Molecular Interaction Between Typhoid Toxin And *N*-linked Glycans: An *in-silico* Investigation

M.Sc. Thesis

By

NAVEEN KUMAR

DEPARTMEMT OF BIOSCIENCES & BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2022

Molecular Interaction Between Typhoid Toxin And *N*-linked Glycans: An *in-silico* Investigation

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree
of
Master of Science

By

NAVEEN KUMAR

DEPARTMEMT OF BIOSCIENCES & BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2022

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Molecular Interaction Between Typhoid Toxin And *N*-linked Glycans: An *in-silico* Investigation** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from August 2020 to May 2022 under the supervision of Dr. Parimal Kar, Associate Professor, Department of Biosciences and Biomedical Engineering.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Naveen Kumar
06/05/2022
Signature of the student
(Naveen Kumar)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Parimal Kar
6/5/2022
Signature of the Supervisor
(Dr. Parimal Kar)

Naveen Kumar has successfully given his M.Sc. Oral Examination held on **05/05/2022**.

Parimal Kar
Signature of Supervisor of MSc thesis
Date: 6/5/2022

P.V. Godgire
Convener, DPGC
Date: 06/05/22

Signature of PSPC Member
Date: 6/5/22

6/5/2022
Signature of PSPC Member
Date:

ACKNOWLEDGEMENTS

I would like to thank first my thesis advisor **Dr. Parimal Kar** for providing me with an opportunity to work in the computational biophysics lab at IIT Indore. I am sincerely thankful to him for providing excellent advice and guidance during the progress of my research. His expertise and insightful feedback have helped me bring my research work to a higher level. He has always inspired me to reach my academic potential. I would like to thank **Dr. Hem Chandra Jha** and **Prof. Amit Kumar** for being on my PG student progress committee. Their valuable feedback and suggestions have helped me to improve my research work. I sincerely thank all the Department of Biosciences and biomedical engineering faculty members, IIT Indore.

Special thanks to my research group members Dr. Nisha Amarnath Jonniya, Mr. Rajarshi Roy, Mr. Md Fulbabu Sk, Mr. Sayan Poddar, Mr. Suman Koirala, Mr. Kapil Ursal, Ms. Subhasmita Mahapatra and Mr. Sunanda Samanta for numerous technical discussions which have enriched my knowledge and understanding on the subject. I am thankful to my colleagues for their contributions to making the work environment quite lively and enriching. I again thank Mr. Rajarshi Roy for guiding me throughout my project, who beautifully and patiently answered all my minor and major doubts.

Finally, I would like to express my heartiest gratitude to my family for their unwavering support. I sincerely thank my parents, **Sri Prem K. Prasad** and **Shrimati Kalpana Devi**, who constantly inspire me. A special thanks to my elder sister, Divya Kumari and my elder brother, Praveen Kumar, for their frequent help. Lastly, I would like to thank my little nephew, Drisha.

(Naveen Kumar)

Dedicated to my

Family

ABSTRACT

Typhoid fever is an enteric fever along with the systemic illness of abdominal pain and fever. The primary causative organism responsible for typhoid fever is *Salmonella enterica* serovar Typhi (*S*. Typhi) bacteria. Typhoid toxin (A_2B_5 configuration) is found to be the key player responsible for developing the clinical symptoms during the Salmonella Typhi human infection. The B subunit is the homopentamer having PltB monomer, which plays the central role during the entry of the toxin into the host cells. Recent experimental studies suggest that each PltB monomer has three glycan-binding sites. These site mainly prefers to bind with multiantennary *N*-linked glycans with a varying degree of affinity among three binding sites. These interactions are mainly governed by the terminal sialic acid residues which can be modifed also. Herein, we studied the conformational dynamics and the binding mechanisms of five different *N*-glycan motifs (combination of modified and unmodified) through all atom molecular dynamics (MD). The conformational variability of the tri-saccharide motif was estimated through analysising the dihedral space of torsional angles as well as the puckering conformations. For elucidating the recognition mechanism, we employed the molecular mechanism generalized Born surface area (MM/GBSA) method. In terms of binding free energy, modifed glycans show better affinity compared to the unmodified carbohydrates. In the case of modified glycans, hydrogen bonding occurs via the Neu5Ac and acetyl modification of the glycans. In contrast, in the case of unmodified glycans, the binding occurs via the Neu5Ac, Gal and GlcNAc. Overall, our study may help to understand the interaction mechanisms of the attachement of typhoid toxin to the cell surface and to design glycomimetic molecules to prevent the disease.

LIST OF PUBLICATIONS

- Kumar, N.; Roy, R.; Kar, P.; Molecular Interaction Between Typhoid Toxin And *N*-linked Glycans: An *in-silico* Investigation (**Manuscript under preparation**)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AMBER	Assisted Model Building with Energy Refineme
PDB	Protein Data Bank
RMSD	Root-mean-square-deviations
RMSF	Root-mean-square-fluctuations
SASA	Solvent-Accessible-Surface-Area
PCA	Principal component Analysis
TIP3P	Transferable Intermolecular Potential with three-Point charges
TI	Thermodynamic Integration
GB	Generalized Born
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics Generalized-Born Surface Area
gg	gauche-gauche
gt	gauche-trans
tg	trans-gauche
BS	Binding Sites
GalNac	N-acetylglucosamine
Gal	Galactose
Neu5Ac	N-Acetyl-Neuraminic acid
MD	Molecular Dynamics
NPT	Isobaric-Isothermal ensemble
PME	Particle mesh Ewald

Chapter 1

Introduction and background

1.1 Enteric Fever

Salmonella is the leading cause of diarrhoea worldwide. *Salmonella enterica* subspecies enterica are rod-shaped and gram-negative bacteria having more than 2500 serovars. *Salmonella* causes two kinds of human diseases: typhoidal diseases and non-typhoidal disease[1]. Typhoidal serovars cause enteric fever known as typhoid and paratyphoid fever. non-typhoidal serovars cause non-invasive gastroenteritis and are usually self-limiting. Typhoid and paratyphoid fever are enteric fevers that received relatively lesser attention globally.

1.2 Causative bacteria

Typhoid fever is commonly known as typhoid, which is a systemic infection[2]. The primary causative organism responsible for typhoid fever is *Salmonella enterica* serovar Typhi (*S.* Typhi) bacteria which belongs to the family Enterobacteriaceae. While paratyphoid fever, known as paratyphoid, is caused by *Salmonella enterica* serovars Paratyphi A, Paratyphi B and Paratyphi C (*S.* Paratyphi A, B and C). The most common serovar is *S.* Paratyphi A. The non-typhoidal *Salmonella* includes *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium). *S. Typhi* caused typhoid is the most common and most threatening enteric disease among all discussed here.

1.3 Disease burden

It is estimated that every year there are 16 million new cases of typhoid, among which the 6,00,000 death take place[3]. It is more threatening in the developing and underdeveloped countries where there is a lack of good sanitary conditions and health infrastructures like in countries of Southeast Asia and Africa. Young children are more susceptible to typhoid because adults develop immunity to repetitive typhoid fever[4,5]. In India, there is a lack of systematic data and data from hospital-based

studies that show that typhoid prevalence is three cases per 1000 of the population per year. Almost one-fourth of the cases are from children under five years of age. The exact figures for typhoid fever remain uncertain because of the lack of diagnostic methods limitations. In 2010, nearly 27 million cases of typhoid fever were reported worldwide[6]. The average fatality rate remains at 1%, but there is high fluctuation in underdeveloped countries like Indonesia and Papua New Guinea, where it can be up to 50%. **Figure 1.1** shows the global incidence of typhoid fever.

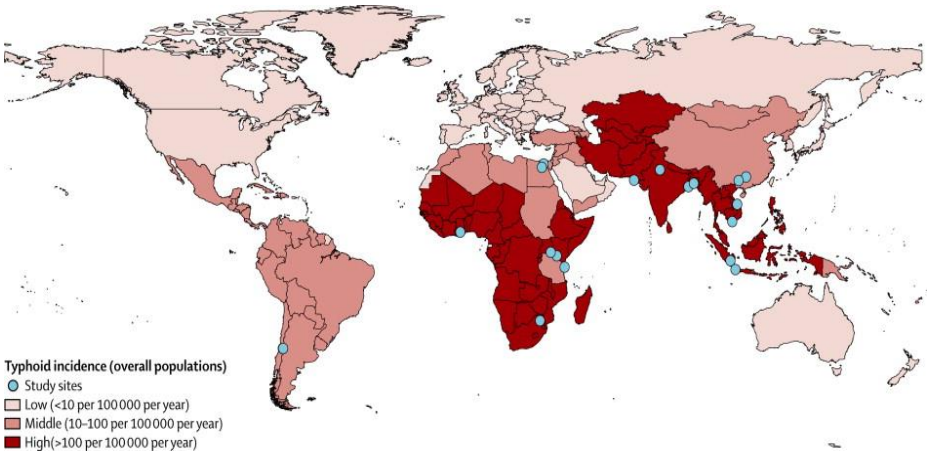

Figure 1.1: Global incidence of typhoid fever caused by S. Typhi[5].

1.4 Transmission and clinical features

Salmonella Typhi is a human restricted pathogenic bacterium, and no other living being is known to be the reservoir of this pathogen[7]. The infection is transmitted via the stool and urine of the infected person, and the essential carriers of the bacteria are contaminated food, water and flies. The faecal-oral route serves as the primary pathway of transmission of Salmonella Typhi into humans. The incubation period of this bacteria is 1 to 14 days. Thus, the areas having good sanitary conditions will have less incidence of typhoid fever.

The clinical features differ among the different age group, and it ranges from mild to severe or fatal[8]. The symptoms may include nausea, fatigue, diarrhoea, prolonged fever and abdominal pain. Late diagnosis

may lead to severe life-threatening complications like systemic infections, leukopenia, and neurological disorders and may be lethal. Re-infection may occur if antibiotics are given at the early onset of primary illness. In 10-15% of cases, complications arise, including typhoid encephalopathy, intestinal perforation, and gastrointestinal bleeding.

1.5 Pathogenesis and host-pathogen interactions

The typhoid bacterium enters the host body via contaminated water and food ingestion. With the help of adhesins, the bacterium attaches to the gut epithelium[9,10]. The Salmonella genome contains pathogenicity islands that encode the virulence factors needed for cellular invasion[11]. The type 3 secretion system protein transports the virulence factors proteins into the host cells[12], which ensures the entry of the *Salmonella* into the cells. It has been found that a dose of 1000 *S. Typhi* bacteria is enough to cause typhoid fever in 55% of individuals. The higher dose of bacteria ingested results in a higher infection rate with a short incubation period. After the ingestion of *Salmonella*, they can tolerate the acidic conditions of the stomach and reach the small intestine. *Salmonella* invades the cells in the intestine by direct uptake by phagocytic (M) cells or via the direct invasion of intestinal epithelium cells. (**Figure 1.2**)

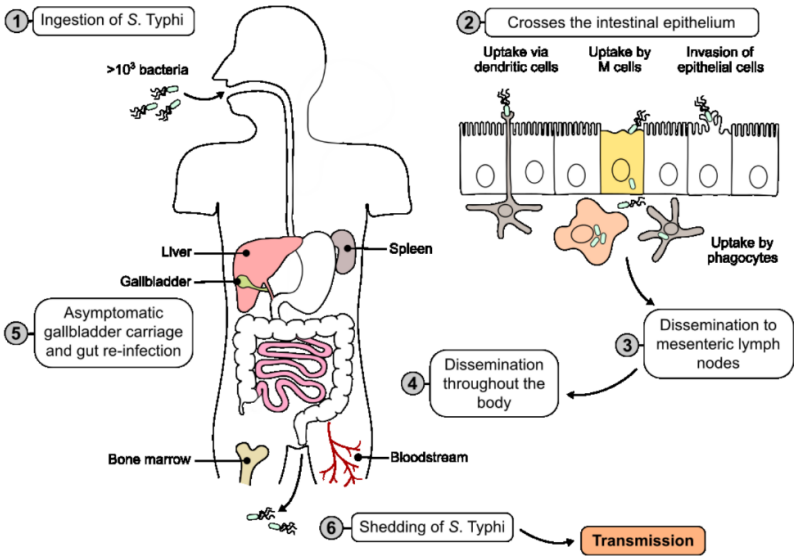

Figure 1.2: *S.* Typhi transmission in humans through the fecal-oral route.

The *S.* Typhi bacterium, when inside the host cell's SCV (Salmonella containing vacuole), synthesises the typhoid holotoxin. This typhoid toxin is then secreted out of the SCV by packing into the small vesicles. These small vesicles are then exported to the host cell membrane, where the vesicles fuse to the cell membrane. The typhoid toxin is released into the extracellular environment with the help of Sec machinery. The toxin recognises the specific glycan receptor on the surface of the cell membrane of the other host cell (**Figure 1.3**).

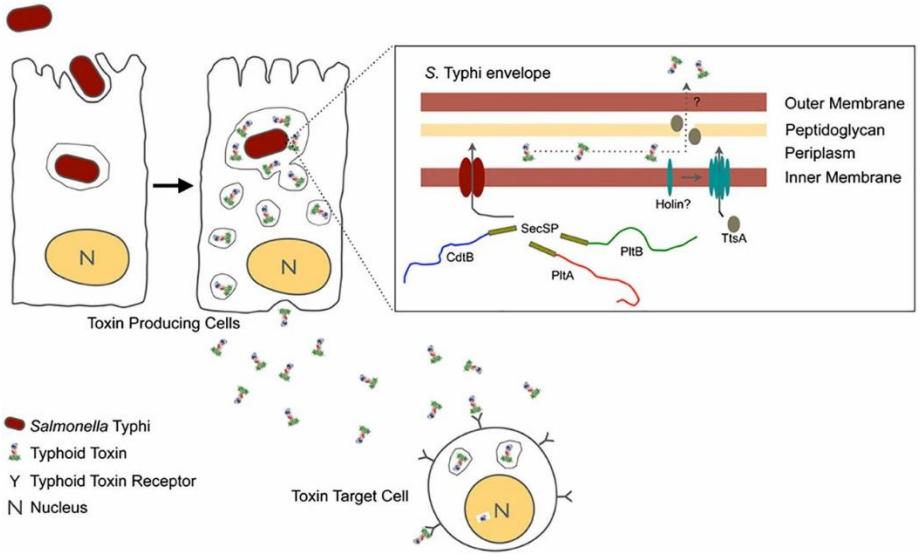

Figure 1.3: Mechanism of infection of typhoid fever[13]

Then it binds to the glycan receptors, and by receptor-mediated endocytosis, the typhoid toxin gets into the host cell, leading to cell intoxication. For entry into the human target cells, the typhoid toxin must bind to the glycosylated surface glycoprotein receptors like CD45 on myelocytic cells and podocalyxin 1 on epithelial cells[14,15]. Humans have sialoglycans on cells terminated with Neu5Ac, while most mammals have sialoglycans on cells terminated with Neu5Gc (N-glycolylneuraminic acid). S. Typhi is a restricted human pathogen. No other hosts are still known. Humans lack the enzyme CMAH (CMP-N-acetyl neuraminic acid hydrolase), which converts Neu5Ac to Neu5Gc. The typhoid toxin cannot bind to the Neu5Gc and becomes human restricted bacterial pathogen[7].

1.6 Typhoid Toxin

Typhoid toxin is the key player responsible for developing the clinical symptoms during the Salmonella Typhi human infection[13,14,16]. Also, it has a central role in the restriction of human pathogenesis. Typhoid toxin belongs to the AB bacterial toxin family. The typhoid toxin has a pyramidal shape and A_2B_5 configuration[17] (**Figure 1.4**). Here, the A_2B_5 configuration refers to two A and five B subunits (**Figure 1.5**). The A subunits are enzymatic ones, and the B subunits are receptor binding ones. The two A subunits are CdtB (Cytolethal distending toxin A subunit) and PltA (Pertussis like toxin A subunit). The five B subunits are the homopentamer of the PltB (pertussis-like toxin B subunit) subunits. Thus, the PltA-CdtB-PltB complex forms the typhoid holotoxin.

The CdtB subunit is a kind of nuclease that causes the host cell cycle, damaging the DNA and cell death[16,18,19]. PltA subunit is a mono ADP-ribosyltransferase, and its role in pathogenesis is still unknown. The PltB has the oligosaccharide (glycan) binding sites. The PltB has a role in the binding with the glycan receptors to ensure the entry of the bacterium inside the host cell. The CdtB subunit is at the apex of the pyramidal holotoxin, PltA is at the centre, and PltB is at the base. The CdtB and the PltB subunits do not interact directly with each other. The PltA interacts with the CdtB via the disulphide linkage between the Cys214 PltA and Cys269 of CdtB[5]. The PltA and CdtB interact via the hydrophobic interactions between the small part of the C-terminal alpha-helix inserted between the PltB. A small part of the C-terminal alpha-helix is inserted between the PltB homopentamer and interacts via the hydrophobic interaction. This hydrophobic interaction stabilises the PltA and PltB complex.

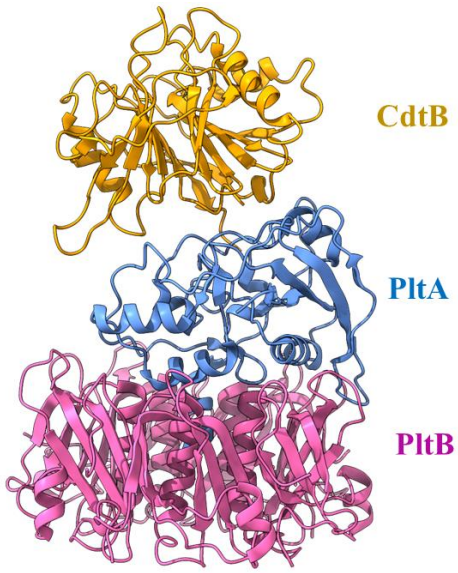

Figure 1.4: The whole typhoid holotoxin. The CdtB and PltA are the enzymatic subunits, and PltB is the receptor-binding subunit.

The PltB is a homopentamer of five PltB molecules, and each of the PltB molecules has three glycan-binding sites, namely BS1, BS2 and BS3[19–21]. Thus, the whole PltB homopentamer has fifteen glycan-binding sites: five BS1, five BS2 and five BS3 sites (**Figure 1.6**). These glycan-binding sites help multivalent interaction with the receptor glycan moieties. The BS1 is located at the lateral side of the PltB monomer, and the BS2 and BS3 are located next to each other.

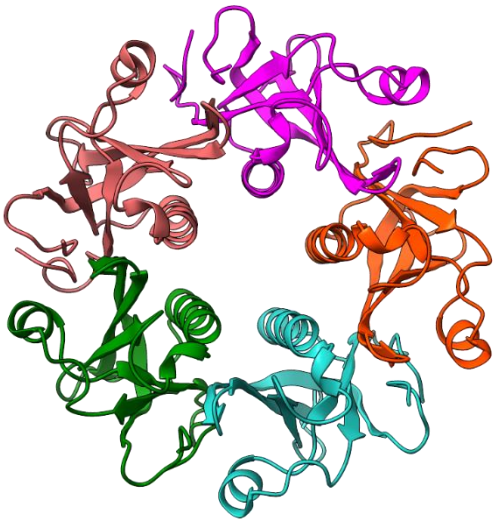

Figure 1.5: Structure of PltB protein of typhoid toxin. It consists of five chains: A, B, C, D and E.

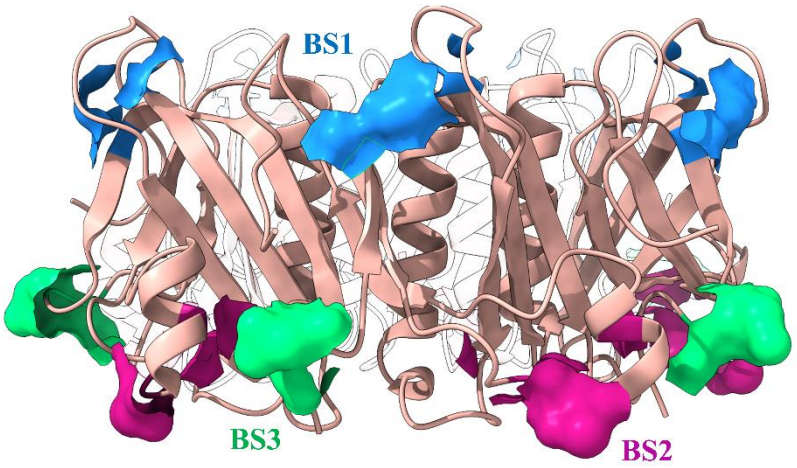

Figure 1.6: The Glycan binding sites (BS): BS1, BS2 and BS3 in the PltB homopentamer. Each monomer of the PltB protein has three glycan-binding sites. Thus, 15 glycan-binding sites in the whole PltB protein.

1.7 Glycans

Glycans can be explained as a glycosidic linked group of monosaccharides covalently bonded with the biomolecule such as proteins (glycoproteins) and lipids (glycolipids).

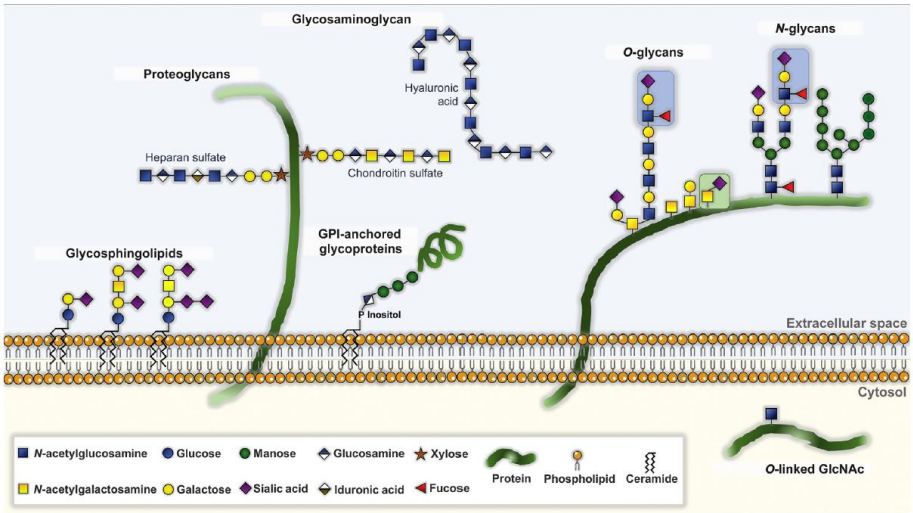

Figure 1.7: Glycan diversity[22]

Glycoproteins are of two types depending on the attachment of oligosaccharide molecule with the amino acid: *N*-linked and *O*-linked glycans. N-linked glycans are attached via the nitrogen atom on the asparagine residue, while O-linked glycans are linked via the oxygen atom on threonine or serine residues. The attachment of the

oligosaccharide can take place in a single residue or at multiple sites. Based on the branching pattern, glycans can be grouped as bi-antennary, tri-antennary, etc (**Figure 1.7**).

In glycolipids, the glycosidic bonds attach polar oligosaccharides to lipophilic lipid chains. Thus, glycolipids are amphiphilic. Glycolipids are commonly found in the eukaryotic cell membrane.

1.7.1 Biological roles of glycans

The role of glycans in the human body can be physical, metabolic, and structural. The proteoglycans help maintain the tissue structure porosity and help in creating binding sites for other glycans. Glycosylation of the protein provides a barrier to recognising protease and antibodies. Also, it plays an essential role in protein folding. Glycans also help in protective storage depot for different biomolecules; for example, many heparin-binding growth factors remain attached to the GAG chains of the extracellular matrix, which prevents the diffusion from its site and protects it from non-specific proteolysis.

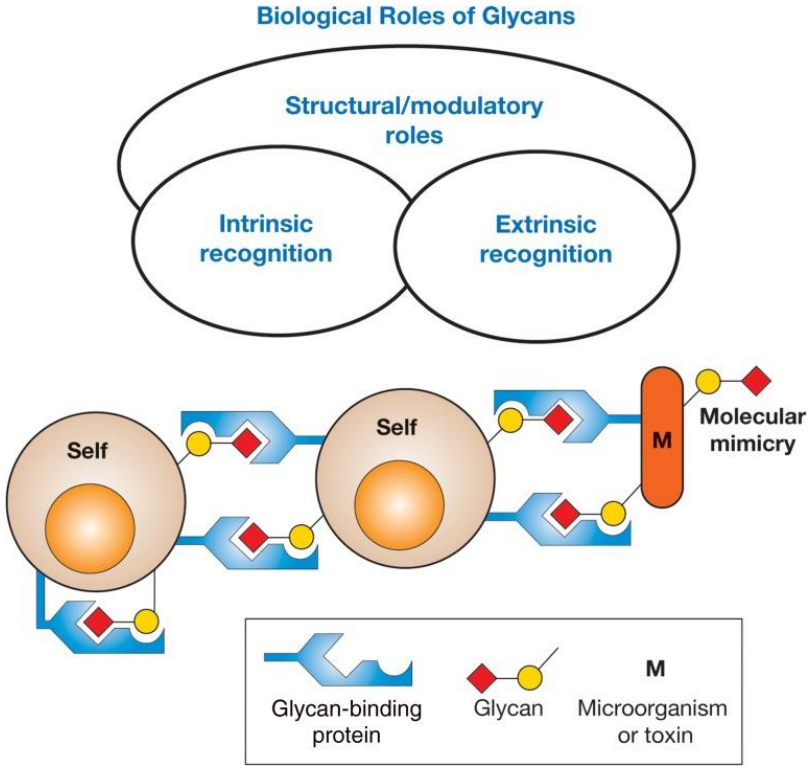

Figure 1.8: Biological role of glycans[23]

The glycans help in the recognition processes of the biomolecules, which are of two types: extrinsic and intrinsic recognition (**Figure 1.8**). In the case of extrinsic recognition, many symbionts and pathogens have evolved in such a way that they recognise the specific host glycans. Glycan-binding proteins mainly govern this kind of interaction. One example is the recognition of Heliobacter by adhesion molecule to the gastric sialoglycans. There are also different glycan-binding proteins and intrinsic proteins present in the cells performing the recognition; examples are cell-matrix interactions[24,25], fertilisation and reproductions[26,27], trafficking of glycoproteins[28,29].

1.8 Protein-glycan interaction

The protein-glycan interactions play a significant role in the molecular recognition processes, essential for biological events (**Figure 1.9**). According to the thermodynamics perspective, protein-glycan interactions are weak. Since the carbohydrate molecules are hydrophilic, it causes an enthalpic penalty (desolvation for interaction with a receptor). Mainly, hydrogen bonding and CH-π interaction occurred, and both have similar interaction strengths. Carbohydrate molecules are highly flexible, and when it binds to protein, their conformational flexibility decreases, making this reaction entropically unfavourable. Thus, the carbohydrate/glycan-protein interactions are enthalpy-entropy compensation reactions. The dissociation constants (K_D) of the carbohydrate and lectins interactions lie in the range of μM-mM. The multivalent mechanism of the lectins increases the binding affinity and helps in selectivity[31]. In the case of the N-glycosylated protein, the

carbohydrate-protein interaction is combined with protein-protein interaction increase the binding affinity[32].

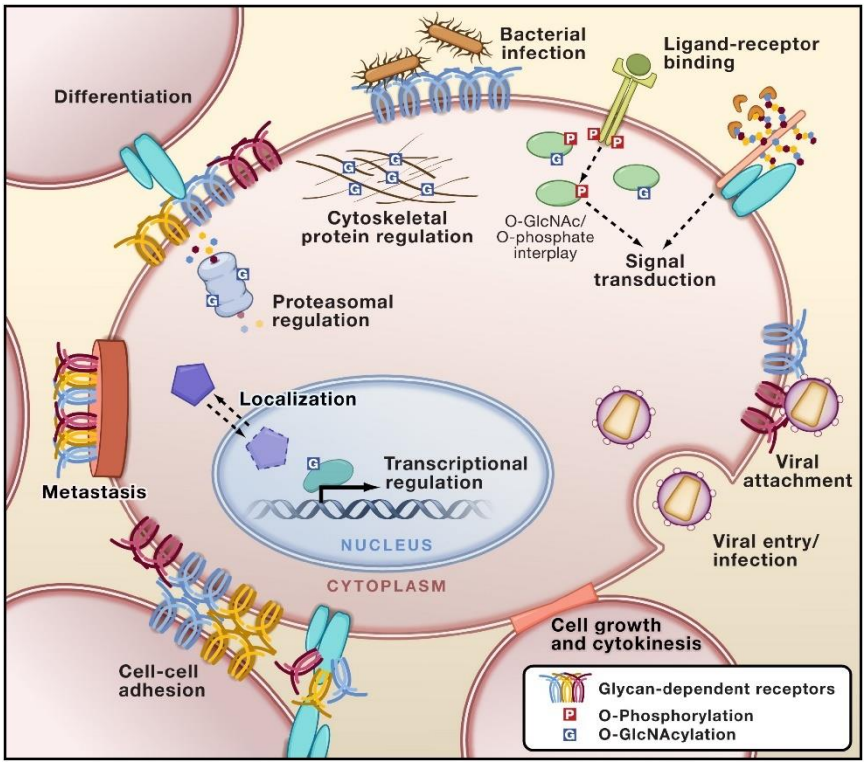

Figure 1.9: Summary of the role of the glycans[30]

1.9 Structure and functions of glycans

The PltB prefers to bind with the Neu5Ac terminated trisaccharide motif sequence (consensus sequence Neu5Aca2-3Galb1-3/b1-4Glc/GlcNAc) attached to the multiantennary N-linked glycans[1,33]. Thus, the typhoid toxin shows tropism towards cells having Neu5Ac (N-acetylneuraminic acid) terminated trisaccharide motif sequence attached to the multiantennary N-linked glycans like gallbladder epithelial, intestinal, brain endothelial cells of arterioles and immune cells[7].

The Neu5Ac is present in two forms: unmodified and modified in humans. The modified Neu5Ac has the most common 9-O acetylation at the C-9 position[23,34]; the other modification includes 4-O acetylation at the C-4 position[35,36], 7,9-O-acetylation in C-7 and C-9 position in human cells. These modifications occur naturally because of homeostasis. There are two O-acetyl transferases known as acetyl-CoA: sialate 7(9)-O-acetyltransferase (CASD1) and acetyl-CoA: sialate 4-O-

acetyltransferase. The other modifications of the additional chemical groups may also include acetylation, sulfonylation, methylation and lactylation.

The modified 9-O-acetylated α2–3 sialoside (Neu5,9Ac2α2-3Galβ1-4GlcNAc) (where Neu5Ac is N-Acetylneuraminic acid, Gal is Galactose and *N*-Acetylglucosamine) have 14 times more binding affinity as compared with the unmodified α2–3 sialosides (Neu5Ac2α2-3Galβ1-4GlcNAc). Also, modified 9-O-acetylated α2–6 sialoside (Neu5,9Ac2α2-6Galβ1-4GlcNAc) have slightly more binding affinity as compared with the unmodified α2–6 sialosides (Neu5Ac2α2-6Galβ1-4GlcNAc). The modified 4-O-acetylated α2–3 sialoside (Neu5,4Ac2α2-3Galβ1-4GlcNAc) has 11 times more binding affinity as compared with the unmodified α2–3 sialosides (Neu5Ac2α2-3Galβ1-4GlcNAc). These results suggest that the modified glycans at the Neu5Ac have a better binding affinity towards the PltB's glycan-binding sites[8].

Chapter 2

Theoretical background

MD simulation is based on Newton's laws of motion, where the movement of each atom is analysed computationally over a period. The coordinates and corresponding velocity of all atoms are calculated, which helps visualise the molecule's dynamics. MD simulations help to study the biological system processes that are dynamic and complex in nature[37–39]. There are biological processes that take place on different time scales, as shown in **Table 2.1**.

Table 2.1: Relationship between simulation time-length and biological systems[40].

Processes	**Time scale (time/distance)**
Atomic fluctuations, short loop, and side-chain motions	ps (10^{-15} to 10^{-1} s)/0.01-5 Å
Domain and helix motions	ns (10^{-9} to 1 s)/1-10 Å
Protein folding and unfolding, interactions	ms (10^{-7} to 104 s)/>5 Å

Thus, MD simulations help to predict details of the dynamic processes such as protein folding/unfolding, conformational changes, protein stability, molecular interactions, and recognition. MD simulation is a computational technique based on Newton's second law of motion. Integrating Newton's equation of motion gives the trajectories of the atoms as they change with time. By using these trajectories, the average value of properties can be predicted.

Quantum mechanics considers the electrons in the system; hence if few of the electrons are left out for calculations, then many electrons must be considered for calculations. It is more time-consuming. The biological systems have large, solvated macromolecules. This is too large for considering quantum mechanics. Force field methods are also known as molecular mechanics; they don't consider the electronic

motions and only consider the system's energy as a function of nuclear positions.

2.1 Force Fields

The molecular modelling force fields can be described in terms of four components of the inter and intra forces within the system. When there is the deviation of bond and angles parameters w.r.t. the equilibrium or reference values, then energy penalties are used. In the molecular mechanics, there is function describing the energy changes when bonds are rotated and the interaction between the two non-bonded components of the system. The total potential energy of the macromolecule system V (r)Total is divided into two components: external interactions and internal interactions, shown in **Figure 2.1**.

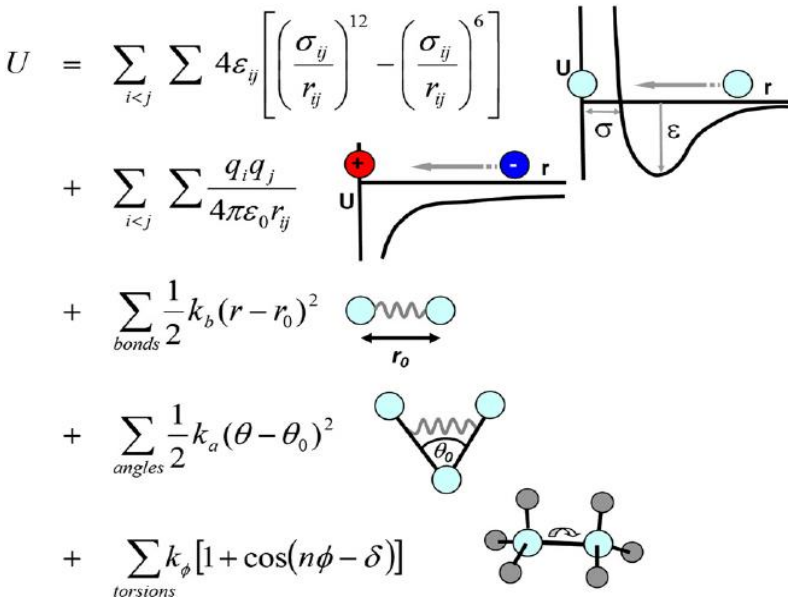

Figure 2.1: The potential energy function components for molecular interactions in force fields approximation[41].

There are different force fields for the protein and carbohydrate force fields. The force fields for protein and carbohydrates are different because of the difference in their bonding patterns and primary structure. The force fields are included in the molecular dynamics simulation packages such as AMBER (Assisted model building and energy refinement)[42], CHARMM (Chemistry at HARvard molecular mechanics)[43], OPLS (Optimised potential for lipid simulations)[44], GROMOS (GROningen MOlecular Simulation)[45] etc.

2.1.1 Protein force field

AMBER (Assisted Model Building with Energy Refinement) simulation package comes with various force fields for proteins such as ff14SB[46], ff19SB[47], CHARMM36m[48], etc. Among all the biological macromolecules, proteins are the most studied macromolecules. AMBER's force field is the commonly used force field for proteins. There are different versions of the AMBER protein force field, with ff14SB being the latest. One limitation of the AMBER force field is charges are fixed; thus, it is less accurate than the polarisable force field. In our work, we have used the ff14SB force field.

2.1.2 Carbohydrate force field

Carbohydrates have more degrees of flexibility, thus challenging developing a force field. There are various carbohydrate force fields such as the GLYCAM family, CHARMM, GROMOS 45a4, OPLS, etc. In our work, we have used GLYCAM family force fields.

GLYCAM_06[49] is the most popular series of the carbohydrates force fields. It includes: GLYCAM_06a, GLYCAM_06b, GLYCAM_06e, GLYCAM_06e, GLYCAM_06EP, GLYCAM_06j[50], etc. Among these, the GLYCAM_06j is the most used and latest version.

2.2 MD Simulations

In molecular dynamics simulation, we can predict the time evolution of the macromolecules by solving the Newton's equation of motion. We get the trajectory having every atom cartesian coordinate in every timeframe. Using different analysis tools, the overall data gives some meaningful information at the thermodynamics and structural levels. The biological processes are complex, and wet lab experiments cannot describe the macromolecular interactions with such details.

In the molecular dynamics simulation, by integrating Newton's law of motion, successive configurations of the system are obtained, which is the trajectory that defines the position and velocities of the particles in that system with time. Newton's second law of motion states that 'Force is equal to the rate of change of momentum.

$$F_i = m_i a_i \quad [2.1]$$

where F_i is the force field acting on the particle i of mass m_i having acceleration a_i . The trajectory thus obtained is put into the following differential equation:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad [2.2]$$

$$m_i \frac{d^2 r_i}{dt^2} = f_{x_i} = -\frac{\partial U\,(r_1, r_2 \ldots\ldots r_N)}{\partial r_i} \quad [2.3]$$

where m_i is the mass of the particle along the x_i coordinate with F_{x_i} force on that particle in that direction. Potential energy $U\,(r_1, r_2 \ldots\ldots r_N)$ depends on the coordinates of the N number of particles.

2.3 Initial Conditions

Initially, there is zero overall momentum. Random distribution of velocities is provided, and its magnitude provides the required temperature and is then corrected.

$$p = \sum_{i=1}^{N} m_i v_i = 0 \quad [2.4]$$

By using the Maxwell-Boltzmann at a specific temperature, the random velocities v_i are chosen, which provides the probability that a particle i has v_x velocity at temperature T in the x-direction.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \quad [2.5]$$

Thus, using the velocity, the temperature can be calculated using the relation:

$$T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{[p_i]}{2m_i} \quad [2.6]$$

where N is the total number of atoms in the system.

2.4 Integration Algorithms

The integration algorithms are used to integrate the equations of motions having the positions and velocities of the atoms in the trajectory file. The trajectory file describes the velocities, positions and acceleration of each atom.

We have seen earlier that those atomic positions of all atoms in the biological macromolecule are the function of the potential energy. There is no analytical solution for the potential energy function; thus, we use the numerical algorithms methods to integrate the equation of motion. The types of integration algorithms include verlet algorithm[51], leapfrog algorithm[52], velocity verlet[53], Beeman's algorithm[54], etc. There are a few disadvantages and advantages associated with all these algorithms, and there are specific criteria for using them. All these algorithms use the Taylor series expansion to approximate the acceleration (a_i), velocity (v_i), and position (r_i) of the atoms.

2.4.1 Verlet algorithm

It is the most used integration algorithm. Loup Verlet first used it in 1960.

By expanding the i^{th} particle's position (r_i) at time $t + \Delta t$ and $t - \Delta t$ and b_i is the third derivative of r_i

$$r_i(t + \Delta t) = r_i(t) + v(t)\Delta t + \left(\frac{1}{2}\right) a_i(t)\Delta t^2 + \left(\frac{1}{6}\right) a_i(t)\Delta t^3 + O(\Delta t^4) \quad [2.7]$$

$$r_i(t - \Delta t) = r_i(t) - v(t)\Delta t - \left(\frac{1}{2}\right) a_i(t)\Delta t^2 - \left(\frac{1}{6}\right) a_i(t)\Delta t^3 - O(\Delta t^4) \quad [2.8]$$

In case of the verlet integrator, $O(\Delta t^4)$ is the local error in the position of the atom. Adding the equation

$$r_i(t + \Delta t) + r_i(t - \Delta t) = 2r(t) + a_i(t)\Delta t^2 + O(\Delta t^4) \quad [2.9]$$

By putting the value, $a_i = \frac{F_i}{m_i}$

$$r_i(t + \Delta t) = 2r(t) - r_i(t - \Delta t) + \frac{F_i}{m_i}(t)\Delta t^2 + O(\Delta t^4) \quad [2.10]$$

The velocities can be written as

$$v_i(t) = \frac{1}{2\Delta t}[r_i(t + \Delta t) - r_i(t - \Delta t)] \quad [2.11]$$

Another way for determining the velocities can be obtained by dividing the $r_i(t + \Delta t)$ and $r_i(t - \Delta t)$ at $(t + \Delta t)$ and $(t - \Delta t)$ by central difference formula $(2\delta t)$.

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} \quad [2.12]$$

2.4.2 Leapfrog algorithm

It is a modification of the Verlet algorithm where the velocities are calculated for the time $t + \frac{1}{2}\Delta t$, then positions are estimated at the $t + \Delta t$. Hence, the positions have leap over the velocities.

$$r(t + \Delta t) = r(t) + v(t + \frac{1}{2}\Delta t)\Delta t \quad [2.13]$$

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + a(t)\Delta t \quad [2.14]$$

2.4.3 Velocity Verlet algorithm

This algorithm calculates the positions, velocities and the acceleration at the same time $(t + \Delta t)$

$$r(t + \Delta t) = r(t) + \Delta t\, v\Delta(t) + \frac{1}{2}\Delta t^2 a(t) \quad [2.15]$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}\Delta t[a(t) + a(t + \Delta t)] \quad [2.16]$$

2.4.4 Beeman's algorithm

This algorithm allows a large number of atoms in MD simulation. This is another modification of the Verlet algorithm. It creates a position similar the Verlet algorithm but with different velocities.

$$r_i(t + \Delta t) = r_i(t) + v(t)\Delta t + \frac{1}{6}[a_i(t + \Delta t) - 2a_i(t)]\Delta t^2 + O(\Delta t^4) \quad [2.17]$$

$$v_i(t + \Delta t) = v_i(t) + \frac{1}{12}[5a_i(t + \Delta t) + 8a_i(t) - a(t + \Delta t)]\Delta t + O(\Delta t^4) \quad [2.18]$$

2.5 Timestep

The time step selection during MD simulations is one of the crucial factors. The timestep refers to the frequency at which the integration is performed. The larger time step will lead to instabilities in the integration algorithm, while the smaller time step will be computationally expensive and only limited phase space will be covered. Generally, 1 fs or lower time step is preferred for the MD simulations for the biological macromolecules to measure the fastest motion of the system accurately. SHAKE[55] or LINCS[56] algorithm is used for constraining all bonds involving hydrogen atoms. This allows us to use a larger time step of 2 fs. Thus, most commonly, 1-2 fs timestep is frequently used with restraint on bonds involving hydrogen atoms.

2.6 Periodic boundary conditions

Periodic boundary conditions are needed for the approximation of the bulk properties. These tend to reduce the effects of the finite simulation

box. Using the periodic boundary conditions, the particles in the simulation box feel placed in the bulk fluid. The simulation box is replicated along all its sides. When any particle leaves the simulation box, the mirror of that particle enters the simulation box as if the simulation box is infinite[57] (**Figure 2.2**).

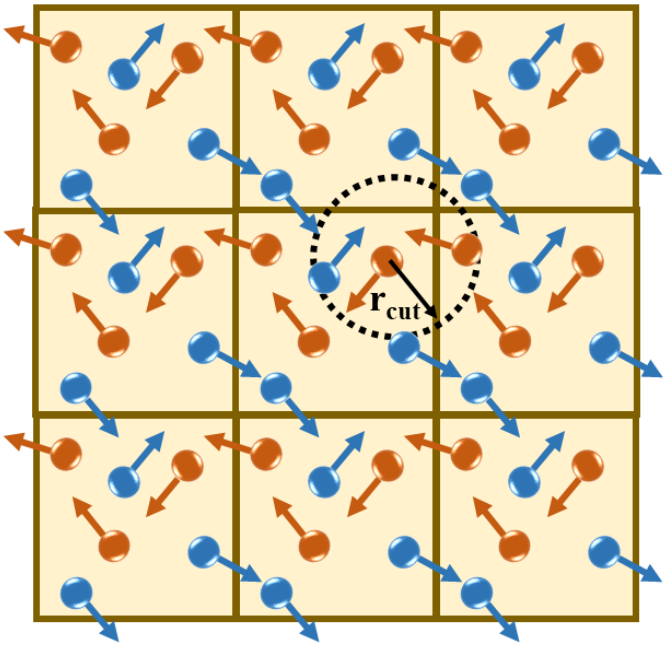

Figure 2.2: 2-D representation of periodic boundary conditions in the simulation box surrounded by the mirror boxes.

2.7 Long-range interactions

There can be two types of interactions, bonded and non-bonded interactions. Calculating the bonded interactions connected via the covalent bonds is computationally less expensive. In contrast, if we consider the nonbonded interactions, there can be many two atoms pairs in the system which have any interaction. Calculating this non-bonded interaction for every two-atom pair will be computationally very expensive. Generally, we ignore the Lennard-Jones and coulombic potentials after the specific distance as their effect is negligible. It is insignificant as compared to the error in simulation. This is the main reason why a cut off scheme is used beyond a specific distance. There are few methods for calculating the long-range forces like Ewald summation[58], the cell multiple methods and the reaction field method.

2.8 Thermostats

A constant temperature ensemble is required to mimic the experiments conducted in laboratory conditions. The temperature is measured using the kinetic energies defined using the equipartition theorem:

$$\frac{3}{2}NK_BT = \langle \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \rangle \quad [2.19]$$

Different thermostat algorithms work by alternating the Newton equations of motions that are by default at constant energy. Thermostats should not be used to calculate dynamical properties like diffusion coefficients. Also, it is recommended to turn off the thermostat after the equilibration of the system, which attended to the desired temperature. Some of the popular thermostats used in the MD are gaussian[59], Berendsen[60], Bussi-Donadio-Parrinello[61], Anderson[62], langevin[63], etc.

2.8.1 Gaussian Thermostat

The primary purpose of the Gaussian thermostat is to keep the instantaneous temperature and desired temperature constant. This is done via:

$$F = F_{interaction} + F_{constraint} \quad [2.20]$$

Where $F_{interaction}$ is the standard interaction interactions and $F_{constraint}$ is the Langrage multiplier, which helps in keeping kinetic energy constant. The Gaussian barostat uses the gaussian principle of least constraint in maintaining the instantaneous temperature.

2.8.2 Langevin Thermostat

Langevin thermostat considers the microcanonical equations of motions along with the Brownian dynamics. Also, the viscosity and the random collision effects of an implicit solvent.

$$F = F_{interaction} + F_{friction} + F_{random} \quad [2.21]$$

Where $F_{friction}$ is the damping parameter and F_{random} is the random collisions with the molecules of solvent. We have used Langevin Thermostat for our work.

2.9 Barostats

The laboratory experiments are generally done in the open air under constant temperature and pressure. These conditions are also maintained during the MD simulations using the isothermal-isobaric ensemble. Very similar to the thermostats, we use Barostats to maintain the constant pressure during MD simulation. Some popular barostats are known as Berendsen[60], Anderson[62], Martyna-Tuckerman-Tobias-Klein[64], etc.

2.9.1 Berendsen barostat

It is based on the simple volume rescaling method. Berendsen barostat should be used up to the equilibration and shouldn't be used in the production run. In Berendsen barostat, the pressure change is modified by adding an additional term to the equations of the motion.

$$\frac{dP}{dt} = \frac{P_o - P}{T_p} \qquad [2.22]$$

Where P_o is the external pressure bath, P is the instantaneous pressure and T_p is the time constant. We have used the Berendsen barostat for our work.

2.10 Protocols of MD Simulation

Every system comes up with its unique considerations and challenges. But the basics steps of molecular dynamics simulation remain the same:

1. System preparation
2. Solvation
3. Minimisation
4. Heating
5. Equilibration
6. Production

2.10.1 System Preparation

The selection of the appropriate structure (protein, ligand, glycans, lipids, etc.) is the first step toward setting up the system. The 3-D conformation of these structures can be downloaded from the RCSB

PDB website or can be modelled. In the case of glycan modelling, the GLYCAM web server is a widely used server[65]. It should also be electrically neutral, and protonation states should be checked.

2.10.2 Solvation

All biological processes take place in the body fluid consisting of water. We provide water (implicit or explicit) to the molecule. There are two types of water models: implicit and explicit. In the implicit water models (also known as the continuum model), the effects of the solvent molecules are given to the system by the reaction field method. While in the case of the explicit water models, the molecular contribution of each and all solvent molecule's atoms is taken into consideration directly. An explicit solvation system is more computationally expensive and more accurate than implicit water models, which are less computationally expensive.

The most used water models are the 3-site model, where each site is for the nuclei of the water molecule. The 3-site model includes TIP3P[66], SPC[67], SPC/E[68] and TIPS[69] models. The TIP3P is the most used model for two reasons: it is computationally efficient and compatible with the current force fields. The TIP5P[70] and TIP6P[71] are computationally expensive. We have used TIP3P as the explicit water model for glycans and the protein-glycan complexes in our work.

2.10.3 Minimisation

The main goal of minimisation is to find a local energy minimum of the starting structure. This is necessary as when we heat the system in the further steps, the system shouldn't blow up, which means the forces on any atom shouldn't be large enough that the atom will move too far. There are various minimisation algorithms like Newton Raphson, steepest descent and conjugate gradient method.

2.10.4 Heating

Since there is no temperature in the inputs structure, there is no energy in the system. Hence, in this step we have to gradually increase the kinetic energy of the system to reach the desired temperature. The

kinetic energy of the system should be gradually increases over a proper time frame to avoid the sudden increase in the kinetic energy, making the simulation unstable. In the NVE ensemble, energy cannot be added, In NPT ensemble the kinetic energy is needed to expand the system to maintain the constant pressure. Based on above reasons, the ensemble for the heating step is NVT.

2.10.5 Equilibration

Since, the production run is performed under the NPT ensemble, a buffer time is required between the heating and production run to switch the ensemble. In equilibration, we make sure that the system remains in thermodynamic equilibrium and equilibrates the density of the solvent. The parameters of both the production and equilibration stages remains same.

2.10.6 Production

After the completion of the equilibration, we start the collection of data for analysis. This step is known as production. The MD trajectory is saved at given intervals of time to capture the whole event. At this stage, the trajectory analysis starts. The typical simulation length for the biological macromolecules ranges from nanoseconds to microseconds.

2.11 Binding free energy calculations

Calculating the binding free energy is the most crucial factor in the protein-ligand/glycan binding. Binding free energy is the sum of all interactions present between the binding of the protein and ligand/glycan. The Free energy perturbation method gives more accurate results while being computationally expensive. While the MM/PB(GB)SA[72,73] method is computationally efficient and precise. We have used MM/GBSA for our work[74,75].

2.11.1 Molecular mechanics Poisson-Boltzmann/generalised Born surface area (MMPB/GBSA)

The MM/PB(GB)SA is one of the most widely used methods for calculating the binding free energy between the protein-ligand/glycan complex. This method uses both the continuum solvation models and

classical force fields to estimate binding free energy. The calculation of the binding free energy is described below[76,77]:

$$\Delta G_{bind} = G(complex) - G(receptor) - G\ (ligand) \qquad [2.23]$$

Each of the complex, receptor and ligand-free energy is given by (represented by *x*):

$$G(x) = E_{MM}(x) + G_{solv}(x) - TS\ (x) \qquad [2.24]$$

$$E_{MM}(x) = E_{bonded}(x) + E_{nonbonded}(x) \qquad [2.25]$$

$$E_{bonded}(x) = E_{bond}(x) + E_{angle}(x) + E_{dihedral}(x) \qquad [2.26]$$

$$E_{nonbonded}(x) = E_{electrostatics}(x) + E_{vdW}(x) \qquad [2.27]$$

where, G_{solv} is the solvation free energy, T is the temperature, S is the entropy, E_{MM} is the total molecular mechanic's energy. The bonded terms include bond (E_{bond}), angle (E_{angle}) and dihedral ($E_{dihedral}$). While the non-bonded term includes electrostatics ($E_{electrostatics}$) and van der Waals (E_{vdW}). Adding all these:

$$\Delta G_{bind} = \Delta G_{MM} - \Delta G_{solv} - T\Delta S \qquad [2.28]$$

$$E_{MM} = \Delta E_{bonded} + \Delta E_{electrostatics} + \Delta E_{vdW} \qquad [2.29]$$

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{vdW} + \Delta G_{elec} = \Delta G_{SASA} + \Delta G_{PB/GB} \qquad [2.30]$$

$$\Delta G_{SASA} = \gamma . SASA + \beta \qquad [2.31]$$

where G_{cav} is the energy needed for making cavity in the solvent for accommodating solute, γ is the surface tension proportionality set at 0.92 Kcal.mol^{-1} Å^{-2} and β is offset value set to 0.92 Kcal.mol^{-1}. The electrostatic contribution can be estimated by either PB[78] or GB[79] model. MMPBSA method is used for protein-ligand/protein, peptide interactions. However, in case of protein-glycan interactions, MMGBSA is used widely used technique[80–85].

Chapter 3

Materials and Methods

3.1. System preparation

We need the crystal PDB structures for the MD simulation, which can be downloaded from the RCSB PDB website. The PDB file consists of every atom's x, y and z coordinates, along with other details. Our study broadly simulated three different types of systems: free glycans, apo PltB protein and complex (PltB protein-glycan complex).

Table 3.1: List of PDBs used in our study.

PDB ID	Resolution	Attached Glycans
4RHR	2.08Å	NA
6P4M	1.80Å	Neu5Ac2α2-3Galβ1-4GlcNAc
6P4N	1.70Å	Neu5Ac2α2-6Galβ1-4GlcNAc
6TYN	2.33Å	Neu5,9Ac2α2-3Galβ1-4GlcNAc
6TYQ	1.88Å	Neu5,9Ac2α2-6Galβ1-4GlcNAc
6TYO	2.04Å	Neu5,4Ac2α2-3Galβ1-4GlcNAc

We divided the five glycans into modified (*Gly1*, *Gly2* and *Gly3*) and unmodified (*Gly4* and *Gly5*) systems, where modification is lying on the Neu5Ac residue. The glycans used in our study are shown in **Table 3.2** & **Figure 3.1**.

Table 3.2: Sequence information and type of modifications of glycans used in our study.

Glycan details	PDB ID	Nature of Modification at Neu5Ac	System name
Neu5,9Ac2α2-3Galβ1-4GlcNAc	6TYN	9-O-acetylation	***Gly1***
Neu5,9Ac2α2-6Galβ1-4GlcNAc	6TYQ	9-O-acetylation	***Gly2***
Neu5,4Ac2α2-3Galβ1-4GlcNAc	6TYO	4-O-acetylation	***Gly3***

Neu5Ac2α2-3Galβ1-4GlcNAc	6P4M	NA	***Gly4***
Neu5Ac2α2-6Galβ1-4GlcNAc	6P4N	NA	***Gly5***

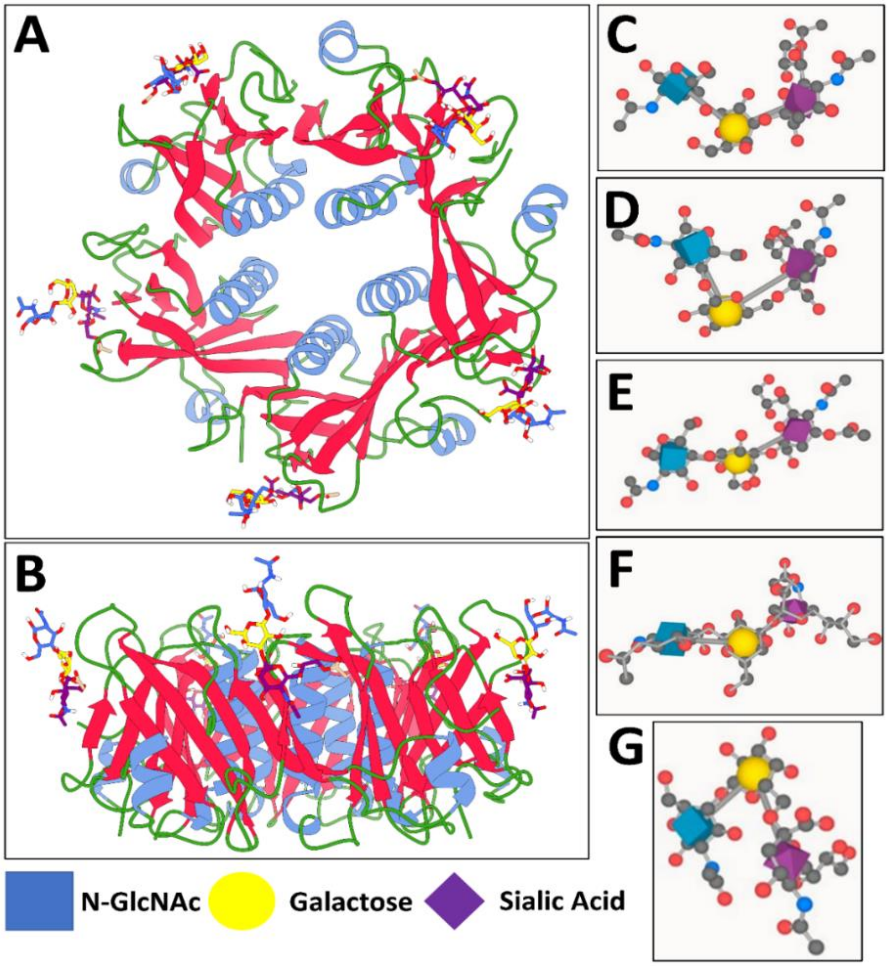

Figure 3.1: Ribbon representation of PltB of Typhoid toxin and glycan molecules. A) Top view, B) Lateral view. SNFG representation of all five glycans used in our study; C) *Gly1*, D) *Gly2*, E) *Gly3*, F) *Gly4*, G) *Gly5*.

To get insights into Typhoid toxin-free state dynamics, we also conduct the simulation of Apo PltB protein. Finally, we constructed the complexes for all five glycans. As for now, we constructed all five complexes by occupying the BS1 site for each monomer. This construction aims to understand each binding site's glycan dynamics and recognition properties. Future work will also include the binding properties in the other two binding sites, namely BS2 and BS3. All five glycans were extracted from their concerned complex crystal structures.

The list of the apoprotein and complex used in our study is shown in **Table 3.3**.

Table 3.3: The details of the apoprotein and the complex systems.

Protein	**PDB ID**	**Glycans attached at BS1 site**	**System name**
PltB	4RHR	NA	**Apo**
PltB	6TYN	*Gly1*	***Com1***
PltB	6TYQ	*Gly2*	***Com2***
PltB	6TYO	*Gly3*	***Com3***
PltB	6P4M	*Gly4*	***Com4***
PltB	6P4N	*Gly5*	***Com5***

For each complex PDB structure, we extracted all the available bound glycans for each system and estimated their dihedral angles and the puckering states using the *cpptraj* module of the AmberTools19[86]. After critical inspection of all the values, the input structure for the free glycans was selected. All the monosaccharides were renamed according to the GYCAM nomenclature to recognise the amber force field. For solvating the glycans in 150 mM salt concentration, receptors and other molecules were removed for each case.

Glycam06j_1 force field was used to estimate the glycan structures in our study[87]. Each glycan system was solvated using TIP3P water molecules in an octahedron box[88]. The distance between the wall and solute was kept as 10 Å. To solvate in 150 mM concentration, the proper amount of Na+ and CL- ions were added to the systems. Details of the box size, water molecules and the number of ions is listed in **Table 3.4**. These parameters were varied because of the initial conformations of each glycan molecule.

Table 3.4: Details of water-box and charge constituent for all the systems.

	Volume (A^3)	**Charge**	**Na+ ions**	**Cl- ions**
Gly1	74304.8	-1	8	7
Gly2	70352.6	-1	7	6
Gly3	73584.5	-1	8	7
Gly4	70052.2	-1	7	6
Gly5	50132.9	-1	7	6
Apo	501335.4	0	45	45
Com1	590794.7	-5	58	53
Com2	767711.4	-5	74	69
Com3	602318.1	-5	59	54
Com4	650845.8	-5	64	59
Com5	686121.1	-5	67	62

For the protein part of our study, ff14SB force field was used[89]. The Apo simulation, i.e., the PltB pentamer structure in the absence of glycan molecules, were downloaded from the rcsb database (PDB ID 4RHR)[90]. A similar protocol was used for the glycan to solvate the Apo structure. Details of the system parameter can be found in **Table 3.4**.

Now, to construct the complex structure, we used three glycans (*Gly1*-3) or the modified glycan molecule. However, the crystal structure for all three complexes was the absence of glycan molecules in their few binding sites. As PltB protein consists of five identical monomer structures, each chain has a binding pocket for the glycan molecules. However, several experimental studies suggest that each monomer consists of three binding sites, namely BS1, BS2 and BS3. However, for this study, we choose to construct our complex molecules with respect to the BS1 site. As the few binding pockets were the absence of glycan molecules, we constructed the whole complex by swapping the geometrical orientation of each monomer. After the construction of each complex, a similar water box was used to solvate

them. As discussed above, complex structures were also solvated in the 150 mM salt concentration. Details of the systems are given in **Table 3.4**. All simulations were conducted using the AMBER18 software suite[42].

3.2. MD simulation protocol

3.2.1 Minimisation

We had done two minimisation steps for all glycan, Apo protein and complex systems. In the first minimisation, we restraint our systems with the help of weak restraint force, followed by minimisation in the free state. In the first step minimisation, initially, (first 500 steps) system was minimised using the steepest descent algorithm, and then we used the conjugate gradient algorithm for the minimisation (next 500 steps). Then, we remove the restrictions on the position of atoms and again optimise the complex using the steepest descent algorithm (first 100 steps) followed by a conjugate gradient algorithm (next 900 steps). The bond, including hydrogen atoms, were kept restrained using the SHAKE algorithm[91]. The long-range electrostatic interactions were estimated using the particle mesh Ewald (PME) method, where we kept 10 Å as a cut-off.

3.2.2 Heating

After minimisation, the heating of all the systems was performed in steps wise. The Langevin thermostat maintained the temperature with a collision frequency of 2 ps^{-1}. It is always better to do heating in stages as it reduces the chances of the blow-up of the whole system after every stage as it equilibrates the system. We ran heating up to 300 K in six stages (0 to 50 K, 50 K to 100 K, 100 K to 150 K, 150 K to 200 K, 200 K to 250 K and 250 K to 300 K). In each stage of the heating, there are 50000 steps. The Timestep for our simulation was chosen two fs to keep match with the vibration of the systems.

3.2.3 Equilibration

After minimisation, we equilibrated the system at 300 K for the 10 ns explicit solvent MD simulation time in the NPT ensemble. At this

equilibration step, the temperature was kept constant to equilibrate the system more at 300 K, and the pressure was kept constant using the Berendsen barostat[92] at 1 atm pressure.

3.2.4 Production

After equilibration, the final production run was performed at the constant 300 K using the Langevin thermostat. For free glycan simulation, we ran each glycan sequence for 2 μs and replicated the same for three replicas yielding 6 µs complete run for each glycan. In this study, the Apo protein and complex structure were simulated up to 500 ns and replicated for three replicas yielding a 1.5 μs complete run for each apo and complex. Thus, we have a total of 30 μs length simulation for glycans, 1.5 μs for apo and 7.5 μs for complexes.

3.3. Analysis

All the trajectories were analysed using the *cpptraj* module of the AmberTools19[86]. For the trajectory analysis, the first 100 ns data from each run were removed to ensure the minimisation of the initial noises. After that, the combined trajectories for three replicas were prepared and subjected to different calculations related to the dynamics of the systems. Initial stability and flexibility were estimated by calculating the root mean squared deviations (RMSD) and the root mean squared fluctuations (RMSF) with respect to the well-equilibrated conformations. Conformational space of the sugar molecules in the free and bound state was also estimated with the help of glycosidic dihedral angle, Cremer-Pople (CP)[93] parameters of ring puckering. Further, free energy surface (FES) was constructed for reaction coordinates estimated in our studies (i.e., dihedral angles, puckering coordinates) by using the Boltzmann equation:

$$\Delta G = -k_B T \, ln(\rho) - k \qquad [3.1]$$

where k_B is the Boltzmann constant, T is the temperature, and ρ is the probability density of the geometric coordinate x. The parameter (k) was selected such that the global minimum was located at 0 kcal/mol. Also, to evaluate the hydrogen bonding, a distance cut-off of ≤ 3.5 Å and an

angle cut-off of ≥120º were used[94]. The last 200 ns of production simulations were taken for the binding free energy calculations of the complexes. Entropy estimation was ignored because of the high computational cost.

Chapter 4

Results and Discussion

4.1. Conformational dynamics of glycans

To evaluate the conformational space, the RMSD analysis of all the glycans was estimated with respect to the initial heavy atom coordinates. The probability distribution of the RMSD values for all cases were estimated using kernel density estimation (KDE) method and shown in **Figure 4.1**. Multipe coexisting conformations were observed in all glycans as observed from the widespread distribution of RMSD. Among all glycans, *Gly2* and *Gly5* show a large conformational space range from 0 to 5 Å. *Gly1* has two peaks corresponding to the RMSD between 1 Å and 2 Å. *Gly2* has a peak at 1 Å and is spread up to 5 Å. *Gly3* has one peak at 3 Å, and it is not as spread as others. *Gly4* has one peak at around 1 Å, and *Gly5* has one peak at 1 Å and is spread to the RMSD 0 Å to 5 Å. The modification of the oxygen atoms of the sialic acid residue does not yield any significant changes in the conformational space for all the glycans. So from the RMSD analysis, characteristic features of the glycans were not obtained.

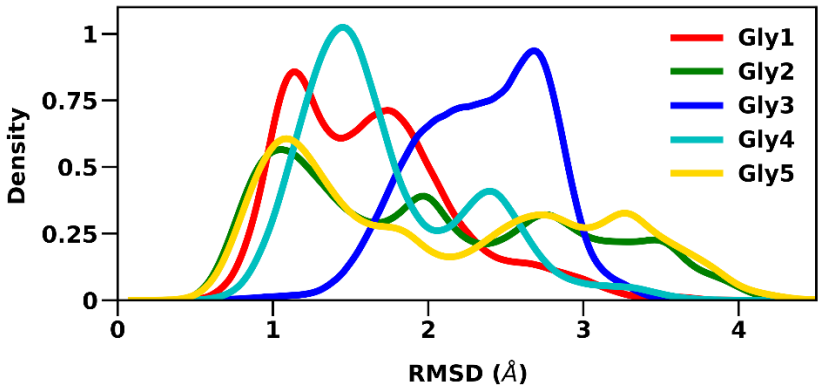

Figure 4.1: Probability distribution plot of root-mean-square-deviations (RMSDs) of all the five glycans with respect to the corresponding well-equilibrated structure.

To estimate the conformational space after boudn to the typhoid toxin, similar calculation was conducted in case of the bound glycans.

Hence, the probability distribution of RMSD for each glycans in free and bounds were estimated using KDE and shown in **Figure 4.2**. The conformational space of free *Gly2* and complex-bound *Gly2* are highest among others. In all cases, the five complex-bound glycans have similar conformational space while free glycans have sightly different conformation. These indicate a stable binding aThe modification of the glycans does not yield any significant changes compared to unmodified glycans in the RMSD analysis of the glycan-protein complex.

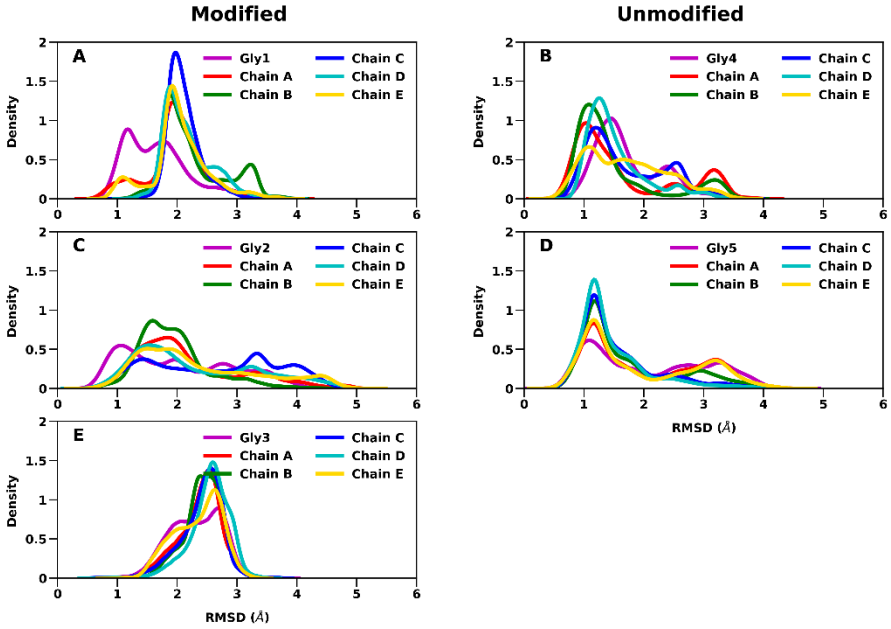

Figure 4.2: Probability distribution plot of Root-mean-square-deviations (RMSDs) of free glycans and the complex bound glycan in each chain. A) *Gly1* and *Com1* B) *Gly4* and *Com4* C) *Gly2* and *Com2* D) *Gly5* and *Com5* E) *Gly5* and *Com5*.

4.2 Dynamics of the sugar ring

To estimate the fluctuation around dihedral angles, we calculate the ϕ/ψ values for each crystal structure of all the five glycans in the BS1, BS2 and BS2 binding sites, as shown in **Table 4.1**. As observed in **Table 4.1**, each glycan has several conformations in the binding site of the pentamer Typhoid pentamer. This estimation further helps us to choose the proper input for MD simulations. The conformations, which occur more frequently, were selected as an input to the MD simulation study.

Table 4.1: Conformation of dihedral angle of all five glycans from the crystal structures.

Gly1						
	ϕ			**ψ**		
Linkage	**BS1 A**	**BS1 C**	**BS1 E**	**BS1 A**	**BS1 C**	**BS1 E**
1-4	-55.34	2.92	-135.50	148.91	140.28	93.65
2-3	101.69	66.08	80.95	-123.04	-132.27	-108.25
Gly2						
	ϕ			**ψ**		
Linkage	**BS1 C**	**BS1 D**	**BS1 E**	**BS1 C**	**BS1 D**	**BS1 E**
1-4	61.08	59.79	58.82	124.14	128.34	125.87
2-6	61.48	73.39	59.63	-166.59	-153.80	-157.60
Gly3						
	ϕ			**Ψ**		
Linkage	**BS1 C**	**BS1 E**	**BS2 A**	**BS1 C**	**BS1 E**	**BS2 A**
1-4	-50.05	-66.03	-39.32	114.87	119.52	-79.28
2-3	62.22	62.59	55.10	-124.88	-139.47	-125.74
Gly4						
	ϕ			**Ψ**		
Linkage	**BS1 A**	**BS1 C**	**BS1 E**	**BS1 A**	**BS1 C**	**BS1 E**
1-4	71.26	-42.02	-51.88	131.67	168.19	134.37
2-3	147.51	59.54	64.30	70.95	-123.58	-131.05
Gly5						
	ϕ			**Ψ**		
Linkage	**BS1 C**	**BS1 D**	**BS1 E**	**BS1 C**	**BS1 D**	**BS1 E**
1-4	-74.54	-43.84	-39.46	101.36	-172.50	140.00
2-6	62.39	61.08	68.06	-177.91	-154.75	-165.37

For estimating the free energy space constructed by dihedral angles of the five free glycans, we used the combined trajectories for all three replicas, as shown in **Figure 4.3**. There are two glycosidic linkages in the trisaccharide glycans; one is Neu5Ac2α2-3Gal (2-3) or Neu5Ac2α2-

6Gal (2-6), and the other is Galβ1-4GlcNAc (1-4) which is present across all five sequences.

For the Galβ1-4GlcNAc (1-4) linkage, principal minima around (-75,120) with minor variation were observed in glycans. So, irrespective of modification, all the glycans show a single prominent minimum. However, the Neu5Ac2α2-3Gal (2-3) linkage in *Gly1*, *Gly3*, and *Gly4* show two equiprobable minima regions located at (65.00,-125.00) and (65.00,-175.00). Also, this value changes when the linkage changes from 2-3 to 2-6 where another minima region appeared (65,150) along with the (65.00,-175.00) conformation. However, this region shrinks in the case of the *Gly5*.

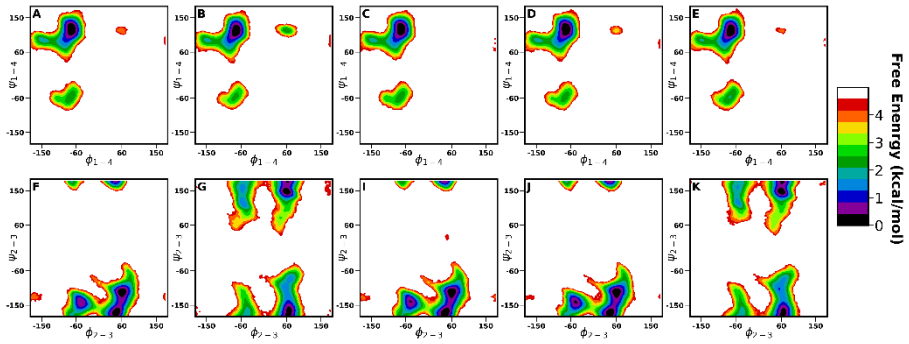

Figure 4.3: Free energy surfaces (kcal/mol) constructed for the free glycans using the φ/ψ glycosidic angles over all three replicas.

To compare the free energy space constructed by dihedral angles of the glycans bounded to the different protein chains, we used the combined trajectories for all three replica runs shown in **Figure 4.4 – Figure 4.8**. We oberserved very similar principal free energy minima for all the complex-bound glycans as compared to the principal free energy minima of free glycans.

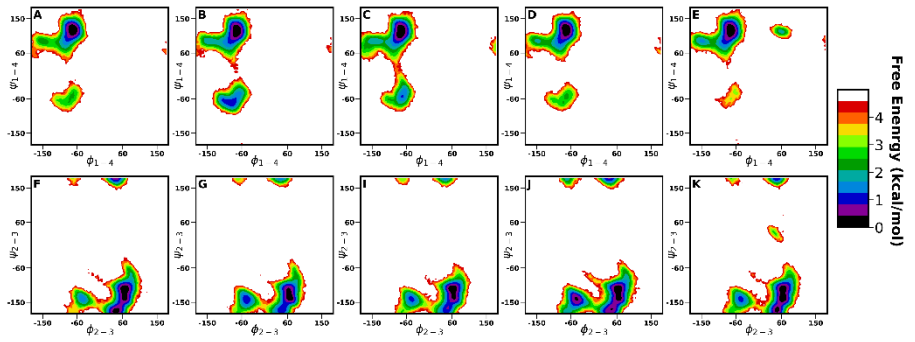

Figure 4.4: Free energy surfaces (kcal/mol) constructed for the complex bound glycans of *Com1* using the φ/ψ glycosidic angles over all three replicas.

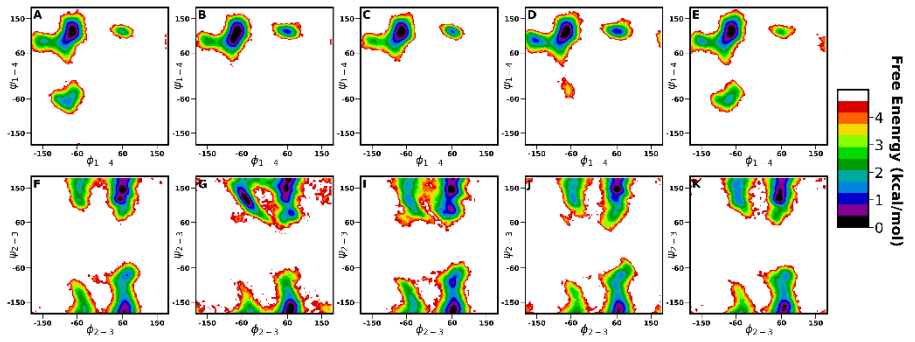

Figure 4.5: Free energy surfaces (kcal/mol) constructed for the complex bound glycans of *Com2* using the φ/ψ glycosidic angles over all three replicas.

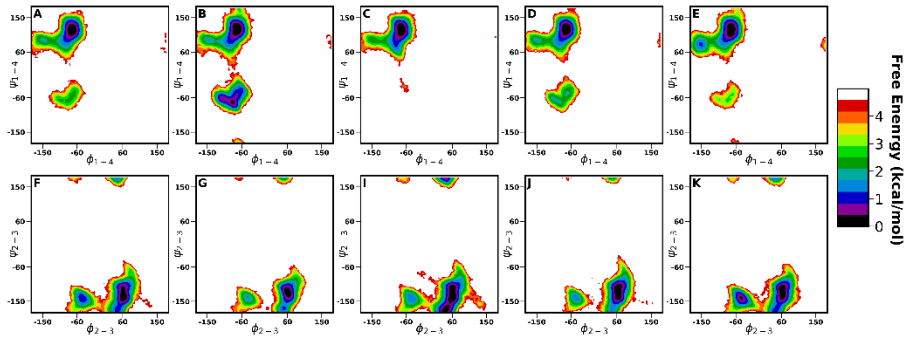

Figure 4.6: Free energy surfaces (kcal/mol) constructed for the complex bound glycans of *Com3* using the φ/ψ glycosidic angles over all three replicas.

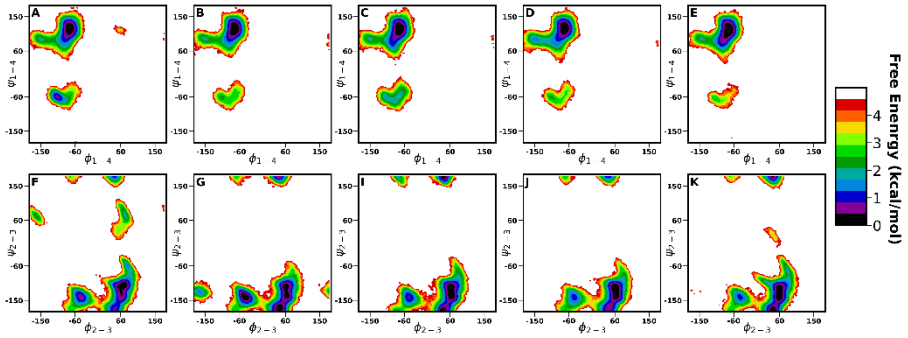

Figure 4.7: Free energy surfaces (kcal/mol) constructed for the complex bound glycans of *Com4* using the φ/ψ glycosidic angles over all three replicas.

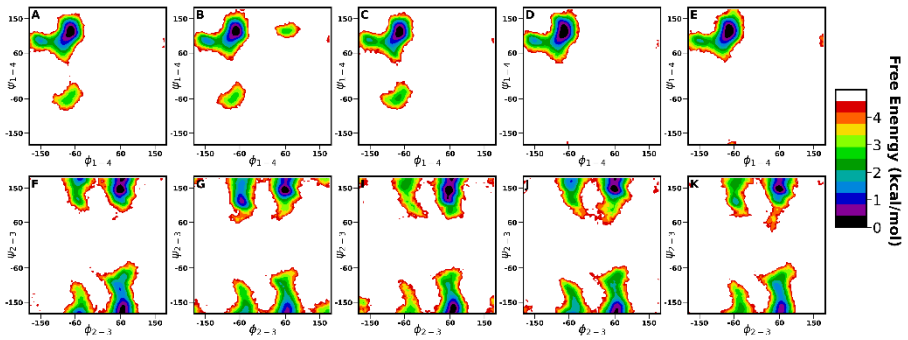

Figure 4.8: Free energy surfaces (kcal/mol) constructed for the complex bound glycans of *Com5* using the φ/ψ glycosidic angles over all three replicas.

To investigate the conformation of the ring around the 150 mM solution, puckering of the carbohydrate rings was estimated using the Cremer-Pople convention, which further divided the puckering space into 38 different and distinct conformations. Among all the conformation, the most probable conformations are $^{4}C_{1}$ and $^{1}C_{4}$ chair forms. So, to investigate the transition between both the chair forms, we constructed a 1-D conformational free energy along with the θ coordinate of the pucker for all the five glycans (in free and bound state), as shown in **Figure 4.9 – Figure 4.13.**

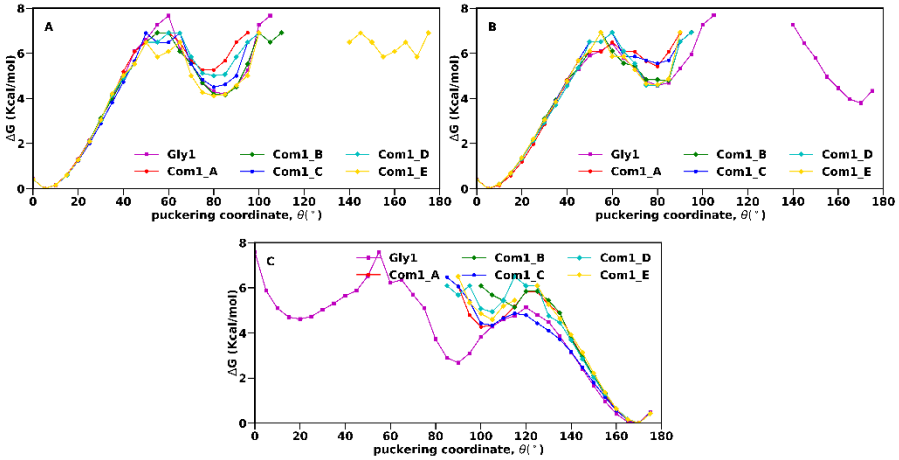

Figure 4.9: Cremer-Pople puckering profiles of *Gly1* and the complex bound glycan *Gly1* to different chains (A, B, C, D and E) of *Com1*. A) For GlcNAc B) For Gal C) For Neu5Ac.

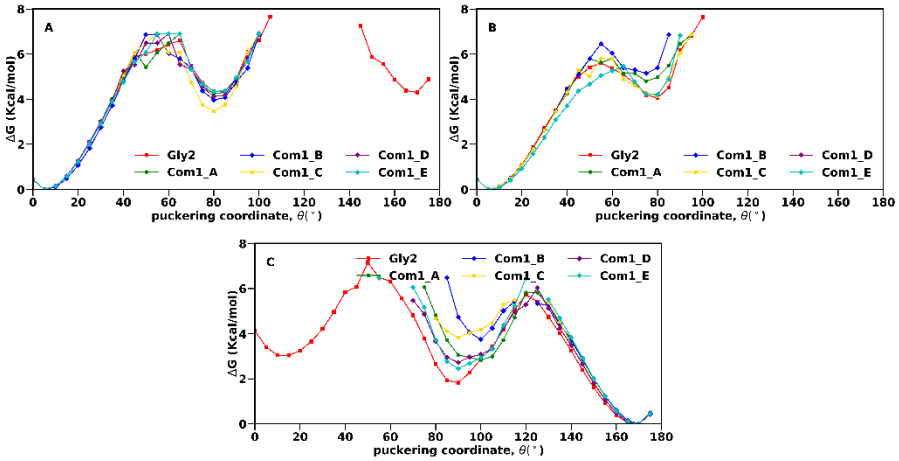

Figure 4.10: Cremer-Pople puckering profiles of *Gly2* and the complex bound glycan *Gly2* to different chains (A, B, C, D and E) of *Com2*. A) For GlcNAc B) For Gal C) For Neu5Ac.

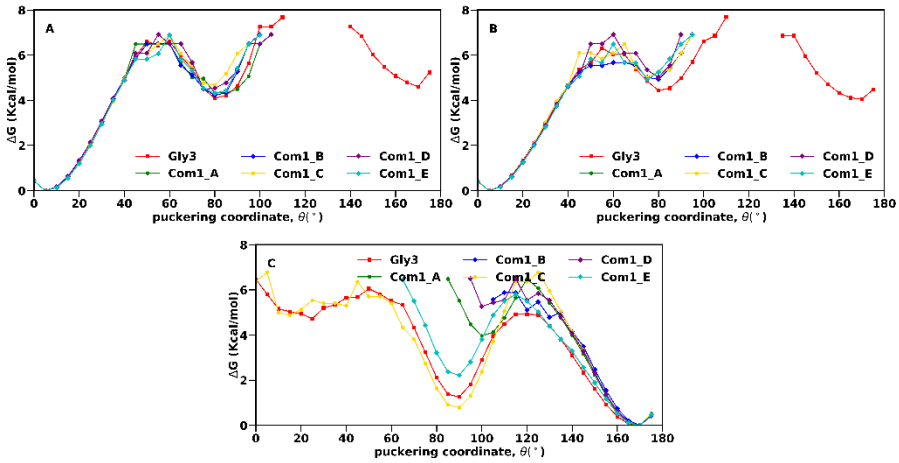

Figure 4.11: Cremer-Pople puckering profiles of *Gly3* and the complex bound glycan *Gly3* to different chains (A, B, C, D and E) of *Com3*. A) For GlcNAc B) For Gal C) For Neu5Ac.

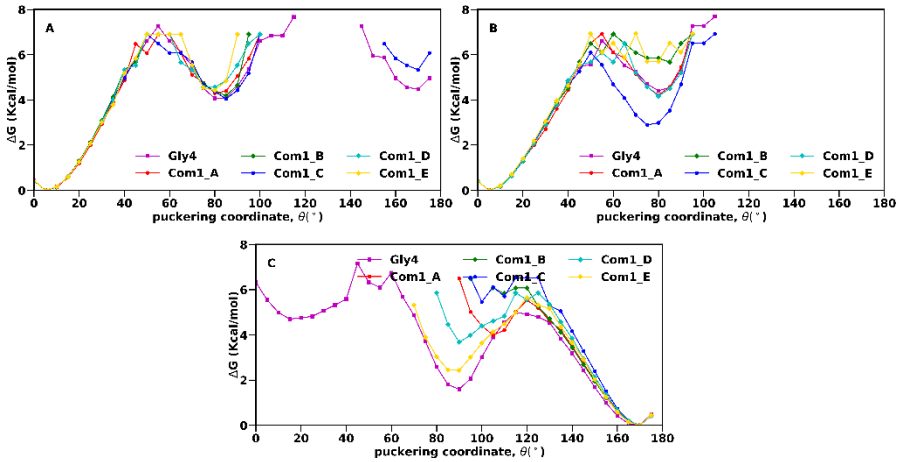

Figure 4.12: Cremer-Pople puckering profiles of *Gly4* and the complex bound glycan *Gly4* to different chains (A, B, C, D and E) of *Com4*. A) For GlcNAc B) For Gal C) For Neu5Ac.

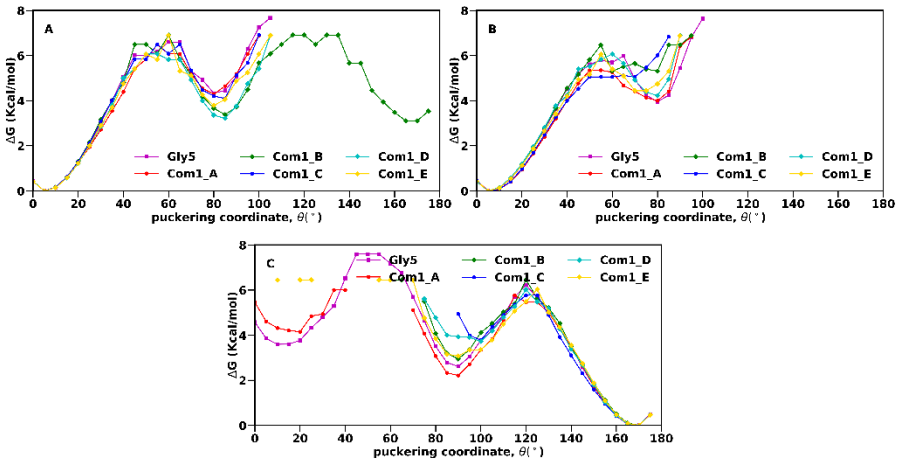

Figure 4.13: Cremer-Pople puckering profiles of *Gly5* and the complex bound glycan *Gly5* to different chains (A, B, C, D and E) of *Com5*. A) For GlcNAc B) For Gal C) For Neu5Ac.

For all the three-carbohydrate residue, GlcNAc, Gal, and Neu5Ac (Modified/Unmodified) shows a similar transition profile from $^4C_1$ to $^1C_4$ in all 5 cases. In our simulation time-length, GlcNac and Gal monosaccharides show $^4C_1$ chair form as its global minimum, while Neu5Ac shows $^1C_4$ chair form as global minimum. Apart from the sialic acid (in free glycans), neither the monosaccharides could not complete the transition from $^4C_1$ to $^1C_4$ state. Also, sialic acid in which the modification was done shows merely any changes in its puckering sampling. To further investigate the other conformers other than the chair forms, we plotted the puckering coordinates (θ,ϕ) according to the Mercator representation shown in **Figure 4.14 – Figure 4.19**. Here we show the Mercator representation for the free glycans and bound glycans. GlcNAc and Galactose residue only samples the $^4C_1$ along with very few conformations in skewed boat structure in both the free glycan and complex-bound glycans. In the case of sialic acid, $^1C_4$ conformational state was sampled in both the free glycans and comple-bound glycans, also few non-chair conofrmations were also sampled in the free glycans. The modification of the sialic acid doesn't bring any change in the conformation, which remains in $^1C_4$ state in all cases. So, only chair form was sampled in all three monosaccharides in complex-bound glycans cases, which play a role in recognising PltB protein.

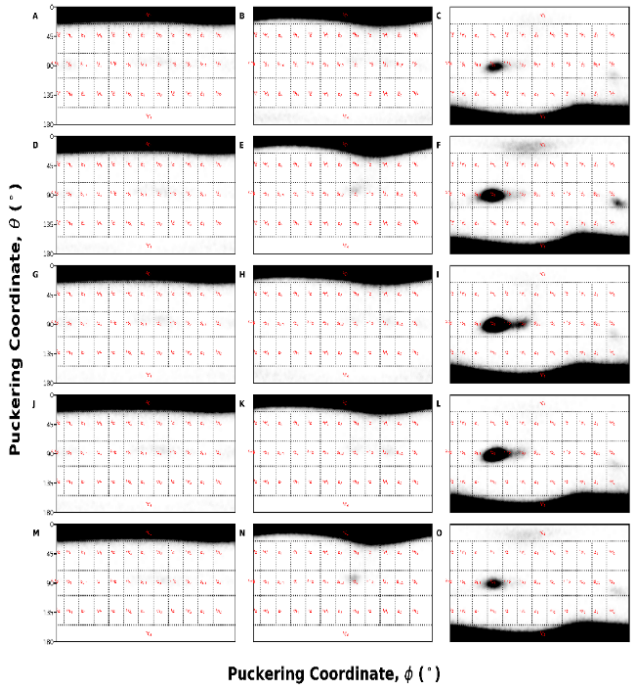

Figure 4.14: Mercator representation of the Cremer−Pople Sphere for free glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

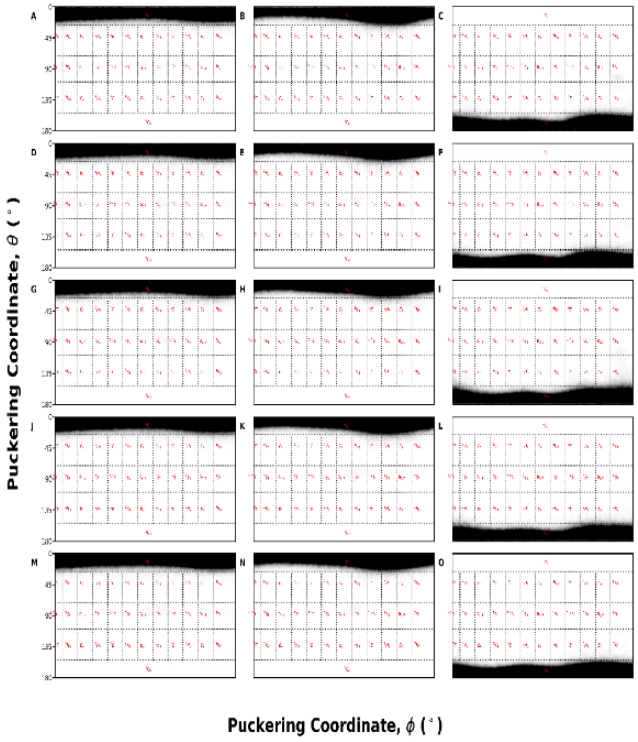

Figure 4.15: Mercator representation of the Cremer−Pople Sphere for *Com1* complex-bound glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

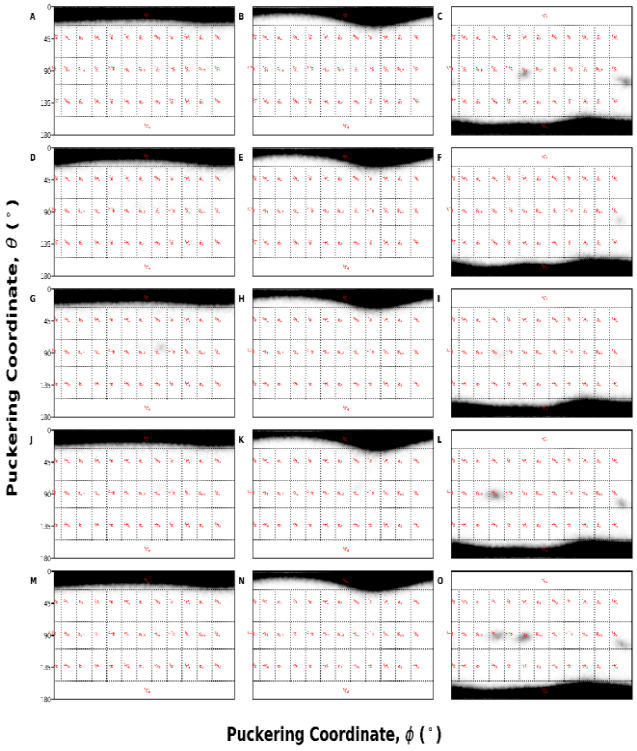

Figure 4.16: Mercator representation of the Cremer–Pople Sphere for *Com2* complex-bound glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

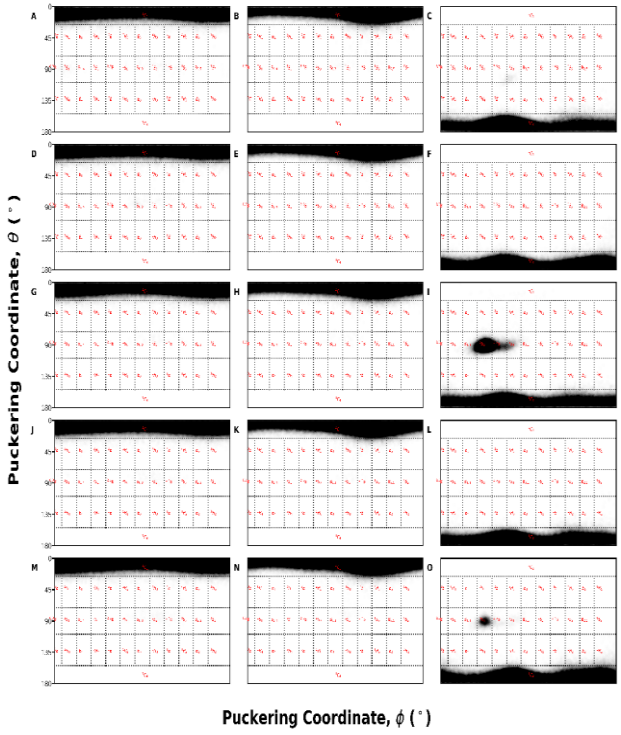

Figure 4.17: Mercator representation of the Cremer–Pople Sphere for *Com3* complex-bound glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

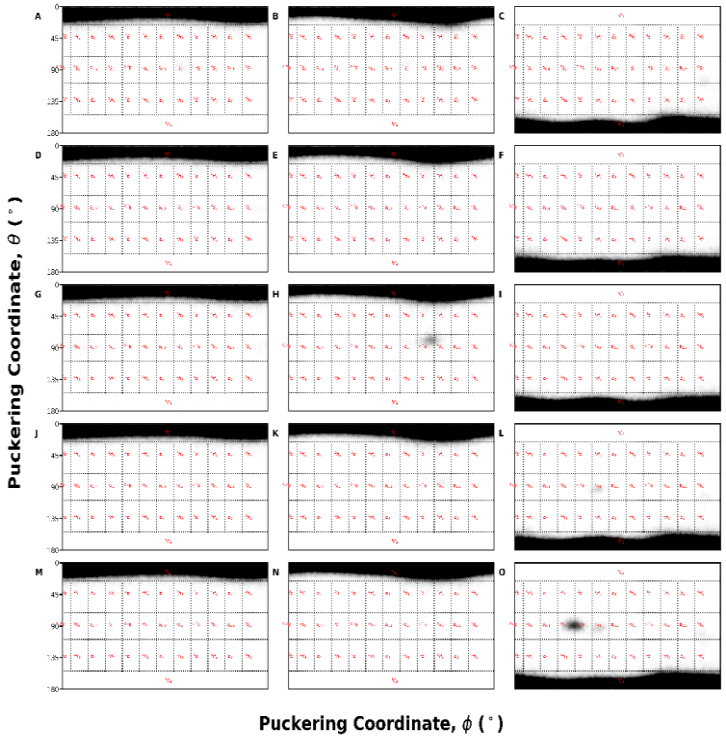

Figure 4.18: Mercator representation of the Cremer−Pople Sphere for *Com4* complex-bound glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

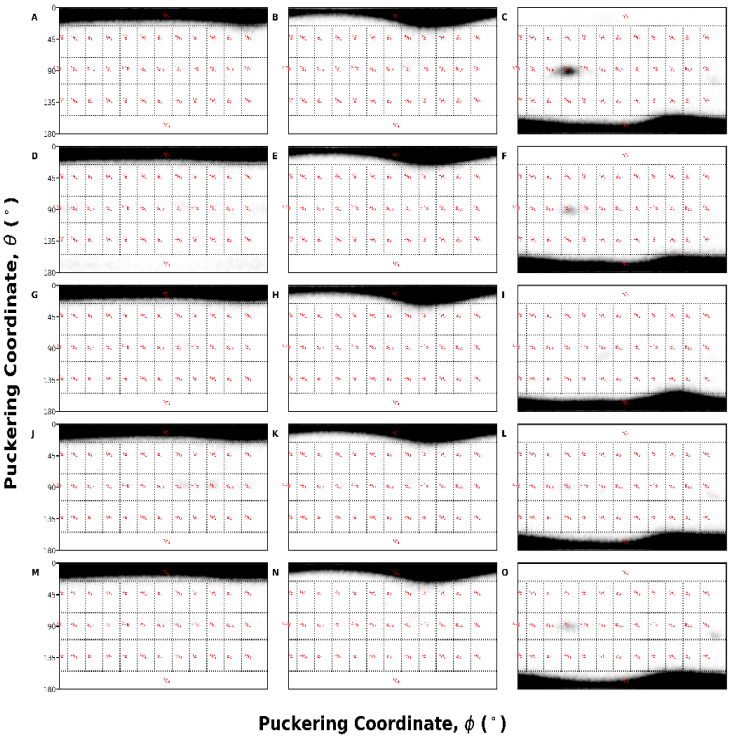

Figure 4.19: Mercator representation of the Cremer−Pople Sphere for *Com2* complex-bound glycans (1st column: GlcNAc; 2nd column: Gal; 3rd column: Neu5Ac). A)-C) *Gly1*, D)-F) *Gly2*, G)-I) *Gly3*, J)-L) *Gly4*, M)-O) *Gly5*.

Table 4.2: Occupancy of hydrogen bonds between residues of each glycan molecule.

		Acceptor	Donor	Distance	Angle	Occupancy (%)
Gly1	**Run1**	Gal@O5	GlcNAc@O3	2.78	155.57	22.74
		Neu5,9Ac@O6	Gal@O2	2.78	154.35	15.6
		Neu5,9Ac@O1B	Gal@O4	2.72	159.47	14.69
		Neu5,9Ac@O1A	Gal@O4	2.72	159.46	13.64
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.80	156.92	12.42
		Gal@O4	Gal@O6	2.77	146.71	3.17
	Run2	Gal@O5	GlcNAc@O3	2.78	155.50	22.6
		Neu5,9Ac@O1B	Gal@O4	2.72	159.29	14.8
		Neu5,9Ac@O6	Gal@O2	2.79	154.44	14.72
		Neu5,9Ac@O1A	Gal@O4	2.72	159.39	13.98
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.79	156.89	12.44
	Run3	Gal@O5	GlcNAc@O3	2.78	155.60	23.01
		Neu5,9Ac@O6	Gal@O2	2.78	154.35	15.14
		Neu5,9Ac@O1A	Gal@O4	2.72	159.44	14.21
		Neu5,9Ac@O1B	Gal@O4	2.72	159.34	14.05
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.79	156.94	12.78
Gly2	**Run1**	Gal@O5	GlcNAc@O3	2.79	155.45	25.91
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.79	157.26	13.13
		Neu5,9Ac@O1A	Gal@O4	2.70	165.06	11.08
		Neu5,9Ac@O1B	Gal@O4	2.70	164.96	10.15
	Run2	Gal@O5	GlcNAc@O3	2.78	155.34	25.75
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.79	157.12	13.09
		Neu5,9Ac@O1B	Gal@O4	2.70	165.00	10.72
	Run3	Gal@O5	GlcNAc@O3	2.79	155.57	26.55
		Neu5,9Ac@O5N	Neu5,9Ac@O7	2.79	156.99	12.77
		Neu5,9Ac@O1B	Gal@O4	2.70	165.00	11
		Neu5,9Ac@O1A	Gal@O4	2.70	165.04	10.53
Gly3	**Run1**	Gal@O5	GlcNAc@O3	2.78	155.48	22.58
		Neu5,4Ac@O1A	Gal@O4	2.72	159.97	15.15
		Neu5,4Ac@O1B	Gal@O4	2.72	159.91	14.21
		Neu5,4Ac@O6	Gal@O2	2.79	154.36	13.49
	Run2	Gal@O5	GlcNAc@O3	2.78	155.54	22.88
		Neu5,4Ac@O1A	Gal@O4	2.72	159.90	14.99
		Neu5,4Ac@O6	Gal@O2	2.78	154.30	14.45
		Neu5,4Ac@O1B	Gal@O4	2.72	159.78	13.95

	Run3	Gal@O5	GlcNAc@O3	2.78	155.60	22.78
		Neu5,4Ac@O1B	Gal@O4	2.72	159.88	16.54
		Neu5,4Ac@O1A	Gal@O4	2.72	159.95	14.46
		Neu5,4Ac@O6	Gal@O2	2.79	154.28	13.48
Gly4	**Run1**	Gal@O5	GlcNAc@O3	2.78	155.63	23.42
		Neu5Ac@O6	Gal@O2	2.79	154.50	14.95
		Neu5Ac@O1B	Gal@O4	2.72	159.81	14.73
		Neu5Ac@O5N	Neu5Ac@O7	2.79	157.19	13.95
		Neu5Ac@O1A	Gal@O4	2.72	159.75	13.2
	Run2	Gal@O5	GlcNAc@O3	2.78	155.62	23.22
		Neu5Ac@O1B	Gal@O4	2.72	159.91	15.51
		Neu5Ac@O5N	Neu5Ac@O7	2.79	157.35	14.88
		Neu5Ac@O1A	Gal@O4	2.72	159.93	14.73
		Neu5Ac@O6	Gal@O2	2.79	154.51	14.54
	Run3	Gal@O5	GlcNAc@O3	2.78	155.56	23.38
		Neu5Ac@O1B	Gal@O4	2.72	159.85	14.84
		Neu5Ac@O6	Gal@O2	2.79	154.50	14.31
		Neu5Ac@O5N	Neu5Ac@O7	2.79	157.13	13.54
		Neu5Ac@O1A	Gal@O4	2.72	159.88	13.37
Gly5	**Run1**	Gal@O5	GlcNAc@O3	2.78	155.34	25.63
		Neu5Ac@O1A	Gal@O4	2.70	164.99	11.07
		Neu5Ac@O1B	Gal@O4	2.70	164.98	11.06
	Run2	Gal@O5	GlcNAc@O3	2.78	155.35	26.67
		Neu5Ac@O5N	Neu5Ac@O7	2.79	157.29	11.9
		Neu5Ac@O1B	Gal@O4	2.70	164.96	11.29
		Neu5Ac@O1A	Gal@O4	2.70	165.01	10.18
	Run3	Gal@O5	GlcNAc@O3	2.78	155.42	26.44
		Neu5Ac@O1B	Gal@O4	2.70	165.01	10.96
		Neu5Ac@O1A	Gal@O4	2.70	164.94	10.39

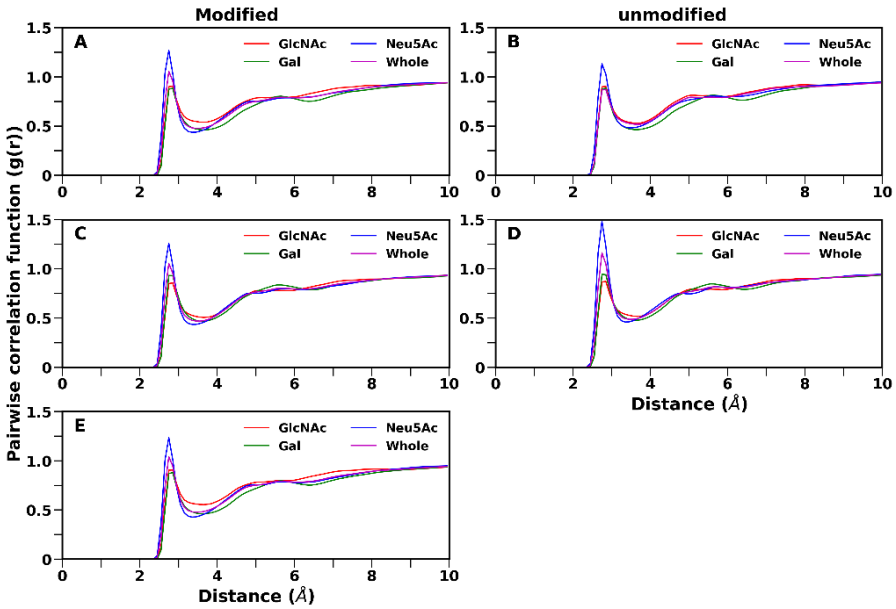

Figure 4.20: Pairwise correlation function, g(r) of water molecules in the simulation as a function of distance from the oxygen atoms of each atom of each monosaccharide residue as well as the whole (red line). A) *Gly1* B) *Gly4* C) *Gly2* D) *Gly5* E) *Gly3*.

Figure 4.15 shows the radial distribution function of the three monosaccharides and the whole glycan for each of the five glycans. In the case of both the modified and unmodified glycans, the pairwise correlation function is around, so the affinity towards the water is not affected by the modification of the glycans.

4.3 Conformational dynamics of Apo protein and complex

4.3.1 Structure stability

The RMSD (Root-Mean-Square-Deviation) time evolution plot comparing the RMSD of the Apo protein with the five complexes is shown in **Figure 4.17.** After the 200 ns, the RMSD of the Apo protein and the complexes get stable, which remains stable up to the last 500 ns. The apoprotein has more deviation than the three complexes showing that the complexes are more stable than the apoprotein.

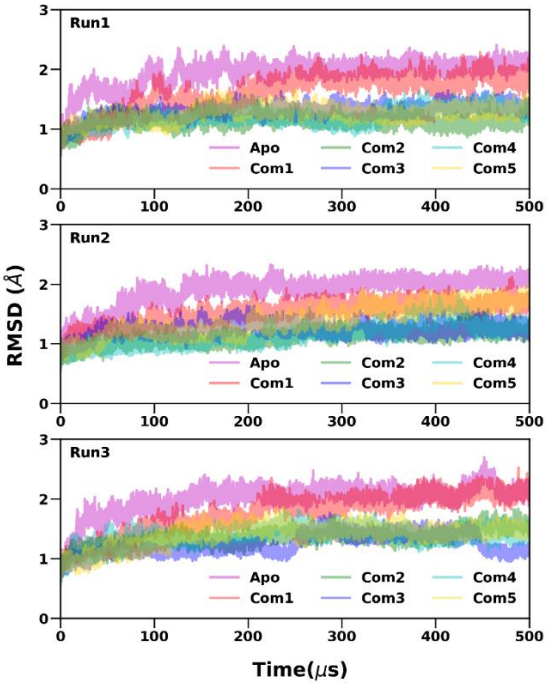

Figure 4.21: Time evolution series of the RMSD of Apo protein and the five complexes for three Runs.

4.3.2 SASA

The solvent-accessible surface area (SASA) is estimated using the KDE (kernel density estimation) method for the apo and the five complexes. The modification on the glycans doesn't show any significant changes in the SASA.

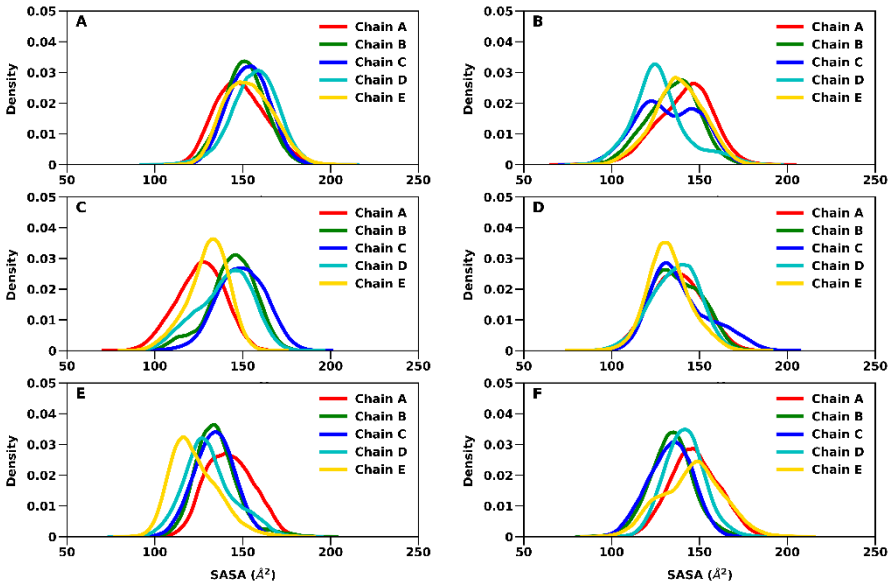

Figure 4.22: Probability distribution plot of solvent-accessible-surface-area (SASA) of the Apo protein and five complexes with respect to the initial structure. A) Apo B) *Com1* C) *Com2* D) *Com3* E) *Com4* F) *Com5*.

4.3.3 Compactness analysis

Figure 4.18 shows the radius of gyration (RoG) plot (KDE) of the Apo protein and the five complexes. The radius of gyration measures the compactness of the system. The more compact the system is after the binding will be more stable complex. There are no significant changes in the radius of gyration of the Apo and the other complexes. The minor fluctuations are because of the different binding patterns of the glycans with the proteins.

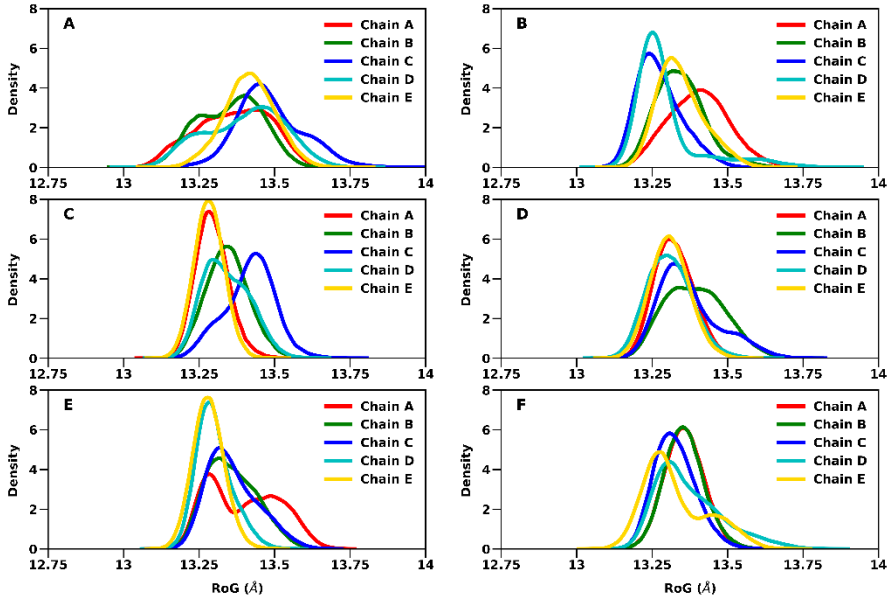

Figure 4.23: Probability distribution plot of the Radius of gyration (RoG) of the Apo protein and five complexes with respect to the initial structure. A) Apo B) *Com1* C) *Com2* D) *Com3* E) *Com4* F) *Com5*

The RMSF (Root mean square fluctuations) per residue plot for each complex and each three-run is shown in **Figure 4.19 – Figure 4.22**. The plot follows a similar fashion for all the complex and the Apo protein, which suggests that the key residues are the same in each case. The key residues are 60-66, which corresponds to the loop that is also responsible for the recognition of the glycans. Similar to **Figure 4.19 – Figure 4.22,** the RMSF of the complexes are less than apoprotein, which shows that the apoprotein becomes stable after binding to glycans. There is also a loop region between the 105-109 residues close to the BS2 site. We found that in each run, we got a different plot for the same complex and same chain. Hence, we considered each run differently and looked

at each case. The higher peak at the 60-66 amino acid regions indicates either the system's lower binding energy or the glycan's detachment from that chain. The lower peak, the 60-66 amino acid residue region, suggests a higher binding affinity of glycans toward the binding site. Also, in some cases, the peak at the 105-109 amino acid residue region indicates that the glycans have some minor interactions with the BS2 site.

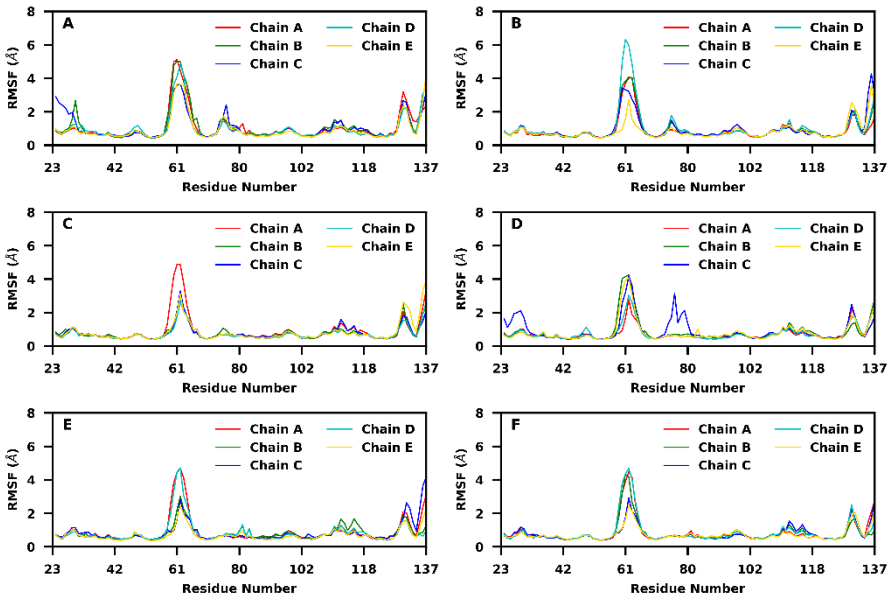

Figure 4.24: Root mean square fluctuations (RMSFs) of C_α atoms for Apo and all five complexes (Run1). A) Apo B) *Com1* C) *Com2* D) *Com3* E) *Com4* F) *Com5*.

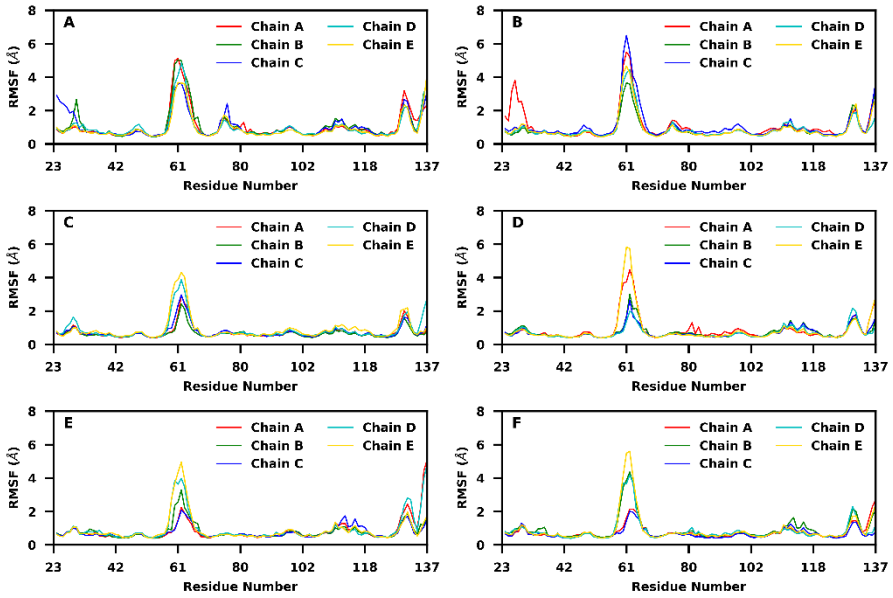

Figure 4.25: Root mean square fluctuations (RMSFs) of C_α atoms for Apo and all five complexes (Run2). A) Apo B) *Com1* C) *Com2* D) *Com3* E) *Com4* F) *Com5*.

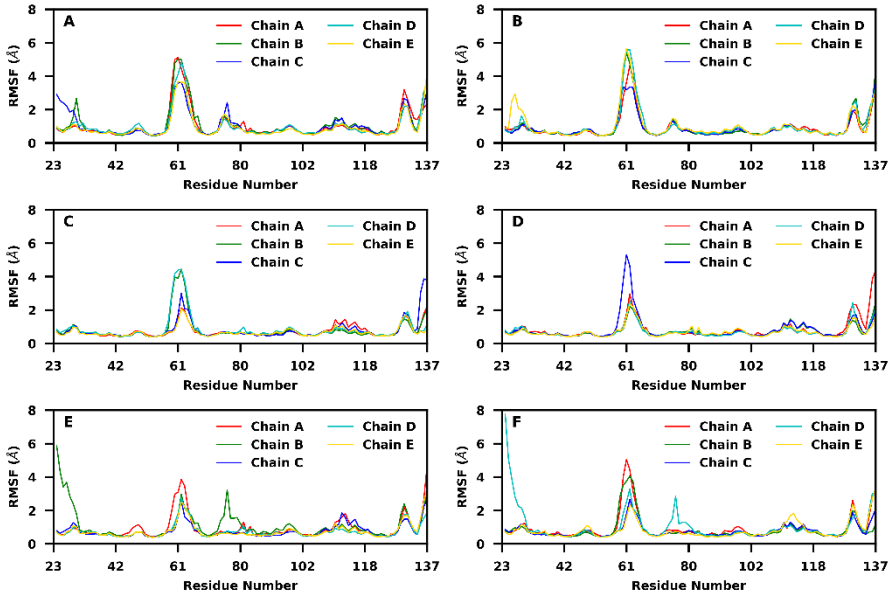

Figure 4.26: Root mean square fluctuations (RMSFs) of C_α atoms for Apo and all five complexes (Run3). A) Apo B) *Com1* C) *Com2* D) *Com3* E) *Com4* F) *Com5*.

4.4 Distance between glycans and key residues

The distance between the centre of the mass of the glycans and the centre of mass of the key residues is shown in the **Figure 4.22 – Figure 4.30**

for each complex and each run. This figure suggests the attachment and the detachment of the glycan from the binding site. In the case where the plot remains stable, the glycan was bounded to the binding site throughout the simulation, while where the distance increased between the glycan and the key residues there, the glycan detached from the binding site and went away. These attachment and detachment is not the same in the three runs of simulations. It is necessary to consider each case individually. The glycans molecules are small and highly flexible, and because of this, we observed a significant difference in the binding patterns of the same glycans at the same chain and complex in three different runs. For example, in the case of the *Com1*, the glycans remained attached to the chain A, C and E in run1; A, B,C, D and E in run2 and A and E in run3. The detailed attachment and the detachment of the glycans with their respective chains for each complex and each run is given in the **Table 4.3 – Table4.5.**

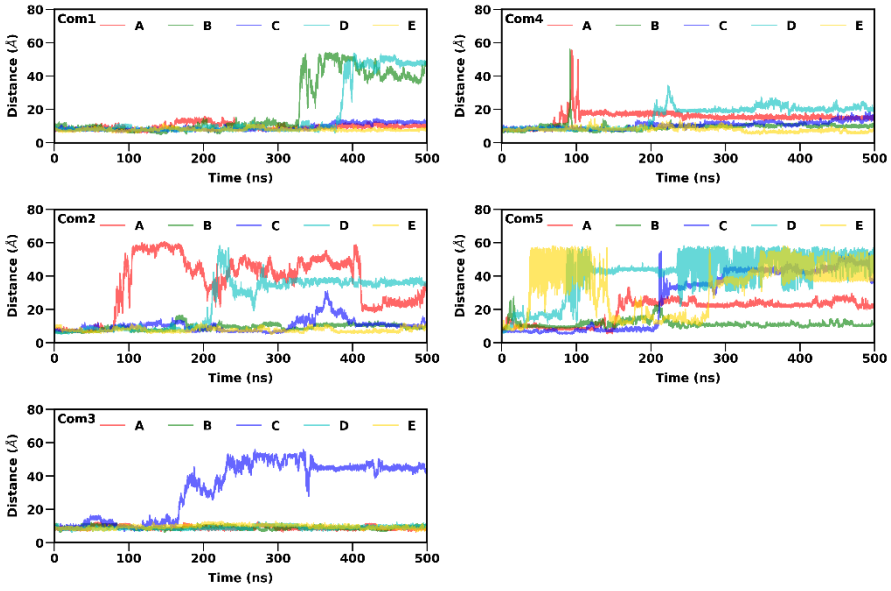

Figure 4.27: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 1 (BS1) of all the five complexes in Run1.

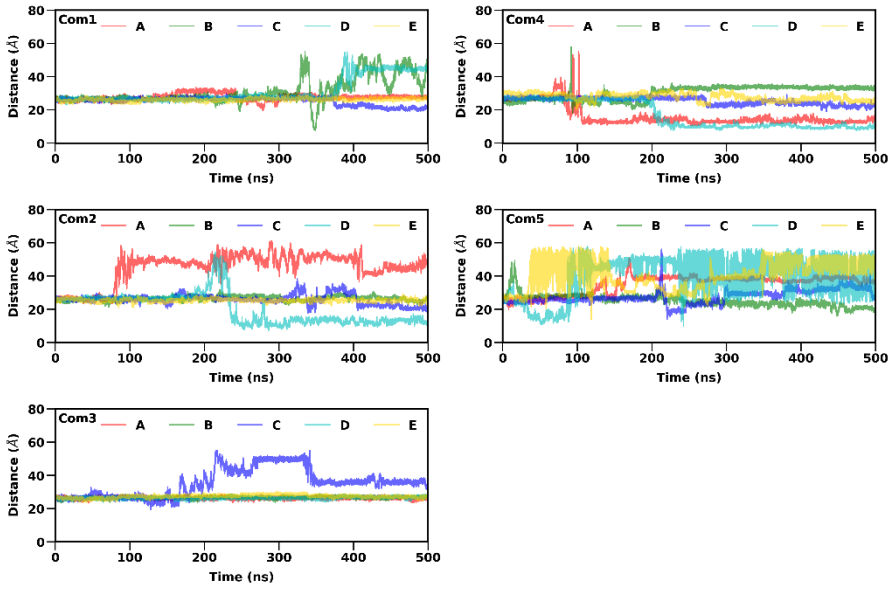

Figure 4.28: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 2 (BS3) of all the five complexes in Run1.

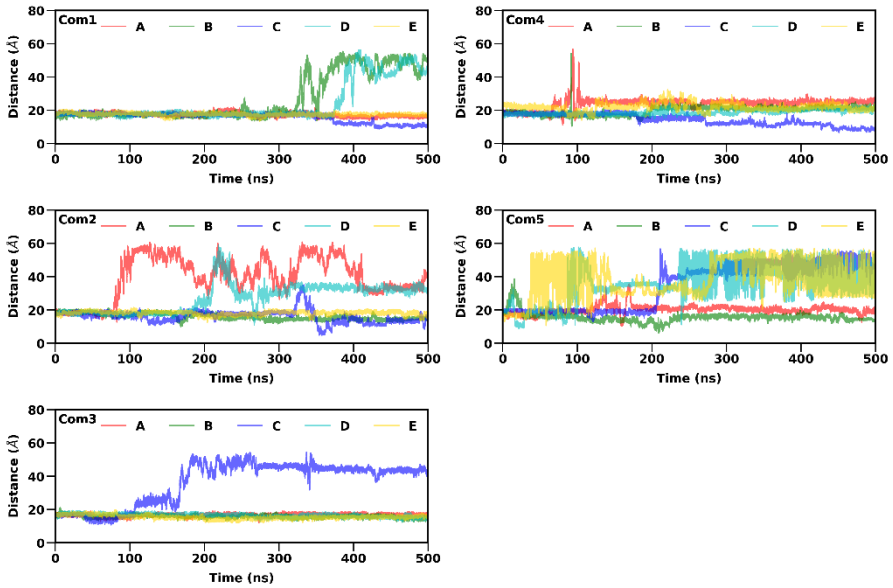

Figure 4.29: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 3 (BS3) of all the five complexes in Run1.

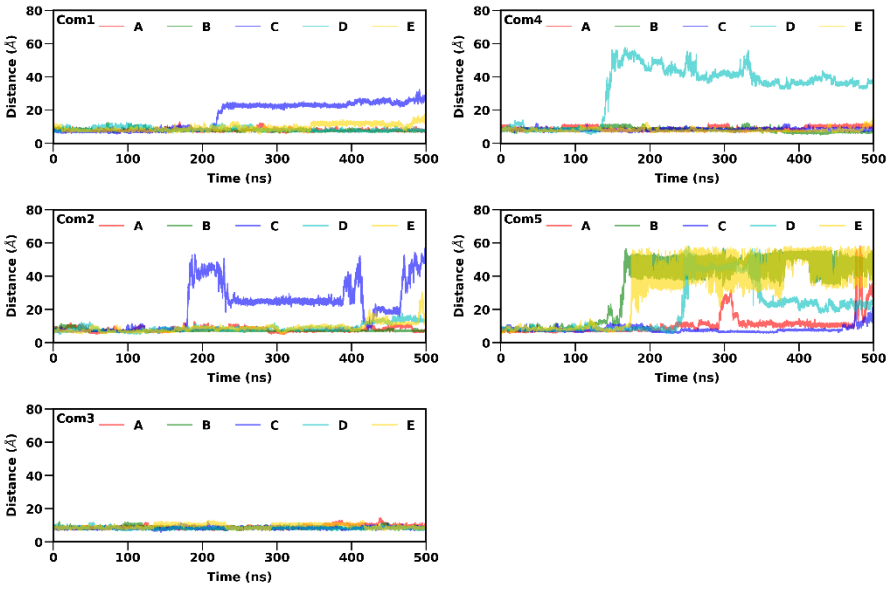

Figure 4.30: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 1 (BS1) of all the five complexes in Run2.

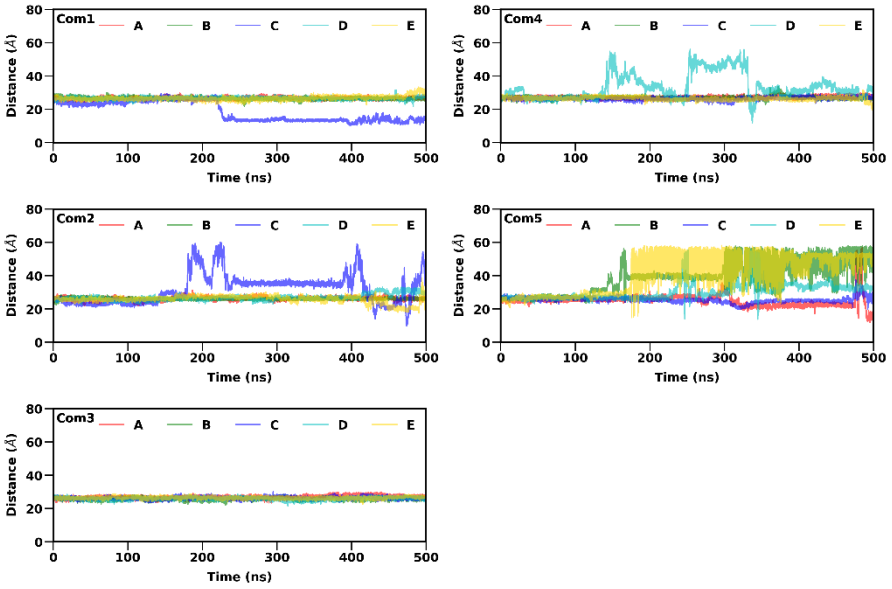

Figure 4.31: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 2 (BS2) of all the five complexes in Run2.

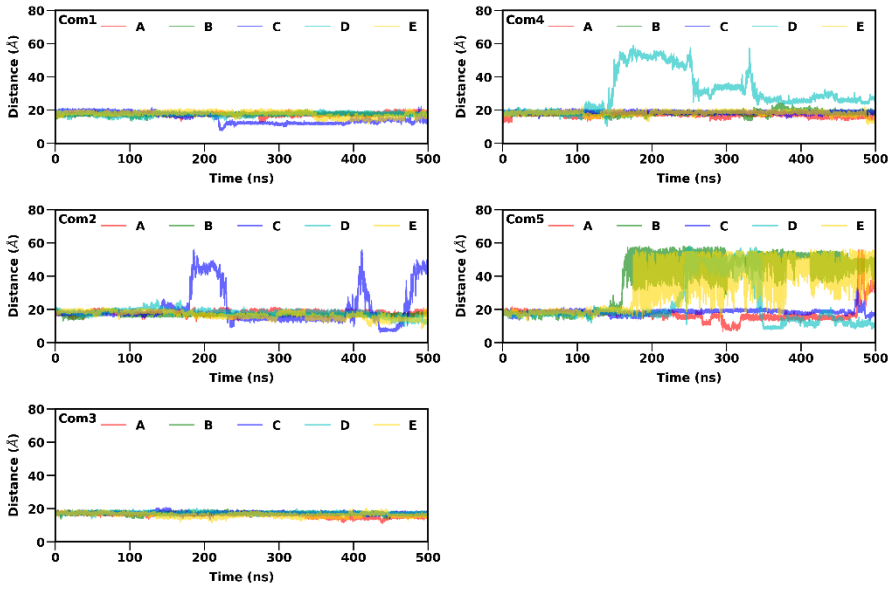

Figure 4.32: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 3 (BS3) of all the five complexes in Run2.

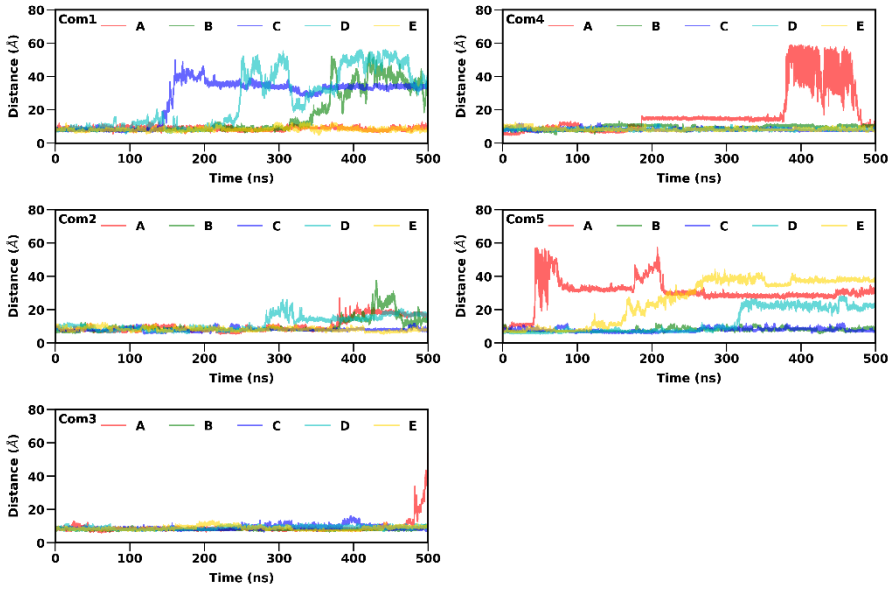

Figure 4.33: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 1 (BS1) of all the five complexes in Run3.

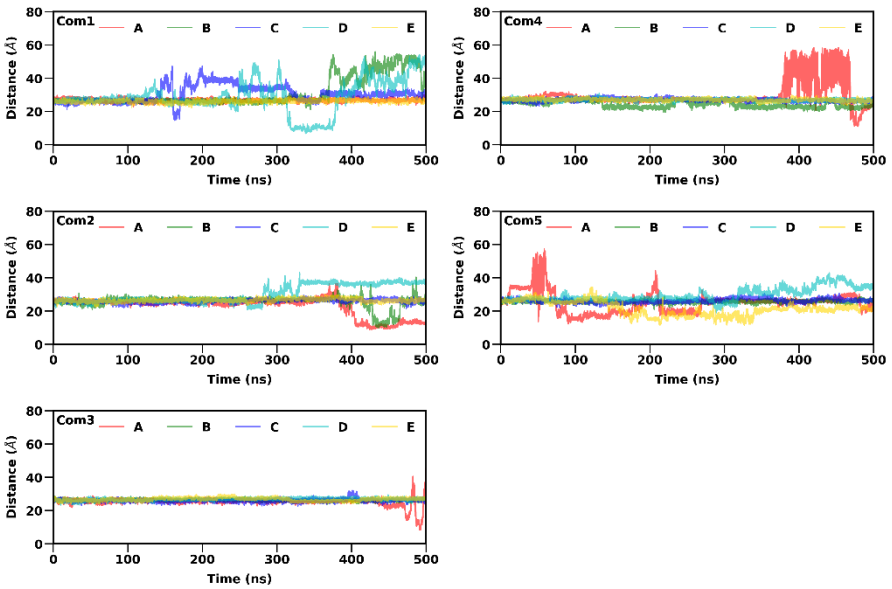

Figure 4.34: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 2 (BS2) of all the five complexes in Run3.

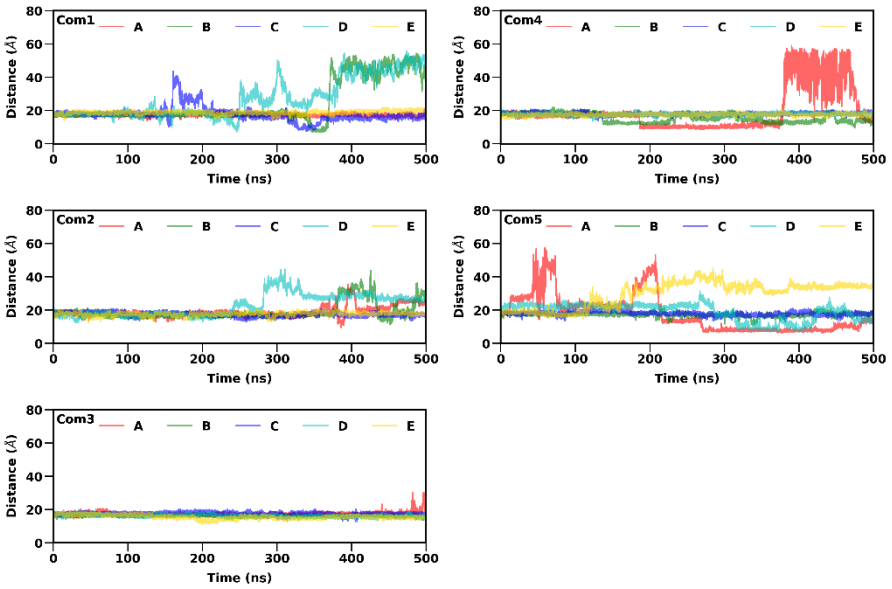

Figure 4.35: The time evolution plot of the distance between the centre of mass of glycans and the key residues of binding sites 2 (BS2) of all the five complexes in Run3.

Table 4.3: The table shows the glycans that remained attached to the binding site after the simulation (Run1). 'A' represents attachment, 'D' represents detachment, and the 'A' represents that the glycan was also bounded to the binding site in the crystal PDB.

	Chain-A	**Chain-B**	**Chain-C**	**Chain-D**	**Chain-E**
Com1	A		A		A
Com2		A	A		A
Com3	A	A		A	A
Com4	A	A	A	A	A
Com5	A	A			

Table 4.4: The table shows the glycans that remained attached to the binding site after the simulation (Run2). 'A' represents attachment, 'D' represents detachment, and the 'A' represents that the glycan was also bounded to the binding site in the crystal PDB.

	Chain-A	**Chain B**	**Chain-C**	**Chain-D**	**Chain-E**
Com1	A	A	D	A	A
Com2	A	A		A	A
Com3	A	A	A	A	A
Com4	A	A	A		A
Com5	A		A	A	

Table 4.5: The table shows the glycans that remained attached to the binding site after the simulation (Run3). 'A' represents attachment, 'D' represents detachment, and the 'A' represents that the glycan was also bounded to the binding site in the crystal PDB.

	Chain-A	**Chain B**	**Chain-C**	**Chain-D**	**Chain-E**
Com1	A				A
Com2	A	A	A	A	A
Com3	A	A	A	A	A
Com4	A	A	A	A	A
Com5		A	A	A	

4.5 PCA

We have performed the principal component analysis (PCA) of the apo and complex PltB, as shown in **Figure 4.37**. For each case, PCA was done for all three independent runs. In the case of apo simulations, similar observations were found in each replica. While in the case of the

complex, each run yields different minima region. In *Com1*, Run2 has the most minima region. In *Com2*, Run1 has the most minima regions. In *Com3*, Run1 has the most minima regions, and the energy barrier between the two minima regions is also less. In *Com4*, Run1 has the most minima regions again, and the energy difference between the minima regions is relatively minor. In *Com5*, Run2 has the most minima regions.

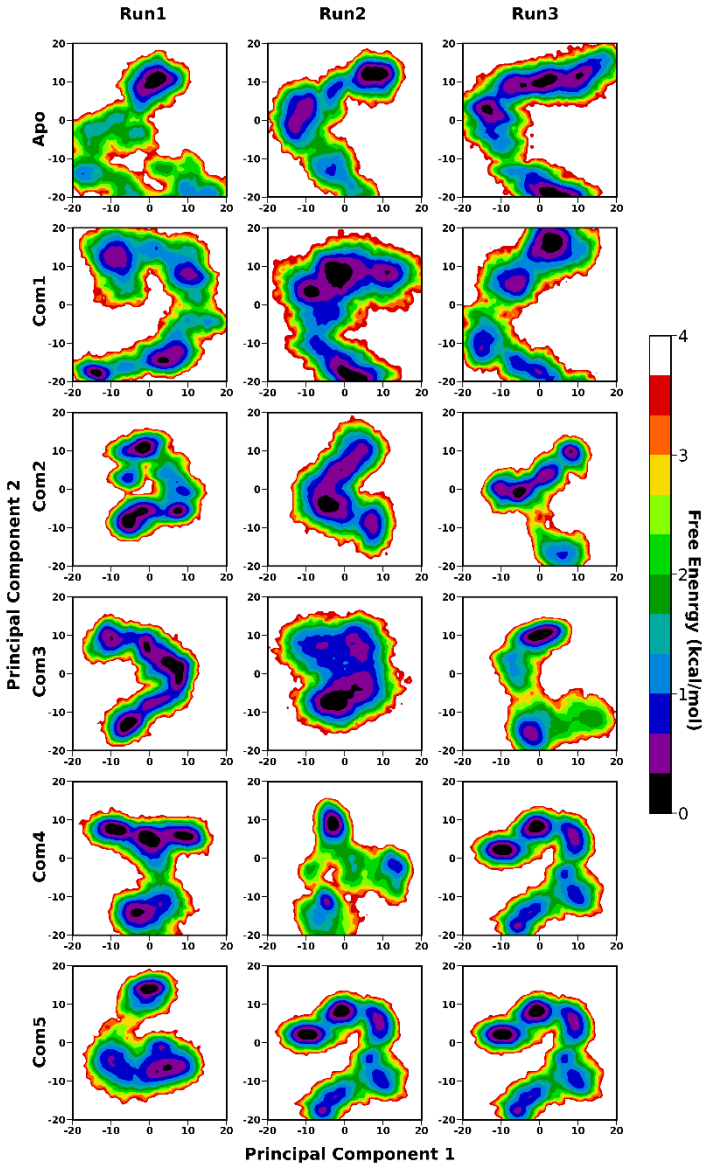

Figure 4.36: Principal component analysis map for each of the Apo protein and five complexes (For three runs).

4.5 Binding Free-energy Calculations

We have used the MM/GBSA (molecular mechanics generalised Born Boltzmann surface area) to calculate the total Binding free energy

(ΔG_{bind}). The components of the total binding free energy include the van der Waals (ΔG_{vdW}), electrostatics (ΔG_{elec}), polar (ΔG_{polar}) and non-polar ($\Delta G_{nonpolar}$); these are listed in **Table 4.6** and shown in **Figure 4.31 – Figure 4.35**. The table shows that intermolecular electrostatic and van der Waals and non-polar interactions favour the formation of the complex, whereas polar solvation disfavors the binding formation. The total binding free energy of the complexes is given in the Table along with the average values. Thus, the glycan-binding in *Com3* is more stable than all other complexes. In contrast, the *Com5* has the lowest binding free energy. The modified glycan-protein complexes have more binding affinity than the unmodified glycan-protein complexes. In the case of *Com3*, van der Waals interactions have the highest contribution, which was compensated by the total polar group contribution (sum of electrostatic and polar solvation energy ~ 14 kcal/mol).

Table 4.6: Binding free energies of all three complexes and the different binding free energy components from the MM//GBSA scheme. (The green shaded number indicates that in these chains, the glycan was already present ain the crystal PDB).

System	**A**	**B**	**C**	**D**	**E**	**Avg**	**Avg** (Run1+Run2 +Run3)
	Run1						
Gly1	-20.95		-20.53		-28.41	-23.30	**-22.27**
Gly2		-28.86	-15.90		-28.74	-24.50	**-23.73**
Gly3	-26.22	-33.20		-27.08	-30.05	-29.14	**-28.03**
Gly4	-13.81	-23.96	-16.63	-17.00	-18.81	-17.99	**-19.63**
Gly5	-19.25	-15.98				-17.62	**-18.59**
	Run2						
Gly1	-26.57	-29.21	-20.53	-26.15	-13.84	-23.26	
Gly2	-30.07	-24.62		-21.60	-18.10	-21.44	
Gly3	-26.91	-28.11	-26.37	-26.62	-28.31	-27.26	
Gly4	-22.30	-18.78	-20.84		-25.15	-21.77	
Gly5	-13.59		-30.61	-13.82		-19.34	
	Run3						
Gly1	-24.39				-16.13	-20.26	
Gly2	-17.28	-28.11	-29.29	-26.62	-24.96	-25.25	
Gly3	-20.90	-35.14	-20.17	-24.75	-37.45	-27.68	
Gly4	-11.52	-17.00	-23.84	-21.82	-21.48	-19.13	

Gly5		-20.62	-22.03	-13.76		-18.80	

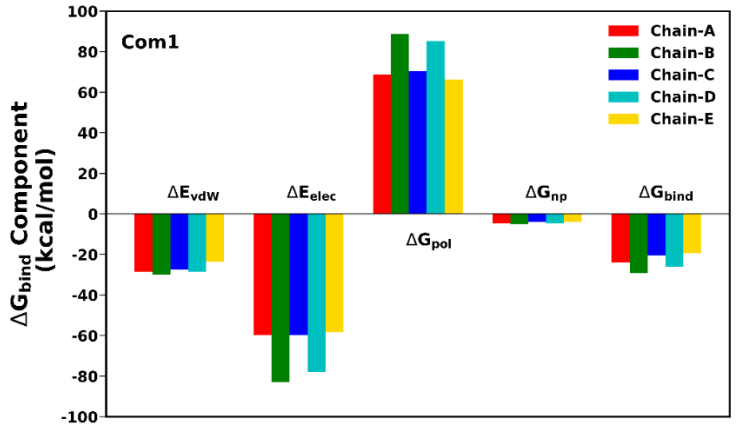

Figure 4.37: Binding free energy and its component for *Com1* (all five chains).

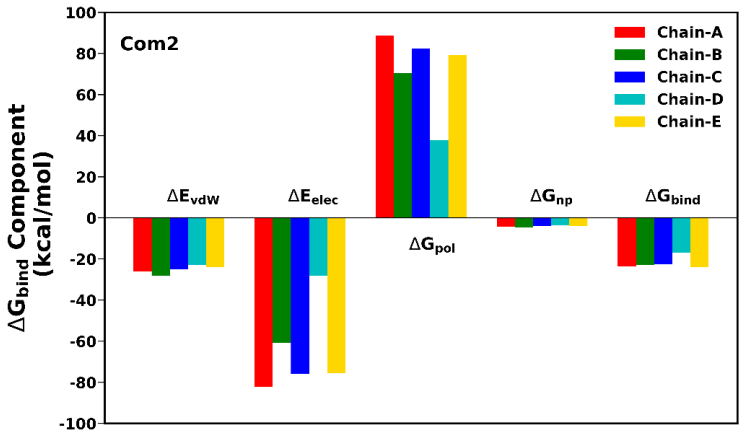

Figure 4.38: Binding free energy and its component for *Com2* (all five chains).

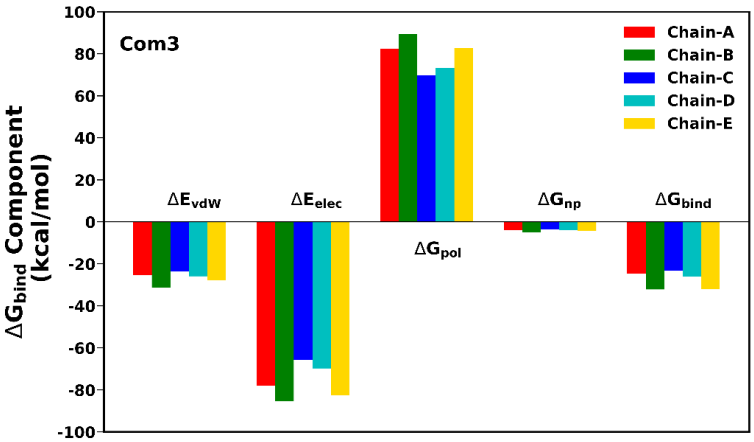

Figure 4.39: Binding free energy and its component for *Com3* (all five chains).

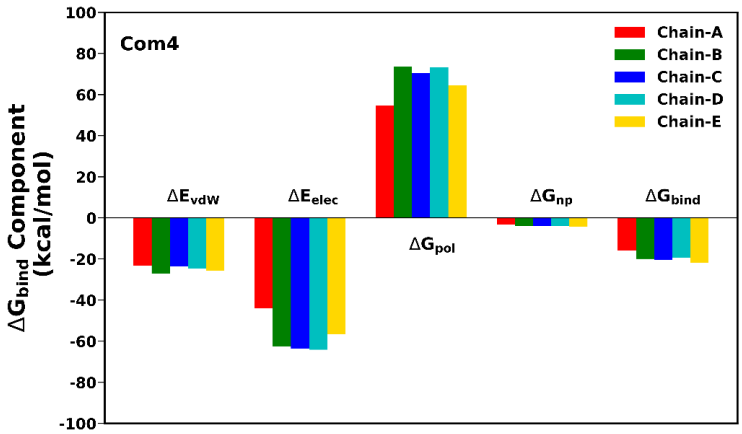

Figure 4.40: Binding free energy and its component for *Com4* (all five chains).

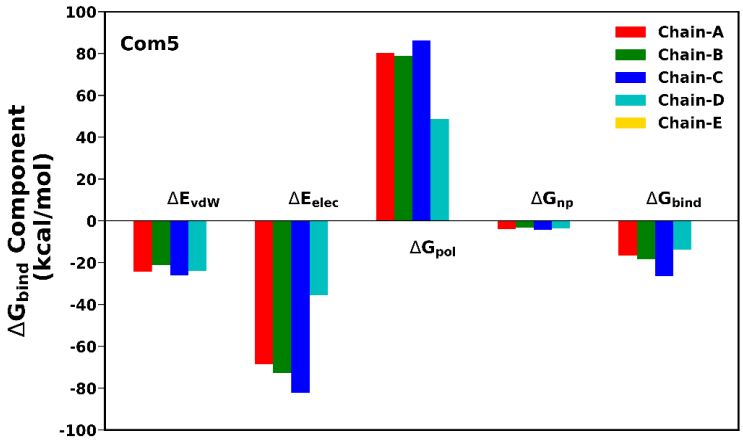

Figure 4.41: Binding free energy and its component for *Com5* (all five chains).

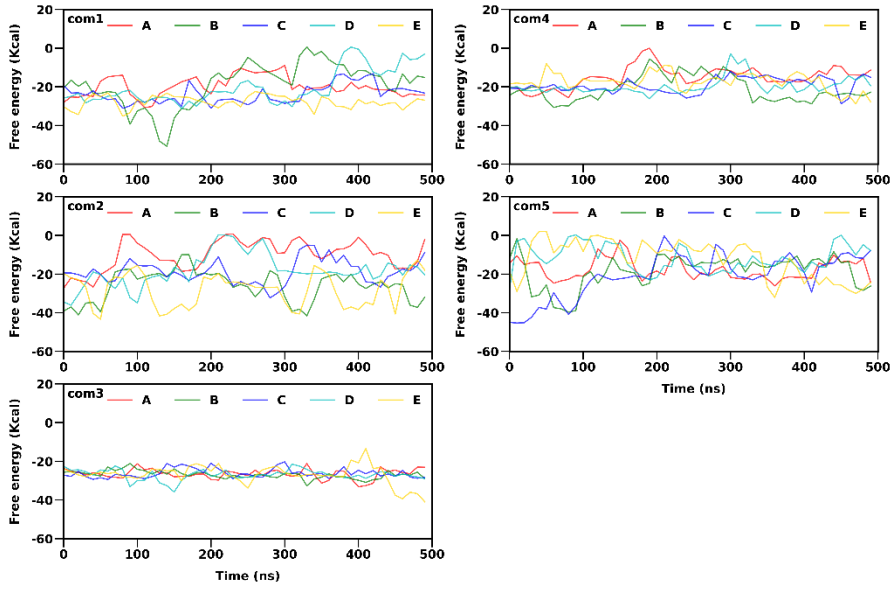

Figure 4.42: Time evolution series of Binding free energy of five complexes during Run1.

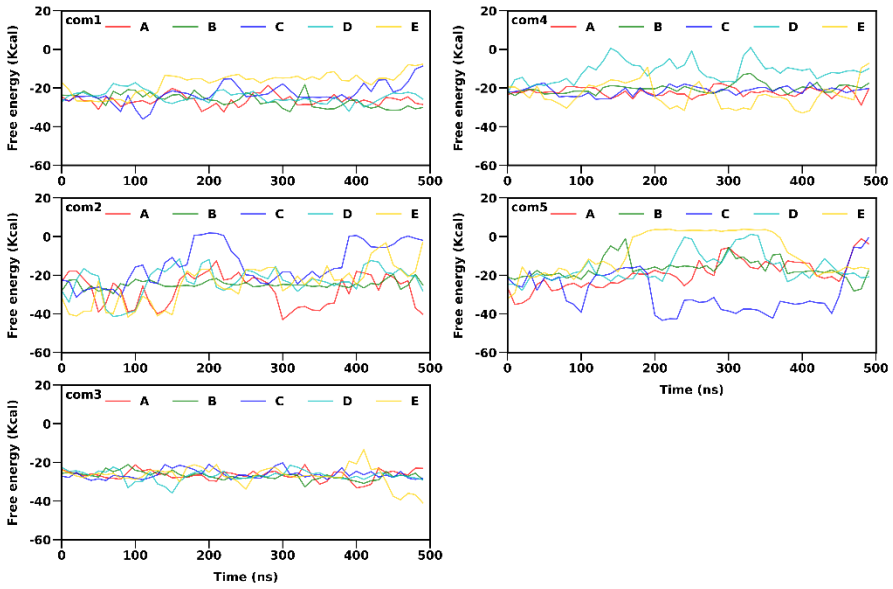

Figure 4.43: Time evolution series of Binding free energy of five complexes during Run2.

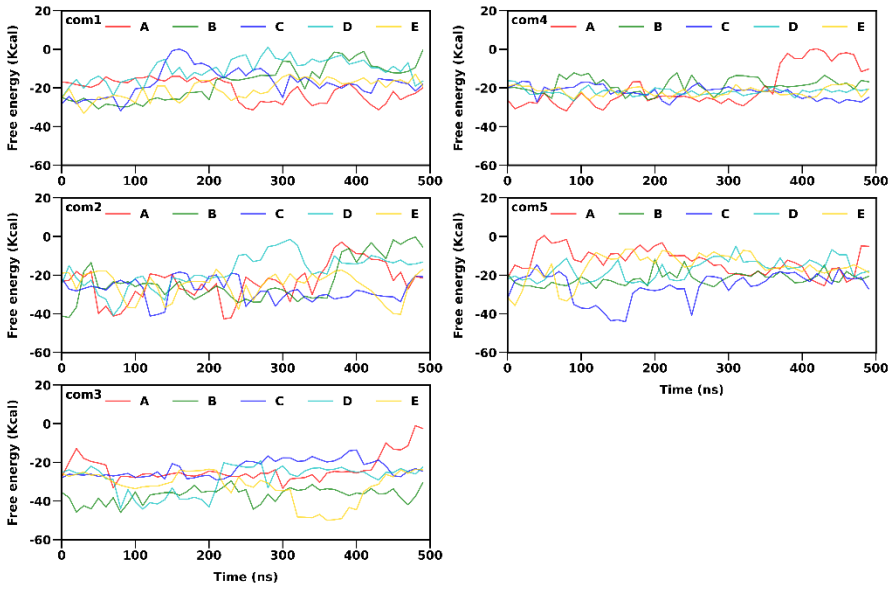

Figure 4.44: Time evolution series of Binding free energy of five complexes during Run3.

Table 4.7: The contribution of different components towards the calculation of binding free energy for all five complexes (Run1, Run2 and Run3).

Systems			**ΔG_{vdW}**	**ΔG_{elec}**	**ΔG_{pol}**	**ΔG_{np}**	**ΔE_{MM}**	**ΔG_{solv}**	**ΔG_{bind}**
Com1	**Chain-A**	**Run1**	-30.52	-63.36	77.54	-4.58	-93.88	72.96	-20.93
		Run2	-28.56	-68.04	74.82	-4.77	-96.59	70.04	-26.55

		Run3	-26.57	-47.88	54.34	-4.28	-74.45	50.06	-24.39
		Avg	-28.55 (1.98)	-59.76 (10.55)	68.9 (12.68)	-4.54 (0.25)	-88.31 (12.04)	64.35 (2.46)	-23.96 (2.83)
	Chain-B	**Run1**							
		Run2	-29.99	-82.83	88.69	-5.06	-112.82	83.64	-29.18
		Run3							
		Avg	-29.99	-82.83	88.69	-5.06	-112.82	83.64	-29.18
	Chain-C	**Run1**	-28.87	-64.61	77.21	-4.23	-93.49	72.98	-20.51
		Run2	-25.94	-54.72	64.01	-3.86	-80.66	60.15	-20.51
		Run3							
		Avg	-27.405 (2.07)	-59.665 (6.99)	70.61 (9.33)	-4.045 (0.26)	-87.075 (9.07)	66.565 (9.07)	-20.51 (0)
	Chain-D	**Run1**							
		Run2	-28.53	-78.05	85.09	-4.64	-106.58	80.45	-26.13
		Run3							
		Avg	-28.53	-78.05	85.09	-4.64	-106.58	80.45	-26.13
	Chain-E	**Run1**	-29.95	-79.81	86.39	-5.02	-109.76	81.37	-28.38
		Run2	-18.98	-47.04	55.26	-3.08	-66.02	52.18	-13.84
		Run3	-21.74	-48.50	57.51	-3.40	-70.24	54.11	-16.13
		Avg	-23.56 (5.71)	-58.45 (18.51)	66.39 (17.36)	-3.83 (1.04)	-82.01 (24.13)	62.56 (16.32)	-19.45 (7.82)
Com2	**Chain-A**	**Run1**							
		Run2	-27.59	-93.02	95.08	-4.53	-120.62	90.55	-30.07
		Run3	-24.46	-71.08	82.15	-3.90	-95.53	78.56	-17.28
		Avg	-26.02 (2.21)	-82.05 (15.51)	88.61 (9.14)	-4.21 (0.45)	-108.07 (17.74)	84.55 (8.47)	-23.67 (9.04)
	Chain-B	**Run1**	-27.08	-93.69	96.39	-4.46	-120.76	91.93	-28.83
		Run2	-37.60	-47.00	65.82	-5.85	-84.60	59.97	-24.62
		Run3	-19.36	-41.92	48.95	-3.21	-61.29	45.74	-15.55
		Avg	-28.01 (9.156)	-60.87 (28.54)	70.39 (24.05)	-4.51 (1.32)	-88.88 (29.97)	65.88 (23.66)	-23 (6.79)
	Chain-C	**Run1**	-22.35	-57.58	67.58	-3.53	-79.93	64.05	-15.88
		Run2							
		Run3	-27.73	-94.26	97.26	-4.56	-121.99	92.70	-29.29
		Avg	-25.04 (3.80)	-75.92 (25.94)	82.42 (20.99)	-4.04 (0.73)	-100.96 (29.74)	78.37 (20.26)	-22.58 (9.48)
	Chain-D	**Run1**							
		Run2	-21.02	-66.17	69.36	-3.77	-87.19	65.59	-21.60
		Run3	-25.00	9.84	6.20	-3.46	-15.16	2.73	-12.43
		Avg	-23.01 (2.81)	-28.16 (53.74)	37.78 (44.66)	-3.61 (0.22)	-51.17 (50.93)	34.16 (44.45)	-17.01 (6.48)
	Chain-E	**Run1**	-25.73	-81.10	82.40	-4.29	-106.8	78.12	-28.72
		Run2	-21.72	-63.01	70.19	-3.56	-84.73	66.63	-18.10

		Run3	-23.88	-82.10	85.03	-4.02	-105.97	81.01	-24.96
		Avg	-23.78 (2.01)	-75.40 (10.74)	79.21 (7.92)	-3.96 (0.37)	-99.17 (12.51)	75.25 (7.61)	-23.93 (5.38)
Com3	**Chain-A**	**Run1**	-26.20	-85.21	89.16	-3.97	-111.42	85.19	-26.22
		Run2	-26.67	-83.56	87.38	-4.06	-110.24	83.32	-26.91
		Run3	-23.26	-64.91	70.93	-3.65	-88.17	67.28	-20.90
		Avg	-25.38 (1.85)	-77.89 (11.27)	82.49 (10.05)	-3.89 (0.215)	-103.28 (13.1)	78.60 (9.84)	-24.68 (3.29)
	Chain-B	**Run1**	-31.44	-93.27	96.35	-4.84	-124.71	91.51	-33.20
		Run2	-27.50	-81.51	85.10	-4.20	-109.01	80.90	-28.11
		Run3	-35.25	-81.53	87.09	-5.45	-116.78	81.64	-35.14
		Avg	-31.40 (3.87)	-85.44 (6.78)	89.51 (6.00)	-4.83 (0.62)	-116.83 (7.85)	84.68 (5.92)	-32.15 (3.63)
	Chain-C	**Run1**							
		Run2	-26.32	-76.59	80.56	-4.02	-102.91	76.54	-26.37
		Run3	-21.12	-54.83	58.91	-3.13	-75.95	55.78	-20.17
		Avg	-23.72 (3.68)	-65.71 (15.39)	69.735 (15.31)	-3.575 (0.63)	-89.43 (19.06)	66.16 (14.68)	-23.27 (4.38)
	Chain-D	**Run1**	-25.42	-67.28	69.43	-3.81	-92.70	65.62	-27.08
		Run2	-26.42	-74.45	78.32	-4.07	-100.87	74.25	-26.62
		Run3	-25.74	-67.58	72.27	-3.70	-93.32	68.57	-24.75
		Avg	-25.86 (0.51)	-69.77 (4.06)	73.34 (4.54)	-3.86 (0.19)	-95.63 (4.55)	69.48 (4.39)	-26.15 (1.23)
	Chain-E	**Run1**	-28.62	-83.26	86.18	-4.36	-111.87	81.82	-30.05
		yun2	-23.90	-74.71	74.18	-3.88	-98.61	70.30	-28.31
		Run3	-30.82	-89.51	87.73	-4.86	-120.33	82.88	-37.45
		Avg	-27.78 (3.54)	-82.49 (7.43)	82.70 (7.42)	-4.37 (0.49)	-110.27 (10.95)	78.33 (6.98)	-31.94 (4.85)
Com4	**Chain-A**	**Run1**	-29.75	-27.89	47.50	-3.68	-57.63	43.83	-13.81
		Run2	-22.11	-70.78	74.37	-3.79	-92.88	70.58	-22.30
		Run3	-17.43	-33.59	41.96	-2.46	-51.02	39.50	-11.52
		Avg	-23.10 (6.22)	-44.09 (23.29)	54.61 (17.34)	-3.31 (0.74)	-67.18 (22.50)	51.30 (16.83)	-15.88 (5.68)
	Chain-B	**Run1**	-32.87	-69.17	82.72	-4.65	-102.03	78.08	-23.96
		Run2	-25.20	-54.67	64.95	-3.86	-79.87	61.09	-18.78
		Run3	-23.07	-64.20	73.56	-3.28	-87.27	70.28	-16.99
		Avg	-27.05 (5.15)	-62.68 (7.37)	73.74 (8.89)	-3.93 (0.69)	-89.72 (11.28)	69.82 (8.50)	-19.91 (3.62)
	Chain-C	**Run1**	-22.08	-59.15	67.87	-3.27	-81.27	64.60	-16.63
		Run2	-22.26	-67.66	72.85	-3.77	-89.92	69.08	-20.84
		Run3	-26.44	-63.90	70.84	-4.34	-90.34	66.50	-23.84
		Avg	-23.59 (2.47)	-63.57 (4.26)	70.52 (2.50)	-3.79 (0.53)	-87.18 (5.12)	66.73 (2.25)	-20.44 (3.62)
	Chain-D	**Run1**	-26.22	-61.02	74.38	-4.14	-87.23	70.24	-17.00
		Run2							
		Run3	-22.99	-67.24	72.29	-3.89	-90.23	68.41	-21.82

		Avg	-24.605 (2.28)	-64.13 (4.40)	73.335 (1.48)	-4.015 (0.18)	-88.73 (2.12)	69.325 (1.29)	-19.41 (3.41)
	Chain-E	**Run1**	-25.45	-41.98	52.54	-3.92	-67.43	48.63	-18.81
		Run2	-26.19	-69.30	74.76	-4.41	-95.50	70.35	-25.15
		Run3	-25.13	-58.82	66.55	-4.09	-83.95	62.47	-21.48
		Avg	-25.59 (0.54)	-56.7 (13.78)	64.62 (11.24)	-4.14 (0.25)	-82.29 (14.11)	60.48 (11)	-21.81 (3.18)
Com5	**Chain-A**	**Run1**	-31.21	-57.99	74.39	-4.43	-89.20	69.96	-19.25
		Run2	-17.37	-79.28	86.27	-3.21	-96.65	83.05	-13.59
		Run3							
		Avg	-24.29 (9.79)	-68.635 (15.05)	80.33 (8.4)	-3.82 (0.86)	-92.925 (5.27)	76.505 (9.26)	-16.42 (4.00)
	Chain-B	**Run1**	-21.26	-59.18	67.50	-3.04	-80.43	64.46	-15.98
		Run2							
		Run3	-21.23	-86.07	90.12	-3.44	-107.30	86.68	-20.62
		Avg	-21.24 (0.02)	-72.62 (19.01)	78.81 (15.99)	-3.24 (0.28)	-93.865 (19.00)	75.57 (15.71)	-18.3 (3.28)
	Chain-C	**Run1**							
		Run2	-26.85	-100.63	101.47	-4.60	-127.48	96.87	-30.61
		Run3	-25.09	-63.99	71.18	-4.13	-89.08	67.05	-22.03
		Avg	-25.97 (1.24)	-82.31 (25.91)	86.32 (21.42)	-4.36 (0.33)	-108.28 (27.15)	81.96 (21.09)	-26.32 (6.07)
	Chain-D	**Run1**							
		Run2	-22.32	-32.85	44.62	-3.27	-55.17	41.35	-13.82
		Run3	-25.28	-37.88	53.10	-3.70	-63.16	49.40	-13.76
		Avg	-23.8 (2.09)	-35.36 (3.56)	48.86 (6.00)	-3.48 (0.30)	-59.16 (5.65)	45.375 (5.69)	-13.79 (0.04)
	Chain-E	**Run1**							
		Run2							
		Run3							
		Avg							

4.6 Residue wise contribution of Protein Chain

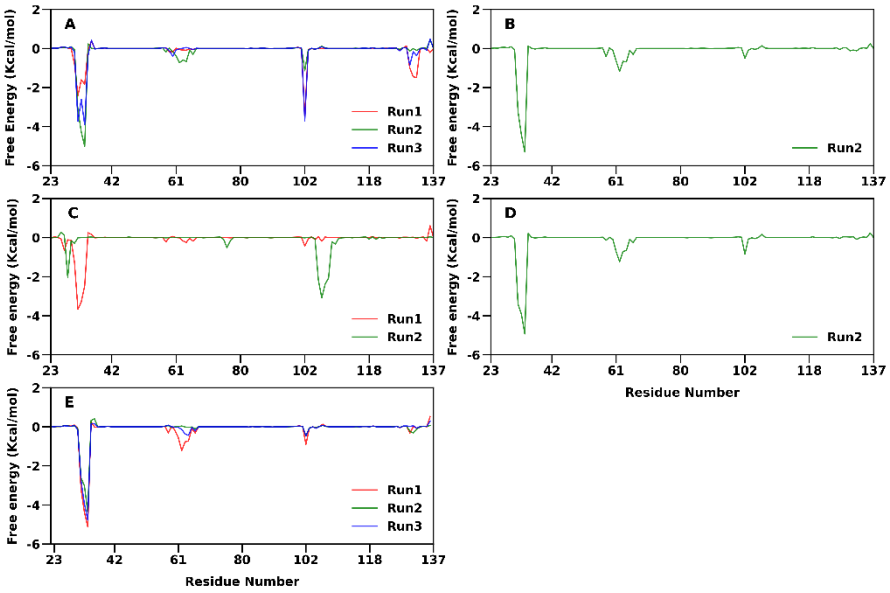

Figure 4.45: The protein chain's amino acid residue wise contribution towards the binding free energy for *Com1*.

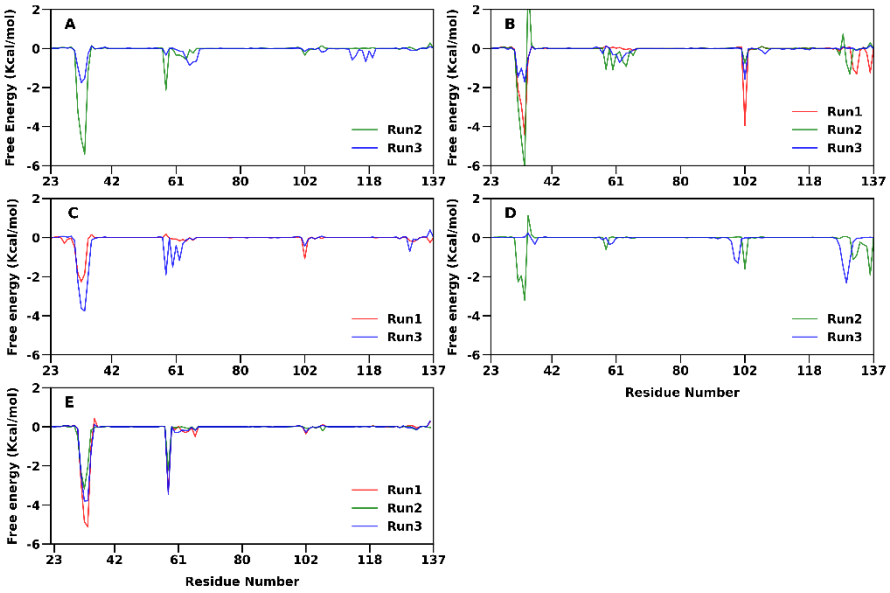

Figure 4.46: The protein chain's amino acid residue wise contribution towards the binding free energy for *Com2*.

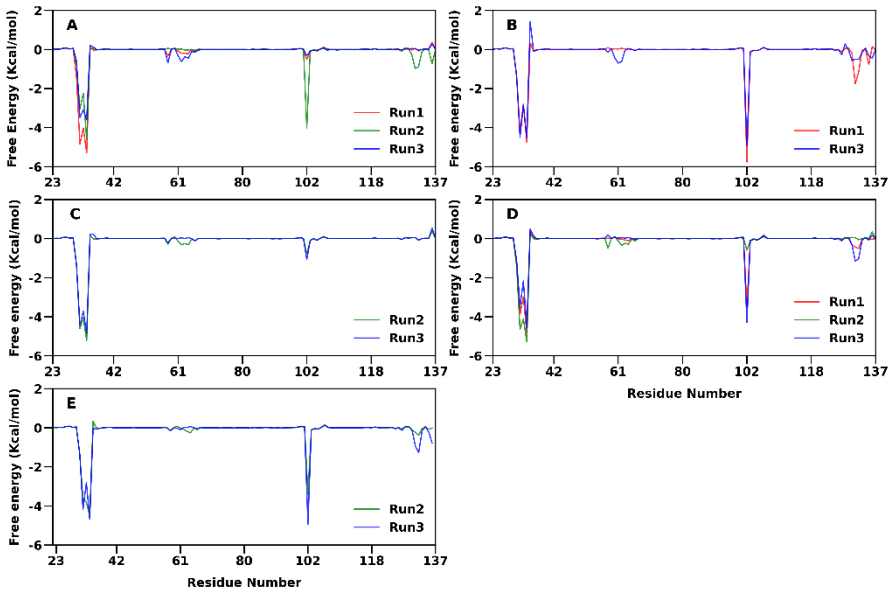

Figure 4.47: The protein chain's amino acid residue wise contribution towards the binding free energy for *Com3*.

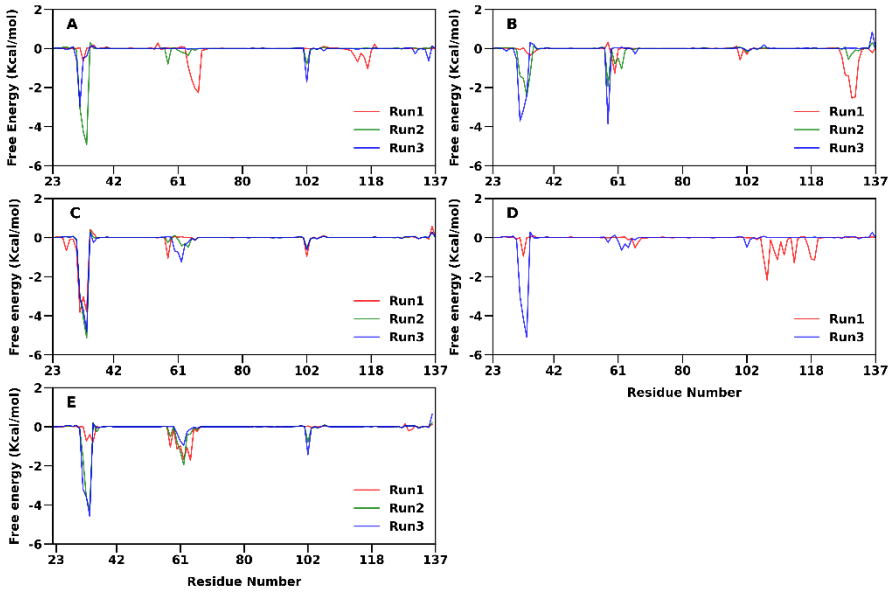

Figure 4.48: The protein chain's amino acid residue wise contribution towards the binding free energy for *Com4*.

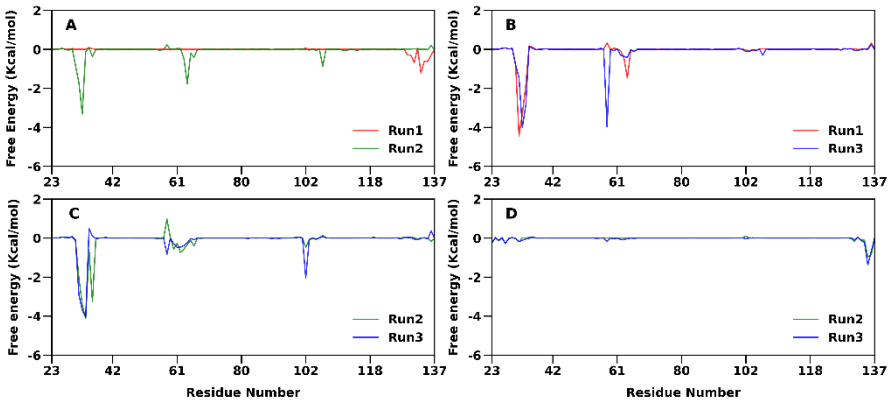

Figure 4.49: The protein chain's amino acid residue wise contribution towards the binding free energy for *Com5*.

The per residue wise contribution of the protein chain in calculating the binding free energy for each chain and each run is shown in **Figures 4.36 – Figure 4.40.** The key residue for the BS1 site is S35 and K59; for BS2, it is W108, T109 and PF113. In each of the cases, we observed a peak at the 35th residue that indicates the glycans are binding with the help of the 35th residue. Also, there is another lower peak at the 59th residue in all the cases, which also indicates the interaction with the BS1 site. The higher peak at the 35th and the 59th residue corresponds to the higher binding free energy in that particular case. The lower peak at the 35th and the 59th residue suggests the lower binding free energy. Also, in cases where there is some lower peak at other sites indicates that the glycan has moved out or interacted with the other residues. The peak at the 109 regions suggests that the glycan interacts with the BS2 site. In a case like *Com4* chain D(run1), there is a very low peak at the 35th residue and have peak at the 108 region, which indicates that this glycan has moved away from the BS1 site and is interacting with the BS2 site.

4.7 Residue wise contribution of monosaccharides

The per residue wise contribution of the monosaccharides in the total binding free energy for all the complexes and all the runs is shown in **Figures 4.41 – Figure 4.45.** Also, the **Table 4.8** shows a similar thing in the tabular form. We found that the sialic acid of both modified and unmodified glycans is the main player responsible for the interactions of the glycans in the binding sites of proteins. The contribution of the

GlcNAc and Gal is very low in the modified glycan-protein complexes. In contrast, the contribution is little higher in the case of the unmodified protein-glycan complexes. Also, in the case of the unmodified glycan-protein complexes, the sialic acid's contribution is lesser than the modified glycan-protein complexes.

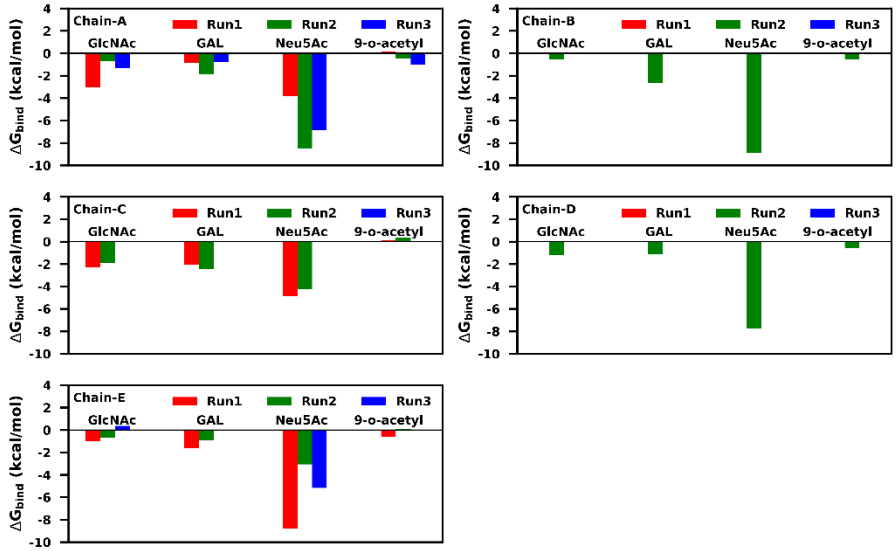

Figure 4.50: Per monosaccharide residue wise contribution of the glycan towards the binding free energy for *Com1*.

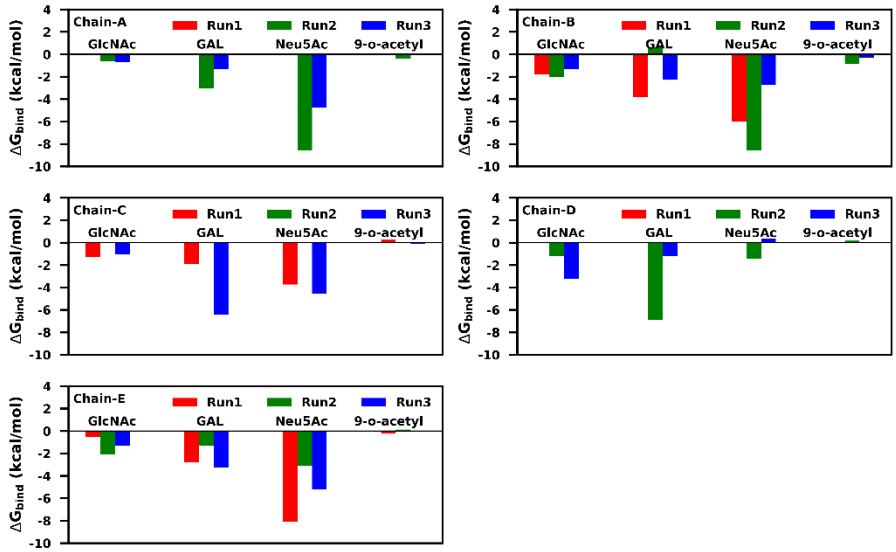

Figure 4.51: Per monosaccharide residue wise contribution of the glycan towards the binding free energy for *Com2*.

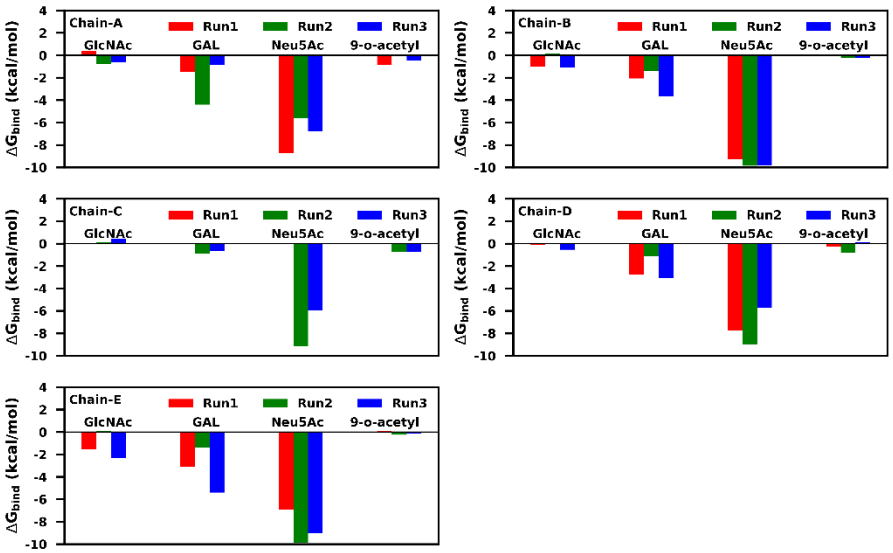

Figure 4.52: Per monosaccharide residue wise contribution of the glycan towards the binding free energy for *Com3*.

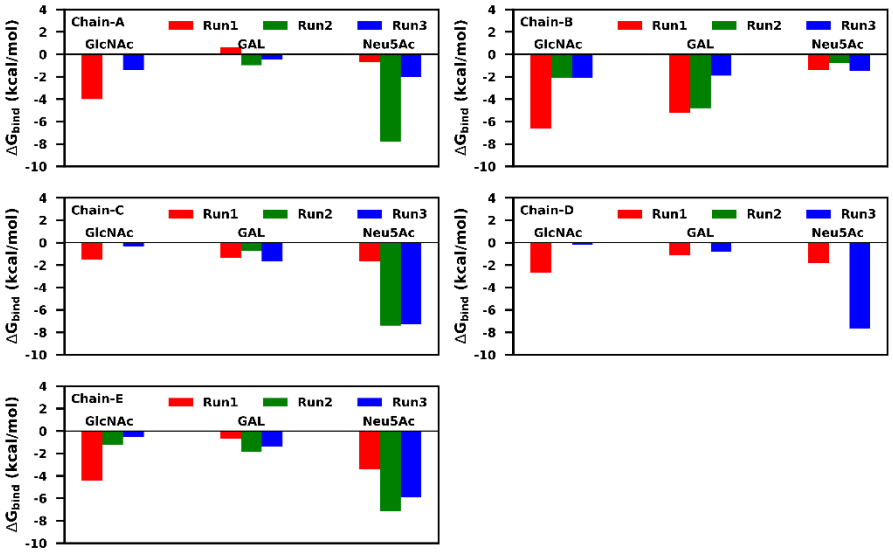

Figure 4.53: Per monosaccharide residue wise contribution of the glycan towards the binding free energy for *Com4*.

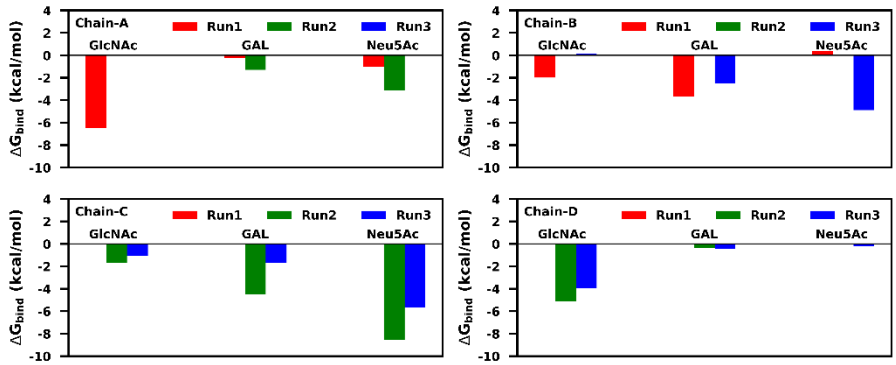

Figure 4.54: Per monosaccharide residue wise contribution of the glycan towards the binding free energy for *Com5*.

Table 4.8: The contribution of each of the monosaccharides towards calculating the binding free energy of five complexes (Run1, Run2 and Run3) in *Com1*.

Monosaccharide			**T_{vdW}**	**T_{ele}**	**T_{GB}**	**T_{np}**	**T_{Total}**
Com1 **Chain-A**	**Reducing end sugar**	**Run1**	-0.23	-0.4	0.68	-0.08	-0.03
		Run2	-0.08	0.52	-0.5	-0.06	-0.13
		Run3	-0.14	2.39	-2.24	-0.1	-0.08
		Avg	-0.15 (0.07)	0.84 (1.42)	-0.69 (1.47)	-0.08 (0.02)	-0.08 (0.05)
	GlcNAc	**Run1**	-4.87	-7.2	9.99	-0.92	-3.01
		Run2	-1.79	-1.78	3.26	-0.37	-0.68
		Run3	-2.34	-4.74	6.19	-0.45	-1.34
		Avg	-3.00 (1.64)	-4.57 (2.71)	6.48 (3.37)	-0.58 (0.30)	-1.68 (1.2)
	Gal	**Run1**	-3.92	-1.1	4.7	-0.55	-0.86
		Run2	-2.32	-6.98	7.93	-0.51	-1.88
		Run3	-3.25	-2.48	5.46	-0.54	-0.81
		Avg	-3.16 (0.80)	-3.52 (3.07)	6.03 (1.69)	-0.53 (0.02)	-1.18 (0.60)
	Neu5Ac	**Run1**	-6.04	-22.76	26.18	-1.25	-3.86
		Run2	-8.13	-25.19	26.37	-1.54	-8.49
		Run3	-5.76	-16.81	16.87	-1.19	-6.88

		Avg	-6.64 (1.29)	-21.59 (4.31)	23.14 (5.43)	-1.33 (0.19)	-6.41 (2.35)
	9-0 acetyl	**Run1**	-0.2	-0.23	0.65	-0.06	0.15
		Run2	-1.96	-0.59	2.7	-0.63	-0.48
		Run3	-1.8	-2.29	3.57	-0.47	-0.99
		Avg	-1.32 (0.98)	-1.04 (1.10)	2.31 (1.50)	-0.39 (0.29)	-0.44 (0.57)
Com1 **Chain-B**	**Reducing end sugar**	**Run1**					
		Run2	-0.2	-0.68	0.5	-0.12	-0.5
		Run3					
		Avg	-0.2	-0.68	0.5	-0.12	-0.5
	GlcNAc	**Run1**					
		Run2	-2.46	-3.12	5.5	-0.48	-0.55
		Run3					
		Avg	-2.46	-3.12	5.5	-0.48	-0.55
	Gal	**Run1**					
		Run2	-2.31	-8.72	8.92	-0.54	-2.65
		Run3					
		Avg	-2.31	-8.72	8.92	-0.54	-2.65
	Neu5Ac	**Run1**					
		Run2	-8.05	-28.63	29.33	-1.52	-8.88
		Run3					
		Avg	-8.05	-28.63	29.33	-1.52	-8.88
	4-0 acetyl	**Run1**					
		Run2	-1.97	-0.26	2.33	-0.63	-0.54
		Run3					
		Avg	-1.97	-0.26	2.33	-0.63	-0.54
Com1 **Chain-C**	**Reducing end sugar**	**Run1**	-0.14	0	0.03	-0.12	-0.23
		Run2	-0.12	-0.41	0.66	-0.05	0.08
		Run3					
		Avg	-0.13 (0.01)	-0.20 (0.29)	0.34 (0.44)	-0.08 (0.05)	-0.07 (0.22)

	GlcNAc	**Run1**	-4.3	-5.09	7.82	-0.74	-2.31
		Run2	-3.7	-2.3	4.72	-0.67	-1.94
		Run3					
		Avg	-4.0 (0.42)	-3.69 (1.98)	6.27 (2.19)	-0.70 (0.05)	-2.12 (0.26)
	Gal	**Run1**	-3.36	-5.17	7	-0.51	-2.04
		Run2	-2.97	-6.35	7.3	-0.47	-2.48
		Run3					
		Avg	-3.16 (0.28)	-5.76 (0.83)	7.15 (0.21)	-0.49 (0.03)	-2.26 (0.31)
	Neu5Ac	**Run1**	-5.71	-22.53	24.49	-1.11	-4.87
		Run2	-5.65	-20.43	22.82	-1.02	-4.28
		Run3					
		Avg	-5.68 (0.04)	-21.48 (1.48)	23.65 (1.18)	-1.06 (0.06)	-4.57 (0.42)
	9-0 acetyl	**Run1**	-0.92	0.48	0.82	-0.28	0.1
		Run2	-0.53	2.13	-1.13	-0.13	0.33
		Run3					
		Avg	-0.72 (0.28)	1.30 (1.17)	-0.15 (1.38)	-0.20 (0.11)	0.21 (0.16)
Com1 **Chain-D**	**Reducing end sugar**	**Run1**					
		Run2	-0.08	-0.79	0.72	-0.09	-0.23
		Run3					
		Avg					
	GlcNAc	**Run1**					
		Run2	-2.35	-1.95	3.53	-0.43	-1.2
		Run3					
		Avg					
	Gal	**Run1**					
		Run2	-2.14	-5.12	6.53	-0.42	-1.15
		Run3					
		Avg					

	Neu5Ac	**Run1**					
		Run2	-7.89	-31.25	32.89	-1.47	-7.73
		Run3					
		Avg					
	9-0 acetyl	**Run1**					
		Run2	-1.8	0.09	1.7	-0.58	-0.59
		Run3					
		Avg					
***Com1* Chain-E**	**Reducing end sugar**	**Run1**	-0.11	-1.24	1.24	-0.09	-0.19
		Run2	-0.03	-0.47	0.51	-0.02	0
		Run3	-0.03	0.43	-0.36	-0.01	0.03
		Avg	-0.06 (0.05)	-0.43 (0.84)	0.46 (0.80)	-0.04 (0.04)	-0.05 (0.12)
	GlcNAc	**Run1**	-2.15	-1.8	3.33	-0.41	-1.03
		Run2	-1.07	-3.57	4.19	-0.24	-0.7
		Run3	-0.59	0.31	0.73	-0.1	0.36
		Avg	-1.27 (0.80)	-1.69 (1.94)	2.75 (1.80)	-0.25 (0.16)	-0.46 (0.73)
	Gal	**Run1**	-2.35	-5.81	7.06	-0.53	-1.64
		Run2	-2.98	-2.87	5.45	-0.52	-0.92
		Run3	-1.73	-1.45	3.45	-0.27	0
		Avg	-2.35 (0.62)	-3.38 (2.22)	5.32 (1.81)	-0.44 (0.15)	-0.85 (0.82)
	Neu5Ac	**Run1**	-8.31	-31.4	32.45	-1.51	-8.77
		Run2	-4.87	-15.62	18.49	-1.07	-3.07
		Run3	-7.33	-23.19	26.91	-1.56	-5.16
		Avg	-6.84 (1.77)	-23.40 (7.89)	25.95 (7.03)	-1.38 (0.27)	-5.67 (2.88)
	9-0 acetyl	**Run1**	-2.05	0.35	1.77	-0.66	-0.6
		Run2	-0.54	-0.99	1.77	-0.18	0.06
		Run3	-1.2	-0.36	1.91	-0.38	-0.02
		Avg	-1.26 (0.76)	-0.33 (0.67)	1.82 (0.08)	-0.41 (0.24)	-0.19 (0.36)

Table 4.9: The contribution of each of the monosaccharides towards calculating the binding free energy of five complexes (Run1, Run2 and Run3) in *Com2*.

Monosaccharide			**T_{vdW}**	**T_{ele}**	**T_{GB}**	**T_{np}**	**T_{Total}**
Com2	**Reducing**	**Run1**					
Chain-A	**end**	**Run2**	-0.05	-0.44	0.41	-0.04	-0.13
	sugar	**Run3**	-0.12	-2.39	2.45	-0.11	-0.17
		Avg	-0.08 (0.05)	-1.41 (1.38)	1.43 (1.44)	-0.07 (0.05)	-0.15 (0.03)
	GlcNAc	**Run1**					
		Run2	-2.27	-1.3	3.41	-0.44	-0.61
		Run3	-2.39	-3.1	5.21	-0.45	-0.73
		Avg	-2.33 (0.08)	-2.2 (1.27)	4.31 (1.27)	-0.44 (0.01)	-0.67 (0.08)
	Gal	**Run1**					
		Run2	-2.09	-12.13	11.59	-0.43	-3.06
		Run3	-2.01	-6.58	7.63	-0.39	-1.35
		Avg	-2.05 (0.06)	-9.35 (3.92)	9.61 (2.8)	-0.41 (0.03)	-2.20 (1.21)
	Neu5Ac	**Run1**					
		Run2	-7.92	-32.82	33.74	-1.54	-8.54
		Run3	-6.82	-24.45	27.85	-1.38	-4.79
		Avg	-7.37 (0.78)	-28.63 (5.92)	30.79 (4.16)	-1.46 (0.11)	-6.665 (2.65)
	9-0 acetyl	**Run1**					
		Run2	-1.46	0.18	1.41	-0.48	-0.36
		Run3	-0.88	0.98	0.27	-0.26	0.1
		Avg	-1.17 (0.41)	0.58 (0.57)	0.84 (0.81)	-0.37 (0.16)	-0.13 (0.32)
Com2		**Run1**	0.26	-3.26	2.15	-0.16	-1.01
Chain-B		**Run2**	-0.57	5.29	-2.54	-0.32	1.86

	Reducing end sugar	**Run3**	-0.15	-0.33	0.52	-0.08	-0.05
		Avg	-0.15 (0.41)	0.57 (4.34)	0.04 (2.38)	-0.19 (0.12)	0.27 (1.46)
	GlcNAc	**Run1**	-3.67	-2.43	4.87	-0.57	-1.8
		Run2	-7.8	1.72	5.27	-1.18	-2
		Run3	-3.01	-1.11	3.31	-0.55	-1.35
		Avg	-4.83 (2.60)	-0.61 (2.12)	4.48 (1.04)	-0.77 (0.36)	-1.72 (0.33)
	Gal	**Run1**	-4.61	-13.08	14.7	-0.79	-3.79
		Run2	-0.92	-1.83	3.49	-0.1	0.64
		Run3	-1.74	-7.87	7.78	-0.43	-2.25
		Avg	-2.42 (1.94)	-7.59 (5.63)	8.66 (5.66)	-0.44 (0.34)	-1.8 (2.25)
	Neu5Ac	**Run1**	-4.85	-27.96	27.83	-1.03	-6.01
		Run2	-7.56	-26.58	27.1	-1.54	-8.59
		Run3	-1.74	-7.87	7.78	-0.43	-2.25
		Avg	-4.72 (2.91)	-20.80 (11.22)	20.90 (11.37)	-1.00 (0.56)	-5.62 (3.19)
	4-0 acetyl	**Run1**	-0.68	-0.11	1.1	-0.22	0.08
		Run2	-1.94	-2.09	3.85	-0.65	-0.84
		Run3	-1.01	-1.41	2.4	-0.27	-0.29
		Avg	-1.21 (0.65)	-1.20 (1.01)	2.45 (1.38)	-0.38 (0.23)	-0.35 (0.46)
***Com2* Chain-C**	**Reducing end sugar**	**Run1**	-0.13	-3.99	3.84	-0.1	-0.38
		Run2					
		Run3	-0.11	0.56	-0.33	-0.08	0.05
		Avg	-0.12 (0.01)	-1.715 (3.22)	1.755 (2.95)	-0.09 (0.01)	-0.165 (0.30)
	GlcNAc	**Run1**	-2.84	-3.65	5.77	-0.56	-1.28
		Run2					
		Run3	-2.32	-2.93	4.72	-0.51	-1.04

		Avg	-2.58 (0.37)	-3.29 (0.51)	5.24 (0.74)	-0.53 (0.03)	-1.16 (0.17)
	Gal	**Run1**	-2.16	-5.84	6.53	-0.42	-1.89
		Run2					
		Run3	-4.39	-19.96	18.71	-0.82	-6.46
		Avg	-3.27 (1.58)	-12.9 (9.98)	12.62 (8.61)	-0.62 (0.28)	-4.17 (3.23)
	Neu5Ac	**Run1**	-5.26	-15.73	18.26	-1	-3.73
		Run2					
		Run3	-6.26	-25.77	28.73	-1.27	-4.56
		Avg	-5.76 (0.71)	-20.75 (7.10)	23.49 (7.4)	-1.13 (0.19)	-4.14 (0.59)
	9-0 acetyl	**Run1**	-0.78	0.42	0.78	-0.18	0.24
		Run2					
		Run3	-0.79	0.97	0	-0.25	-0.08
		Avg	-0.79 (0.01)	0.69 (0.39)	0.39 (0.55)	-0.22 (0.05)	0.08 (0.23)
***Com2* Chain-D**	**Reducing end sugar**	**Run1**					
		Run2	-0.09	0.41	-0.27	-0.03	0.02
		Run3					
		Avg					
	GlcNAc	**Run1**					
		Run2	-2.83	-3.29	5.45	-0.55	-1.22
		Run3					
		Avg					
	Gal	**Run1**					
		Run2	-3.72	-17.05	14.77	-0.89	-6.89
		Run3					
		Avg					
	Neu5Ac	**Run1**					
		Run2	-3.51	-12.87	15.71	-0.78	-1.45
		Run3					

		Avg					
	9-0 acetyl	**Run1**					
		Run2	-0.36	-0.28	0.92	-0.11	0.17
		Run3					
		Avg					
***Com2* Chain-E**	**Reducing end sugar**	**Run1**	-0.06	0.07	-0.01	-0.02	-0.02
		Run2	0.01	-2.87	1.99	-0.17	-1.04
		Run3					
		Avg	-0.03 (0.05)	-1.4 (2.08)	0.99 (1.41)	-0.09 (0.11)	-0.53 (0.72)
	GlcNAc	**Run1**	-1.87	-1.34	3.08	-0.42	-0.56
		Run2	-3.94	-3.71	6.25	-0.73	-2.12
		Run3					
		Avg	-2.90 (1.46)	-2.52 (1.68)	4.66 (2.24)	-0.57 (0.22)	-1.34 (1.10)
	Gal	**Run1**	-1.79	-10.96	10.33	-0.35	-2.77
		Run2	-2.1	-5.05	6.15	-0.32	-1.33
		Run3					
		Avg	-1.945 (0.22)	-8.00 (4.18)	8.24 (2.96)	-0.33 (0.02)	-2.05 (1.02)
	Neu5Ac	**Run1**	-7.4	-29.22	29.95	-1.45	-8.12
		Run2	-4.33	-20.48	22.63	-0.94	-3.12
		Run3					
		Avg	-5.865 (2.17)	-24.85 (6.18)	26.29 (5.18)	-1.19 (0.36)	-5.62 (3.54)
	9-0 acetyl	**Run1**	-1.74	0.9	1.19	-0.58	-0.23
		Run2	-0.51	0.61	0.19	-0.15	0.13
		Run3					
		Avg	-1.12 (0.87)	0.75 (0.20)	0.69 (0.71)	-0.36 (0.30)	-0.05 (0.26)

Table 4.10: The contribution of each of the monosaccharides towards calculating the binding free energy of five complexes (Run1, Run2 and Run3) in *Com3*.

Monosaccharide			**T_{vdW}**	**T_{ele}**	**T_{GB}**	**T_{np}**	**T_{Total}**
Com3 **Chain-A**	**Reducing end sugar**	**Run1**	-0.03	-0.3	0.35	-0.01	0.01
		Run2	-0.12	0.02	0.13	-0.03	-0.01
		Run3	-0.11	-0.73	0.83	-0.07	-0.09
		Avg	-0.09 (0.05)	-0.34 (0.38)	0.44 (0.36)	-0.04 (0.03)	-0.03 (0.05)
	GlcNAc	**Run1**	-0.76	-0.43	1.74	-0.16	0.39
		Run2	-2.7	-2.64	5.12	-0.55	-0.78
		Run3	-1.91	-1.21	2.82	-0.33	-0.64
		Avg	-1.79 (0.98)	-1.43 (1.12)	3.23 (1.7)	-0.35 (0.20)	-0.34 (0.64)
	Gal	**Run1**	-2.02	-5.85	6.88	-0.47	-1.46
		Run2	-4.07	-11.48	11.92	-0.76	-4.4
		Run3	-1.54	-4.7	5.68	-0.33	-0.89
		Avg	-2.54 (1.34)	-7.34 (3.63)	8.16 (3.31)	-0.52 (0.22)	-2.25 (1.89)
	Neu5Ac	**Run1**	-8.22	-33.46	34.56	-1.61	-8.74
		Run2	-5.1	-25.81	26.25	-0.95	-5.61
		Run3	-7.05	-24.53	26.24	-1.45	-6.78
		Avg	-6.79 (1.58)	-27.93 (4.83)	29.02 (4.80)	-1.34 (0.34)	-7.04 (1.58)
	9-0 acetyl	**Run1**	-2.06	-2.57	4.19	-0.43	-0.87
		Run2	-1.35	-1.87	3.64	-0.33	0.1
		Run3	-1.02	-1.28	2.08	-0.23	-0.44
		Avg	-1.48 (0.53)	-1.91 (0.65)	3.30 (1.09)	-0.33 (0.1)	-0.40 (0.49)
Com3 **Chain-B**	**Reducing end sugar**	**Run1**	-0.08	2.01	-1.86	-0.04	0.04
		Run2	-0.04	0.73	-0.64	-0.02	0.03
		Run3	-0.25	1.92	-1.72	-0.17	-0.22

		Avg	-0.12 (0.11)	1.55 (0.71)	-1.41 (0.67)	-0.08 (0.08)	-0.05 (0.15)
	GlcNAc	**Run1**	-2.38	-2.64	4.49	-0.51	-1.04
		Run2	-0.92	-0.15	1.36	-0.17	0.12
		Run3	-3.51	-5.73	8.79	-0.67	-1.12
		Avg	-2.27 (1.3)	-2.84 (2.8)	4.88 (3.73)	-0.45 (0.26)	-0.68 (0.69)
	Gal	**Run1**	-5.04	-4.92	8.68	-0.8	-2.07
		Run2	-1.95	-4.74	5.63	-0.33	-1.39
		Run3	-5.16	-7.25	9.56	-0.82	-3.68
		Avg	-4.05 (1.82)	-5.64 (1.4)	7.96 (2.06)	-0.65 (0.28)	-2.38 (1.18)
	Neu5Ac	**Run1**	-6.87	-41	39.98	-1.35	-9.25
		Run2	-7.71	-30.91	30.42	-1.66	-9.86
		Run3	-7.32	-27.34	26.26	-1.41	-9.81
		Avg	-7.3 (0.42)	-33.08 (7.08)	32.22 (7.03)	-1.47 (0.16)	-9.64 (0.34)
	4-0 acetyl	**Run1**	-1.35	-0.08	1.74	-0.35	-0.05
		Run2	-1.34	-2.29	3.71	-0.33	-0.25
		Run3	-1.39	-2.36	3.84	-0.34	-0.25
		Avg	-1.36 (0.03)	-1.58 (1.3)	3.10 (1.18)	-0.34 (0.01)	-0.18 (0.12)
Com3 **Chain-C**	**Reducing end sugar**	**Run1**					
		Run2	-0.06	0.23	-0.15	-0.03	0
		Run3	-0.01	0.8	-0.76	-0.01	0.02
		Avg	-0.03 (0.03)	0.51 (0.40)	-0.45 (0.43)	-0.02 (0.01)	0.01 (0.01)
	GlcNAc	**Run1**					
		Run2	-1.01	-0.55	1.87	-0.21	0.1
		Run3	-0.42	0	0.92	-0.05	0.44
		Avg	-0.71 (0.42)	-0.27 (0.39)	1.39 (0.67)	-0.13 (0.11)	0.27 (0.24)
	Gal	**Run1**					

		Run2	-1.81	-4.95	6.25	-0.39	-0.9
		Run3	-1.43	-4.3	5.39	-0.31	-0.64
		Avg	-1.62 (0.27)	-4.62 (0.46)	5.82 (0.61)	-0.35 (0.06)	-0.77 (0.18)
	Neu5Ac	**Run1**					
		Run2	-8.34	-29.85	30.71	-1.64	-9.12
		Run3	-6.84	-20.67	22.94	-1.37	-5.94
		Avg	-7.59 (1.06)	-25.26 (6.49)	26.82 (5.49)	-1.50 (0.19)	-7.53 (2.25)
	9-0 acetyl	**Run1**					
		Run2	-1.94	-3.19	4.76	-0.41	-0.77
		Run3	-1.85	-3.25	4.79	-0.4	-0.72
		Avg	-1.89 (0.06)	-3.22 (0.04)	4.77 (0.02)	-0.40 (0.01)	-0.74 (0.04)
***Com3* Chain-D**	**Reducing end sugar**	**Run1**	-0.05	1.36	-1.3	-0.02	-0.01
		Run2	-0.05	0.12	-0.06	-0.03	-0.02
		Run3	-0.08	2	-1.86	-0.02	0.04
		Avg	-0.06 (0.02)	1.16 (0.96)	-1.07 (0.92)	-0.02 (0.01)	0.00 (0.03)
	GlcNAc	**Run1**	-1.37	-1.36	2.9	-0.3	-0.13
		Run2	-1.02	-0.79	2.1	-0.22	0.07
		Run3	-1.81	-3.5	5.19	-0.44	-0.55
		Avg	-1.4 (0.40)	-1.88 (1.43)	3.40 (1.60)	-0.32 (0.11)	-0.20 (0.32)
	Gal	**Run1**	-3.36	-7.23	8.5	-0.65	-2.74
		Run2	-1.7	-5.33	6.3	-0.39	-1.12
		Run3	-4.61	-7.64	9.98	-0.83	-3.11
		Avg	-3.22 (1.46)	-6.73 (1.23)	8.26 (1.85)	-0.62 (0.22)	-2.32 (1.06)
	Neu5Ac	**Run1**	-6.09	-23.81	23.32	-1.16	-7.74
		Run2	-8.42	-27.69	28.77	-1.66	-9
		Run3	-4.63	-22.49	22.17	-0.78	-5.73

		Avg	-6.38 (1.91)	-24.66 (2.70)	24.75 (3.53)	-1.2 (0.44)	-7.49 (1.65)
	9-0 acetyl	**Run1**	-1.84	-2.6	4.59	-0.44	-0.29
		Run2	-2.02	-3.54	5.18	-0.42	-0.79
		Run3	-1.74	-2.16	4.5	-0.46	0.04
		Avg	-1.87 (0.14)	-2.77 (0.70)	4.76 (0.37)	-0.44 (0.02)	-0.35 (0.42)
***Com3* Chain-E**	**Reducing end sugar**	**Run1**	-0.13	1.23	-1.01	-0.04	0.04
		Run2	-0.04	0.73	-0.64	-0.02	0.03
		Run3	-0.09	2.59	-2.46	-0.02	0.02
		Avg	-0.09 (0.05)	1.52 (0.96)	-1.37 (0.96)	-0.03 (0.01)	0.03 (0.01)
	GlcNAc	**Run1**	-2.89	-4.94	6.91	-0.6	-1.52
		Run2	-0.92	-0.15	1.36	-0.17	0.12
		Run3	-2.71	-5.65	6.68	-0.62	-2.3
		Avg	-2.17 (1.09)	-3.58 (2.99)	4.98 (3.14)	-0.46 (0.25)	-1.23 (1.24)
	Gal	**Run1**	-4.51	-7.81	9.95	-0.74	-3.11
		Run2	-1.95	-4.74	5.63	-0.33	-1.39
		Run3	-4.5	-11.17	11.09	-0.83	-5.4
		Avg	-3.65 (1.48)	-7.91 (3.22)	8.89 (2.88)	-0.63 (0.27)	-3.3 (2.01)
	Neu5Ac	**Run1**	-5.55	-28.1	27.79	-1.08	-6.93
		Run2	-7.71	-30.91	30.42	-1.66	-9.86
		Run3	-6.67	-28.11	26.92	-1.19	-9.06
		Avg	-6.64 (1.08)	-29.04 (1.62)	28.38 (1.82)	-1.31 (0.31)	-8.62 (1.51)
	9-0 acetyl	**Run1**	-1.23	-2.01	3.63	-0.31	0.08
		Run2	-1.34	-2.29	3.71	-0.33	-0.25
		Run3	-1.44	-2.41	4.08	-0.36	-0.13
		Avg	-1.34 (0.10)	-2.24 (0.21)	3.81 (0.24)	-0.33 (0.03)	-0.1 (0.17)

Table 4.11: The contribution of each of the monosaccharides towards calculating the binding free energy of five complexes (Run1, Run2 and Run3) in *Com4*.

Monosaccharide			**T_{vdW}**	**T_{ele}**	**T_{GB}**	**T_{np}**	**T_{Total}**
Com4	**Reducing**	**Run1**	-1.01	-2.8	2.94	-0.31	-1.18
Chain-A	**end sugar**	**Run2**	-0.04	0.51	-0.43	-0.02	0.02
		Run3	-0.21	-0.74	0.81	-0.12	-0.26`
		Avg	-0.42 (0.52)	-1.01 (1.67)	1.11 (1.70)	-0.15 (0.15)	-0.47 (0.63)
	GlcNAc	**Run1**	-9.04	-1.41	7.78	-1.34	-4.01
		Run2	-0.85	-0.62	1.65	-0.17	0.02
		Run3	-3.35	-0.87	3.27	-0.47	-1.41
		Avg	-4.41 (4.20)	-0.97 (0.40)	4.23 (3.18)	-0.66 (0.61)	-1.8 (2.04)
	Gal	**Run1**	-1.25	-0.54	2.56	-0.14	0.63
		Run2	-1.55	-5.33	6.22	-0.32	-0.98
		Run3	-1.89	-1.83	3.52	-0.27	-0.48
		Avg	-1.56 (0.32)	-2.57 (2.48)	4.1 (1.90)	-0.24 (0.09)	-0.27 (0.82)
	Neu5Ac	**Run1**	-3.57	-9.19	12.74	-0.71	-0.73
		Run2	-8.62	-29.94	32.66	-1.91	-7.82
		Run3	-3.26	-13.36	15.37	-0.79	-2.04
		Avg	-5.15 (3.01)	-17.50 (10.98)	20.26 (10.82)	-1.14 (0.67)	-3.53 (3.77)
Com4	**Reducing**	**Run1**	-0.58	4.41	-3.55	-0.16	0.11
Chain-B	**end sugar**	**Run2**	-0.06	1.08	-1.16	-0.12	-0.25
		Run3	-0.17	1.13	-0.81	-0.09	0.06
		Avg	-0.27 (0.27)	2.21 (1.91)	-1.84 (1.49)	-0.12 (0.03)	-0.03 (0.19)
	GlcNAc	**Run1**	-8.92	-7.85	11.38	-1.23	-6.63
		Run2	-3.99	-5.9	8.46	-0.68	-2.12
		Run3	-3.86	-5.82	8.2	-0.66	-2.14

		Avg	-5.59 (2.88)	-6.52333 (1.15)	9.35 (1.77)	-0.86 (0.32)	-3.63 (2.60)
	Gal	**Run1**	-3.48	-13.15	11.94	-0.53	-5.23
		Run2	-3.86	-11.01	10.76	-0.69	-4.8
		Run3	-3.4	-4.48	6.43	-0.45	-1.91
		Avg	-3.58 (0.25)	-9.55 (4.52)	9.71 (2.90)	-0.56 (0.12)	-3.98 (1.81)
	Neu5Ac	**Run1**	-3.45	-17.99	20.89	-0.83	-1.38
		Run2	-4.96	-11.51	16.34	-0.94	-0.8
		Run3	-4.1	-22.93	26.48	-0.95	-1.5
		Avg	-4.17 (0.76)	-17.48 (5.73)	21.24 (5.08)	-0.91 (0.07)	-1.23 (0.37)
***Com4* Chain-C**	**Reducing end sugar**	**Run1**	-0.1	-0.37	0.38	-0.09	-0.18
		Run2	-0.06	0.08	0	-0.03	-0.01
		Run3	-0.06	1.08	-1.03	-0.03	-0.03
		Avg	-0.07 (0.02)	0.26 (0.74)	-0.22 (0.73)	-0.05 (0.03)	-0.07 (0.09)
	GlcNAc	**Run1**	-3.34	-4.23	6.62	-0.59	-1.53
		Run2	-1.06	-0.96	2.3	-0.23	0.05
		Run3	-1.77	-0.55	2.28	-0.33	-0.37
		Avg	-2.06 (1.17)	-1.91 (2.02)	3.73 (2.50)	-0.38 (0.19)	-0.62 (0.82)
	Gal	**Run1**	-3.21	-3.97	6.24	-0.46	-1.39
		Run2	-1.53	-4.38	5.52	-0.32	-0.72
		Run3	-2.95	-5.16	6.9	-0.48	-1.69
		Avg	-2.56 (0.90)	-4.50 (0.60)	6.22 (0.69)	-0.42 (0.09)	-1.27 (0.50)
	Neu5Ac	**Run1**	-4.39	-21.01	24.73	-1.04	-1.7
		Run2	-8.48	-28.56	31.54	-1.9	-7.4
		Run3	-8.45	-27.31	30.32	-1.87	-7.31
		Avg	-7.11 (2.35)	-25.63 (4.05)	28.86 (3.63)	-1.60 (0.49)	-5.47 (3.26)

***Com4* Chain-D**	**Reducing end sugar**	**Run1**	-0.31	-2.24	2.5	-0.19	-0.24
		Run2					
		Run3	-0.07	0.45	-0.37	-0.04	-0.02
		Avg	-0.19 (0.17)	-0.89 (1.90)	1.065 (2.03)	-0.11 (0.11)	-0.13 (0.16)
	GlcNAc	**Run1**	-5.19	-3.12	6.47	-0.83	-2.67
		Run2					
		Run3	-1.19	-1.22	2.53	-0.33	-0.81
		Avg	-3.19 (2.83)	-2.17 (1.34)	4.5 (2.79)	-0.58 (0.35)	-1.74 (1.32)
	Gal	**Run1**	-2.08	-4.66	5.95	-0.32	-1.1
		Run2					
		Run3	-1.63	-4.52	5.68	-0.33	-0.81
		Avg	-1.86 (0.32)	-4.59 (0.10)	5.81 (0.19)	-0.32 (0.01)	-0.96 (0.21)
	Neu5Ac	**Run1**	-5.53	-20.5	25.47	-1.29	-1.85
		Run2					
		Run3	-8.6	-28.33	31.19	-1.89	-7.64
		Avg	-7.06 (2.17)	-24.41 (5.54)	28.33 (4.04)	-1.59 (0.42)	-4.74 (4.09)
***Com4* Chain-E**	**Reducing end sugar**	**Run1**	-0.03	0.57	-1.35	-0.22	-1.02
		Run2	-0.13	0.24	-0.2	-0.07	-0.15
		Run3	-0.1	0.42	-0.32	-0.05	-0.06
		Avg	-0.09 (0.05)	0.41 (0.16)	-0.62 (0.63)	-0.11 (0.09)	-0.41 (0.53)
	GlcNAc	**Run1**	-5.45	-8.1	9.97	-0.87	-4.46
		Run2	-2.83	-2.23	4.28	-0.45	-1.23
		Run3	-1.93	-0.9	2.59	-0.32	-0.55
		Avg	-3.40 (1.83)	-3.74 (3.83)	5.61 (3.87)	-0.55 (2.9)	-2.08 (2.09)
	Gal	**Run1**	-2.15	-3.06	4.78	-0.27	-0.71
		Run2	-3.24	-6.47	8.39	-0.51	-1.83
		Run3	-2.54	-5.94	7.59	-0.49	-1.38

		Avg	-2.64 (0.55)	-5.16 (1.83)	6.92 (1.90)	-0.42 (0.13)	-1.31 (0.56)
	Neu5Ac	**Run1**	-5.1	-10.4	13.09	-1.05	-3.45
		Run2	-6.9	-26.19	27.57	-1.64	-7.17
		Run3	-8	-23	26.8	-1.75	-5.94
		Avg	-6.7 (1.46)	-19.86 (8.35)	22.49 (8.15)	-1.48 (0.38)	-5.52 (1.90)

Table 4.12: The contribution of each of the monosaccharides towards calculating the binding free energy of five complexes (Run1, Run2 and Run3) in *Com3*.

Monosaccharide			T_{vdW}	T_{ele}	T_{GB}	T_{np}	T_{Total}
Com5	**Reducing**	**Run1**	-0.62	-4.38	4.65	-0.33	-0.67
Chain-A	**end sugar**	**Run2**	-0.05	-0.48	0.57	-0.01	0.03
		Run3					
		Avg	-0.33 (0.40)	-2.43 (2.76)	2.61 (2.89)	-0.17 (0.23)	-0.32 (0.49)
	GlcNAc	**Run1**	-8.27	-11.85	15.03	-1.39	-6.47
		Run2	-1.04	-1.51	2.79	-0.16	0.08
		Run3					
		Avg	-4.66 (5.11)	-6.68 (7.31)	8.91 (8.66)	-0.77 (0.87)	-3.19 (4.63)
	Gal	**Run1**	-2.24	-2.33	4.64	-0.32	-0.24
		Run2	-2.95	-7.37	9.65	-0.63	-1.3
		Run3					
		Avg	-2.59 (0.50)	-4.85 (3.56)	7.14 (3.54)	-0.47 (0.22)	-0.77 (0.75)
	Neu5Ac	**Run1**	-4.48	-10.45	14.89	-1.01	1.04
		Run2	-4.65	-30.27	33.01	-1.21	-3.12
		Run3					
		Avg	-4.56 (0.12)	-20.36 (14.01)	23.95 (12.81)	-1.11 (0.14)	-1.04 (2.94)
Com5		**Run1**	-0.01	-0.8	0.71	-0.05	-0.15

Chain-B	**Reducing end sugar**	**Run2**					
		Run3	-0.05	0.2	-0.09	-0.02	0.04
		Avg	-0.03 (0.03)	-0.3 (0.71)	0.31 (0.57)	-0.03 (0.02)	-0.06 (0.13)
	GlcNAc	**Run1**	-4.2	-4.16	7.03	-0.66	-2
		Run2					
		Run3	-1.16	-1.23	2.77	-0.23	0.15
		Avg	-2.68 (2.15)	-2.69 (2.07)	4.9 (3.01)	-0.44 (0.30)	-0.92 (1.52)
	Gal	**Run1**	-4.32	-10.69	12.08	-0.77	-3.71
		Run2					
		Run3	-2.91	-9.69	10.65	-0.59	-2.54
		Avg	-3.61 (1.0)	-10.19 (0.71)	11.36 (1.01)	-0.68 (0.13)	-3.12 (0.83)
	Neu5Ac	**Run1**	-2.09	-13.94	16.99	-0.58	0.38
		Run2					
		Run3	-6.5	-32.31	35.31	-1.39	-4.89
		Avg	-4.29 (3.12)	-23.12 (13.00)	26.15 (12.95)	-0.98 (0.57)	-2.26 (3.73)
Com5 **Chain-C**	**Reducing end sugar**	**Run1**					
		Run2	-0.12	-1.59	1.56	-0.04	-0.2
		Run3	-0.12	1.07	-0.94	-0.03	-0.03
		Avg	-0.12 (0)	-0.26 (1.88)	0.31 (1.77)	-0.03 (0.01)	-0.11 (0.12)
	GlcNAc	**Run1**					
		Run2	-3.09	-2.05	3.89	-0.43	-1.68
		Run3	-2.71	-3.1	5.26	-0.54	-1.09
		Avg	-2.9 (0.27)	-2.58 (0.74)	4.58 (0.97)	-0.49 (0.08)	-1.39 (0.42)
	Gal	**Run1**					
		Run2	-3.19	-12.65	11.97	-0.61	-4.47
		Run3	-2	-6.96	7.72	-0.43	-1.69

		Avg	-2.59 (0.84)	-9.80 (4.02)	9.84 (3.00)	-0.52 (0.13)	-3.08 (1.97)
	Neu5Ac	**Run1**					
		Run2	-7.02	-34.02	34.11	-1.65	-8.57
		Run3	-7.72	-23	26.72	-1.69	-5.68
		Avg	-7.37 (0.49)	-28.51 (7.79)	30.41 (5.23)	-1.67 (0.03)	-7.12 (2.04)
***Com5* Chain-D**	**Reducing end sugar**	**Run1**					
		Run2	-0.07	-1.63	0.98	-0.21	-0.92
		Run3	-0.46	-1.22	1.91	-0.19	0.04
		Avg	-0.26 (0.28)	-1.42 (0.29)	1.45 (0.66)	-0.2 (0.01)	-0.44 (0.68)
	GlcNAc	**Run1**					
		Run2	-6.44	-5.99	8.36	-1.09	-5.15
		Run3	-5.75	-5.11	7.86	-0.93	-3.93
		Avg	-6.09 (0.49)	-5.55 (0.62)	8.11 (0.35)	-1.01 (0.11)	-4.54 (0.86)
	Gal	**Run1**					
		Run2	-2.19	-3.06	5.2	-0.33	-0.37
		Run3	-3.03	-2.82	5.87	-0.5	-0.48
		Avg	-2.61 (0.59)	-2.94 (0.17)	5.53 (0.47)	-0.41 (0.12)	-0.42 (0.08)
	Neu5Ac	**Run1**					
		Run2	-2.46	-5.75	8.63	-0.5	-0.08
		Run3	-3.41	-9.79	13.78	-0.77	-0.19
		Avg	-2.93 (0.67)	-7.77 (2.86)	11.20 (3.64)	-0.63 (0.19)	-0.13 (0.08)
***Com5* Chain-E**	**Reducing end sugar**	**Run1**					
		Run2					
		Run3					
		Avg					
	GlcNAc	**Run1**					
		Run2					

		Run3					
		Avg					
	Gal	**Run1**					
		Run2					
		Run3					
		Avg					
	Neu5Ac	**Run1**					
		Run2					
		Run3					
		Avg					

4.8 Hydrogen bonds and protein-glycan interaction

We chose the glycans and proteins with the best binding free energy for analysing the protein glycan interactions shown in **Figures 4.46**. For *Com1*, the selected glycan was bound with the chain B in run2 (-29.21). For *Com2*, the selected glycan was bound with the chain A in run2 (-30.07). For *Com3*, the selected glycan was bound with the chain E in run3 (-37.45). For *Com4*, the selected glycan was bound with the chain E in run2 (-25.15). For *Com5*, the selected glycan was bound with the chain C in run2 (-30.61). The interaction diagram shows that both hydrogen bonding and hydrophobic interactions are responsible for stable binding. Also, we observed that the modified glycans interact with the binding sites via the sialic acid and the modification of the sialic acid. While in the case of the unmodified glycans, they interact with the binding sites via all the three monosaccharides, i.e. GlcNAc, Gal and Neu5Ac. The key residues in the BS1 sites are S35, and the K59; these two resides has been found to take part in the hydrogen bonding in all five cases. Also, S35 residue has higher hydrogen bonding occupancy than K59. In the case of the modified glycans, the hydrogen bond is also

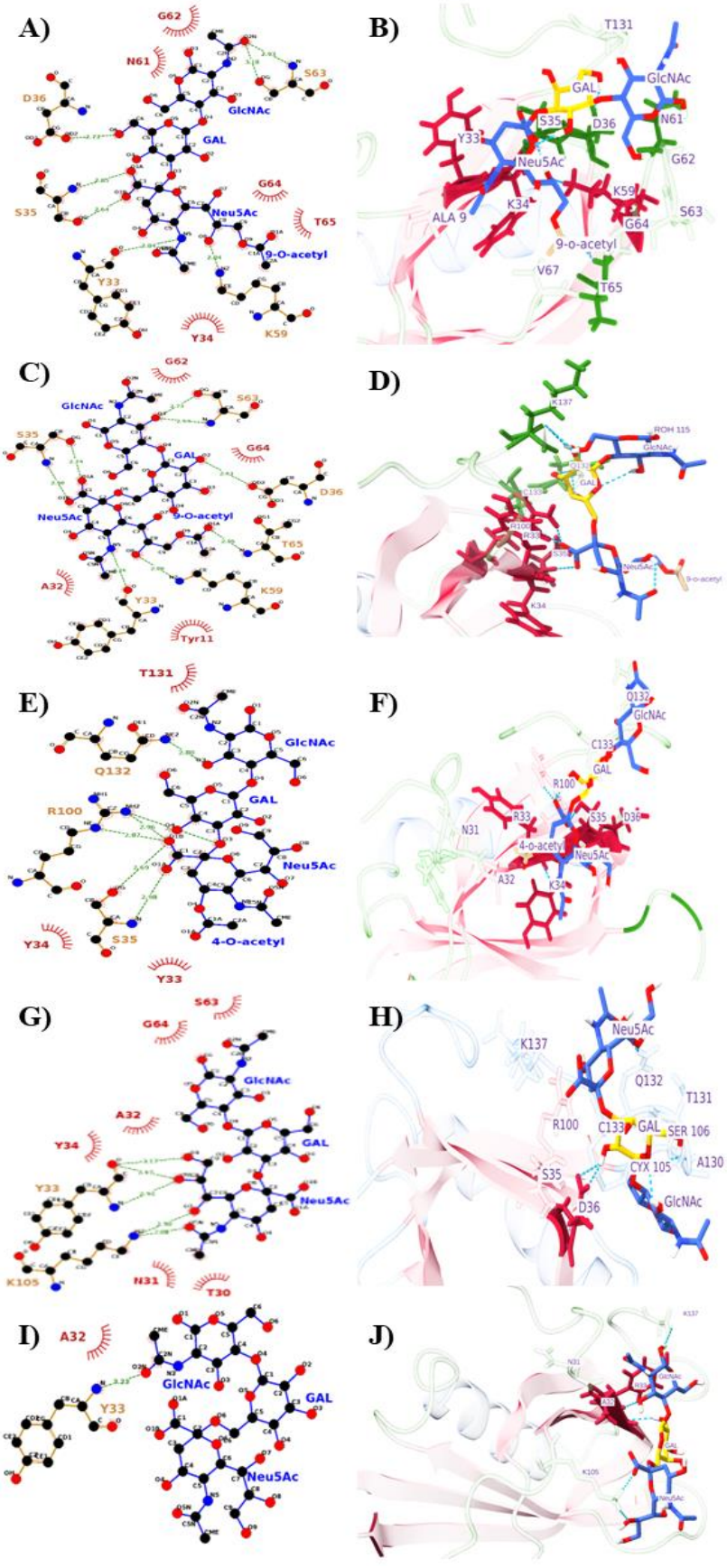

Figure 4.55: Protein-glycan interaction A) and B) *Com1*; C) and D) *Com2*; E) and F) *Com3*; G and H) *Com4*; I) and J) *Com5*

formed because of the modification on the sialic acid, while it is absent in the case of unmodified glycans. The detailed table of the hydrogen bonding between the glycans and the protein for all the chains and runs is given in Table 11.

Table 4.13: Occupancy of hydrogen bonds between glycan and protein during MD simulation.

		Acceptor	**Donor**	**Distance**	**Angle**	**Occupancy (%)**
***Com1* Chain-A**	**Run1**	Neu5Ac@O1A	R100@NH2	2.79	158.79	38.01
		Neu5Ac@O1A	S35@OG	2.63	162.26	35.01
		Neu5Ac@O1B	R100@NH2	2.79	159.76	33.47
		Neu5Ac@O1B	S35@OG	2.63	161.92	29.27
		S35@O	Gal@O2	2.78	158.77	34.06
		T131@O	Neu5Ac@O4	2.78	159.11	29.89
		Y33@O	GlcNAc@O3	2.70	159.47	19.26
		T131@O	Neu5Ac@N5	2.88	158.09	12.58
	Run2	Neu5Ac@O1A	S35@OG	2.67	162.72	41.27
		Neu5Ac@O1B	S35@N	2.84	160.16	34.79
		Neu5Ac@O1B	S35@OG	2.69	161.51	27.7
		Neu5Ac@O1A	S35@N	2.84	159.52	16.14
		9-O-Acetyl@O1A	T65@N	2.88	156.40	12.24
		Neu5Ac@O8	K59@NZ	2.86	155.50	11.05
		Y33@O	Neu5Ac@N5	2.85	155.31	42.2
		D36@OD2	Gal@O6	2.70	163.67	17.87
		D36@OD1	Gal@O6	2.70	163.44	14.74
	Run3	Neu5Ac@O1B	S35@OG	2.63	161.73	51.74
		Neu5Ac@O1A	S35@OG	2.63	161.38	44.04
		Neu5Ac@O8	S35@N	2.89	161.17	40.52
		Neu5Ac@O1B	R100@NH2	2.82	160.00	33.5
		Neu5Ac@O1A	R100@NH2	2.82	159.39	28.96
		Neu5Ac@O1A	S35@N	2.83	161.72	15.51
		Y33@O	Neu5Ac@O8	2.78	156.08	57.6
		Y33@O	Neu5Ac@N5	2.85	157.21	19.18
		Y33@OH	Neu5Ac@N5	2.91	159.88	17.2

		D36@OD1	Gal@O6	2.74	158.02	14.4
		D36@OD2	Gal@O6	2.73	158.73	13.29
		S35@O	GlcNAc@O6	2.76	161.83	12.45
Com1	**Run1**	-	-	-	-	-
Chian-B	**Run2**	Neu5Ac@O1B	S35@OG	2.66	163.16	50.15
		Neu5Ac@O1A	S35@OG	2.67	162.92	46.44
		Neu5Ac@O1A	S35@N	2.83	160.87	36.71
		Neu5Ac@O1B	S35@N	2.83	161.10	32.82
		9-O-Acetyl@O1A	T65@N	2.88	156.71	17.43
		Neu5Ac@O8	K59@NZ	2.86	155.78	15.8
		Neu5Ac@O8	K59@NZ	2.86	155.87	15.7
		Neu5Ac@O8	K59@NZ	2.86	155.48	14.21
		Y33@O	Neu5Ac@N5	2.86	155.08	56.48
		D36@OD2	Gal@O6	2.70	163.39	24.94
		D36@OD1	Gal@O6	2.71	163.04	22.84
	Run3	-	-	-	-	-
Com1	**Run1**	Neu5Ac@O1A	S35@OG	2.66	163.69	34.55
Chain-C		Neu5Ac@O1B	S35@N	2.83	161.14	31.95
		Y33@O	Neu5Ac@N5	2.86	154.92	25.89
		Y33@O	Gal@O2	2.75	155.51	13.85
		D36@OD2	Gal@O6	2.70	163.77	12.76
	Run2	Neu5Ac@O1A	S35@N	2.83	161.67	14.08
		Neu5Ac@O7	N106@ND2	2.85	163.81	11.97
		Neu5Ac@O1B	S35@OG	2.65	162.88	11.44
		Y33@O	Neu5Ac@N5	2.86	157.77	15.81
		K105@O	Gal@O2	2.74	158.96	15.58
	Run3	-	-	-	-	-
Com1	**Run1**	-	-	-	-	-
Chain-D	**Run2**	Neu5Ac@O1B	S35@OG	2.67	162.39	63.31
		Neu5Ac@O1A	S35@N	2.84	161.49	47.89
		Neu5Ac@O1A	S35@OG	2.69	160.02	34.96
		Neu5Ac@O1B	S35@N	2.84	161.50	17.92
		9-O-Acetyl@O1A	T65@N	2.88	156.25	12.98
		Y23@O	Neu5Ac@N5	2.85	157.28	56.92
		Y23@O	Neu5Ac@O8	2.79	156.38	10.51
		D36@OD2	3LB_588@O6	2.71	163.50	10.03
	Run3	-	-	-	-	-
Com1	**Run1**	Neu5Ac@O1A	S35@OG	2.68	162.06	54.39
Chain-E		Neu5Ac@O1B	S35@N	2.83	160.30	50.14

		Neu5Ac@O1B	S35@OG	2.69	161.49	48.67
		Neu5Ac@O1A	S35@N	2.83	160.44	34.65
		9-O-ACETYL@O1A	T65@N	2.88	154.96	23.96
		Neu5Ac@O8	K59@NZ	2.86	155.24	20.03
		Neu5Ac@O8	K59@NZ	2.86	155.20	19.98
		Neu5Ac@O8	K59@NZ	2.86	155.14	19.55
		Y33@O	Neu5Ac@N5	2.86	155.19	70.64
		D36@OD2	Gal@O6	2.71	162.66	21.07
		D36@OD1	Gal@O6	2.71	162.87	19.71
		S63@O	4YB_582@O6	2.71	161.42	13.75
	Run2	Neu5Ac@O1A	S35@OG	2.68	161.54	26.38
		Neu5Ac@O1B	S35@OG	2.69	160.98	21.31
		Neu5Ac@O1B	S35@N	2.85	160.43	19.7
		Neu5Ac@O1A	S35@N	2.84	160.33	15.12
		Y33@O	Neu5Ac@N5	2.85	158.14	26.45
	Run3	Neu5Ac@O1B	S35@OG	2.70	160.74	41.96
		Neu5Ac@O1A	S35@OG	2.70	160.87	38.01
		Neu5Ac@O1A	S35@N	2.84	161.47	33.31
		Neu5Ac@O1B	S35@N	2.84	161.10	29.63
		Y33@O	Neu5Ac@N5	2.85	158.29	59.34
Com2	**Run1**	-	-	-	-	-
Chain-A	**Run2**	Neu5Ac@O1A	S35@OG	2.68	162.20	64.84
		Neu5Ac@O1B	S35@N	2.83	162.21	63.84
		Neu5Ac@O1A	S35@N	2.84	160.22	23.56
		Neu5Ac@O1B	S35@OG	2.68	161.54	23.14
		9-O-Acetyl@O1A	T65@N	2.87	157.01	14.98
		Neu5Ac@O8	K59@NZ	2.86	155.22	11.23
		Neu5Ac@O8	K59@NZ	2.86	154.31	10.92
		Neu5Ac@O8	K59@NZ	2.86	154.61	10.18
		Y33@O	Neu5Ac@N5	2.86	158.02	58.81
		D36@OD1	6LB_593@O3	2.66	163.67	27.87
		D36@OD2	6LB_593@O3	2.66	164.06	27.6
		D36@OD1	6LB_593@O2	2.70	159.13	13.05
		D36@OD2	6LB_593@O2	2.70	158.87	11.65
		D36@OD2	6LB_593@O4	2.69	161.95	10.76
		D36@OD1	6LB_593@O4	2.68	162.13	10.35
	Run3	Neu5Ac@O4	S118@OG	2.74	163.15	45.19
		Neu5Ac@O5N	V68@N	2.87	161.37	33.75

		Neu5Ac@O1A	S35@OG	2.67	162.23	20.99
		Neu5Ac@O1B	S35@N	2.84	161.58	17.94
		Neu5Ac@O1B	S35@OG	2.68	161.48	13.8
		Neu5Ac@O1A	S35@N	2.84	161.83	11.4
		9-O-Acetyl@O1A	T65@N	2.87	157.71	11.17
		P68@O	Neu5Ac@O7	2.75	162.75	46.81
		Y33@O	Neu5Ac@N5	2.86	156.85	24.78
		G135@O	4YB_592@O6	2.76	158.25	16.72
Com2 **Chain-B**	**Run1**	Neu5Ac@O1A	S35@OG	2.67	162.45	47.66
		Neu5Ac@O1B	S35@OG	2.69	162.95	35.72
		Neu5Ac@O1B	R100@NE	2.84	152.32	31.78
		Neu5Ac@O1B	S35@N	2.85	161.12	31.66
		Neu5Ac@O1B	R100@NH2	2.83	149.74	29.54
		Neu5Ac@O1A	S35@N	2.87	161.23	29.01
		Neu5Ac@O1A	R100@NH2	2.81	154.69	28.2
		Neu5Ac@O1A	R100@NE	2.86	151.75	17.49
		K23@O	Gal@O3	2.67	165.25	33.2
		K23@O	Gal@O4	2.69	158.84	23.01
		Y33@O	Neu5Ac@N5	2.85	155.94	19.8
		K23@OXT	Gal@O4	2.71	158.89	19.3
		T131@O	GlcNAc@N2	2.86	156.42	17.48
		K23@OXT	Gal@O3	2.67	164.94	16.14
	Run2	Neu5Ac@O1A	S35@OG	2.64	163.29	55.09
		Neu5Ac@O1B	S35@N	2.83	162.31	47
		Neu5Ac@O1B	S35@OG	2.67	163.09	46.28
		Neu5Ac@O1A	S35@N	2.84	162.48	39.56
		9-O-Acetyl@O1A	T65@N	2.87	158.48	31.93
		Neu5Ac@O8	K59@NZ	2.87	157.46	16.1
		Neu5Ac@O8	K59@NZ	2.87	157.29	14.89
		Neu5Ac@O8	K59@NZ	2.87	156.93	14.68
		Y33@O	Neu5Ac@N5	2.85	156.90	73.6
		A130@O	ROH@O1	2.75	159.30	35.77
	Run3	Neu5Ac@O1A	S35@OG	2.62	162.36	18.84
		Neu5Ac@O1B	S35@OG	2.62	162.21	16.59
		Neu5Ac@O1B	R100@NH2	2.81	159.66	13.98
		Neu5Ac@O1A	R100@NH2	2.81	160.80	13.3
		Neu5Ac@O8	S35@N	2.90	158.66	10.85
		Y33@O	Neu5Ac@O8	2.80	156.27	22.84
		D36@OD2	Gal@O4	2.64	163.70	19.01

		D36@OD1	Gal@O4	2.63	163.22	13.91
		Y33@OH	Neu5Ac@N5	2.90	159.78	11.86
Com2 **Chain-C**	**Run1**	Neu5Ac@O1A	S35@OG	2.69	161.95	16.26
		Neu5Ac@O1B	S35@OG	2.69	161.83	13.19
		Neu5Ac@O1B	S35@N	2.85	160.61	11.72
		Y33@O	Neu5Ac@N5	2.84	158.04	16.89
	Run2	-	-	-	-	-
	Run3	Neu5Ac@O1A	S35@N	2.84	161.64	45.42
		Neu5Ac@O1A	S35@OG	2.66	162.85	44.67
		Neu5Ac@O1B	S35@OG	2.68	161.32	42.8
		Neu5Ac@O1B	S35@N	2.85	161.70	33.52
		Y23@O	Neu5Ac@N5	2.86	157.02	16.15
Com2 **Chain-D**	**Run1**	-	-	-	-	-
	Run2	Neu5Ac@O1B	S35@OG	2.66	163.43	35.04
		Neu5Ac@O1A	S35@N	2.86	163.02	19.64
		6LB_583@O3	S35@OG	2.75	163.20	15.86
		6LB_583@O3	Y23@OH	2.79	160.86	12.62
		Neu5Ac@O7	Y23@OH	2.76	164.80	11.93
	Run3	Neu5Ac@O1A	S35@OG	2.67	162.55	16.74
		Neu5Ac@O1B	S35@N	2.84	162.80	15.23
		Y23@O	Neu5Ac@N5	2.86	157.02	16.15
Com2 **Chain-E**	**Run1**	Neu5Ac@O1B	S35@OG	2.67	162.10	29.13
		Neu5Ac@O1A	S35@N	2.84	162.08	25.17
		Neu5Ac@O1A	S35@OG	2.68	161.82	16.29
		Neu5Ac@O1B	S35@N	2.85	159.89	15.69
		Y33@O	Neu5Ac@N5	2.86	157.82	33.74
	Run2	Neu5Ac@O1B	S35@N	2.84	162.59	26.71
		Neu5Ac@O1A	S35@N	2.84	161.69	11.42
		Neu5Ac@O1A	S35@OG	2.67	161.95	10.09
		E84@OE1	Gal@O3	2.68	165.28	21.39
		E84@OE1	Gal@O4	2.65	163.37	20.31
		Y33@O	Neu5Ac@O4	2.74	159.51	17.78
		P80@O	Gal@O2	2.75	161.01	15.21
		D36@OD1	Gal@O3	2.67	163.43	13.02
		Y33@O	Neu5Ac@N5	2.87	158.09	12.21
		D36@OD2	Gal@O3	2.67	163.45	11.17
	Run3	Neu5Ac@O1B	S35@OG	2.68	162.01	42.24
		Neu5Ac@O1A	S35@OG	2.68	161.62	41
		Neu5Ac@O1A	S35@N	2.83	161.32	39.25

		Neu5Ac@O1B	S35@N	2.83	160.75	37.18
		9-O-Acetyl@O1A	T65@N	2.87	157.46	21.57
		Neu5Ac@O8	K59@NZ	2.86	158.80	13.91
		Neu5Ac@O8	K59@NZ	2.86	158.33	11.58
		Neu5Ac@O8	K59@NZ	2.86	158.29	10.86
		Y33@O	Neu5Ac@N5	2.85	156.34	48.43
		D36@OD1	Gal@O2	2.66	165.66	20.3
		N61@O	Gal@O3	2.75	157.38	19.85
		T131@O	ROH@O1	2.73	158.12	14.74
		S63@O	Gal@O4	2.78	158.03	14.56
		D36@OD1	Gal@O3	2.67	163.10	12.83
		Y33@O	Neu5Ac@O4	2.74	159.83	12.05
		D36@OD2	Gal@O3	2.67	163.54	12.01
		D36@OD2	GlcNAc@O6	2.68	164.92	11.98
Com3 **Chain-A**	**Run1**	Neu5Ac@O1A	S35@OG	2.66	163.32	83.97
		Neu5Ac@O1B	S35@N	2.83	162.15	77.42
		4-O-Acetyl@O1A	Y33@N	2.86	160.36	69.29
		Neu5Ac@O1B	S35@OG	2.71	161.40	16.44
		Neu5Ac@O8	K59@NZ	2.86	157.81	16.34
		Neu5Ac@O8	K59@NZ	2.86	157.90	16.33
		Neu5Ac@O8	K59@NZ	2.86	158.02	15.99
		Neu5Ac@O1A	R100@NH2	2.80	155.73	14.55
		Y33@O	Neu5Ac@N5	2.86	150.30	55.84
		D36@OD1	Gal@O6	2.69	163.70	30.08
		D36@OD2	Gal@O6	2.70	163.57	22.03
	Run2	4-O-Acetyl@O1A	Y33@N	2.86	160.22	52.02
		Neu5Ac@O1B	S35@OG	2.68	162.37	37.43
		Neu5Ac@O1B	R100@NH2	2.80	151.94	34.86
		Neu5Ac@O1A	S35@OG	2.68	163.52	33.65
		Neu5Ac@O1A	S35@N	2.85	161.08	31.12
		Neu5Ac@O1B	R100@NE	2.84	150.33	24.58
		Gal@O6	CYX_110@N	2.89	159.20	14.94
		Neu5Ac@O1B	S35@N	2.84	161.40	11.59
		Y33@O	Neu5Ac@N5	2.86	150.88	27.51
	Run3	Neu5Ac@O1B	S35@OG	2.66	163.98	35.06
		Neu5Ac@O1A	S35@N	2.84	162.44	31.93
		4-O-Acetyl@O1A	Y33@N	2.86	160.26	22.6
		Neu5Ac@O8	K59@NZ	2.85	158.00	10.95
		Y33@O	Neu5Ac@N5	2.86	152.73	25.87

		D36@OD1	Gal@O6	2.69	163.81	11.04
Com3 **Chain-B**	**Run1**	Neu5Ac@O1B	R100@NH2	2.78	153.62	81.06
		4-O-ACETYL@O1A	Y33@N	2.86	160.12	65.51
		Neu5Ac@O1A	S35@OG	2.68	163.67	62.24
		Neu5Ac@O1B	R100@NE	2.86	148.47	45.3
		Gal@O6	CYX_224@N	2.89	161.01	36.66
		Neu5Ac@O1B	S35@OG	2.69	159.78	34.18
		Neu5Ac@O1A	S35@N	2.86	157.31	23.33
		Neu5Ac@O8	S35@N	2.88	155.31	13.84
		Y33@O	Neu5Ac@N5	2.85	152.59	36.86
		S35@O	Neu5Ac@O8	2.80	155.89	11.83
	Run2	4-O-acetyl@O1A	Y33@N	2.86	160.14	68.65
		Neu5Ac@O1B	S35@OG	2.66	163.46	60.67
		Neu5Ac@O1A	S35@N	2.84	161.82	45.88
		Neu5Ac@O1A	S35@OG	2.67	163.36	24.83
		Neu5Ac@O8	K59@NZ	2.86	158.27	15.99
		Neu5Ac@O8	K59@NZ	2.86	158.19	14.14
		Neu5Ac@O1A	R100@NH1	2.79	156.35	13.88
		Neu5Ac@O8	K59@NZ	2.86	158.08	13.39
		Neu5Ac@O1B	R100@NH1	2.80	155.85	13.36
		Neu5Ac@O1B	R100@NH2	2.78	156.54	11.83
		Neu5Ac@O1A	R100@NH2	2.82	156.55	11.29
		Neu5Ac@O1B	S35@N	2.85	161.98	10.77
		Y33@O	Neu5Ac@N5	2.85	152.21	62.04
		D36@OD1	Gal@O6	2.69	164.01	16.63
		D36@OD2	Gal@O6	2.70	163.53	12.76
	Run3	Neu5Ac@O1A	R100@NH2	2.77	155.58	90.73
		Neu5Ac@O1B	S35@OG	2.66	164.22	86.2
		4-O-ACETYL@O1A	Y33@N	2.86	159.75	66.34
		Neu5Ac@O1A	R100@NE	2.86	148.53	39.47
		Gal@O6	CYX_224@N	2.89	160.39	30.13
		Neu5Ac@O1B	S35@N	2.85	149.64	21.1
		Neu5Ac@O8	S35@N	2.88	155.41	16.03
		Y33@O	Neu5Ac@N5	2.84	151.75	53.94
		K23@OXT	Gal@O4	2.70	162.03	22.49
		K23@O	Gal@O4	2.69	162.47	21.4
		S35@O	Neu5Ac@O8	2.81	155.69	14.45

Com3	**Run1**	-	-	-	-	-
Chain-C	**Run2**	Neu5Ac@O1A	S35@OG	2.66	163.91	76.9
		Neu5Ac@O1B	S35@N	2.83	162.53	70.43
		9-O-ACETYL@O1A	Y33@N	2.86	161.26	69.15
		Neu5Ac@O1B	S35@OG	2.68	162.45	21.23
		Neu5Ac@O8	K59@NZ	2.86	158.04	19.24
		Neu5Ac@O8	K59@NZ	2.86	157.71	19.05
		Neu5Ac@O8	K59@NZ	2.86	157.72	18.24
		Neu5Ac@O1A	S35@N	2.85	162.79	12.27
		Y33@O	Neu5Ac@N5	2.86	150.83	63.6
		D36@OD2	Gal@O6	2.69	163.99	22.05
		D36@OD1	Gal@O6	2.69	163.78	17.56
	Run3	Neu5Ac@O1B	S35@OG	2.66	163.32	76.75
		ACX_585@O1A	Y33@N	2.86	162.17	67.36
		Neu5Ac@O1A	S35@N	2.84	162.18	67.01
		Neu5Ac@O1A	S35@OG	2.72	160.99	24.33
		Y33@O	Neu5Ac@N5	2.85	152.28	65.69
		D36@OD1	Gal@O6	2.68	164.55	25.19
		D36@OD2	Gal@O6	2.68	164.63	21.35
Com3	**Run1**	4-O-Acetyl@O1A	Y23@N	2.86	161.43	67.57
Chain-D		Neu5Ac@O1A	S35@OG	2.67	162.96	51.64
		Neu5Ac@O1B	S35@OG	2.69	159.53	43.57
		Neu5Ac@O1B	R77@NH2	2.80	152.48	42.31
		Neu5Ac@O1B	S35@N	2.83	161.83	36.41
		Neu5Ac@O1B	R77@NE	2.86	149.48	27.1
		Neu5Ac@O1A	S35@N	2.88	159.59	18.4
		Gal@O6	C133@N	2.89	159.27	14.49
		Neu5Ac@O1A	R77@NH2	2.80	157.81	13.23
		Neu5Ac@O8	K59@NZ	2.86	159.56	10.03
		Y23@O	Neu5Ac@N5	2.86	150.44	25.4
		K23@OXT	Gal@O4	2.66	160.76	15.95
		K23@O	Gal@O4	2.66	160.93	15.69
		D36@OD2	Gal@O6	2.70	163.19	11.48
		D36@OD1	Gal@O6	2.70	163.28	11.23
	Run2	Neu5Ac@O1B	S35@OG	2.67	164.00	48.08
		4-O-Acetyl@O1A	Y23@N	2.86	160.39	46.2
		Neu5Ac@O1A	S35@N	2.83	163.12	42.69
		Neu5Ac@O1A	S35@OG	2.69	161.92	19.34

		Neu5Ac@O1B	S35@N	2.83	161.42	16.03
		Neu5Ac@O8	K59@NZ	2.86	158.50	15.11
		Neu5Ac@O8	K59@NZ	2.86	158.81	14.69
		Neu5Ac@O8	K59@NZ	2.86	158.47	13.44
		Y23@O	Neu5Ac@N5	2.86	150.92	41.6
		D36@OD2	Gal@O6	2.69	164.00	18.61
		D36@OD1	Gal@O6	2.70	163.93	13.98
	Run3	Neu5Ac@O1B	R77@NH2	2.80	153.31	77.44
		4-O-Acetyl@O1A	Y23@N	2.86	161.70	66.85
		Neu5Ac@O1B	S35@OG	2.70	158.56	52.4
		Neu5Ac@O1B	R77@NE	2.86	150.10	48.05
		Gal@O6	C133@N	2.89	160.23	38.29
		Neu5Ac@O1A	S35@N	2.88	158.21	34.26
		Neu5Ac@O1A	S35@OG	2.67	164.34	34.18
		Y23@O	Neu5Ac@N5	2.84	151.16	22.42
		K23@O	Gal@O4	2.68	160.96	13.36
		K23@OXT	Gal@O4	2.69	160.94	12.05
Com3 **Chain-E**	**Run1**	Neu5Ac@O1B	R100@NH2	2.78	153.86	63.8
		9-O-Acetyl@O1A	Y33@N	2.86	160.11	57.26
		Neu5Ac@O1A	S35@OG	2.67	163.75	48.8
		Neu5Ac@O1B	R100@NE	2.86	149.41	33.13
		3LB_588@O6	C133@N	2.89	159.62	32.57
		Neu5Ac@O1B	S35@OG	2.72	159.93	28.01
		Neu5Ac@O1A	S35@N	2.87	157.14	22.76
		Neu5Ac@O1B	S35@N	2.84	160.28	12.48
		Neu5Ac@O1A	R100@NH2	2.81	153.29	10.62
		Neu5Ac@O1A	R100@NE	2.83	153.47	10.5
		Y33@O	Neu5Ac@N5	2.84	150.18	27.49
	Run2	9-O-Acetyl@O1A	Y33@N	2.86	159.91	68.18
		Neu5Ac@O1A	R100@NH2	2.79	156.02	51.4
		Neu5Ac@O1B	S35@OG	2.68	162.38	51.29
		Neu5Ac@O1A	S35@OG	2.70	159.90	47.06
		Neu5Ac@O1B	S35@N	2.85	157.86	35.84
		3LB_588@O6	C133@N	2.89	159.91	23.63
		Neu5Ac@O1A	R100@NE	2.88	148.92	18.97
		Neu5Ac@O1A	S35@N	2.81	160.00	17.95
		Neu5Ac@O1B	R100@NH2	2.79	162.93	11.98
		Y33@O	Neu5Ac@N5	2.85	150.13	47.71
	Run3	9-O-Acetyl@O1A	Y33@N	2.86	160.04	65.5

		Neu5Ac@O1B	R100@NH2	2.80	153.83	62.5
		Neu5Ac@O1B	S35@OG	2.70	158.40	48.77
		Neu5Ac@O1A	S35@OG	2.69	162.48	44.89
		Neu5Ac@O1B	R100@NE	2.87	149.80	35.35
		3LB_588@O6	C133@N	2.89	159.95	32.14
		Neu5Ac@O1A	S35@N	2.87	156.51	24.33
		Neu5Ac@O1B	S35@N	2.83	159.16	14.37
		Neu5Ac@O1A	R100@NH2	2.79	155.30	13.82
		Y33@O	Neu5Ac@N5	2.84	150.20	28.49
Com4 **Chain-A**	**Run1**	ROH_575@O1	V68@N	2.91	160.25	19.51
		P68@O	4YB_576@N2	2.87	151.13	20.25
		F117@O	ROH_575@O1	2.76	154.26	12.73
	Run2	Neu5Ac@O1A	S35@OG	2.68	162.53	43.38
		Neu5Ac@O1B	S35@N	2.83	160.71	37.59
		Neu5Ac@O1B	S35@OG	2.69	162.15	23.33
		Neu5Ac@O1A	S35@N	2.84	160.71	21.57
		Neu5Ac@O8	K59@NZ	2.85	158.82	15.17
		Neu5Ac@O8	K59@NZ	2.85	158.97	14.97
		Neu5Ac@O8	K59@NZ	2.85	158.63	13.67
		Y33@O	Neu5Ac@N5	2.85	154.78	48.96
		D36@OD1	Gal@O6	2.69	163.59	14.39
	Run3	Neu5Ac@O1A	R100@NH2	2.78	161.88	15.81
		Neu5Ac@O8	Y33@OH	2.80	163.06	12.42
		Neu5Ac@O1B	R100@NH1	2.85	159.96	10.06
Com4 **Chain-B**	**Run1**	Neu5Ac@O1A	S35@N	2.84	162.80	12.79
		Neu5Ac@O1B	S35@OG	2.66	163.46	12.69
		SER_220@O	GlcNAc@O6	2.72	161.15	61.89
		D36@OD1	Gal@O2	2.71	161.11	40.1
		D36@OD2	Gal@O2	2.72	160.01	26.3
		Y33@O	Neu5Ac@N5	2.86	156.42	13.25
		A60@O	GlcNAc@N2	2.89	158.33	12.63
		T131@O	Neu5Ac@O9	2.76	159.76	10.72
	Run2	Neu5Ac@O1B	S35@OG	2.66	163.29	20.15
		Neu5Ac@O1A	S35@N	2.84	161.79	18.69
		Neu5Ac@O1B	S35@N	2.85	161.03	14.46
		Neu5Ac@O1A	S35@OG	2.67	162.77	13.03
		D36@OD1	Gal@O2	2.68	164.79	61.12
		Y33@O	Neu5Ac@O4	2.75	156.28	38.87

		A130@O	ROH@O1	2.75	159.78	28.19
		Y33@O	Neu5Ac@N5	2.85	156.83	22.9
	Run3	Neu5Ac@O1B	S35@N	2.86	162.47	20.89
		Neu5Ac@O1A	S35@N	2.85	162.64	14.82
		Y33@O	Gal@O2	2.71	158.98	23.49
		K23@O	GlcNAc@O6	2.68	163.43	14.62
		K23@O	Neu5Ac@O9	2.69	163.72	13.84
		K23@O	Neu5Ac@O8	2.71	164.31	12.29
Com4 **Chain-C**	**Run1**	Neu5Ac@O1B	S35@OG	2.66	162.69	30.44
		Neu5Ac@O1A	S35@N	2.85	160.45	27.26
		Neu5Ac@O5N	S35@N	2.86	159.56	17.78
		Neu5Ac@O1A	S35@OG	2.71	161.59	17.16
		Gal@O2	Y33@N	2.90	156.54	15.52
		Gal@H2O	Y33@N	2.85	142.84	12.26
		Y33@O	Gal@O2	2.77	152.22	30.97
		Y33@O	Neu5Ac@O7	2.79	158.71	21.26
		Y33@O	Neu5Ac@N5	2.85	157.15	21.08
		N29@O	GlcNAc@O3	2.73	159.43	15.23
		G27@O	ROH@O1	2.74	161.91	12.01
		D28@O	GlcNAc@O6	2.78	159.98	11.86
		T30@O	GlcNAc@N2	2.89	156.69	10.79
	Run2	Neu5Ac@O1A	S35@OG	2.68	161.81	62.49
		Neu5Ac@O1B	S35@N	2.84	160.00	47.04
		Neu5Ac@O1B	S35@OG	2.71	159.95	37.3
		Neu5Ac@O1A	S35@N	2.83	157.90	18.43
		Neu5Ac@O8	K59@NZ	2.86	158.55	14.01
		Neu5Ac@O8	K59@NZ	2.86	158.60	12.26
		Neu5Ac@O8	K59@NZ	2.86	158.53	12.23
		Y33@O	Neu5Ac@N5	2.85	158.01	69.54
		D36@OD2	Gal@O6	2.69	163.67	16.27
		N61@O	Gal@O6	2.76	157.00	15.59
		S23@O	Neu5Ac@O9	2.75	158.93	14.47
		D36@OD1	Gal@O6	2.69	163.38	12.01
	Run3	Neu5Ac@O1A	S35@OG	2.69	161.91	54.34
		Neu5Ac@O1B	S35@OG	2.68	161.55	43.38
		Neu5Ac@O1A	S35@N	2.84	157.74	27.18
		Neu5Ac@O1B	S35@N	2.83	159.17	26.64
		Neu5Ac@O8	K59@NZ	2.86	158.60	12.18
		Neu5Ac@O8	K59@NZ	2.86	158.64	12.08

		Neu5Ac@O8	K59@NZ	2.86	158.38	12.05
		Y33@O	Neu5Ac@N5	2.84	156.88	64.58
		N61@O	Gal@O6	2.76	158.19	21.31
		D36@OD1	Gal@O6	2.69	163.72	18.98
		S23@O	Neu5Ac@O9	2.75	158.92	12.62
		D36@OD2	Gal@O6	2.69	163.65	12.6
Com4 **Chain-D**	**Run1**	Neu5Ac@O1A	S35@OG	2.67	163.12	19.58
		Neu5Ac@O1B	S35@N	2.84	160.68	17.53
		Y23@O	Neu5Ac@N5	2.85	156.01	22.05
		T109@O	ROH@O1	2.73	161.44	19.92
		W108@O	GlcNac@O6	2.77	160.71	16.7
		N106@OD1	Gal@O2	2.72	159.16	12.16
	Run2	-	-	-	-	-
	Run3	Neu5Ac@O1A	S35@OG	2.67	163.27	36.99
		Neu5Ac@O1B	S35@N	2.84	161.07	32.98
		Neu5Ac@O1B	S35@OG	2.68	162.26	16.8
		Neu5Ac@O8	K59@NZ	2.86	158.50	11.88
		Neu5Ac@O8	K59@NZ	2.85	158.63	11.63
		Neu5Ac@O8	K59@NZ	2.86	158.57	11.16
		Y23@O	Neu5Ac@N5	2.85	155.81	39.93
		D36@OD1	Gal@O6	2.69	163.52	13.73
		D36@OD2	Gal@O6	2.69	163.54	13.18
Com4 **Chain-E**	**Run1**	Neu5Ac@O5N	T65@N	2.87	158.40	10.16
		S63@O	Gal@O2	2.77	153.94	15.85
		D36@OD1	ROH@O1	2.68	165.17	14.07
		D36@O	GlcNAc@O6	2.77	160.28	13.47
		S63@O	Neu5Ac@O7	2.76	157.03	12.18
	Run2	GlcNAc@O3	C133@N	2.89	161.74	22.18
		Neu5Ac@O1B	S35@N	2.85	158.12	20.29
		Neu5Ac@O1B	S35@OG	2.71	161.70	18.61
		Neu5Ac@O1A	S35@OG	2.70	161.69	18.5
		Neu5Ac@O1A	S35@N	2.86	159.70	13.33
		K23@OXT	Gal@O6	2.70	163.87	35.29
		Y33@O	Neu5Ac@N5	2.86	157.20	26.28
		N61@O	Gal@O6	2.74	158.10	22.1
		S63@O	Neu5Ac@O9	2.76	157.85	17.32
		T131@O	GlcNAc@N2	2.89	160.62	13.59
		K23@O	Neu5Ac@O9	2.72	163.56	12.09
	Run3	Neu5Ac@O1B	S35@OG	2.67	162.34	39.33

		Neu5Ac@O1A	S35@N	2.84	159.76	29.62
		Neu5Ac@O1A	S35@OG	2.68	162.24	25.67
		Neu5Ac@O1B	S35@N	2.83	159.41	10.54
		Y33@O	Neu5Ac@N5	2.85	156.07	46.28
		D36@OD2	Gal@O6	2.69	163.66	15.59
		D36@OD1	Gal@O6	2.69	163.48	11.53
Com5 **Chain-A**	**Run1**	T116@O	ROH@O1	2.73	161.04	23.9
		F117@O	GlcNAc@O6	2.71	159.46	13.95
	Run2	Neu5Ac@O1A	S35@OG	2.68	161.62	35.88
		Neu5Ac@O1B	S35@N	2.84	161.54	25.19
		Neu5Ac@O1B	S35@OG	2.69	160.48	22.1
		Neu5Ac@O1A	Y34@OH	2.71	162.21	11.7
		Gal@O3	Y33@N	2.90	161.59	11.69
		Y33@O	Gal@O4	2.73	158.09	26.93
		Y33@O	Neu5Ac@N5	2.87	154.85	24.87
		N61@O	GlcNAc@N2	2.87	161.19	17.78
		E37@OE1	Neu5Ac@O9	2.68	162.31	12.04
		T65@O	Neu5Ac@O8	2.76	157.19	12.02
	Run3	-	-	-	-	-
Com5 **Chain-B**	**Run1**	Neu5Ac@O1B	Y34@OH	2.70	160.33	23.84
		Neu5Ac@O1A	Y34@OH	2.70	160.30	21.82
		Gal@O2	S35@OG	2.82	163.63	11.48
		Y33@O	Gal@O2	2.70	162.43	66.42
		K23@OXT	ROH@O1	2.68	160.88	29.72
		K23@O	ROH@O1	2.69	160.72	19.09
	Run2	-	-	-	-	-
	Run3	Neu5Ac@O1B	S35@N	2.85	162.29	24.56
		Neu5Ac@O1A	S35@OG	2.70	160.54	13.43
		Neu5Ac@O1A	S35@N	2.86	161.13	12.8
		Neu5Ac@O4	Y33@N	2.87	156.13	11.54
		Y33@O	Neu5Ac@N5	2.86	153.91	13.04
		N61@O	GlcNAc@N2	2.87	161.32	10.81
Com5 **Chain-C**	**Run1**	-	-	-	-	-
	Run2	Neu5Ac@O1A	S35@OG	2.69	161.83	71.23
		Neu5Ac@O1A	S35@N	2.84	157.03	23.45
		Neu5Ac@O1B	S35@OG	2.70	160.66	22.68
		Neu5Ac@O1B	S35@N	2.84	160.36	17.59
		GlcNAc@O2N	T65@OG1	2.74	160.26	10.66
		Y33@O	Neu5Ac@N5	2.86	159.66	60.68

		E37@OE1	Neu5Ac@O8	2.72	166.22	40.86
		E37@OE1	Neu5Ac@O9	2.64	159.81	40.79
		D36@OD2	Gal@O4	2.68	163.85	25.88
		D36@OD1	Gal@O4	2.68	163.57	20.12
	Run3	Neu5Ac@O1B	S35@OG	2.67	162.19	60.99
		Neu5Ac@O1A	S35@N	2.84	161.46	46.06
		Neu5Ac@O1A	S35@OG	2.69	160.92	27.2
		Neu5Ac@O1B	S35@N	2.85	159.29	18.94
		Neu5Ac@O8	K59@NZ	2.86	156.37	10.58
		Y33@O	Neu5Ac@N5	2.84	156.61	63.52
		D36@OD2	Gal@O2	2.66	165.19	14.34
		D36@OD2	Gal@O3	2.66	163.73	11.15
		D36@OD1	Gal@O3	2.66	163.64	10.66
Com5 **Chain-D**	**Run1**	-	-	-	-	-
	Run2	Neu5Ac@O1A	S35@N	2.84	162.86	22.11
		Neu5Ac@O1B	S35@N	2.84	161.75	13.51
		Neu5Ac@O1B	S35@OG	2.67	163.01	13.08
		Y23@O	Neu5Ac@O4	2.73	161.18	25.55
		K23@O	ROH_575@O 1	2.65	165.65	11.22
	Run3	-	-	-	-	-
Com5 **Chain-E**	**Run1**	-	-	-	-	-
	Run2	-	-	-	-	-
	Run3	S76@O	GlcNac@O6	2.73	160.74	11.71

Chapter 5

Conclusions and scope for future work

5.1 Conclusions

In this study, we have elucidated the conformational propensity of the free tri-saccharides of *N*-glycan with and without modifications (4-O and 9-O) using multiple replica molecular dynamics (MRMD). Conformation of the similar glycans were also estimated in complexed with the typhoid toxin (PltB) .Conformational dynamics of glycans were characterized by estimating the glycosdic torsional angles, distortion of ring geometry. Overall flexibility of tri-saccharides was independent of the modification on sialic acids. The nature of similar linkages (i.e. 1-4, 2-3 and 2-6) were remain same in all five glycans (free and complex state) used in our study. However, in terms of ring conformations, each carbohydrates samplex more number of ring conformations compared to the complex state. Allthough in few cases, free and complexed carbohydrates shared the same path in sampling between both the major chair conformations. Apart from the $4C_1$ and $1C_4$ chair forms, several non-chair conformations like $^{5}S_1$, $^{2,5}B$ and $^{2}S_0$ were observed in both modified and unmodified sialic acids, which diminishes after complex formation.

Multiple replica simulations of apo and glycan PltB illuminate the conformational changes as observed from RMSD distribution. Other key parameters, like radius of gyration (RoG), solvent accessible surface area (SASA) shows stable conformations of the whole pentamer as well as in the individual chains. This stability in the conformations support the estimation of binding free energy. The recognition between the typhoid toxin and *N*-glycans were estimated by the molecular mechanics generalized Born surface area (MM/GBSA) method. *Com3* (4-O-acetylated α2–3 Sialosides) among all five complexes shows the highest affinity . While the *Com5* (α2–6 Sialosides) has the lowest binding free

energy. The modified glycan-protein complex has higher binding free energy than the unmodified glycan-protein complex. The van der Waals contribution plays a key role in the recognition process. However, the higher contribution in the electrostatic contribution compensates by the increase contribution of the polar group. Also, several key hydrogen bonds were also found, which plays a significant contribution in the binding. Thus, we concluded that the modification on the sialic acid increases the affinity towards the binding sites of PltB protein.

Also, we found that there exists a dynamic equilibrium between different conformation in the protein. This was the reason behind the movement of the glycans from BS1 site to BS2 site. Also, in case of the modified glycans the interaction takes place by the help of sialic acid and the modification present on it, while in case of the unmodified glycans all the three monosaccharides take part in the interaction. We also analyses the hydrogen bonding between the glycans and binding sites for each complex and found that the hydrogen bonding is the main player behind the interaction.

So, our study provides a detailed understanding of the recognition process of these *N*-glycans by the PltB toxin, which may further help to understand the disease pathology of the century-old disease of typhoid.

5.2 Future Work

In our current study we have studied the glycans that were binded to the BS1 sites only. In future the binding mechanism at the BS2 and BS3 sites can also be studied. Also, the binding mechamism and conformation can be studied when the glycans are attached to the BS1, BS2 and BS3 sites simultaneously. As have considered three residues in *N*-glycans, so the long different *N*-glycans can be used to elucidate the recognition mechanism. Thus the effect of the whole multiantennary chain towards binding with the typhoid toxin can be studied in future.

References:

(1) Lee, S.; Yang, Y.-A.; Milano, S. K.; Nguyen, T.; Ahn, C.; Sim, J. H.; Thompson, A. J.; Hillpot, E. C.; Yoo, G.; Paulson, J. C.; Song, J. Salmonella Typhoid Toxin PltB Subunit and Its Nontyphoidal Salmonella Orthologue Confer Differential Host Adaptation and Virulence. *Cell Host Microbe* **2020**, *27* (6), 937-949.e6. https://doi.org/10.1016/j.chom.2020.04.005.

(2) Raffatellu, M.; Wilson, R. P.; Winter, S. E.; Bäumler, A. J. Clinical Pathogenesis of Typhoid Fever. *J. Infect. Dev. Ctries.* **2008**, *2* (4), 260–266. https://doi.org/10.3855/jidc.219.

(3) Parry, C. M.; Hien, T. T.; Dougan, G.; White, N. J.; Farrar, J. J. Typhoid Fever. *N. Engl. J. Med.* **2002**, *347* (22), 1770–1782. https://doi.org/10.1056/NEJMra020201.

(4) Stanaway, J. D.; Reiner, R. C.; Blacker, B. F.; Goldberg, E. M.; Khalil, I. A.; Troeger, C. E.; Andrews, J. R.; Bhutta, Z. A.; Crump, J. A.; Im, J.; Marks, F.; Mintz, E.; Park, S. E.; Zaidi, A. K. M.; Abebe, Z.; Abejie, A. N.; Adedeji, I. A.; Ali, B. A.; Amare, A. T.; Atalay, H. T.; Avokpaho, E. F. G. A.; Bacha, U.; Barac, A.; Bedi, N.; Berhane, A.; Browne, A. J.; Chirinos, J. L.; Chitheer, A.; Dolecek, C.; Zaki, M. E. S.; Eshrati, B.; Foreman, K. J.; Gemechu, A.; Gupta, R.; Hailu, G. B.; Henok, A.; Hibstu, D. T.; Hoang, C. L.; Ilesanmi, O. S.; Iyer, V. J.; Kahsay, A.; Kasaeian, A.; Kassa, T. D.; Khan, E. A.; Khang, Y.-H.; Razek, H. M. A. E.; Melku, M.; Mengistu, D. T.; Mohammad, K. A.; Mohammed, S.; Mokdad, A. H.; Nachega, J. B.; Naheed, A.; Nguyen, C. T.; Nguyen, H. L. T.; Nguyen, L. H.; Nguyen, N. B.; Nguyen, T. H.; Nirayo, Y. L.; Pangestu, T.; Patton, G. C.; Qorbani, M.; Rai, R. K.; Rana, S. M.; Ranabhat, C. L.; Roba, K. T.; Roberts, N. L. S.; Rubino, S.; Safiri, S.; Sartorius, B.; Sawhney, M.; Shiferaw, M. S.; Smith, D. L.; Sykes, B. L.; Tran, B. X.; Tran, T. T.; Ukwaja, K. N.; Vu, G. T.; Vu, L. G.; Weldegebreal, F.; Yenit, M. K.; Murray, C. J. L.; Hay, S. I. The Global Burden of Typhoid and Paratyphoid Fevers: A Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet Infect. Dis.* **2019**, *19* (4), 369–381. https://doi.org/10.1016/S1473-3099(18)30685-6.

(5) Mogasale, V.; Maskery, B.; Ochiai, R. L.; Lee, J. S.; Mogasale, V. V.; Ramani, E.; Kim, Y. E.; Park, J. K.; Wierzba, T. F. Burden of Typhoid Fever in Low-Income and Middle-Income Countries: A Systematic, Literature-Based Update with Risk-Factor Adjustment. *Lancet Glob. Health* **2014**, *2* (10), e570–e580. https://doi.org/10.1016/S2214-109X(14)70301-8.

(6) Buckle, G. C.; Walker, C. L. F.; Black, R. E. Typhoid Fever and Paratyphoid Fever: Systematic Review to Estimate Global Morbidity

and Mortality for 2010. *J. Glob. Health* **2012**, *2* (1), 010401. https://doi.org/10.7189/jogh.02.010401.

(7) Deng, L.; Song, J.; Gao, X.; Wang, J.; Yu, H.; Chen, X.; Varki, N.; Naito-Matsui, Y.; Galán, J. E.; Varki, A. Host Adaptation of a Bacterial Toxin from the Human Pathogen Salmonella Typhi. *Cell* **2014**, *159* (6), 1290–1299. https://doi.org/10.1016/j.cell.2014.10.057.

(8) Azmatullah, A.; Qamar, F. N.; Thaver, D.; Zaidi, A. K.; Bhutta, Z. A. Systematic Review of the Global Epidemiology, Clinical and Laboratory Profile of Enteric Fever. *J. Glob. Health 5* (2), 020407. https://doi.org/10.7189/jogh.05.020407.

(9) Gal-Mor, O.; Boyle, E. C.; Grassl, G. A. Same Species, Different Diseases: How and Why Typhoidal and Non-Typhoidal Salmonella Enterica Serovars Differ. *Front. Microbiol.* **2014**, *5*, 391. https://doi.org/10.3389/fmicb.2014.00391.

(10) Gerlach, R. G.; Jäckel, D.; Stecher, B.; Wagner, C.; Lupas, A.; Hardt, W.-D.; Hensel, M. Salmonella Pathogenicity Island 4 Encodes a Giant Non-Fimbrial Adhesin and the Cognate Type 1 Secretion System. *Cell. Microbiol.* **2007**, *9* (7), 1834–1850. https://doi.org/10.1111/j.1462-5822.2007.00919.x.

(11) Marcus, S. L.; Brumell, J. H.; Pfeifer, C. G.; Finlay, B. B. Salmonella Pathogenicity Islands: Big Virulence in Small Packages. *Microbes Infect.* **2000**, *2* (2), 145–156. https://doi.org/10.1016/s1286-4579(00)00273-2.

(12) Lorkowski, M.; Felipe-López, A.; Danzer, C. A.; Hansmeier, N.; Hensel, M. Salmonella Enterica Invasion of Polarized Epithelial Cells Is a Highly Cooperative Effort. *Infect. Immun.* **2014**, *82* (6), 2657–2667. https://doi.org/10.1128/IAI.00023-14.

(13) Galán, J. E. Typhoid Toxin Provides a Window into Typhoid Fever and the Biology of Salmonella Typhi. *Proc. Natl. Acad. Sci.* **2016**, *113* (23), 6338–6344. https://doi.org/10.1073/pnas.1606335113.

(14) Spanò, S.; Ugalde, J. E.; Galán, J. E. Delivery of a Salmonella Typhi Exotoxin from a Host Intracellular Compartment. *Cell Host Microbe* **2008**, *3* (1), 30–38. https://doi.org/10.1016/j.chom.2007.11.001.

(15) Nguyen, T.; Lee, S.; Yang, Y.-A.; Ahn, C.; Sim, J. H.; Kei, T. G.; Barnard, K. N.; Yu, H.; Millano, S. K.; Chen, X.; Parrish, C. R.; Song, J. The Role of 9-O-Acetylated Glycan Receptor Moieties in the Typhoid Toxin Binding and Intoxication. *PLOS Pathog.* **2020**, *16* (2), e1008336. https://doi.org/10.1371/journal.ppat.1008336.

(16) Haghjoo, E.; Galán, J. E. Salmonella Typhi Encodes a Functional Cytolethal Distending Toxin That Is Delivered into Host

Cells by a Bacterial-Internalization Pathway. *Proc. Natl. Acad. Sci.* **2004**, *101* (13), 4614–4619. https://doi.org/10.1073/pnas.0400932101.

(17) Song, J.; Gao, X.; Galán, J. E. Structure and Function of the Salmonella Typhi Chimaeric A2B5 Typhoid Toxin. *Nature* **2013**, *499* (7458), 350–354. https://doi.org/10.1038/nature12377.

(18) A Bacterial Toxin That Controls Cell Cycle Progression as a Deoxyribonuclease I-Like Protein https://www.science.org/doi/10.1126/science.290.5490.354?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed (accessed 2022 -03 -21).

(19) Lara-Tejero, M.; Galán, J. E. Cytolethal Distending Toxin: Limited Damage as a Strategy to Modulate Cellular Functions. *Trends Microbiol.* **2002**, *10* (3), 147–152. https://doi.org/10.1016/s0966-842x(02)02316-8.

(20) Guerra, L.; Cortes-Bratti, X.; Guidi, R.; Frisan, T. The Biology of the Cytolethal Distending Toxins. *Toxins* **2011**, *3* (3), 172–190. https://doi.org/10.3390/toxins3030172.

(21) Locht, C.; Coutte, L.; Mielcarek, N. The Ins and Outs of Pertussis Toxin. *FEBS J.* **2011**, *278* (23), 4668–4682. https://doi.org/10.1111/j.1742-4658.2011.08237.x.

(22) Costa, A. F.; Campos, D.; Reis, C. A.; Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends Cancer* **2020**, *6* (9), 757–766. https://doi.org/10.1016/j.trecan.2020.04.002.

(23) *Essentials of Glycobiology*, 3rd ed.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2015.

(24) Zhuo, L.; Kanamori, A.; Kannagi, R.; Itano, N.; Wu, J.; Hamaguchi, M.; Ishiguro, N.; Kimata, K. SHAP Potentiates the CD44-Mediated Leukocyte Adhesion to the Hyaluronan Substratum*. *J. Biol. Chem.* **2006**, *281* (29), 20303–20314. https://doi.org/10.1074/jbc.M506703200.

(25) de la Motte, C. A.; Hascall, V. C.; Calabro, A.; Yen-Lieberman, B.; Strong, S. A. Mononuclear Leukocytes Preferentially Bind via CD44 to Hyaluronan on Human Intestinal Mucosal Smooth Muscle Cells after Virus Infection or Treatment with Poly(I·C)*. *J. Biol. Chem.* **1999**, *274* (43), 30747–30755. https://doi.org/10.1074/jbc.274.43.30747.

(26) Wassarman, P. M.; Litscher, E. S. Towards the Molecular Basis of Sperm and Egg Interaction during Mammalian Fertilization. *Cells*

Tissues Organs **2001**, *168* (1–2), 36–45. https://doi.org/10.1159/000016804.

(27) Diekman, A. B. Glycoconjugates in Sperm Function and Gamete Interactions: How Much Sugar Does It Take to Sweet-Talk the Egg? *Cell. Mol. Life Sci. CMLS* **2003**, *60* (2), 298–308. https://doi.org/10.1007/s000180300025.

(28) Itin, C.; Roche, A. C.; Monsigny, M.; Hauri, H. P. ERGIC-53 Is a Functional Mannose-Selective and Calcium-Dependent Human Homologue of Leguminous Lectins. *Mol. Biol. Cell* **1996**, *7* (3), 483–493. https://doi.org/10.1091/mbc.7.3.483.

(29) Zhang, B.; Cunningham, M. A.; Nichols, W. C.; Bernat, J. A.; Seligsohn, U.; Pipe, S. W.; McVey, J. H.; Schulte-Overberg, U.; de Bosch, N. B.; Ruiz-Saez, A.; White, G. C.; Tuddenham, E. G. D.; Kaufman, R. J.; Ginsburg, D. Bleeding Due to Disruption of a Cargo-Specific ER-to-Golgi Transport Complex. *Nat. Genet.* **2003**, *34* (2), 220–225. https://doi.org/10.1038/ng1153.

(30) Hart, G. W.; Copeland, R. J. Glycomics Hits the Big Time. *Cell* **2010**, *143* (5), 672–676. https://doi.org/10.1016/j.cell.2010.11.008.

(31) Measuring the forces involved in polyvalent adhesion of uropathogenic Escherichia coli to mannose-presenting surfaces | PNAS https://www.pnas.org/doi/full/10.1073/pnas.230451697 (accessed 2022 -04 -28).

(32) Santos, J. I.; Carvalho de Souza, A.; Cañada, F. J.; Martín-Santamaría, S.; Kamerling, J. P.; Jiménez-Barbero, J. Assessing Carbohydrate–Carbohydrate Interactions by NMR Spectroscopy: The Trisaccharide Epitope from the Marine Sponge Microciona Prolifera. *ChemBioChem* **2009**, *10* (3), 511–519. https://doi.org/10.1002/cbic.200800548.

(33) Yang, Y.-A.; Lee, S.; Zhao, J.; Thompson, A. J.; McBride, R.; Tsogtbaatar, B.; Paulson, J. C.; Nussinov, R.; Deng, L.; Song, J. In Vivo Tropism of Salmonella Typhi Toxin to Cells Expressing a Multiantennal Glycan Receptor. *Nat. Microbiol.* **2018**, *3* (2), 155–163. https://doi.org/10.1038/s41564-017-0076-4.

(34) Li, W.; Xiao, A.; Li, Y.; Yu, H.; Chen, X. Chemoenzymatic Synthesis of Neu5Ac9NAc-Containing A2–3- and A2–6-Linked Sialosides and Their Use for Sialidase Substrate Specificity Studies. *Carbohydr. Res.* **2017**, *451*, 51–58. https://doi.org/10.1016/j.carres.2017.09.003.

(35) Klein, A.; Roussel, P. O-Acetylation of Sialic Acids. *Biochimie* **1998**, *80* (1), 49–57. https://doi.org/10.1016/S0300-9084(98)80056-4.

(36) Schauer, R. Sialic Acids as Regulators of Molecular and Cellular Interactions. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 507–514. https://doi.org/10.1016/j.sbi.2009.06.003.

(37) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

(38) Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int. J. Mol. Sci.* **2020**, *21* (17), 6339. https://doi.org/10.3390/ijms21176339.

(39) Feig, M.; Sugita, Y. Whole-Cell Models and Simulations in Molecular Detail. *Annu. Rev. Cell Dev. Biol.* **2019**, *35* (1), 191–211. https://doi.org/10.1146/annurev-cellbio-100617-062542.

(40) Hernández-Rodríguez, M.; Rosales-Hernández, M. C.; Mendieta-Wejebe, J. E.; Martínez-Archundia, M.; Basurto, J. C. Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design. *Curr. Med. Chem. 23* (34), 3909–3924.

(41) Notman, R.; Anwar, J. Breaching the Skin Barrier--Insights from Molecular Simulation of Model Membranes. *Adv. Drug Deliv. Rev.* **2013**, *65* (2), 237–250. https://doi.org/10.1016/j.addr.2012.02.011.

(42) D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber 2021.

(43) CHARMM: The biomolecular simulation program - Brooks - 2009 - Journal of Computational Chemistry - Wiley Online Library https://onlinelibrary.wiley.com/doi/10.1002/jcc.21287 (accessed 2022 -04 -26).

(44) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666. https://doi.org/10.1021/ja00214a001.

(45) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(46) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(47) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(48) CHARMM36m: an improved force field for folded and intrinsically disordered proteins | Nature Methods https://www.nature.com/articles/nmeth.4067 (accessed 2022 -04 -26).

(49) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. https://doi.org/10.1002/jcc.20820.

(50) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell Jr., A. D. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929–1943. https://doi.org/10.1002/jcc.21777.

(51) Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Mol. Simul.* **1991**, *6* (1–3), 121–142. https://doi.org/10.1080/08927029108022142.

(52) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1* (3), 173–185. https://doi.org/10.1080/08927028808080941.

(53) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76* (1), 637–649. https://doi.org/10.1063/1.442716.

(54) Beeman, D. Some Multistep Methods for Use in Molecular Dynamics Calculations. *J. Comput. Phys.* **1976**, *20* (2), 130–139. https://doi.org/10.1016/0021-9991(76)90059-0.

(55) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(56) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(57) Andrew R Leach - Molecular Modelling - Principles and Applications-Prentice Hall (2001) | PDF https://www.scribd.com/document/449103078/Andrew-R-Leach-Molecular-modelling-principles-and-applications-Prentice-Hall-2001-pdf (accessed 2022 -04 -26).

(58) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(59) Evans, D. J.; Baranyai, A. The Gaussian Thermostat, Phase Space Compression and the Conjugate Pairing Rule. *Mol. Phys.* **1992**, *77* (6), 1209–1216. https://doi.org/10.1080/00268979200103081.

(60) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(61) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(62) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384–2393. https://doi.org/10.1063/1.439486.

(63) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65* (6), 1409–1419. https://doi.org/10.1080/00268978800101881.

(64) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87* (5), 1117–1157. https://doi.org/10.1080/00268979600100761.

(65) GLYCAM-Web | Utilities for molecular modeling of carbohydrates https://glycam.org/ (accessed 2022 -05 -01).

(66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(67) Interaction Models for Water in Relation to Protein Hydration | SpringerLink https://link.springer.com/chapter/10.1007/978-94-015-7658-1_21 (accessed 2022 -04 -26).

(68) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271. https://doi.org/10.1021/j100308a038.

(69) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103* (2), 335–340. https://doi.org/10.1021/ja00392a016.

(70) Improved general-purpose five-point model for water: TIP5P/2018: The Journal of Chemical Physics: Vol 149, No 22 https://aip.scitation.org/doi/10.1063/1.5070137 (accessed 2022 -04 -26).

(71) Nada, H.; van der Eerden, J. P. J. M. An Intermolecular Potential Model for the Simulation of Ice and Water near the Melting Point: A Six-Site Model of H2O. *J. Chem. Phys.* **2003**, *118* (16), 7401–7413. https://doi.org/10.1063/1.1562610.

(72) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119* (16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055.

(73) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2018**, *4*. https://doi.org/10.3389/fmolb.2017.00087.

(74) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* **2003**, *68* (1), 16–34. https://doi.org/10.1002/bip.10270.

(75) Martins, S. A.; Perez, M. A. S.; Moreira, I. S.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Computational Alanine Scanning Mutagenesis: MM-PBSA vs TI. *J. Chem. Theory Comput.* **2013**, *9* (3), 1311–1319. https://doi.org/10.1021/ct4000372.

(76) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897. https://doi.org/10.1021/ar000033j.

(77) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401–9409. https://doi.org/10.1021/ja981844+.

(78) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson–Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–392. https://doi.org/10.1002/jmr.577.

(79) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129. https://doi.org/10.1021/ja00172a038.

(80) Mishra, S. K.; Koča, J. Assessing the Performance of MM/PBSA, MM/GBSA, and QM–MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contributions. *J. Phys. Chem. B* **2018**, *122* (34), 8113–8121. https://doi.org/10.1021/acs.jpcb.8b03655.

(81) Sood, A.; Gerlits, O. O.; Ji, Y.; Bovin, N. V.; Coates, L.; Woods, R. J. Defining the Specificity of Carbohydrate–Protein Interactions by Quantifying Functional Group Contributions. *J. Chem. Inf. Model.* **2018**, *58* (9), 1889–1901. https://doi.org/10.1021/acs.jcim.8b00120.

(82) Hu, Z.; Bie, L.; Gao, J.; Wang, X. Insights into Selectin Inhibitor Design from Endogenous Isomeric Ligands of SLea and SLex. *J. Chem. Inf. Model.* **2021**, *61* (12), 6085–6093. https://doi.org/10.1021/acs.jcim.1c01356.

(83) Wang, X.; Bie, L.; Fei, J.; Gao, J. Insights into the Loop at the E-Selectin Binding Site: From Open to Close Conformation. *J. Chem. Inf. Model.* **2020**, *60* (10), 5153–5161. https://doi.org/10.1021/acs.jcim.0c00939.

(84) Samsonov, S. A.; Pisabarro, M. T. Computational Analysis of Interactions in Structurally Available Protein–Glycosaminoglycan Complexes. *Glycobiology* **2016**, *26* (8), 850–861. https://doi.org/10.1093/glycob/cww055.

(85) Topin, J.; Arnaud, J.; Sarkar, A.; Audfray, A.; Gillon, E.; Perez, S.; Jamet, H.; Varrot, A.; Imberty, A.; Thomas, A. Deciphering the Glycan Preference of Bacterial Lectins by Glycan Array and Molecular

Docking with Validation by Microcalorimetry and Crystallography. *PLOS ONE* **2013**, *8* (8), e71149. https://doi.org/10.1371/journal.pone.0071149.

(86) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(87) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. https://doi.org/10.1002/jcc.20820.

(88) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.

(89) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(90) Deng, L.; Song, J.; Gao, X.; Wang, J.; Yu, H.; Chen, X.; Varki, N.; Naito-Matsui, Y.; Galán, J. E.; Varki, A. Host Adaptation of a Bacterial Toxin from the Human Pathogen Salmonella Typhi. *Cell* **2014**, *159* (6), 1290–1299. https://doi.org/10.1016/j.cell.2014.10.057.

(91) Kräutler, V.; Gunsteren, W. F. van; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.* **2001**, *22* (5), 501–508. https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

(92) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(93) Cremer, D.; Pople, J. A. General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97* (6), 1354–1358. https://doi.org/10.1021/ja00839a011.

(94) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Topics in physical chemistry; Oxford University Press: New York Oxford, 1997.