldstein, I. J., So, L. L., Yang, Y. and Callies, Q. C. (1969) Protein-carbohydrate interaction. XIX. The interaction of concanavalin A with IgM and the glycoprotein phytohemagglutinins of the waxbean and the soybean. *J Immunol*, **103**, 695-698.

Goldstein, I. J. and Hayes, C. E. (1978) The lectins: carbohydrate-binding proteins of plants and animals. *Adv Carbohydr Chem Biochem*, **35**, 127-340.

Gong, Z. and Xiao, Y. (2010) RNA stability under different combinations of amber force fields and solvation models. *J Biomol Struct Dyn*, **28**, 431-441.

Greer, J., Kaufman, H. W. and Kalb, A. J. (1970) An x-ray crystallographic study of concanavalin A. *J Mol Biol*, **48**, 365-366.

Gundertofte, K., Liljefors, T., Norrby, P.-o. and Pettersson, I. (1996) A comparison of conformational energies calculated by several molecular mechanics methods. *Journal of Computational Chemistry*, **17**, 429-449.

Gupta, D., Dam, T. K., Oscarson, S. and Brewer, C. F. (1997) Thermodynamics of lectin-carbohydrate interactions. Binding of the core trimannoside of asparagine-linked carbohydrates and deoxy analogs to concanavalin A. *J Biol Chem*, **272**, 6388-6392.

Gupta, G., Vishveshwara, S. and Surolia, A. (2009) Stability of dimeric interface in banana lectin: Insight from molecular dynamics simulations. *IUBMB Life*, **61**, 252-260.

Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W. and Mackerell, A. D., Jr. (2008) Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem*, **29**, 2543-2564.

Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. and Mackerell, A. D. (2009) CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J Chem Theory Comput*, **5**, 2353-2370.

Halgren, T. A. (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, **17**, 490-519.

Hamdaoui, A., Wataleb, S., Devreese, B., Chiou, S. J., Vanden Broeck, J., Van Beeumen, J., De Loof, A. and Schoofs, L. (1998) Purification and characterization of a group of five novel peptide serine protease inhibitors from ovaries of the desert locust, Schistocerca gregaria. *FEBS Lett*, **422**, 74-78.

Hansia, P., Dev, S., Surolia, A. and Vishveshwara, S. (2007) Insight into the early stages of thermal unfolding of peanut agglutinin by molecular dynamics simulations. *Proteins*, **69**, 32-42.

Hardin, C., Eastwood, M. P., Luthey-Schulten, Z. and Wolynes, P. G. (2000) Associative memory Hamiltonians for structure prediction without homology: Alpha-helical proteins *Proceedings of the National Academy of Sciences of the United States of America* **97** 14235-14240

Harding, L., Scott, R. H., Kellenberger, C., Hietter, H., Luu, B., Beadle, D. J. and Bermudez, I. (1995) Inhibition of high voltage-activated Ca2+ currents from cultured sensory neurones by a novel insect peptide. *J Recept Signal Transduct Res*, **15**, 355-364.

Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B. and Ainsworth, C. F. (1971) Structure of concanavalin A at 4.25-angstrom resolution. *Proc Natl Acad Sci U S A*, **68**, 1393-1397.

Hardman, K. D. and Ainsworth, C. F. (1972) Structure of concanavalin A at 2.4-A resolution. *Biochemistry*, **11**, 4910-4919.

Hardman, K. D., Agarwal, R. C. and Freiser, M. J. (1982) Manganese and calcium binding sites of concanavalin A. *J Mol Biol*, **157**, 69-86.

Hassing, G. S. and Goldstein, I. J. (1970) Ultraviolet difference spectral studies on concanavalin A. Carbohydrate interaction. *Eur J Biochem*, **16**, 549-556.

Hassing, G. S., Goldstein, I. J. and Marini, M. (1971) The role of protein carboxyl groups in carbohydrate-concanavalin A interaction. *Biochim Biophys Acta*, **243**, 90-97.

Herman, E. M., Shannon, L. M. and Chrispeels, M. J. (1985) Concanavalin A is synthesized as a glycoprotein precursor. *Planta*, **165**, 23-29.

Hermann, R. B. (1972) Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *The Journal of Physical Chemistry*, **76**, 2754-2759.

Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, **4**, 435-447.

Hirsch, A. M. (1999) Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. *Curr Opin Plant Biol*, **2**, 320-326.

Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144-1149.

Horejsi, V., Ticha, M., Novotny, J. and Kocourek, J. (1980) Studies on lectins. XLVII. Some properties of D-galactose binding lectins isolated from the seeds of Butea frondosa, Erythrina indica and Momordica charantia. *Biochim Biophys Acta*, **623**, 439-448.

Hotchkiss, M. and Tauber, H. (1931) Studies on Crystalline Urease: II. Hemagglutinating Properties of Urease. *J Immunol*, **21**, 287-292.

Huang, X., Smith, M. C., Berzofsky, J. A. and Barchi, J. J., Jr. (1996) Structural comparison of a 15 residue peptide from the V3 loop of HIV-1IIIb and an O-glycosylated analogue. *FEBS Lett*, **393**, 280-286.

Hubbard, S. J., Thornton, J.M. (1993) NACCESS version 2.1.1, Computer Program, Department of Biochemistry and Molecular Biology, University College London.

Iglesias, J. L., Lis, H. and Sharon, N. (1982) Purification and properties of a D-galactose/N-acetyl-D-galactosamine-specific lectin from Erythrina cristagalli. *Eur J Biochem*, **123**, 247-252.

Imler, J. L. and Hoffmann, J. A. (2000) Signaling mechanisms in the antimicrobial host defense of Drosophila. *Curr Opin Microbiol*, **3**, 16-22.

Inbar, M. and Sachs, L. (1969) Interaction of the carbohydrate-binding protein concanavalin A with normal and transformed cells. *Proc Natl Acad Sci U S A*, **63**, 1418-1425.

Iwanaga, S. (1993) The limulus clotting reaction. *Curr Opin Immunol*, **5**, 74-82.

Jack, A., Weinzierl, J. and Kalb, A. J. (1971) An X-ray crystallographic study of demetallized concanavalin A. *Journal of Molecular Biology*, **58**, 389-390.

Janosi, L., Prakash, A. and Doxastakis, M. (2010) Lipid-modulated sequence-specific association of glycophorin A in membranes. *Biophys J*, **99**, 284-292.

Jayaram, B., Sprous, D. and Beveridge, D. L. (1998) Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field. *The Journal of Physical Chemistry B*, **102**, 9571-9576.

Jiang, S. Y., Ma, Z. and Ramachandran, S. (2010) Evolutionary history and stress regulation of the lectin superfamily in higher plants. *BMC Evol Biol*, **10**, 79.

Joao, H. C., Scragg, I. G. and Dwek, R. A. (1992) Effects of glycosylation on protein conformation and amide proton exchange rates in RNase B. *FEBS Lett*, **307**, 343-346.

Jones, D. B. and Johns, C. O. (1916) SOME PROTEINS FROM THE JACK BEAN, CANAVALIA ENSIFORMIS
Journal of Biological Chemistry, **28**, 67-75.

Jones, D. B., Gersdorff, C. E. F. and Moeller, O. (1924) THE TRYPTOPHANE AND CYSTINE CONTENT OF VARIOUS PROTEINS
Journal of Biological Chemistry, **62**, 183-195.

Jones, G. (1972) Lymphocyte activation. II. Kinetics and specificity of the activation process with phytohaemagglutinin and concanavalin A. *Clin Exp Immunol*, **12**, 403-407.

Jones, J. E. (1924) On the Determination of Molecular Fields. II. From the Equation of State of a Gas
10.1098/rspa.1924.0082. *Proceedings of the Royal Society of London. Series A*, **106**, 463-477.

Jorgensen, W. L. (1981) Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society*, **103**, 335-340.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926-935.

Jorgensen, W. L. and Tirado-Rives, J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, **110**, 1657-1666.

Kadirvelraj, R., Foley, B. L., Dyekjaer, J. D. and Woods, R. J. (2008) Involvement of water in carbohydrate-protein binding: concanavalin A revisited. *J Am Chem Soc*, **130**, 16933-16942.

Kalb, A. J. and Levitzki, A. (1968) Metal-binding sites of concanavalin A and their role in the binding of alpha-methyl d-glucopyranoside. *Biochem J*, **109**, 669-672.

Kalb, A. J. and Lustig, A. (1968) The molecular weight of concanavalin A. *Biochim Biophys Acta*, **168**, 366-367.

Kalyan, N. K. and Bahl, O. P. (1983) Role of carbohydrate in human chorionic gonadotropin. Effect of deglycosylation on the subunit interaction and on its in vitro and in vivo biological properties. *J Biol Chem*, **258**, 67-74.

Kaneko, Y., Nimmerjahn, F. and Ravetch, J. V. (2006) Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science*, **313**, 670-673.

Kanellopoulos, P. N., Tucker, P. A., Pavlou, K., Agianian, B. and Hamodrakas, S. J. (1996) A Triclinic Crystal Form of the Lectin Concanavalin A. *J Struct Biol*, **117**, 16-23.

Kantardjieff, K. A., Hochtl, P., Segelke, B. W., Tao, F. M. and Rupp, B. (2002) Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility. *Acta Crystallogr D Biol Crystallogr*, **58**, 735-743.

Karplus, M., Gelin, B. R. and McCammon, J. A. (1980) Internal dynamics of proteins. Short time and long time motions of aromatic sidechains in PTI. *Biophys J*, **32**, 603-618.

Kaushik, S., Mohanty, D. and Surolia, A. (2009) The role of metal ions in substrate recognition and stability of concanavalin A: a molecular dynamics study. *Biophys J*, **96**, 21-34.

Kaushik, S., Mohanty, D. and Surolia, A. (2011) Role of glycosylation in structure and stability of Erythrina corallodendron lectin (EcorL): A molecular dynamics study. *Protein Sci*, **20**, 465-481.

Kay, C. M. (1970) The presence of beta-structure in concanavalin A. *FEBS Lett*, **9**, 78-80.

Kellenberger, C., Boudier, C., Bermudez, I., Bieth, J. G., Luu, B. and Hietter, H. (1995) Serine protease inhibition by insect peptides containing a cysteine knot and a triple-stranded beta-sheet. *J Biol Chem*, **270**, 25514-25519.

Kirnarsky, L., Prakash, O., Vogen, S. M., Nomoto, M., Hollingsworth, M. A. and Sherman, S. (2000) Structural effects of O-glycosylation on a 15-residue peptide from the mucin (MUC1) core protein. *Biochemistry*, **39**, 12076-12082.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L. and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem*, **29**, 622-655.

Koenig, S. H., Brown, R. D., 3rd and Brewer, C. F. (1973) Solvent proton magnetic relaxation dispersion in solutions of concanavalin A. *Proc Natl Acad Sci U S A*, **70**, 475-479.

Kollman, P. (1993) Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, **93**, 2395-2417.

Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem*, **46**, 331-358.

Krautler, V., van Gunsteren, W. F. and Hunenberger, P. H. (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, **22**, 501-508.

Kriss, C. T., Lou, B.-S., Szabò, L. Z., Mitchell, S. A., Hruby, V. J. and Polt, R. (2000) Enkephalin-based drug design: conformational analysis of O-linked glycopeptides by NMR and molecular modeling. *Tetrahedron: Asymmetry*, **11**, 9-25.

Kromer, E., Nakakura, N. and Lagueux, M. (1994) Cloning of a Locusta cDNA encoding a precursor peptide for two structurally related proteinase inhibitors. *Insect Biochem Mol Biol*, **24**, 329-331.

Kuhn, B. and Kollman, P. A. (2000) A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning. *Journal of the American Chemical Society*, **122**, 3909-3916.

Kulkarni, K. A., Srivastava, A., Mitra, N., Sharon, N., Surolia, A., Vijayan, M. and Suguna, K. (2004) Effect of glycosylation on the structure of Erythrina corallodendron lectin. *Proteins*, **56**, 821-827.

Lagarda-Diaz, I., Guzman-Partida, A. M., Urbano-Hernandez, G., Ortega-Nieblas, M. M., Robles-Burgueno, M. R., Winzerling, J. and Vazquez-Moreno, L. (2009) Insecticidal action of PF2 lectin from Olneya tesota (Palo Fierro) against Zabrotes subfasciatus larvae and midgut glycoconjugate binding. *J Agric Food Chem*, **57**, 689-694.

Laine, E., Martinez, L., Blondel, A. and Malliavin, T. E. (2010) Activation of the Edema Factor of Bacillus anthracis by Calmodulin: Evidence of an Interplay between the EF-Calmodulin Interaction and Calcium Binding. *Biophys J*, **99**, 2264-2272.

Lam, S. K. and Ng, T. B. (2010) Isolation and characterization of a French bean hemagglutinin with antitumor, antifungal, and anti-HIV-1 reverse transcriptase activities and an exceptionally high yield. *Phytomedicine*, **17**, 457-462.

Lam, S. K. and Ng, T. B. (2011) Lectins: production and practical applications. *Appl Microbiol Biotechnol*, **89**, 45-55.

Lange, O. F., van der Spoel, D. and de Groot, B. L. (2010) Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys J*, **99**, 647-655.

Laskowski, M., Jr. (1986) Protein inhibitors of serine proteinases--mechanism and classification. *Adv Exp Med Biol*, **199**, 1-17.

Lawrence, D. A., Ginsburg, D., Day, D. E., Berkenpas, M. B., Verhamme, I. M., Kvassman, J. O. and Shore, J. D. (1995) Serpin-protease complexes are trapped as stable acyl-enzyme intermediates. *J Biol Chem*, **270**, 25309-25312.

Leon, M. A. (1967) Concanavalin A reaction with human normal immunoglobulin G and myeloma immunoglobulin G. *Science*, **158**, 1325-1326.

Li, C. Y., Xu, H. L., Liu, B. and Bao, J. K. (2010) Concanavalin A, from an Old Protein to Novel Candidate Anti-Neoplastic Drug. *Curr Mol Pharmacol*, **3**, 123-128.

Li, Z. and Lazaridis, T. (2004) The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *The Journal of Physical Chemistry B*, **109**, 662-670.

Liang, G., Schmidt, R. K., Yu, H.-A., Cumming, D. A. and Brady, J. W. (1996) Free Energy Simulation Studies of the Binding Specificity of Mannose-Binding Protein. *The Journal of Physical Chemistry*, **100**, 2528-2534.

Liang, Z., Sottrup-Jensen, L., Aspan, A., Hall, M. and Soderhall, K. (1997) Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferrin chain. *Proc Natl Acad Sci U S A*, **94**, 6682-6687.

Light-Wahl, K. J., Winger, B. E. and Smith, R. D. (1993) Observation of the multimeric forms of concanavalin A by electrospray ionization mass spectrometry. *Journal of the American Chemical Society*, **115**, 5869-5870.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950-1958.

Lis, H., Joubert, F. J. and Sharon, N. (1985) Isolation and properties of N-acetyllactosamine-specific lectins from nine erythrina species. *Phytochemistry*, **24**, 2803-2809.

Liu, X., Sejbal, J., Kotovych, G., Koganty, R. R., Reddish, M. A., Jackson, L., Gandhi, S. S., Mendonca, A. J. and Longenecker, B. M. (1995) Structurally defined synthetic cancer vaccines: analysis of structure, glycosylation and recognition of cancer associated mucin, MUC-1 derived peptides. *Glycoconj J*, **12**, 607-617.

Lopez-Jaramillo, F. J., Gonzalez-Ramirez, L. A., Albert, A., Santoyo-Gonzalez, F., Vargas-Berenguel, A. and Otalora, F. (2004) Structure of concanavalin A at pH 8: bound solvent and crystal contacts. *Acta Crystallogr D Biol Crystallogr*, **60**, 1048-1056.

Loris, R., Stas, P. P. and Wyns, L. (1994) Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family. *J Biol Chem*, **269**, 26722-26733.

Mandal, D. K., Kishore, N. and Brewer, C. F. (1994) Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry*, **33**, 1149-1156.

Manoj, N., Srinivas, V. R., Surolia, A., Vijayan, M. and Suguna, K. (2000) Carbohydrate specificity and salt-bridge mediated conformational change in acidic winged bean agglutinin. *J Mol Biol*, **302**, 1129-1137.

Markowitz, H. (1969) Interaction of concanavalin A with polysaccharides of Histoplasma capsulatum. *J Immunol*, **103**, 308-318.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. and de Vries, A. H. (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B*, **111**, 7812-7824.

Martin, A. C. R. a. P., C.T. (2008) ProFit v2.6 : Fitting performed using the McLachlan algorithm. London.

Matheson, N. R., van Halbeek, H. and Travis, J. (1991) Evidence for a tetrahedral intermediate complex during serpin-proteinase interactions. *J Biol Chem*, **266**, 13489-13491.

Mazumder, P. and Mukhopadhyay, C. (2010) Molecular modeling and NMR studies of benzyl substituted mannosyl trisaccharide binding to two mannose-specific lectins: Allium sativam agglutinin I and Concanavalin A. *Biopolymers*, **93**, 952-967.

McKenzie, G. H., Sawyer, W. H. and Nichol, L. W. (1972) The molecular weight and stability of concanavalin A. *Biochim Biophys Acta*, **263**, 283-293.

McKenzie, G. H. and Sawyer, W. H. (1973) The binding properties of dimeric and tetrameric concanavalin A. Binding of ligands to noninteracting macromolecular acceptors. *J Biol Chem*, **248**, 549-556.

McPhalen, C. A., Schnebli, H. P. and James, M. N. (1985a) Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg. *FEBS Lett*, **188**, 55-58.

McPhalen, C. A., Svendsen, I., Jonassen, I. and James, M. N. (1985b) Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo. *Proc Natl Acad Sci U S A*, **82**, 7242-7246.

Medema, J. P., de Jong, J., Peltenburg, L. T., Verdegaal, E. M., Gorter, A., Bres, S. A., Franken, K. L., Hahne, M., Albar, J. P., Melief, C. J. *et al.* (2001) Blockade of the granzyme B/perforin pathway through overexpression of the serine protease inhibitor PI-9/SPI-6 constitutes a mechanism for immune escape by tumors. *Proc Natl Acad Sci U S A*, **98**, 11515-11520.

Meersman, F., Atilgan, C., Miles, A. J., Bader, R., Shang, W., Matagne, A., Wallace, B. A. and Koch, M. H. (2010) Consistent picture of the reversible thermal unfolding of hen egg-white lysozyme from experiment and molecular dynamics. *Biophys J*, **99**, 2255-2263.

Megyeri, P., Pabst, K. M. and Pabst, M. J. (1995) Serine protease inhibitors block priming of monocytes for enhanced release of superoxide. *Immunology*, **86**, 629-635.

Menasche, G., Feldmann, J., Fischer, A. and de Saint Basile, G. (2005) Primary hemophagocytic syndromes point to a direct link between lymphocyte cytotoxicity and homeostasis. *Immunol Rev*, **203**, 165-179.

Mer, G., Hietter, H., Kellenberger, C., Renatus, M., Luu, B. and Lefevre, J. F. (1996a) Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors. *J Mol Biol*, **258**, 158-171.

Mer, G., Hietter, H. and Lefevre, J. F. (1996b) Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated proteinase inhibitor. *Nat Struct Biol*, **3**, 45-53.

Meynier, C., Feracci, M., Espeli, M., Chaspoul, F., Gallice, P., Schiff, C., Guerlesquin, F. and Roche, P. (2009) NMR and MD investigations of human galectin-1/oligosaccharide complexes. *Biophys J*, **97**, 3168-3177.

Min, W., Dunn, A. J. and Jones, D. H. (1992) Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. *Embo J*, **11**, 1303-1307.

Mitra, N., Sharon, N. and Surolia, A. (2003) Role of N-linked glycan in the unfolding pathway of Erythrina corallodendron lectin. *Biochemistry*, **42**, 12208-12216.

Mitra, N., Sinha, S., Ramya, T. N. and Surolia, A. (2006) N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem Sci*, **31**, 156-163.

Miyamoto, S. and Kollman, P. A. (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, **13**, 952-962.

Mobley, D. L., Bayly, C. I., Cooper, M. D., Shirts, M. R. and Dill, K. A. (2009) Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J Chem Theory Comput*, **5**, 350-358.

Montreuil, J. (1980) Primary structure of glycoprotein glycans: basis for the molecular biology of glycoproteins. *Adv Carbohydr Chem Biochem*, **37**, 157-223.

Moothoo, D. N., Canan, B., Field, R. A. and Naismith, J. H. (1999) Man alpha1-2 Man alpha-OMe-concanavalin A complex reveals a balance of forces involved in carbohydrate recognition. *Glycobiology*, **9**, 539-545.

Naidoo, K. J., Denysyk, D. and Brady, J. W. (1997) Molecular dynamics simulations of the N-linked oligosaccharide of the lectin from Erythrina corallodendron. *Protein Eng*, **10**, 1249-1261.

Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. and Yariv, J. (1994) Refined structure of concanavalin A complexed with methyl alpha-D-mannopyranoside at 2.0 A resolution and comparison with the saccharide-free structure. *Acta Crystallogr D Biol Crystallogr*, **50**, 847-858.

Naismith, J. H. and Field, R. A. (1996) Structural basis of trimannoside recognition by concanavalin A. *J Biol Chem*, **271**, 972-976.

Nakakura, N., Hietter, H., Van Dorsselaer, A. and Luu, B. (1992) Isolation and structural determination of three peptides from the insect Locusta migratoria. Identification of a deoxyhexose-linked peptide. *Eur J Biochem*, **204**, 147-153.

Nalivaeva, N. N. and Turner, A. J. (2001) Post-translational modifications of proteins: acetylcholinesterase as a model system. *Proteomics*, **1**, 735-747.

Nangia-Makker, P., Conklin, J., Hogan, V. and Raz, A. (2002) Carbohydrate-binding proteins in cancer, and their ligands as therapeutic agents. *Trends Mol Med*, **8**, 187-192.

Narahari, A. and Swamy, M. J. (2010) Rapid affinity-purification and physicochemical characterization of pumpkin (Cucurbita maxima) phloem exudate lectin. *Biosci Rep*, **30**, 341-349.

Northrup, S. H., Pear, M. R., Lee, C. Y., McCammon, J. A. and Karplus, M. (1982) Dynamical theory of activated processes in globular proteins. *Proc Natl Acad Sci U S A*, **79**, 4035-4039.

Page, M. J. and Di Cera, E. (2008) Serine peptidases: classification, structure and function. *Cell Mol Life Sci*, **65**, 1220-1236.

Parkin, S. and Craig, G. (2002) Crystallization and X-ray diffraction data for a new form of concanavalin A. *Acta Crystallogr D Biol Crystallogr*, **58**, 1032-1033.

Pashov, A., MacLeod, S., Saha, R., Perry, M., VanCott, T. C. and Kieber-Emmons, T. (2005) Concanavalin A binding to HIV envelope protein is less sensitive to mutations in glycosylation sites than monoclonal antibody 2G12. *Glycobiology*, **15**, 994-1001.

Pathiaseril, A. and Woods, R. J. (1999) Relative Energies of Binding for Antibodyâ''Carbohydrate-Antigen Complexes Computed from Free-Energy Simulations. *Journal of the American Chemical Society*, **122**, 331-338.

Person, D. A. and Markowitz, H. (1969) Immunochemical studies of the phytohemagglutinin, concanavalin A. *J Immunol*, **103**, 225-232.

Petrescu, A. J., Petrescu, S. M., Dwek, R. A. and Wormald, M. R. (1999) A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology*, **9**, 343-352.

Peumans, W. J. and Van Damme, E. J. (1995) Lectins as plant defense proteins. *Plant Physiol*, **109**, 347-352.

Pflumm, M. N., Wang, J. L. and Edelman, G. M. (1971) Conformational changes in concanavalin A. *J Biol Chem*, **246**, 4369-4370.

Plotnick, M. I., Mayne, L., Schechter, N. M. and Rubin, H. (1996) Distortion of the active site of chymotrypsin complexed with a serpin. *Biochemistry*, **35**, 7586-7590.

Poretz, R. D. and Goldstein, I. J. (1968) Protein--carbohydrate interaction. XI. A study of turbidity as it relates to concanavalin A-glycan interaction. *Immunology*, **14**, 165-174.

Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. and Suguna, K. (1998) Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 A resolution. *J Mol Biol*, **276**, 787-796.

Pratap, J. V., Bradbrook, G. M., Reddy, G. B., Surolia, A., Raftery, J., Helliwell, J. R. and Vijayan, M. (2001) The combination of molecular dynamics with crystallography for elucidating protein-ligand interactions: a case study involving peanut lectin complexes with T-antigen and lactose. *Acta Crystallogr D Biol Crystallogr*, **57**, 1584-1594.

Project, E., Nachliel, E. and Gutman, M. (2010) Force field-dependent structural divergence revealed during long time simulations of Calbindin d9k. *J Comput Chem*, **31**, 1864-1872.

Qasba, P. K., Balaji, P. V. and Rao, V. S. (1994) Molecular dynamics simulations of oligosaccharides and their conformation in the crystal structure of lectin-carbohydrate complex: importance of the torsion angle psi for the orientation of alpha 1,6-arm. *Glycobiology*, **4**, 805-815.

Quiocho, F. A., Reeke, G. N., Jr., Becker, J. W., Lipscomb, W. N. and Edelman, G. M. (1971) Structure of soncanavalin A at 4 A resolution. *Proc Natl Acad Sci U S A*, **68**, 1853-1857.

Raabe, G. and Maginn, E. J. (2010) A force field for 3,3,3-fluoro-1-propenes, including HFO-1234yf. *J Phys Chem B*, **114**, 10133-10142.

Rahman, A. (1964) Correlations in the Motion of Atoms in Liquid Argon. *Physical Review*, **136**, A405 LP - A411.

Rahman, A. and Stillinger, F. H. (1971) Molecular Dynamics Study of Liquid Water. *The Journal of Chemical Physics*, **55**, 3336-3359.

Rahman, M. M., Bosch, L. V., Baggerman, G., Clynen, E., Hens, K., Hoste, B., Meylaers, K., Vercammen, T., Schoofs, L., De Loof, A. *et al.* (2002) Search for peptidic molecular markers in hemolymph of crowd-(gregarious) and isolated-reared (solitary) desert locusts, Schistocerca gregaria. *Peptides*, **23**, 1907-1914.

Ramachandraiah, G. and Chandra, N. R. (2000) Sequence and structural determinants of mannose recognition. *Proteins*, **39**, 358-364.

Ramis, C., Gomord, V., Lerouge, P. and Faye, L. (2001) Deglycosylation is necessary but not sufficient for activation of proconcanavalin A. *J Exp Bot*, **52**, 911-917.

Redondo, M. J. and Alvarez-Pellitero, P. (2010) The effect of lectins on the attachment and invasion of Enteromyxum scophthalmi (Myxozoa) in turbot (Psetta maxima L.) intestinal epithelium in vitro. *Exp Parasitol*, **126**, 577-581.

Reeke, G. N., Becker, J. W. and Edelman, G. M. (1975a) The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure.
. *Journal of Biological Chemistry*, **250**, 1525-1547.

Reeke, G. N., Jr., Becker, J. W. and Edelman, G. M. (1975b) The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. *J Biol Chem*, **250**, 1525-1547.

Reeke, G. N., Jr., Becker, J. W. and Edelman, G. M. (1978) Changes in the three-dimensional structure of concanavalin A upon demetallization. *Proc Natl Acad Sci U S A*, **75**, 2286-2290.

Reeke, G. N., Jr. and Becker, J. W. (1986) Three-dimensional structure of favin: saccharide binding-cyclic permutation in leguminous lectins. *Science*, **234**, 1108-1111.

Reiser, M. and Timm, J. (2009) Serine protease inhibitors as anti-hepatitis C virus agents. *Expert Rev Anti Infect Ther*, **7**, 537-547.

Reynolds, J. A., Gilbert, D. B. and Tanford, C. (1974) Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc Natl Acad Sci U S A*, **71**, 2925-2927.

Ricci, C. G., de Andrade, A. S., Mottin, M. and Netz, P. A. (2010) Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions. *J Phys Chem B*, **114**, 9882-9893.

Rini, J. M. (1995a) Lectin structure. *Annu Rev Biophys Biomol Struct*, **24**, 551-577.

Rini, J. M. (1995b) X-ray crystal structures of animal lectins. *Curr Opin Struct Biol*, **5**, 617-621.

Rini, J. M. and Lobsanov, Y. D. (1999) New animal lectin structures. *Curr Opin Struct Biol*, **9**, 578-584.

Rizzo, R. C. and Jorgensen, W. L. (1999) OPLS All-Atom Model for Amines:â€‰ Resolution of the Amine Hydration Problem. *Journal of the American Chemical Society*, **121**, 4827-4836.

Robles, A., Medeiros, A., Berois, N., Balter, H. S., Pauwels, E. K. and Osinaga, E. (2010) In-site interaction evaluation of Tn density by inhibition/competition assays. *Nucl Med Biol*, **37**, 453-458.

Rodriguez, A., Tablero, M., Arreguin, B., Hernandez, A., Arreguin, R., Soriano-Garcia, M., Tulinsky, A., Park, C. H. and Seshadri, T. P. (1988) Preliminary x-ray investigation of an orthorhombic crystal of hevein. *J Biol Chem*, **263**, 4047-4048.

Roitberg, A. and Elber, R. (1991) Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *The Journal of Chemical Physics*, **95**, 9277-9287.

Rouge, P., Culerrier, R., Granier, C., Rance, F. and Barre, A. (2010) Characterization of IgE-binding epitopes of peanut (Arachis hypogaea) PNA lectin allergen cross-reacting with other structurally related legume lectins. *Mol Immunol*, **47**, 2359-2366.

Roussel, A., Mathieu, M., Dobbs, A., Luu, B., Cambillau, C. and Kellenberger, C. (2001) Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity. *J Biol Chem*, **276**, 38893-38898.

Rubin, H. (1996) Serine protease inhibitors (SERPINS): where mechanism meets medicine. *Nat Med*, **2**, 632-633.

Ruscetti, F. W. and Chervenick, P. A. (1975) Regulation of the release of colony-stimulating activity from mitogen-stimulated lymphocytes. *J Immunol*, **114**, 1513-1517.

Salzet, M. (2002) Leech thrombin inhibitors. *Curr Pharm Des*, **8**, 493-503.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M. (1996) A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold. *Nat Struct Biol*, **3**, 596-603.

Saper, M. A., Lis, H., Sharon, N. and Shaanan, B. (1987) Crystallization and preliminary X-ray diffraction studies of the lectin from Erythrina corallodendron. *J Mol Biol*, **193**, 823-824.

Scanlon, M. J., Morley, S. D., Jackson, D. E., Price, M. R. and Tendler, S. J. (1992) Structural and computational investigations of the conformation of antigenic peptide fragments of human polymorphic epithelial mucin. *Biochem J*, **284 (Pt 1)**, 137-144.

Schulke, N. and Schmid, F. X. (1988) The stability of yeast invertase is not significantly influenced by glycosylation. *J Biol Chem*, **263**, 8827-8831.

Schuman, J., Qiu, D., Koganty, R. R., Longenecker, B. M. and Campbell, A. P. (2000) Glycosylations versus conformational preferences of cancer associated mucin core. *Glycoconjugate Journal*. Springer Netherlands, vol. 17, pp. 835-848.

Sela, B. A., Wang, J. L. and Edelman, G. M. (1975a) Isolation of lectins of different specificities on a single affinity adsorbent. *J Biol Chem*, **250**, 7535-7538.

Sela, B. A., Wang, J. L. and Edelman, G. M. (1975b) Antibodies reactive with cell surface carbohydrates. *Proc Natl Acad Sci U S A*, **72**, 1127-1131.

Shaanan, B., Lis, H. and Sharon, N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862-866.

Sharma, A., Sekar, K. and Vijayan, M. (2009) Structure, dynamics, and interactions of jacalin. Insights from molecular dynamics simulations examined in conjunction with results of X-ray studies. *Proteins*, **77**, 760-777.

Sharma, A. and Vijayan, M. (2010) Influence of glycosidic linkage on the nature of carbohydrate binding in beta-prism I fold lectins: an X-ray and molecular dynamics investigation on banana lectin-carbohydrate complexes. *Glycobiology*, **21**, 23-33.

Sharma, V. and Surolia, A. (1997) Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J Mol Biol*, **267**, 433-445.

Sharon, N. and Lis, H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, **14**, 53R-62R.

Sheldon, P. S. and Bowles, D. J. (1992) The glycoprotein precursor of concanavalin A is converted to an active lectin by deglycosylation. *Embo J*, **11**, 1297-1301.

Sheldon, P. S., Keen, J. N. and Bowles, D. J. (1996) Post-translational peptide bond formation during concanavalin A processing in vitro. *Biochem J*, **320 (Pt 3)**, 865-870.

Shoham, M., Sussman, J. L., Yonath, A., Moult, J., Traub, W. and Kalb, A. J. (1978) The effect of binding of metal ions on the three-dimensional structure of demetallized concanavalin A. *FEBS Lett*, **95**, 54-56.

Shoham, M., Yonath, A., Sussman, J. L., Moult, J., Traub, W. and Kalb, A. J. (1979) Crystal structure of demetallized concanavalin A: the metal-binding region. *J Mol Biol*, **131**, 137-155.

Shur, B. D. and Hall, N. G. (1982) Sperm surface galactosyltransferase activities during in vitro capacitation. *J Cell Biol*, **95**, 567-573.

Simmerling, C. and Elber, R. (1994) Hydrophobic "Collapse" in a Cyclic Hexapeptide: Computer Simulations of CHDLFC and CAAAAC in Water. *Journal of the American Chemical Society*, **116**, 2534-2547.

Simonet, G., Claeys, I. and Broeck, J. V. (2002) Structural and functional properties of a novel serine protease inhibiting peptide family in arthropods. *Comp Biochem Physiol B Biochem Mol Biol*, **132**, 247-255.

Simonet, G., Claeys, I., Franssens, V., De Loof, A. and Broeck, J. V. (2003) Genomics, evolution and biological functions of the pacifastin peptide family: a conserved serine protease inhibitor family in arthropods. *Peptides*, **24**, 1633-1644.

Simonet, G., Breugelmans, B., Proost, P., Claeys, I., Van Damme, J., De Loof, A. and Vanden Broeck, J. (2005) Characterization of two novel pacifastin-like peptide precursor isoforms in the desert locust (Schistocerca gregaria): cDNA cloning, functional analysis and real-time RT-PCR gene expression studies. *Biochem J*, **388**, 281-289.

Singh, K., Kaur, M., Rup, P. J. and Singh, J. (2009) Effects of indian coral tree, Erythrina indica lectin on eggs and larval development of melon fruit fly, Bactrocera cucurbitae. *J Environ Biol*, **30**, 509-514.

Sinha, S., Mitra, N., Kumar, G., Bajaj, K. and Surolia, A. (2005) Unfolding studies on soybean agglutinin and concanavalin a tetramers: a comparative account. *Biophys J*, **88**, 1300-1310.

Sinha, S., Gupta, G., Vijayan, M. and Surolia, A. (2007) Subunit assembly of plant lectins. *Curr Opin Struct Biol*, **17**, 498-505.

Sinha, S. and Surolia, A. (2007) Attributes of glycosylation in the establishment of the unfolding pathway of soybean agglutinin. *Biophys J*, **92**, 208-216.

Sitkoff, D., Sharp, K. A. and Honig, B. (1994) Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *The Journal of Physical Chemistry*, **98**, 1978-1988.

So, L. L. and Goldstein, I. J. (1967) Protein-carbohydrate interaction. IV. Application of the quantitative precipitin method to polysaccharide-concanavalin A interaction. *J Biol Chem*, **242**, 1617-1622.

So, L. L. and Goldstein, I. J. (1968a) Protein-carbohydrate interaction. 13. The interaction of concanavalin A with alpha-mannans from a variety of microorganisms. *J Biol Chem*, **243**, 2003-2007.

So, L. L. and Goldstein, I. J. (1968b) Protein-carbohydrate interaction. XX. On the number of combining sites on concanavalin A, the phytohemagglutinin of the jack bean. *Biochim Biophys Acta*, **165**, 398-404.

Sola, R. J. and Griebenow, K. (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci*, **98**, 1223-1245.

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. and Case, D. A. (1998) Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidateâ''DNA Helices. *Journal of the American Chemical Society*, **120**, 9401-9409.

Steen, P. V. d., Rudd, P. M., Dwek, R. A. and Opdenakker, G. (1998) Concepts and Principles of O-Linked Glycosylation doi:10.1080/10409239891204198. *Critical Reviews in Biochemistry and Molecular Biology*, **33**, 151-208.

Stehle, T., Yan, Y., Benjamin, T. L. and Harrison, S. C. (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature*, **369**, 160-163.

Stillinger, F. H. and Rahman, A. (1974) Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, **60**, 1545-1557.

Stubbs, H. J., Lih, J. J., Gustafson, T. L. and Rice, K. G. (1996) Influence of core fucosylation on the flexibility of a biantennary N-linked oligosaccharide. *Biochemistry*, **35**, 937-947.

Sultan, N. A., Kavitha, M. and Swamy, M. J. (2009) Purification and physicochemical characterization of two galactose-specific isolectins from the seeds of Trichosanthes cordata. *IUBMB Life*, **61**, 457-469.

Sumner, J. B. (1919) THE GLOBULINS OF THE JACK BEAN, CANAVALIA ENSIFORMIS *Journal of Biological Chemistry*, **37**, 137-142.

Sumner, J. B. and Graham, V. A. (1925) THE GLOBULINS OF THE JACK BEAN (CANAVALIA ENSIFORMIS) *Journal of Biological Chemistry*, **64**, 257-261.

Sumner, J. B. and Howell, S. F. (1935) The Non-Identity of Jack Bean Agglutinin with Crystalline Urease. *J Immunol*, **29**, 133-134.

Sumner, J. B., Howell, S. F. and Zeissig, A. (1935) Concanavalin a and Hemagglutination. *Science*, **82**, 65-66.

Sumner, J. B. and Howell, S. F. (1936a) Identification of Hemagglutinin of Jack Bean with Concanavalin A. *J. Bacteriol.*, **32**, 227-237.

Sumner, J. B. and Howell, S. F. (1936b) THE ROLE OF DIVALENT METALS IN THE REVERSIBLE INACTIVATION OF JACK BEAN HEMAGGLUTININ. *Journal of Biological Chemistry*, **115**, 583-588.

Sumner, J. B., GralÃ©n, N. and Eriksson-Quensel, I.-B. (1938) THE MOLECULAR WEIGHTS OF CANAVALIN, CONCANAVALIN A, AND CONCANAVALIN B *Journal of Biological Chemistry*, **125**, 45-48.

Surolia, A., Sharon, N. and Schwarz, F. P. (1996) Thermodynamics of monosaccharide and disaccharide binding to Erythrina corallodendron lectin. *J Biol Chem*, **271**, 17697-17703.

Swaminathan, C. P., Surolia, N. and Surolia, A. (1998) Role of Water in the Specific Binding of Mannose and Mannooligosaccharides to Concanavalin A. *Journal of the American Chemical Society*, **120**, 5153-5159.

Tams, J. W. and Welinder, K. G. (1998) Glycosylation and thermodynamic versus kinetic stability of horseradish peroxidase. *FEBS Lett*, **421**, 234-236.

Teixeira-Sa, D. M., Reicher, F., Braga, R. C., Beltramini, L. M. and de Azevedo Moreira, R. (2009) Isolation of a lectin and a galactoxyloglucan from Mucuna sloanei seeds. *Phytochemistry*, **70**, 1965-1972.

Teneberg, S., Angstrom, J., Jovall, P. A. and Karlsson, K. A. (1994) Characterization of binding of Gal beta 4GlcNAc-specific lectins from Erythrina cristagalli and Erythrina corallodendron to glycosphinogolipids. Detection, isolation, and characterization of a novel glycosphinglipid of bovine buttermilk. *J Biol Chem*, **269**, 8554-8563.

Tkacz, J. S., Cybulska, E. B. and Lampen, J. O. (1971) Specific staining of wall mannan in yeast cells with fluorescein-conjugated concanavalin A. *J Bacteriol*, **105**, 1-5.

Topf, M., Varnai, P., Schofield, C. J. and Richards, W. G. (2002) Molecular dynamics simulations of the acyl-enzyme and the tetrahedral intermediate in the deacylation step of serine proteases. *Proteins*, **47**, 357-369.

Trapani, J. A. (2001) Granzymes: a family of lymphocyte granule serine proteases. *Genome Biol*, **2**, REVIEWS3014.

Trautwein, C., Rakemann, T., Malek, N. P., Plumpe, J., Tiegs, G. and Manns, M. P. (1998) Concanavalin A-induced liver injury triggers hepatocyte proliferation. *J Clin Invest*, **101**, 1960-1969.

Treesuwan, W., Wittayanarakul, K., Anthony, N. G., Huchet, G., Alniss, H., Hannongbua, S., Khalaf, A. I., Suckling, C. J., Parkinson, J. A. and Mackay, S. P. (2009) A detailed binding free energy study of 2:1 ligand-DNA complex formation by experiment and simulation. *Phys Chem Chem Phys*, **11**, 10682-10693.

Upreti, R. K., Kumar, M. and Shankar, V. (2003) Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics*, **3**, 363-379.

van der Kamp, M. W. and Daggett, V. (2010) Influence of pH on the Human Prion Protein: Insights into the Early Steps of Misfolding. *Biophys J*, **99**, 2289-2298.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, **26**, 1701-1718.

van Gunsteren, W. F. and Berendsen, H. J. C. (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, **34**, 1311 - 1327.

van Gunsteren, W. F. and Mark, A. E. (1992) On the interpretation of biochemical data by molecular dynamics computer simulation. *Eur J Biochem*, **204**, 947-961.

Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, **3**, 97-130.

Vellore, N. A., Yancey, J. A., Collier, G., Latour, R. A. and Stuart, S. J. (2010) Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions. *Langmuir*, **26**, 7396-7404.

Verkhivker, G., Elber, R. and Nowak, W. (1992) Locally enhanced sampling in free energy calculations: Application of mean field approximation to accurate calculation of free energy differences. *The Journal of Chemical Physics*, **97**, 7838-7841.

Verlet, L. (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, **159**, 98 LP - 103.

Vorontsov, II and Miyashita, O. (2009) Solution and crystal molecular dynamics simulation study of m4-cyanovirin-N mutants complexed with di-mannose. *Biophys J*, **97**, 2532-2540.

Vyakarnam, A., Dagher, S. F., Wang, J. L. and Patterson, R. J. (1997) Evidence for a role for galectin-1 in pre-mRNA splicing. *Mol Cell Biol*, **17**, 4730-4737.

Vymetal, J. and Vondrasek, J. (2010) Metadynamics as a tool for mapping the conformational and free-energy space of peptides--the alanine dipeptide case study. *J Phys Chem B*, **114**, 5632-5642.

Wang, C., Eufemi, M., Turano, C. and Giartosio, A. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry*, **35**, 7299-7307.

Wang, J., Cieplak, P. and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, **21**, 1049-1074.

Wang, J., Tan, C., Tan, Y.-H., Lu, Q. and Luo, R. (2008) Poisson-Boltzmann solvents in molecular dynamics simulations. *Commun. Comput. Phys.*, **3**, 1010-1031.

Wang, J. L., Cunningham, B. A. and Edelman, G. M. (1971) Unusual fragments in the subunit structure of concanavalin A. *Proc Natl Acad Sci U S A*, **68**, 1130-1134.

Wang, J. L., McClain, D. A. and Edelman, G. M. (1975) Modulation of lymphocyte mitogenesis. *Proc Natl Acad Sci U S A*, **72**, 1917-1921.

Weerapana, E. and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology*, **16**, 91R-101R.

Wei, A., Rubin, H., Cooperman, B. S. and Christianson, D. W. (1994) Crystal structure of an uncleaved serpin reveals the conformation of an inhibitory reactive loop. *Nat Struct Biol*, **1**, 251-258.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, **106**, 765-784.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. (1986) An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, **7**, 230-252.

Wilfred F. van Gunsteren , H. J. C. B. (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English*, **29**, 992 - 1023.

Williams, B. A., Chervenak, M. C. and Toone, E. J. (1992) Energetics of lectin-carbohydrate binding. A microcalorimetric investigation of concanavalin A-oligomannoside complexation. *J Biol Chem*, **267**, 22907-22911.

Williams, R. L., Greene, S. M. and McPherson, A. (1987) The crystal structure of ribonuclease B at 2.5-A resolution. *J Biol Chem*, **262**, 16020-16031.

Witvrouw, M., Fikkert, V., Hantson, A., Pannecouque, C., O'Keefe B, R., McMahon, J., Stamatatos, L., de Clercq, E. and Bolmstedt, A. (2005) Resistance of human immunodeficiency virus type 1 to the high-mannose binding agents cyanovirin N and concanavalin A. *J Virol*, **79**, 7777-7784.

Woods, R. J., Edge, C. J. and Dwek, R. A. (1994) Protein surface oligosaccharides and protein function. *Nat Struct Biol*, **1**, 499-501.

Woods, R. J., Dwek, R. A., Edge, C. J. and Fraser-Reid, B. (1995) Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development. *J. Phys. Chem.*, **99**, 3832-3846.

Wormald, M. R. and Dwek, R. A. (1999) Glycoproteins: glycan presentation and protein-fold stability. *Structure*, **7**, R155-160.

Wright, C. S. (1997) New folds of plant lectins. *Curr Opin Struct Biol*, **7**, 631-636.

Wu, J. H., Wu, A. M., Yang, Z., Chen, Y. Y., Singha, B., Chow, L. P. and Lin, J. Y. (2010) Recognition intensities of submolecular structures, mammalian glyco-structural units, ligand cluster and polyvalency in abrin-a-carbohydrate interactions. *Biochimie*, **92**, 147-156.

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L. and Wagner, G. (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science*, **269**, 1273-1278.

Wyss, D. F. and Wagner, G. (1996) The structural role of sugars in glycoproteins. *Curr Opin Biotechnol*, **7**, 409-416.

Yahara, I. and Edelman, G. M. (1972) Restriction of the mobility of lymphocyte immunoglobulin receptors by concanavalin A. *Proc Natl Acad Sci U S A*, **69**, 608-612.

Yan, Q., Li, Y., Jiang, Z., Sun, Y., Zhu, L. and Ding, Z. (2009) Antiproliferation and apoptosis of human tumor cell lines by a lectin (AMML) of Astragalus mongholicus. *Phytomedicine*, **16**, 586-593.

Yang, R. Y., Hsu, D. K. and Liu, F. T. (1996) Expression of galectin-3 modulates T-cell growth and apoptosis. *Proc Natl Acad Sci U S A*, **93**, 6737-6742.

Yang, R. Y. and Liu, F. T. (2003) Galectins in cell growth and apoptosis. *Cell Mol Life Sci*, **60**, 267-276.

Yang, Y., Chakravorty, D. K. and Merz, K. M. (2010) Finding a Needle in the Haystack: Computational Modeling of Mg2+ Binding in the Active Site of Protein Farnesyltransferase. *Biochemistry*.

Yariv, J., Kalb, A. J. and Levitzki, A. (1968) The interaction of concanavalin A with methyl alpha-D-glucopyranoside. *Biochim Biophys Acta*, **165**, 303-305.

Young, N. M., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. K. and Thibault, P. (1995) C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in Escherichia coli. Characterization of "ragged ends" by mass spectrometry. *J Biol Chem*, **270**, 2563-2570.

Yuasa, H., Honma, H., Hashimoto, H., Tsunooka, M. and Kojima-Aikawa, K. (2007) Pentamer is the minimum structure for oligomannosylpeptoids to bind to concanavalin A. *Bioorg Med Chem Lett*, **17**, 5274-5278.

Zakharova, E., Horvath, M. P. and Goldenberg, D. P. (2009) Structure of a serine protease poised to resynthesize a peptide bond. *Proc Natl Acad Sci U S A*, **106**, 11034-11039.

Zand, R., Agrawal, B. B. and Goldstein, I. J. (1971) pH-dependent conformational changes of concanavalin A. *Proc Natl Acad Sci U S A*, **68**, 2173-2176.

Zhang, L. W. and Monteiro-Riviere, N. A. (2010) Lectins modulate multi-walled carbon nanotubes cellular uptake in human epidermal keratinocytes. *Toxicol In Vitro*, **24**, 546-551.

Zhang, Y., Lu, M., Cheng, Y. and Li, Z. (2010) H-NOX domains display different tunnel systems for ligand migration. *J Mol Graph Model*, **28**, 814-819.

Zhong, L. (2010) Exposure of Hydrophobic Core in Human Prion Protein Pathogenic Mutant H187R. *J Biomol Struct Dyn*, **28**, 355-361.

Zhong, Y. and Patel, S. (2010) Nonadditive empirical force fields for short-chain linear alcohols: methanol to butanol. Hydration free energetics and Kirkwood-Buff analysis using charge equilibration models. *J Phys Chem B*, **114**, 11076-11092.