`

Molecular Dynamics Simulations of Glycosylated Prion Protein
By
Rohith Vedhthaanth Sekar

A thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science

Department of Physics
University of Alberta

Abstract

While PrP is almost always found in the glycosylated state, there has been very little study attempting to explore the effects of the size of glycans, the composition of glycans, and the site of glycosylation on the stability of PrP. These effects can potentially play an important role in the misfolding of PrP (conversion of PrP^C to PrP^{Sc}), the cause of neurodegenerative prion diseases, either directly or indirectly. Glycosylation patterns found in the brain are different in different regions of the brain, and the disease vulnerability also varies across different parts of the brain, suggesting that glycans could play a role in the infectivity. Glycosylation patterns are also a key characteristic of infectious prion strains. Previous studies on other glycoprotein systems suggest that glycans impact the stability, structure, aggregation rates, and other properties of the protein they are attached to. These properties play a crucial role in the context of PrP misfolding. The difficulty associated with carrying out site specific chemical glycosylation, the large number of glycans that can attach to PrP, and multiple sites of glycosylation in PrP (N181 and N197) make it hard to carry out experiments to systematically explore the effects of the size of glycans, the composition of glycans and the site of glycosylation on PrP.

To explore these effects systematically, we performed molecular dynamics simulations of mono glycosylated and diglycosylated PrP using glycans that differ in size and composition. The behaviour of glycosylated PrP across all the simulations were compared. Our computational simulations provide insights into possible glycan-PrP interactions. Our results show that the glycan-PrP interaction is affected

minimally by the bulkiness of the glycans, but the presence of sialic acid groups strengthens the interaction significantly. We also show that a glycan attached at a site of glycosylation, can take on multiple conformations and can affect the stability of different parts of the protein depending on the conformation it adopts. More importantly, we report events where glycans induce unfolding within the PrP and we identify the segments of PrP that are vulnerable to such unfolding effects.

Preface

The research conducted in this thesis is the original work of Rohith Vedhthaanth Sekar, completed at the Department of Physics, University of Alberta.

All the molecular dynamics simulations discussed in this thesis were generated and analyzed by Rohith Vedhthaanth Sekar. This thesis was composed by Rohith Vedhthaanth Sekar along with contributions from Michael Woodside. All the molecular dynamics simulations discussed in this thesis were performed using the GPUs available in the Compute Canada resources. Access to Compute Canada resources and Molecular Operating Environment software was provided by Jack Tuszynski.

Dedication

I dedicate this thesis to my mom, dad, and sister. Without their love, support, and encouragement, it would not have been possible for me to work towards my dream. I will always be thankful to my family for being a constant source of strength.

Acknowledgement

Firstly, I would like to express my gratitude to my thesis advisor, Michael T. Woodside, for providing me with the opportunity to work as a part of his esteemed group. It was only with his support and patient guidance could this thesis be completed in less than a year and a half from the time I joined the Woodside Lab. The continuous feedback and insightful discussions through regular meetings helped me make quick progress in the project.

I would like to thank Jack Tuszynski for providing access to Compute Canada resources and Molecular operating environment software. These resources were highly utilized in the project. I would also like to thank Phillip and Sara of Jack's research group for helping me begin with molecular dynamics. I am thankful to my committee members Michael T. Woodside, Jack Tuszynski and Lindsay LeBlanc, for providing useful feedback during the committee meetings and helping me make the right decisions for the progress of the project.

I am thankful to all the members of the Woodside Lab for always extending help and patiently clarifying my doubts and queries especially during the early stages of the project. I would like to thank Aaron for proofreading the thesis and providing valuable suggestions. I would also like to thank Juan Pablo for being very kind in allowing me to make a smooth transfer between the research groups.

Finally, I would like to thank my family for their constant encouragement. Their love and support have always been my source of strength.

Contents

List of Tables

List of Figures

List of symbols, nomenclatures and abbreviations

AA   Amino acid

aMD   Accelerated Molecular Dynamics

Asn   Asparagine

CD   Circular dichroism

CJD   Creutzfeldt-Jakob's disease

Cryo-em   Cryo-electron microscopy

Cys   Cysteine

Cα   Alpha carbon atom of an amino acid residue

ECorl   Erythrina corallodendron lectin

EPO   Erythropoietin

FF   Force field

FTIR   Fourier transform infrared spectroscopy

Fuc   Fucose

Gal   Galactose

GalNAc   N-acetylgalactosamine

Glc   Glucose

GLCA   Glucuronic acid

GlcNAc   N-Acetylglucosamine

IdoA   Iduronic acid

IgG   Immunoglobulin G

k_B	Boltzmann constant
Man	Mannose
MC	Monte Carlo
MD	Molecular dynamics
MOE	Molecular operating environment
NMR	Nuclear Magnetic Resonance
PEG	Polyethylene glycol
Phy	Peniophora lycii phytase
QM/MM	Quantum mechanics/molecular mechanics model
recPrP	Recombinant PrP
REMD	Replica exchange molecular dynamics
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SA	Sialic acid
SBA	Soybean Agglutinin
SMFS	Single molecule force spectroscopy
TSE	Transmissible spongiform encephalopathies
vCJD	Variant Creutzfeldt-Jakob Disease
WBA	Winged bean agglutin
Xyl	Xylose

1. Introduction

1.1 Overview and Aim of the project

Prion diseases are progressive neurodegenerative diseases that can affect humans and animals. The loss of brain cells in these diseases can result in loss of balance, malfunctioning of the heart, memory loss, and other pathologies, eventually leading to death. These diseases are unusual because the disease-causing infectious particles are not external organisms such as virus and bacteria, but naturally produced proteins within the body itself that misfolds to take up an infectious form. Therefore, traditional methods of treating infectious diseases cannot be applied to neurodegenerative diseases, making it very challenging to devise a cure.

In prion diseases, protein PrP takes up an infectious abnormal isoform and aggregates in the brain [1]–[5]. The misfolded PrP, PrP^{Sc}, performs templated seeded conversion of PrP to PrP^{Sc} [6]. Mechanisms by which misfolding, conversion, and aggregation occur are poorly understood [4]. Currently, there are only supportive treatments for prionopathies and the survival duration is typically 24 months. Investigating the mechanism of PrP misfolding, its causes and studying any factors that can potentially contribute to the misfolding process will help in the understanding of the underlying cause of the disease, which can, in turn, be helpful in the efforts to devise medical therapeutics.

In the case of PrP, one factor that could play a very crucial role in the folding/misfolding mechanism and the stability of the three-dimensional (3D)

structure is the presence of glycans [7]. Glycans are carbohydrate-based polymers. They can be chemically attached to proteins through a post-translational modification called glycosylation [8]. Proteins that are modified through glycosylation are called glycoproteins. PrP being a glycoprotein, is almost always found in its chemically modified glycosylated state [7], [9]. The bulkiness of the glycans, site of glycosylation and composition of the glycans can dictate the interaction between glycans and the protein. This interaction can have specific effects on the stability of the glycoprotein and its folding mechanism. Previous studies of glycosylated proteins have shown that glycans can have both stabilizing and destabilizing effects on the protein they are attached to [10]–[15]. Glycans can also affect a protein's folding pathway and tendency to aggregate [16]–[19]. All these properties are of immense importance in the context of PrP misfolding. Therefore, studying how these effects manifest in the PrP molecule is essential to the complete understanding of the mechanism of disease spread.

The difficulty in carrying out experimental studies to understand the impact of glycans on PrP is mainly due to the high diversity of the glycoforms [20]. The number of possible glycans that can attach to PrP and the availability of multiple sites of glycosylation is the cause of the diversity [9]. There has been little quantitative study outside simulations and fibrillization studies. Fibrils of un-glycosylated recombinant PrP (recPrP) form parallel in-register beta sheets [21], [22], unlike the structure of fibrils of naturally glycosylated PrP^C that is similar to brain-derived PrP^{Sc} [23] (a beta solenoid [24]), whereas polyethylene glycol (PEG) chains attached to recPrP as glycan mimics prevent fibrillization [25]. Computational studies can provide insights

into the atomistic interaction between glycans and the proteins they are attached to. Glycosylation computational studies in the past have explored the effects of sugar molecules on the stability of various proteins [14], [26]–[29], and the results from these studies strongly suggest that glycans could play an important role in the stability of a much larger protein like PrP. However, there has been very little computational studies that attempt to explore the above-mentioned effects in atomistic scales in the case of PrP.

In this thesis, we use computational methods to systematically explore the effects of specific glycosylation on PrP. We attempt to understand the effects of size, composition, and the site of glycosylation on hamster PrP using all-atom explicit-water molecular dynamics (MD) simulation. This project is a preparatory work for the future study of specific effects of chemical glycosylation on PrP using single molecule force spectroscopy (SMFS) techniques. The results from this study would help decide the choice of glycans for chemical glycosylation, and hint at the stability of the glycoforms.

1.2 Protein Structure

The synthesis of proteins is through translation of RNA by ribosomes [30]. Proteins are essential building blocks of life that perform several essential functions in the body including transportation, storage, cell signalling, and providing mechanical support. Proteins are composed of long sequences of amino acids linked in a chain like fashion, one after the other. An amino acid (AA) is chemically composed of an amino group(-NH2), a carboxyl group(-COOH) and a side chain.

The side chain is unique to each AA. There are 20 different AAs that can make up proteins. These organic compounds form peptide bonds between each other allowing them to arrange in long linear chains which form the primary structure of a protein. The AAs in this sequence are numbered from N-terminus (the end with free amino acid residue) to the C-terminus (the end with free carboxyl group). Interactions between the amino acids in the primary sequence lead to the formation of secondary structures, and the interaction between these secondary structures drives the formation of tertiary structure, the overall 3D structure of the protein. This native 3D structure of the protein plays an important role in determining the protein's function [31].
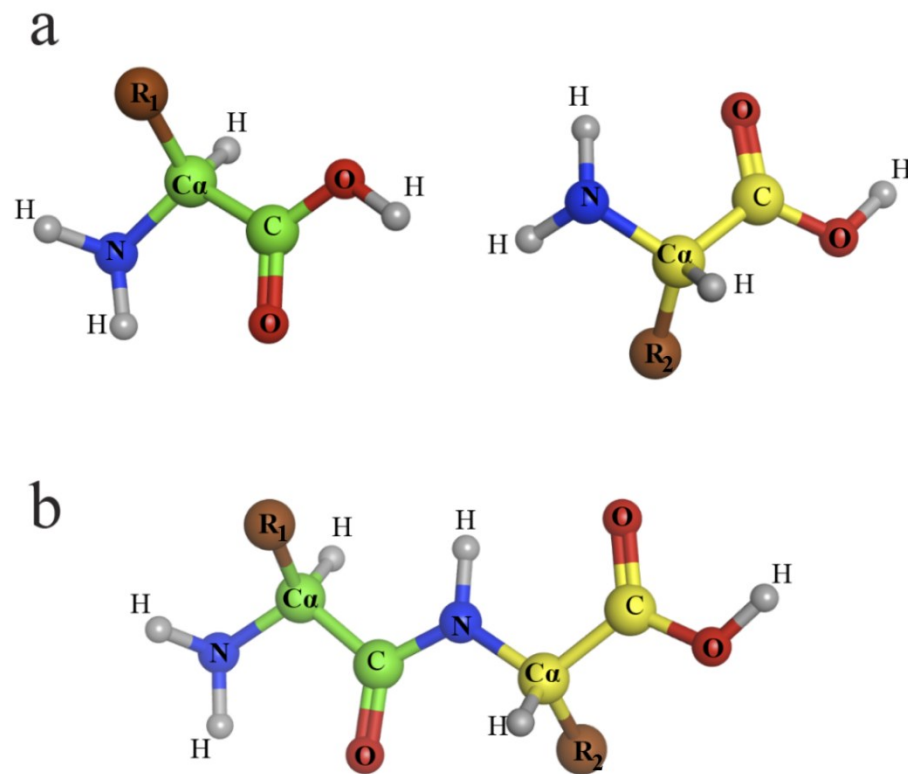
Figure 1-1Stucture of amino acid and Peptide bond (a) Two monomer structures and (b) the monomers connected through a peptide bond
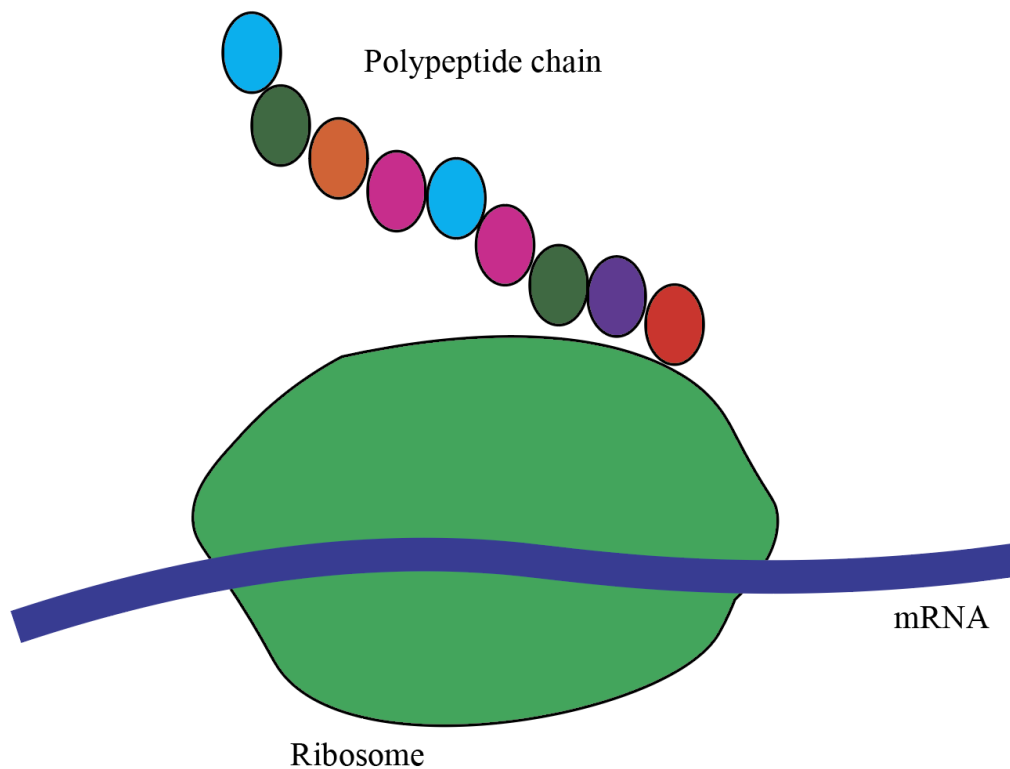
Figure 1-2. A cartoon illustrating protein synthesis

Interaction between the atoms of the backbone can give rise to local folds within the polypeptide creating regular recurring arrangements of adjacent amino acids called secondary structures in the protein [32]. Alpha helices and beta sheets are the most common types of secondary structures found in proteins. In an alpha helix, carbonyl oxygen forms a hydrogen bond with the amide hydrogen of the 4th amino acid further toward the C-terminal. This restricts the chain to form a local helical structure. Hydrogen bonding between carbonyl oxygens and amide hydrogens in parallel chains, i.e., chains lined next to each other such the direction in which N-terminus and C-terminus are facing in one chain matches with that of the other, allows

the formation of parallel a beta pleated sheet. If the N-terminus of one chain aligns with the C-terminus of the other chain, it is called an anti-parallel beta sheet. Hydrogen bonds in a beta sheet are formed nearly perpendicular to the protein chain. These secondary structures come together through further interactions between the side chains of AA from different secondary structures, disulfide bonds between cysteine residues, hydrophilic and hydrophobic interactions with the surrounding water molecules to orient themselves to form the tertiary structure which is the unique stable three-dimensional structure of the protein.
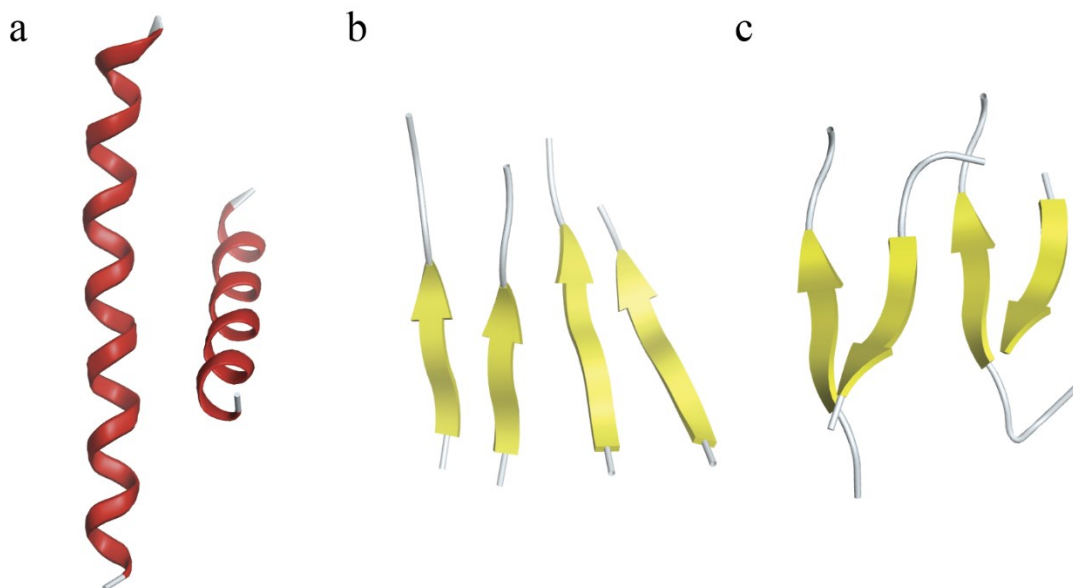
Figure 1-3 Ribbon diagram of secondary structures: (a) alpha helix (b) parallel beta sheets (c) anti-parallel beta sheets

For a protein to accomplish its functions in the cell, it is essential for it to fold into its native three-dimensional structure. Folding into the native structure enables

specific functional groups to be present on its outer surface, thereby achieving its unique function. This folding process is a complex mechanism and has been studied extensively [33]. The energy landscape of folding helps in understanding this folding process. The landscape is a representation of energies of all possible conformations of the protein in a given environment [34], [35]. On this free energy landscape, the native structure is the global minima, the structure with minimum free energy [33], [36]. Finding the native folded state through a random search among all possible configurations can take a very long time, but proteins fold in seconds or less. This paradox was stated by Levinthal [37]. It was later understood that proteins fold through stochastic searching of conformations accessible to the polypeptide chain [35], [38]. The lowest free energy state is achieved through a diffusive search and is not entirely random. It is guided by the stability of local interactions [31], [35]. There are multiple paths in the landscape through which proteins can fold [39]. However, some of the paths can end up trapped in metastable misfolded states [40].

If a protein misfolds into a different structure, the protein is not only unable to carry out its regular function, but it is also possible for it to malfunction leading to lethal effects like in the case of prion diseases [41]. There are mechanisms that exist within the cell to prevent misfolding. Chaperone proteins guide proteins to fold along the proper folding pathway [42]. They shield proteins from other proteins that might hinder the folding process. Proteins with the incorrect shape are unfolded by chaperones to allow them to be re-folded into their native structure [43], [44]. There are also mechanisms to break down and dispose of the misfolded protein that exists within the cells, and it is carried out by protease enzymes. Protease enzymes are a

group of enzymes that catalytically hydrolyze peptide bonds to break down the misfolded proteins [45]. The misfolded proteins that evade these processes, can cause complications. In the case of prion diseases, the misfolded PrP molecules aggregate and eventually form stable fibrils which are resistant to the breakdown mechanism of the protease enzymes [46].
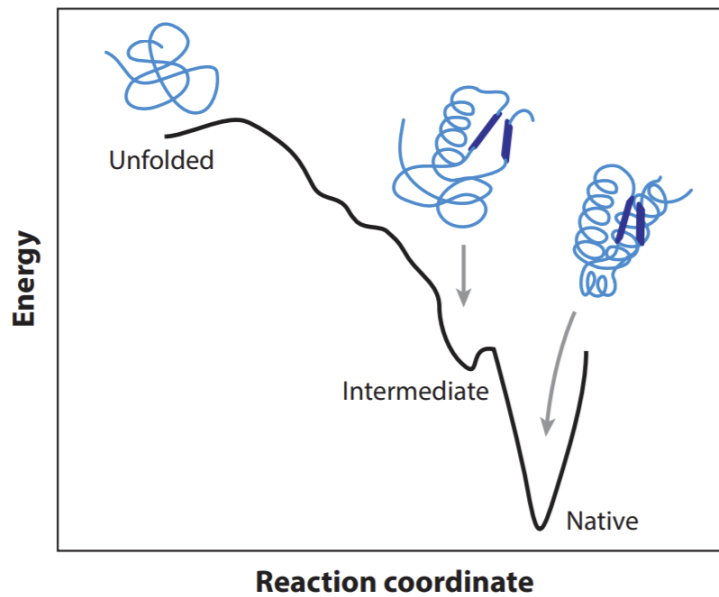
Figure 1-4 Energy landscape: Notational Cartoon of energy landscape in one dimension (adapted from [47])

1.3 Prion Misfolding

PrP is expressed highly in the tissues of the brain and the spine, but its biological activity is unclear [48]. It may also be found in other organs but in much

lower levels of expression [49]. Studies in the past have explored the physiological roles of natively folded PrP. These studies suggest that PrP^C can affect the homeostasis of copper, neuroprotection, stem cell renewal, memory mechanisms, and other functions [48][50][51].

The structure of PrP in its natively folded state, PrP^C, contains three alpha helices and an antiparallel beta sheet. The N-terminal domain is unstructured and flexible, while the C-terminal domain is structured. During the conversion from PrP^C to PrP^{Sc}, the C-terminal region undergoes structural changes and turns protease resistant but the N-terminal region remains protease sensitive [52], [53]. This structured core is 100-110 AAs long depending on the species. The structure of PrP^C has been resolved for various species [54]–[56] and the coordinates of the atoms for the structures can be found in the Protein Data Bank (https://www.rcsb.org). Some snapshots of this native structure from various species are shown below (Figure 1-5). PrP in this native state is non-infectious. Prion diseases are caused by the misfolded form of PrP, PrP^{Sc}. The structure of this infectious PrP^{Sc} has not been resolved because the purification of PrP^{Sc} is difficult as it forms insoluble aggregates. However, studies that used Fourier transform infrared spectroscopy (FTIR) and circular dichroism (CD) to understand the structure of misfolded PrP indicate that PrP^{Sc} is found in an aggregated state and its composition is dominated by beta sheets [57], [58]. Brain derived PrP^{Sc} are thought to have a beta-helical structure, but fibrils of un-glycosylated recombinant PrP (recPRP) form parallel in-register beta sheets unlike brain derived PrP^{Sc} [21]–[23]. Structure of sections of in vitro formed beta rich fibrils are available in the protein data bank [24], [59]–[61].
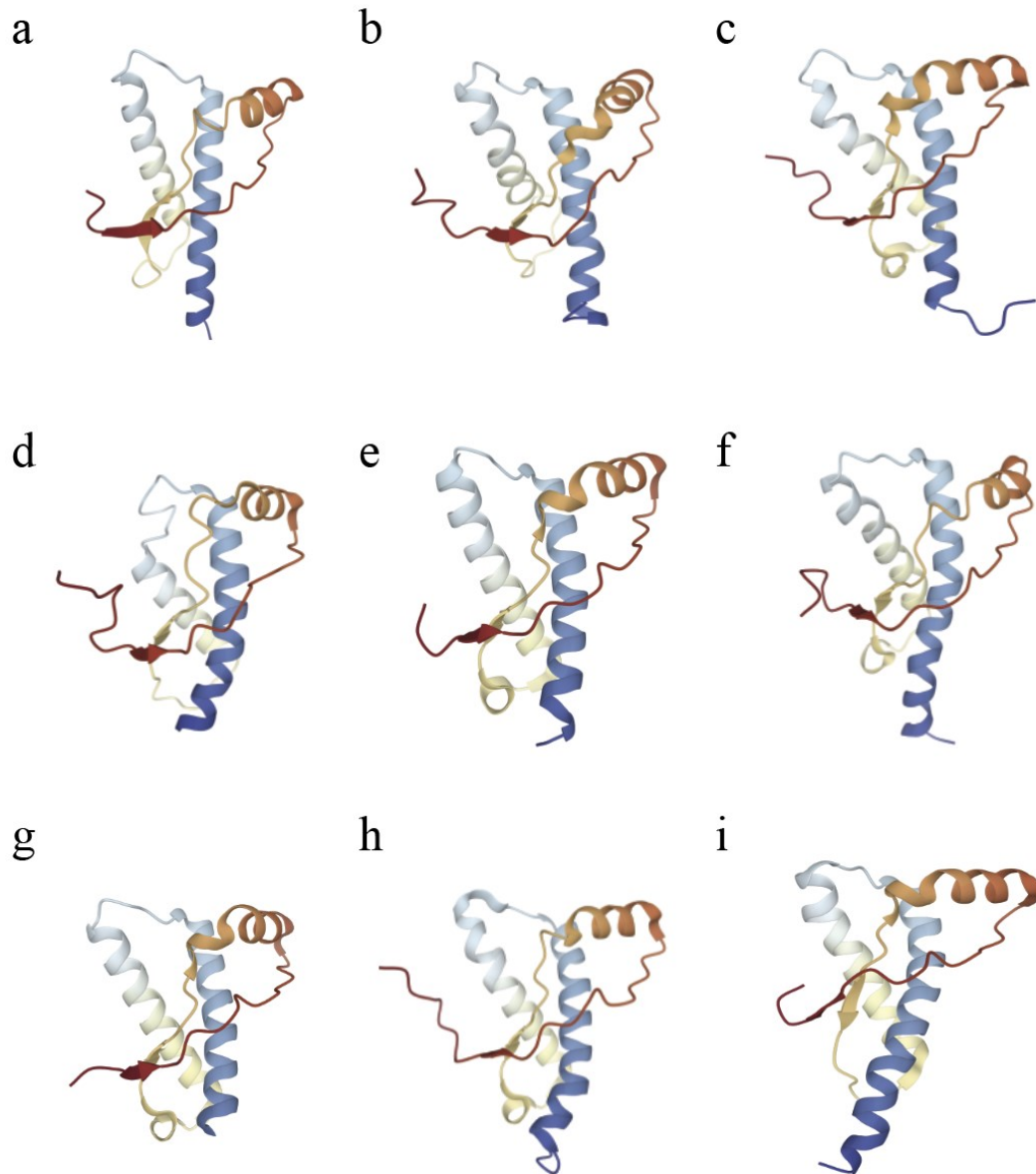
Figure 1-5 Snapshots of PrP molecules (available in RCSB) of some selected species in their native alpha rich state: (a) Bovine, PDB 1dwy [62], (b)Cat, PDB 1xyj [63], (c) elk, PDB 1xyw [56], (d) turtle, PDB 1u5l [64], (e) Deer, PDB 4yxh [65], (f) Sheep , PDB 1xyu [63], (g)Wild-type Rabbit, PDB 3o79 [66], (h) Bank vole Prion, PDB 2k56 [67], and (i) Human Prion (mutant e200k), PDB fkc [68]). All of them share similar architecture rich in alpha helix.
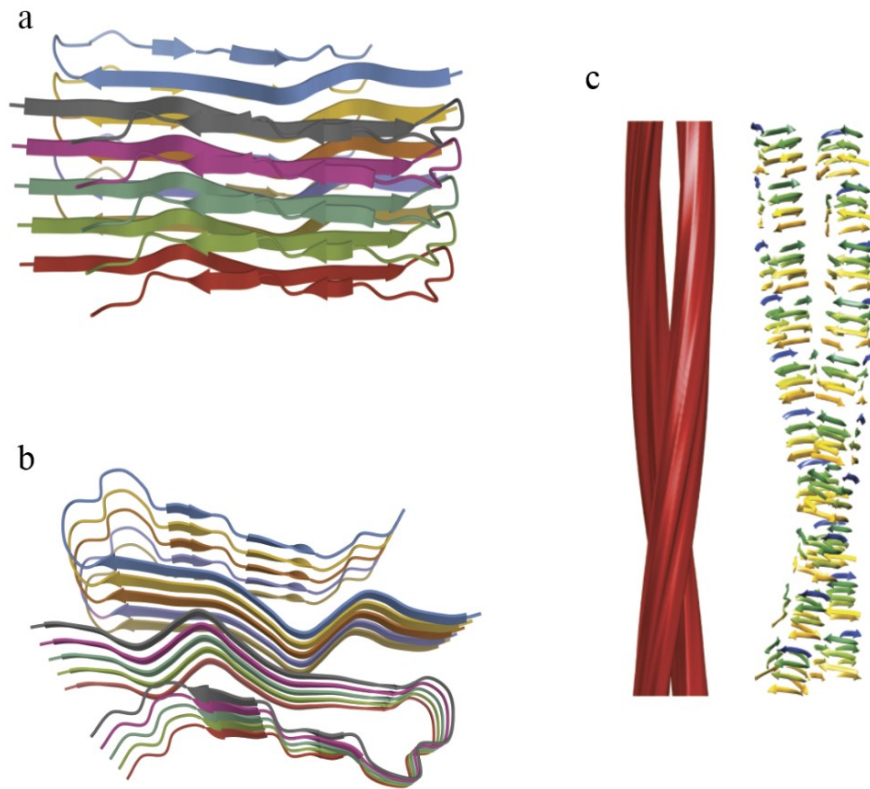
Figure 1-6 Fibril cryo-em: (a,b) Cryo-em structure of a segment of fibrils formed by human PrP in vitro, showcasing in-register beta sheets [61]. (c) View of possible four rung beta solenoid structure of brain derived PrP^{Sc} drawn to approximate 3D reconstruction (adapted from [24])

Interestingly, PrP^{Sc} can act as a seed in the conversion of PrP^C to PrP^{Sc}, i.e., the infectious form can convert the natively folded PrP into the misfolded form, initiating a chain reaction [4]. However, the mechanism of conversion of PrP^C to PrP^{Sc} is controversial and poorly understood. Evidence suggests that PrP^{Sc} could act as a template in the conversion and recruitment of PrP^C [46]. This is the fundamental cause of the propagation of prion disease. Studies also show the existence of

structurally distinct prion strains [69], [70]. Glycosylation patterns of these strains are also different. Findings from experiments show evidence of strain-specific uptake of glycoforms into PrP^{Sc} [71] and certain glycoforms are easier to misfold [72]. Glycosylation patterns in the brain are different in different regions, and disease vulnerability is also different across different regions of the brain [73]–[75]. This suggests that glycosylation could play a crucial role in the misfolding of PrP. The mechanism by which the neurons are killed by misfolded PrP is also unclear. The neurodegenerative diseases caused by PrP misfolding are called transmissible spongiform encephalopathies (TSEs), because the degradation of neurons creates vacuoles in the tissues of the brain, leaving behind a sponge-like structure. These diseases affect both humans and animals with different species having different levels of susceptibility to the disease. Creutzfeldt-Jakob disease (CJD), Variant Creutzfeldt-Jakob disease (vCJD), Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia and kuru are human prion diseases, bovine spongiform encephalopathy (BSE), chronic wasting disease (CWD), scrapie, transmissible mink encephalopathy, feline spongiform encephalopathy and ungulate spongiform encephalopathy are animal prion diseases identified to date by The Centers for Disease Control and Prevention (CDC). The structure of PrP^{Sc} and the mechanism of conversion of alpha rich PrP^{C} to the infectious beta rich PrP^{Sc}, factors affecting this conversion are important unknowns that are required to be solved to completely understand the mechanism and the underlying cause of the disease spread. This understanding is crucial to devising therapeutics.

1.4 PrP - a glycoprotein

A notable feature of PrP is that PrP is a glycoprotein, i.e., it is almost always modified by the chemical addition of glycans through post translational modifications. N181 and N197 are two sites that allow for N-glycosylation in PrP [7]. The glycans that can attach to these sites are bulky oligosaccharides, and when attached to proteins, they are known to affect the stability of the protein both at the local site of glycosylation and globally [10], [12]. Studies conducted in the past to investigate the glycan effects suggest that sugar molecules can affect the folding mechanism by changing the populations of intermediate states, final structures, aggregation tendency, and so on [19], [76]. The properties of proteins that are known to have an impact due to the presence of glycan are of significant importance in the case of PrP misfolding. Therefore, glycans can potentially play an important role in PrP disease spread mechanisms. PrP^{Sc} is almost always found to be in the glycosylated state [77] even though glycosylation is not a necessary requirement for PrP to misfold [78]. Previous studies strongly suggest that glycosylation plays a significant role in the prion disease itself [7]. Also, by changing glycosylation, the infectious properties of some prion strains could be altered [72]. This is supported by experimental findings of strain specific uptake of glycoforms into PrP^{Sc} [71]. Different strains of human prion diseases have been found to contain different patterns of glycosylation [79]. PrP from different tissues can have different glycosylation patterns in the brain [73]. The vulnerability of different regions in the

brain also differs [74], [75], [80]. All this strongly suggests that glycans could play an important role in the misfolding process.

a

b
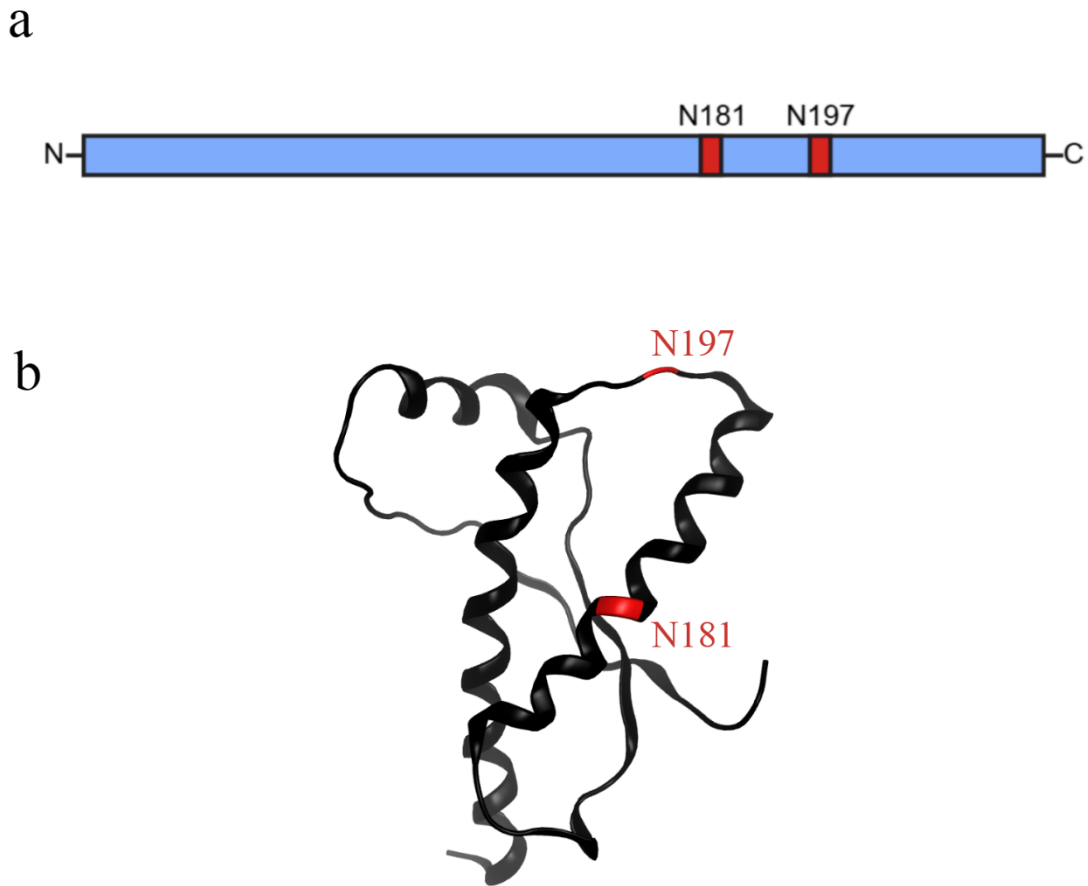
Figure 1-7 N-glycosylation sites on Hamster PrP: a) linear sequence and b) 3D image of Hamster PrP (PDB 1b10) pointing out the glycosylation sites in red [54].

All the above-mentioned effects of glycans on the properties of proteins, like the native structure, its stability, folding mechanisms, and folding pathways, combined with the results from studies suggesting evidence of high levels of glycosylated PrP^{Sc} strains and modified infectivity in glycosylated prions, suggest that glycans could have a significant impact on the stability of PrP structure, folding/misfolding mechanisms, rates, and energy barriers. However, despite the

above-mentioned efforts, the role of glycans in the misfolding of PrP is still poorly understood. This is due to the diversity in the prion glycoforms. Mass spectroscopy studies have revealed that there are more than 50 possible glycan chains detected in the PrP from brain cells [9], [20]. Compared to PrP^C, PrP^{Sc} contains increased levels of tri-antennary and tetra-antennary glycans, i.e., glycans with three are four branches [20]. Each of the two possible sites of glycosylation, N181 and N197, can be occupied or vacant[9], [20]. The number of glycans that can be attached to PrP and the possibility of PrP being mono- or di-glycosylated gives rise to 3000 different glycoforms [20]. This diversity in glycoforms limits the types of experiments and systematic analysis that needs to be performed. Lack of control in producing the desired glycoform, i.e., limiting a glycan chain to link to a selected site between N181 and N197, makes it difficult to synthesise and isolate single molecules with specific glycosylation. This makes performing single molecules studies on specific glycosylated variants hardly possible. Such studies are necessary to understand the specific effects of glycans on PrP's structure and the folding pathway. Therefore, studies aiming to understand the specific effects of glycans on PrP in the past were limited to computational approaches [26]–[28]. While these studies have yielded insights into sight specific stabilizing effects and effects of glycans on dimerization, these computational studies have had several limitations. They either had short simulation times to notice significant effects due to bulky glycans or did not consider the possibility of heterogeneous glycosylation, also the current force fields have seen a lot of improvement over the force fields used in the past studies.

1.5 Outline of the study

This study aims to explore the impact of glycans on prion structure and stability by studying the effects of size of glycans, site of glycosylation, and composition of glycans through all-atom, explicit solvent MD simulations using the molecular dynamics software AMBER.

To investigate these effects systematically, we modelled three different glycans that differed in structural complexity and composition. The effects of the glycans were studied on a model of Hamster PrP (PDB 1b10 [54]). Three different glycoforms using each glycan were built, i.e., mono glycosylation at N181, mono glycosylation at N197 and homo diglycosylation. Each of the glycoforms (including an unglycosylated PrP) were simulated for 1 microsecond. Three such simulations were produced for every glycoform. The dynamics of these glycoforms and that of unglycosylated PrP were studied and compared.

The rest of the thesis is organised into four chapters. Chapter 2 describes glycosylation and discusses the various known effects of glycosylation on proteins. Chapter 3 introduces molecular dynamics, discusses the construction of glycoforms, and the choice of force fields used to simulate PrP glycoforms in this study. Chapter 4 presents the results and observations from all the simulations, and Chapter 5 outlines the results and discusses possible future studies that are motivated by the results.

2. Glycosylation

Proteins are modified through covalent addition of specific chemical groups. These enzymatic modifications can either occur post translationally or co-translationally. Glycosylation, phosphorylation, methylation, acetylation and addition of lipids are some commonly seen modifications to proteins[81]. These modifications can directly impact the structure, stability, folding, solubility, catalytic function and biological activity of the proteins. The probability of proteins existing with no modifications in physiological systems is low, most proteins generally undergo multiple post translational modifications [82], [83].

Glycosylation is the chemical modification of proteins through covalent attachment of glycan groups, i.e., oligosaccharides/carbohydrate groups [8]. It is also known to be the most complex form of covalent modification to proteins due to the large number of enzymatic steps involved in the process [82]. Glycosylation is known to affect protein solubility and protein functions like adhesion, motility, and cell signaling [84], [85]. Depending on the atom that forms the link between the protein and the glycans, glycosylation is divided into O-glycosylation and N-glycosylation [8]. Since PrP undergoes N-glycosylation only, we focus only on the mechanism of N-glycosylation in this thesis. The nitrogen atom in the amide group of asparagine (Asn) forms the link between protein and glycans in N-glycosylation. For N-glycosylation to occur, it is necessary that the asparagine present in the AA sequence of the protein falls in the pattern, Asn-X-Ser/Thr, where X is any amino acid but proline [8]. N-glycosylation begins in the endoplasmic reticulum and can occur co-

translationally (if a protein is secreted into the ER co-translationally) or post-translationally. In spite of the fact that glycans are often structurally highly flexible, they can provide stability to glycoproteins at the local site of glycosylation [86], [87]. In most cases, they are also known to help proteins fold into their native 3D structure and provide additional stability, but sometimes, they can have destabilizing effects on the structure[13], [15]. Presence of glycans can also affect the enzymatic activity of proteins [88], [89].

Several studies in the past have attempted to understand how glycans affect the structure and functioning of proteins. Even though a lot has been learned, yet there are so many things that are still unclear. Glycans are generated by linking multiple monosaccharide units through glycosidic bonds and there is a large variety of chemically different monosaccharide units that can take part in this process. These units can link in any possible order to form the glycan. The monosaccharide units can also link to form multiple branches within the glycan. Due to this, the glycans can be seen to have diverse possible configurations. This diversity is the main challenge faced in studying its effects, as it makes isolation and characterization difficult. The bulkiness also makes it challenging to crystalize the glycoprotein.

The next sub-section discusses the mechanism of N-glycosylation, the type of glycosylation PrP molecules can undergo.

2.1 N-glycosylation

N-glycosylation begins in the Endoplasmic Reticulum (ER) and the glycans continue to get modified in the Golgi apparatus [90], [91]. The synthesis of glycans

can begin even before the translation of the protein. It is estimated that mammalian glycans have more than 7000 structures which are assembled only using 10 monosaccharide units, namely, Fucose (Fuc), galactose (Gal), glucose (Glc), N-acetyl galactosamine (GalNAc), N-acetylglucosamine (GlcNAc), glucuronic acid (GLCA), iduronic acid (IdoA), mannose (Man), sialic acid (SA), and xylose (Xyl) [92], [93]. Biosynthesis of mammalian glycans occurs through complicated pathways. It begins with the synthesis of dolichol-linked precursor oligosaccharide [94], [95]. N-acetylglucoseamine (GlcNAc) attaches to dolichol phosphate on the cytoplasm side of ER [96]. This is followed by the building of the core of the glycan unit through the addition of mannose units in the ER. More mannose units can be further added to this core to initiate branching in the glycan structure. Then, glucosyltransferase attaches more glucose molecules to the mannose units in the glycan. Oligosaccharyl transferase transfers this glycan to the protein by recognizing the consensus sequence at the N-glycosylation site [97]. The glucose residues present in the glycans at this stage are known to attract chaperones which help in protein folding[95], [98]. After which, glucosidase removes the glucose units from the sugar molecule. The glycans are further enzymatically edited through the addition and deletion of various monosaccharide units through various stages in the Golgi apparatus [91].
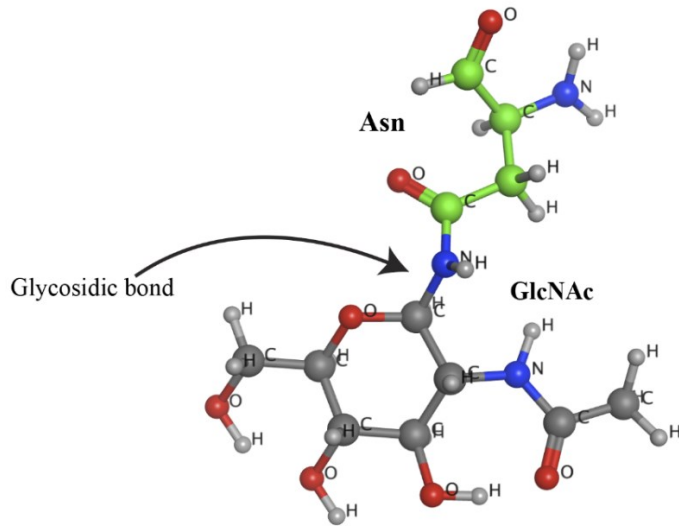
Figure 2-1 N-glycosidic linkage
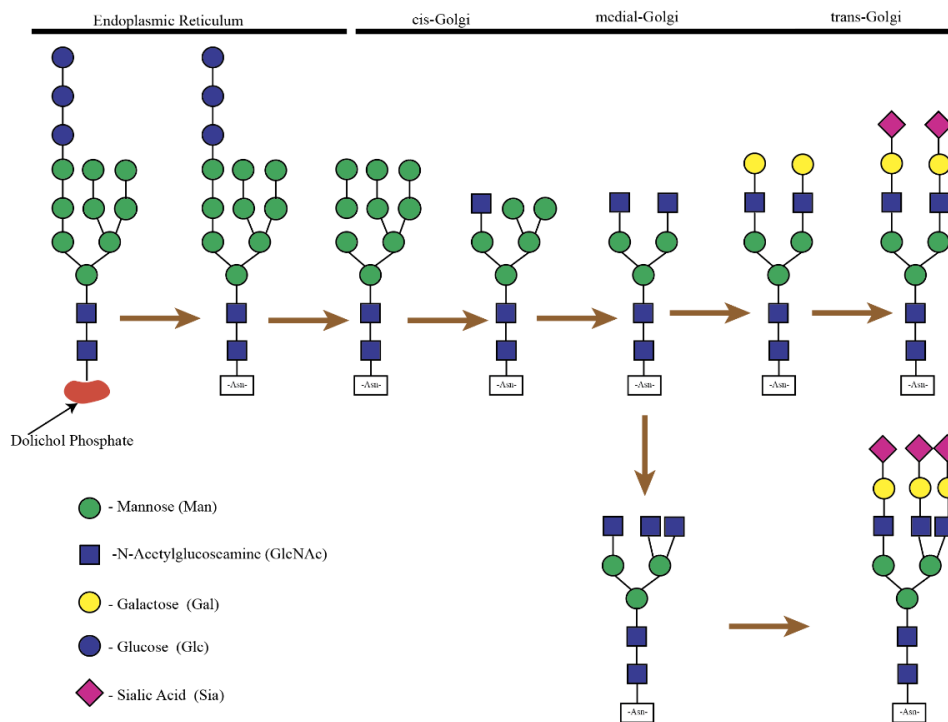
Figure 2-2 Illustration of glycan development through ER and Golgi apparatus:

A cartoon showcasing development of glycans in different stages through ER and

Golgi apparatus ([99], [100])

The core of the glycan that links to the protein is a common sugar sequence for all N-glycans [90]. It is generally made up of two GlcNAc units attached linearly to the Asn residue to form a disaccharide structure, followed by a mannose unit which is then followed by two more mannose units linked to the first mannose unit creating two branches. This core can further be extended through linear or branched chains of monosaccharide units, and there is no necessary template to this construction. N-glycans can be high mannose, i.e, the rest of the glycan apart from the core can be rich in mannose units. They can also be complex, i.e., after the core, the rest of the glycan is rich in a variety of different monosaccharide units like GlcNac, galactose, fucose, sialic acid, and other monosaccharides. There is also a hybrid type wherein some branches can be mannose rich while other branches have complex compositions [93], [101]. The glycans can grow to a size that is comparable or even much larger than the protein they are attached to. Since N-glycosylation may occur co-translationally, the glycans could attach to the protein even before its complete 3D structure is formed [10]. This can result in altering the stability of the intermediate state to which the glycan attaches and could potentially change the protein folding pathway.
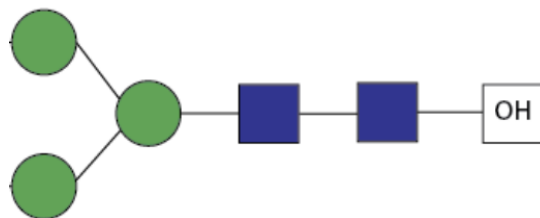
Figure 2-3 Structure of the core of N-glycans
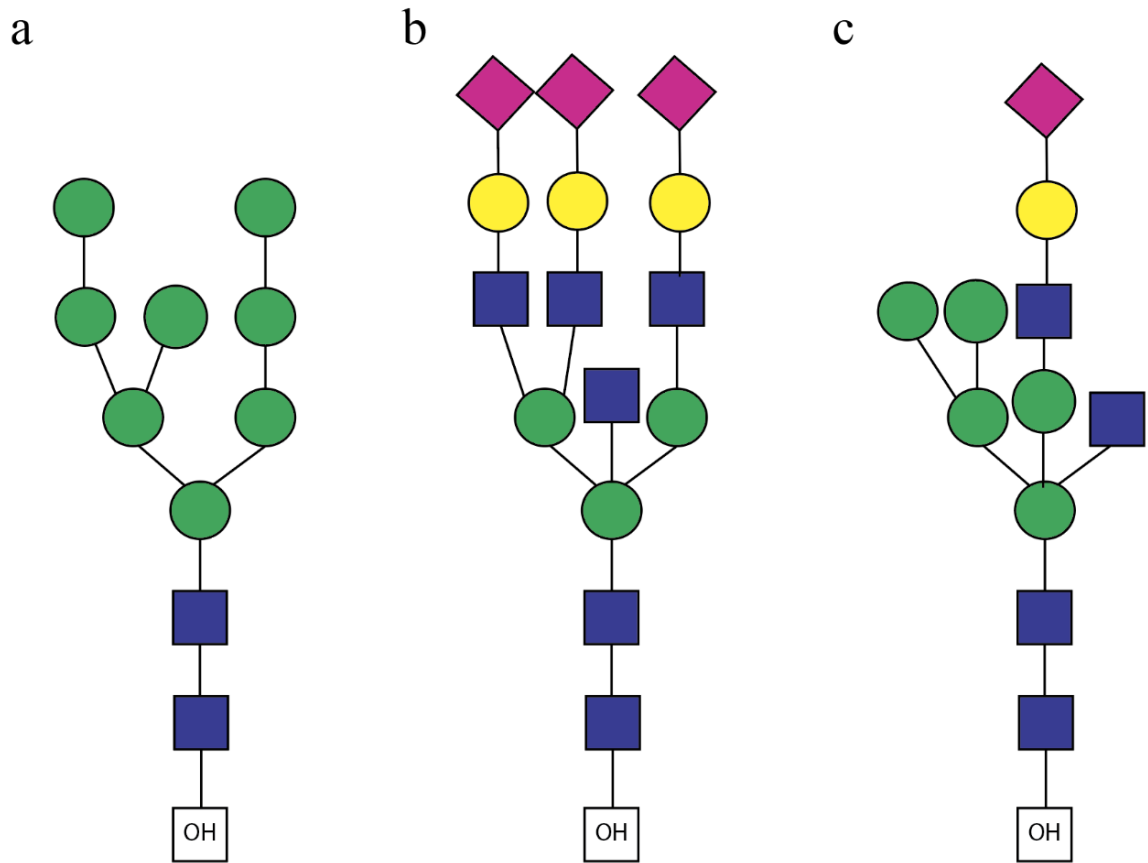
Figure 2-4 Illustration of types of glycans: examples of a) mannose rich glycan, b)

a complex glycan, c) a hybrid glycan represented through block diagrams.

2.2 Impact of glycosylation on proteins

Glycans can have varied impacts on the stability of the protein's structure.

Several in silico and in vitro studies have been conducted in the past to explore these

effects [14], [15], [17], [19], [29], [102]–[104]. Glycans usually increase the thermal

and kinetic stability of the proteins they are attached to by reducing dynamic

fluctuations, and this increase in the stability is manifested as an increase in the unfolding temperature and can be experimentally measured. There are also studies that report destabilizing effects of glycans on the glycoproteins, this was mainly due to decreased long range interactions within the protein. The in-silico studies conducted to explore the effects of glycans on the stability of glycoproteins shed lights on the interaction between glycan and protein at a molecular level. These studies suggest that glycans can increase the stability of the local site of glycosylation. Effects on the stability of the glycoprotein due to N-glycosylation is relatively higher than O-glycosylation [12], [105], [106]. A recent study summarizes and reviews the role of glycosylation in protein folding [10]. Some results from past studies on the glycan effects on protein structure and stability, protein folding, and other properties, along with studies discussing effects of bulkiness, branching and presence of sialic acid units in the glycans are mentioned below.

Glycosylation of ECorL (lectin from Erythrina corallodendron) resulted in no changes in the overall 3-dimensional structure, but it was seen to reduce the non-polar solvent accessible area when compared to its non-glycosylated variant [76]. This result supports the experimental observation of higher thermodynamic stability of glycosylated variants [107]. The study also shows that glycans' dynamic view of interactions with protein residues is very different from what can be observed in the static picture in the crystal structure. It is also suggested that glycans help proteins to form long range interactions between the AAs, thereby helping the protein in its folding process. In another study using glycosylated form of SBA (soybean agglutinin), a beta rich protein, showed higher stability in thermal unfolding at normal

and elevated temperatures due to formation of non-covalent interaction between glycan and protein [108]–[111]. In fact, the folded structure of legume lectin was held by the glycans through the non-covalent interactions. The dynamics of glycosylated Ribonuclease B showed that glycosylation resulted in increased stability by reducing dynamic fluctuation [11]. The global structures are however not affected through glycosylation. However, NMR measurements showed that N-glycosylated variants of the protein showed protection of amide-proton resonances both at local sites of glycosylation and far away residues from solvent exchange. This was not seen in un-glycosylated protein, thereby resulting in its reduced stability. While all the above-mentioned studies provide evidence for increased stability of glycoproteins due to glycosylation, it is not the case always. Glycosylation in MM1 protein resulted in destabilization of the structure. This was because of the loss of long-range interaction due to the formation of short-range interaction with the attached monosaccharide [15]. Im7 has a glycosylation site on a short loop between ordered secondary structure and on a helix. When glycosylated at the compact turn motif, it is seen to positively promote folding. But when the glycosylation site is at the center of a helix, it negatively affects its folding [29]. The core trisaccharide of an N glycan has shown to accelerate the folding by 4-fold and stabilize the beta sandwich structure in adhesion domain of the human immune cell receptor cluster of differentiation (hCD2ad) [103]. In another study exploring the effects of glycan, this time by introducing a glycosylation site in PinWW domain, it was observed that the excluded volume effects due to glycan destabilizes the structured part of the proteins [102]. Folding kinetics can also be hugely affected by the presence of glycans. In the

absence of glycans in erythropoietin (EPO), the intermediate states were seen to have reduced stability. It also changed the rate limiting step of the folding reaction and the folding kinetics were accelerated for both the intermediate and the final native 3D structure. Glycosylation was seen to increase the stability of the intermediate species greatly than the native structure. This suggested that glycosylation does not accelerate the folding by biasing folding pathways, but instead, it slows down the process. The study also provided evidence to show glycosylation destabilizes the unfolded state thus contributing to greater equilibrium stability of EPO [19].

Thermodynamic and kinetic stability is also affected by glycosylation. The increase in the thermodynamic stability also results in an increase in the melting point, and this can be experimentally detected. A computational study on SH3 domain variants showed that increase in thermodynamic stability is the result of destabilization of the unfolded state rather than the stabilization of the folded state. It was also noted that the polysaccharide chains forced the unfolded systems to adopt more extending conformations[14]. But in MM1 the thermodynamic stability was reduced due to conformational distortion induced by trading off long range interaction to form short range interactions with glycans [15].

Presence of glycans can also affect protein aggregation [112]. This is important in the context of PrP misfolding as PrP^{Sc} prefers its aggregated state. The steric effects of the bulky sugar molecules often lead to inhibition of aggregation. The un-glycosylated forms of Peniophora lycii phytase (Phy) showed 200 times faster aggregation compared to the glycosylated form[17]. In the variable domain of HIV1 Envelope protein, glycosylation prevented the formation of secondary structures by

stabilizing pre-existing conformations and provided resistance against thermal unfolding. The simulations from the same study showed that glycan-glycan interactions along with disulfide linkages provided further stability and increased the formation of beta sheets [113]. In a study of dimerization of human carcino embryonic antigen-related cell adhesion, the addition of an N-GlcNAc at potential glycosylation sites inhibits dimer formation[18]. In winged bean agglutin lectins (WBA) it was observed that dimerization was independent of glycosylation, similar to the study on ECorL, as the glycosylation site was located far away from the site involved in dimerization. In acidic WBA (WBAII) glycosylation reduced the structural integrity [114]–[117].

Some of the monosaccharides that make up the glycans, carry an electronegative charge. Sialic acids are one such diverse group of molecules with nine carbon backbone and are usually attached to the end of the glycans. Due to their negative charge and hydrophilicity, sialic acids play a variety of roles in normal and pathological processes[118]–[120]. Specific changes in the oligosaccharide attached to glycoproteins are known to be involved in disease progression and they have been identified as biomarkers for several diseases. There are many studies in recent years attempting to understand the role of glycosylation in cancer, inflammation, and other human diseases. Glycans are also known to play an active role in the immune system. Immunoglobulin G (IgG) is a glycoprotein and glycosylation is important for its binding with the receptors and un-glycosylated IgG antibodies are unable to mediate the inflammatory response. Glycosylation of IgG is extensively studied in rheumatoid arthritis. It was observed that decreased terminal sialylation and galactosylation of

IgG were the common factors in autoimmune disorders [121], [122]. Expression of certain sialic acid rich antigens such as cancer driving hallmarks in advanced stages of cancers [123]. To increase the affinity to the carbohydrate-binding proteins, such as galectins and selectins, these antigens crowd the surface of cancer cells. This interaction of glycans with carbohydrate-binding proteins is a crucial event for neoplastic progression and the formation of cancer metastases [124]. Even though SA is seen to play an important role in many diseases, very little has been explored about the effects of sialic acid on the dynamics of the glycan and its effects on the stability of the glycoprotein. One computational study on biantennary and tri-antennary glycans suggests that SA can affect the flexibility of the glycans and may change the protein function they are attached to [125]. Since sialylation is seen commonly in glycans in prion strains, it is also suspected that it could play a role in animal to human transmission, it is therefore essential to understand the effects of the presence of SA on PrP [126].
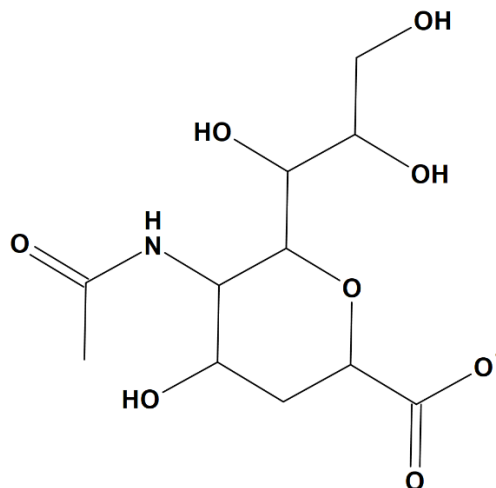
Figure 2-5 Structure of Sialic acid: Chemical structure of the negatively charged N-Acetylneuraminic acid, the most common sialic acid found in humans.

2.3 Possible Effects of Glycosylation on PrP

All of the above-mentioned studies strongly suggest that glycosylation can affect the protein structure by stabilizing it in most cases and destabilizing it in some cases. Both long-range interactions and short-range interactions in the glycoprotein can be influenced by the presence of a glycan. They can also have an impact on the folding mechanisms by altering the stability of the intermediate species and changing the rate limiting steps. Glycans can also impact the aggregation, oligomerization properties of the glycoprotein they are attached to. All these properties are also crucial in the case of PrP misfolding. Prion diseases, which currently have no cure, are purely caused by the misfolding of the PrP glycoprotein and past studies have shown that glycans can have a significant role in the process [7]. Strains of PrP^{Sc} also indicate glycans could have definitive roles since they are known to have a characteristic preference to patterns of glycosylation. As mentioned in the previous sections, the vulnerability of different parts of the brain varies, so does the pattern of glycans observed in different tissues of the brain. The above-mentioned list of studies suggests that it is possible that glycans may affect the process of folding/misfolding by altering the stability, folding pathway, and aggregating tendency of prions proteins. Therefore, understanding the effects of glycans on the structure and the stability of PrP despite the complexity involved becomes crucial. It is extremely important to investigate how the effects of glycans manifest in the case of PrP to get a better understanding of the molecular mechanism of the disease spread.

PrP's N-glycosylation sites, N181 and N197, are present on an alpha helix and on a short loop connecting two helices respectively [7], which is similar to the case in Im7 where glycosylation affected the stability based on the secondary structure of the site of glycosylation[29]. Even though glycosylation is not a requirement for PrP misfolding, PrP^{Sc} is mostly found to be glycosylated. In fact, certain glycoforms are known to convert PrP^{C} to PrP^{Sc} with more ease [72]. This is supported by the experiments finding strain-specific uptake of glycoforms into PrP^{Sc} and by the link between tissue dependent glycosylation patterns in the brain suggesting that the composition plays an important role in the misfolding [71]. The distinctive patterns of phenotypic prions that are propagated allow glycans to be used as strain markers [69], [70], [127]. Studies also suggest that species barrier for prion infection may also depend on the glycoform [128]. All of this motivated us to investigate the effects of the site of glycosylation, size of the glycans and the presence of sialic acids in glycans attached to PrP. We aim to explore these questions through MD simulations of various glycosylated forms of PrP. The techniques and models used are discussed in the next section.

3. Molecular Dynamics Methods

Molecular dynamics (MD) has seen a lot of development in recent years and people rely on MD simulations a lot to help understand and interpret results from experiments [129]. Experimental techniques used to visualize biomolecules can capture static molecules but are limited by their inability to directly visualize the dynamics of the molecules. The functioning of biomolecules depends greatly on the dynamics of these molecules, therefore along with the knowledge of the static structures, understanding the dynamics also becomes crucial [104]. To understand how glycans affect PrP, we are interested in "watching" the dynamics of the single molecules at atomic scales, which is not yet possible through experiments, although it can be done in computational simulations. MD provides a way to study the dynamics by generating time evolution trajectories of these systems. Conformational change, ligand binding, protein folding, the flexibility of different segments of protein, interaction within the molecule and interaction of the molecule with the solvent, are explorable through MD. This is also the reason why MD is being used extensively in drug discovery [129], [130]. In another project that I am actively involved in, which aims to find therapeutic solutions to Covid-19, we used MD extensively to model and identify different conformations of the frameshift stimulatory pseudoknot of the Sars-Cov-2, the coronavirus which caused the COVID-

19 pandemic [131]. This structure could be a fruitful target for therapeutics treating COVID-19.

Molecular dynamics was used for the first time in 1957 to perform simulations of simple gas [132]. Twenty years later, the first molecular dynamics of protein was performed on bovine pancreatic trypsin inhibitor [133]. From then on, the efficiency, modelling capability, and accuracy to replicate real biophysical environments have gone up. MD simulations are also being used to interpret experimental results and design further experiments. X-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy (cryo-EM), and other experimental techniques, within the limitations of their resolution, provide information on the static structures, i.e., relative coordinates of the atomic positions. This information is used as the starting structure in MD simulation. However, these structures can often have atom clashes, missing atoms, missing residues, crystallographic water molecules, and other defects. These problems arise due to the resolution of the experiments and the difficulty in crystalizing large protein structures. All these need to be resolved before setting up the simulation. As a first step of the correction, the missing residues and atoms are added at the right location, unwanted crystallographic water molecules are removed, and the protonation states are assigned. Water molecules and the boundaries of the box in which protein is placed to run the simulations, if used, can be defined at this stage. Next, the steric clashes, bad bond angles and bond lengths are removed through an energy minimization step. By doing this, the system's energy will reach a nearest local minimum, i.e., it reaches a stable conformer that is structurally closest to the starting structure.

Next, the system is gradually heated to bring it to the desired temperature. Then, equilibration is performed in which the force, velocity and atomistic positions of all the atoms are calculated at every step to evolve the system in time. This generates the system trajectory. The accuracy of the description of the forces on the system in an MD simulation is essential to determine how realistic the simulations are.

Molecular force fields models are used to compute forces used in the calculation of the time evolution of the system. There are several models of force fields optimized for different systems, for example, there are protein force fields specific to proteins, nucleic acid force fields to describe nucleic acids, and so on. A protein force field, for instance, would have terms accounting for electrostatic interactions, spring-like terms for covalent bonds, torsional potentials, bond angle and bond lengths, specific to standard amino acids [134]. These force fields along with any additional external forces that are defined on the system are accounted for in the Hamiltonian used to calculate the force on individual entities of the system. Over the years there has been a lot of refinement in the force fields, specifically, parameters defining the side chain interactions have been refined extensively. From a study involving extensive comparison of simulation results with experimental data, it is conclusive that over the years, protein force fields have seen a great improvement in their ability to accurately describe systems in spite of their inherent approximations [135].

The atomic coordinates of the atoms in MD simulations evolve according to the laws of Newtonian Physics. Force F_i, on the ith atom, with mass m_i and acceleration a_i is given by,

$$\vec{F}_i = m_i \vec{a}_i \qquad\qquad 3.1$$

$$\vec{F}_i = -\vec{\nabla}_i (V) \qquad\qquad 3.2$$

If the initial coordinates, initial velocities, and the potential energy of the system are known, the coordinates and velocities for the consecutive time steps can be generated through numerical integration. The interaction potential energy can be written as sum of two terms, a term representing covalent interactions and a term representing non-covalent interactions, as described below,

$$E_{total} = E_{covalent} + E_{non-covalent} \qquad\qquad 3.3$$
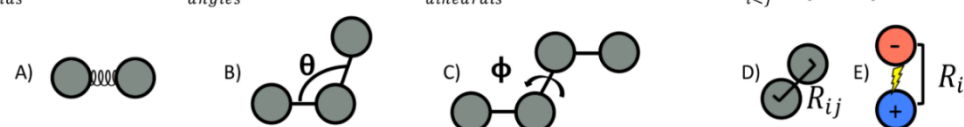
$$E_{tot} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - r\theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

Figure 3-1 A typical force field (FF) model: The total potential energy written as a sum of bonded, and non-bonded energy terms. Terms (A-C) are the "covalent" energy terms they refer to harmonic potentials of bond lengths, bond angles and dihedral angles. Terms (D, E) are the "non-covalent" terms, and they include electrostatic interaction energy and Lennard-Jones potential. The Figure adapted from [129]

Time scales for the events explored through MD like protein folding, ligand binding, and other events, can range from a few picoseconds to microseconds. Events like aggregation can take several days. But the time step used in MD simulations can only be a few femtoseconds, one order of magnitude lesser than the time taken for hydrogen bond vibration (the fastest atomistic movement), to ensure numerical stability in the time evolution integration. Given the power of modern-day computers and GPUs [136], [137], it is still very difficult to run millisecond long simulations within reasonable time limits. Few hundred nanoseconds to a few microseconds of simulation time are still the optimistic time scales to explore through MD simulations, while keeping track of all the atoms in the system, using today's technology [129], [134]. To reduce the running time of the simulation, approximations can be introduced. For example, implicit water models, which replaces water molecules by a continuum medium with dielectric properties of water and coarse-grained models for proteins, where an entire AA is replaced with one, two, four or six beads [138], [139]. Even though both of the aforementioned approximation models produce less accurate systems, they drastically reduce the computational cost by decreasing the number of entities to keep track of. Such approximations can be useful to study mesoscopic behaviour of large molecules, which are currently very difficult to study using the all-atom models.

3.1 Explicit and Implicit solvation models

To create a realistic environment for the biomolecules, they are usually placed in a water filled box with well-defined periodic boundaries. The periodic boundary ensures that the water molecules exiting the box from one side are compensated by incoming water molecules from the opposite side to maintain a constant density. The size of the box is selected such that the least distance between the biomolecules in the system and the sides of the box is above 8-10 Å, the general cut- off distance used to compute electrostatic interactions. For bigger and more dynamic molecules, a box with much a larger dimension is chosen. This would also increase the number of water molecules included in the box many times. Keeping track of all the oxygen and hydrogen atoms of all the water molecules along with the biomolecule and computing their interaction for every timestep is required for the system to progress in time. Therefore, an explicit description of water molecules naturally increases the computational cost. An alternative is an approximation that defines water as a continuum medium with dielectric properties of water [140]. Implicit solvation models have been continuously refined over the years, yet they are far behind in catching up to the accuracy of the explicit solvents [141]. The stochastic nature of explicit water molecules is key to solvent-protein interactions. It is for this reason, we use TIP3P water model [142], a widely used explicit solvent model, to study glycosylated PrP.
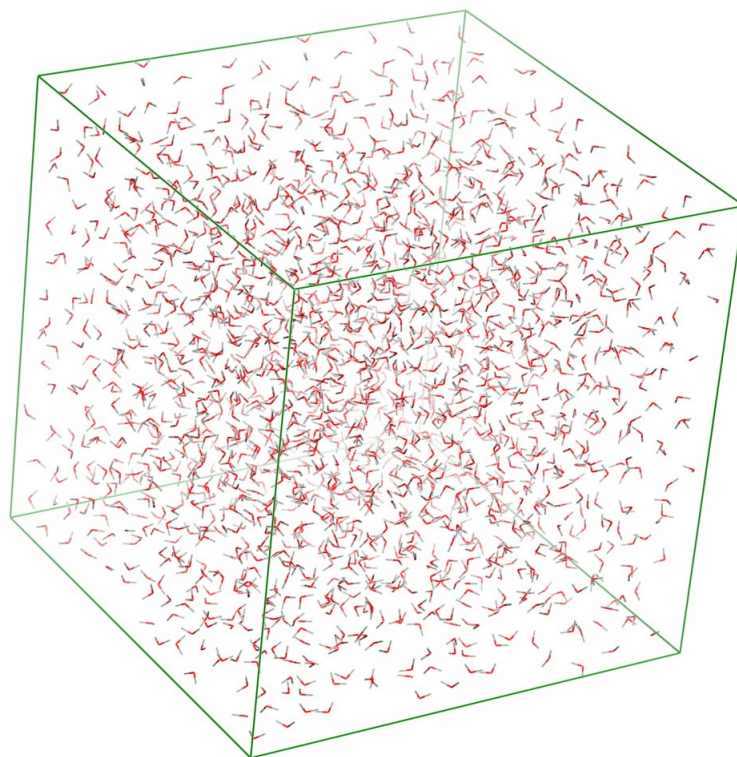
Figure 3-2 Cartoon representation of explicit water molecules: Water molecules of TIP3P water box.

3.2 Representation of the system

The approach in coarse grained models is to represent a system, say amino acids, by pseudo atoms which are usually referred to as 'Beads' [143]. These pseudo atoms are constructed to replace sets of atoms in the system. The number of beads used to represent the AA is chosen depending on the timescale of the events of interest and accuracy requirements [143]. A protein, therefore, would be represented by a chain of connected beads. Representing the AAs through beads can decrease the

number of degrees of freedom. While the atomistic interactions are not captured, coarse grained force fields aim at capturing the mesoscopic behaviour [139], [144] . Folding and structure predictions of large proteins, the interaction between large proteins and protein membrane systems, aggregation, and other events involving large number of atoms or longer time scales, are studied using coarse grained approximations as it can be very expensive in the all-atom representation. It can also be used to run multiple long simulations for an extensive sampling of the phase space within a very reasonable running time. However, even the six-bead coarse grain model has higher accuracy than the single bead, Cα model, but it still fails to capture the side chain dynamics accurately unlike all-atom models.

a

b

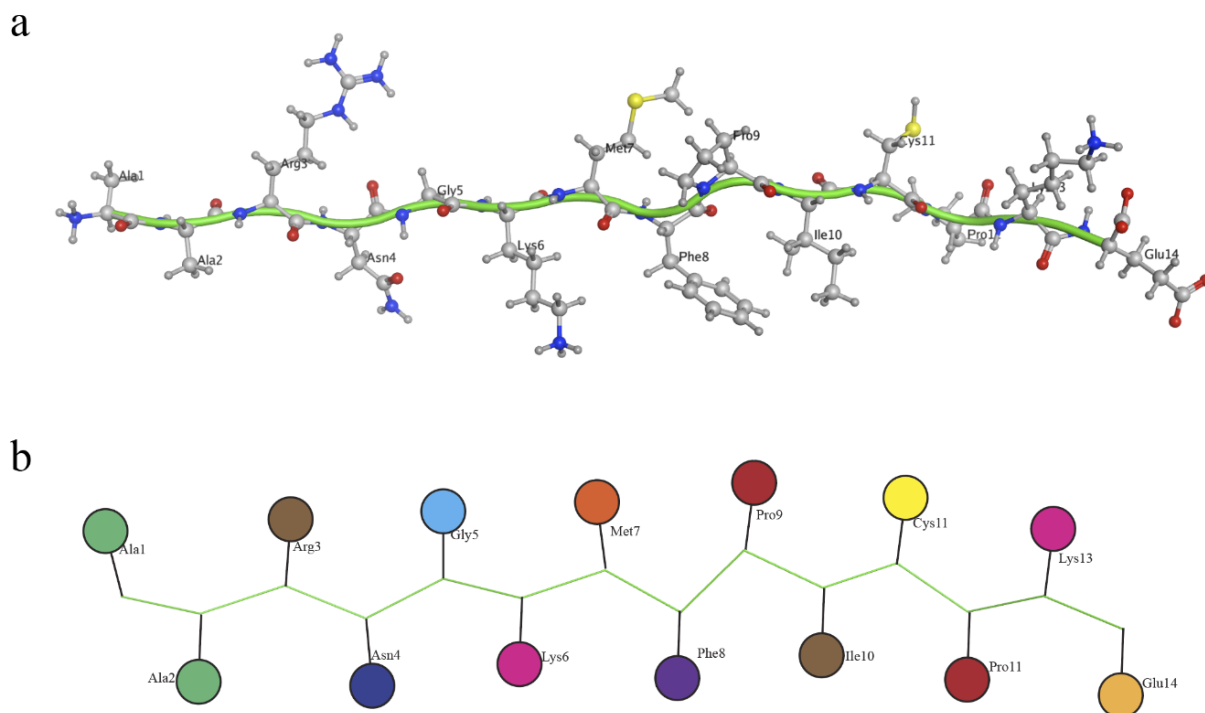**Figure 3-3- All-atom and coarse-grained representation:** A peptide chain represented by a) all-atom model and b) one-bead coarse grained model.

The all-atom representation describes every atom in the system uniquely providing higher accuracy in the description of the systems [145]. Conformational dynamics, interactions between secondary structures, the effect of perturbations, mutations, and ligand binding are captured with great detail through the all-atom description [129], [134]. Apart from the computational cost, another disadvantage in using the all-atom model is that it does not allow for breaking or making of covalent bonds during the running of the simulation. This drawback is present in both all-atom and coarse-grained representation. Quantum mechanical (QM) model is designed to overcome this problem [146]. In the QM model, the system is modelled using quantum mechanical approaches. It is usually employed to study chemical reactions. But due to the extensive calculations, the computations involved need a lot of time which is why they are used to model systems with few 100 atoms. Most often, a hybrid model, QM/MM (quantum mechanics/molecular mechanics) is used to describe a part of the system using QM while the rest of the system is described using the all-atom model [147]. This is because quantum mechanical calculations can increase the computational cost by many times.

Therefore, the choice of the description of the system depends on the system of interest, the properties that are being studied on the system, computational resource availability, i.e., both speed of the processors and the memory availability to store data, the nature of the analysis, and the timeline availability for the study.

In this project, we aim to understand the effects of attaching a glycan to the prion protein by studying the conformational changes, differences in the interaction

within the protein, and other physical properties. The glycans that are used in this study differ in composition and size, ranging from two units of monosaccharides to much larger branched structure that is comparable to the size of the protein itself. This motivated us to employ the all-atom model for our description of the PrP glycoforms. Coarse grained simulations have not been able to describe glycan interactions in PrP accurately [15].

3.3 Computer simulations for higher sampling

Molecular dynamics simulations, despite its convenience of use, suffers from certain disadvantages due to its dependence on computational power. For instance, if the folding time of a protein is estimated to be 500 ns, it is not always likely that it would take the same time or even a microsecond for it to fold. It is possible that the protein gets trapped in a local minimum on its folding pathway and remains there for an extended period of time. Spending a long time in the trap would consume computational time and resources while not exploring much of the phase-space. The difficulty in overcoming a huge free energy barrier for folding can pose challenges to the MD simulation. To overcome this problem and to explore more of the phase-space, other computational simulation methods like Monte Carlo (MC), Replica exchange molecular dynamics (REMD), accelerated molecular dynamics (aMD) were developed. Historically MC simulations were performed before MD. In MC, a random gaussian step is added to the positions of the atom and algorithms exist to keep or reject the new step based on the energy of the new system [148], [149].

Systems with lesser degrees of freedom can benefit from MC. REMD is aimed at exploring larger configurational space in lesser computational times [150], [151]. Multiple MD simulations are performed at different temperatures simultaneously, and a pair-wise exchange is performed at regular intervals depending on temperature and energy differences between selected simulations. By allowing the systems of similar potential energies to sample conformations at different temperatures, systems can overcome energy barriers on the potential energy surface thereby producing a large ensemble of states [152]. The exchange attempts are accepted with a Metropolis probability,

$$P_{REMD}(i \leftrightarrow j) = \min \{1, \exp [(\{\beta_j\} - \{\beta_i\})(U_j - U_i)]\} \qquad 3.4$$

where, U_i is the potential energy of ith replica, $\beta_i = 1/(k_B T_i)$, k_B is Boltzmann constant and T_i is the temperature of the ith replica.

In accelerated MD (aMD), a non-negative boost potential, $\Delta V(r)$, is added to the system when the true potential energy of the system, $V(r)$, is below a chosen threshold, E, to decrease the energy barrier [153], [154]. The aMD simulation is performed using a modified potential, $V^*(r)$. The boost to the potential, $\Delta V(r)$, is computed using E, $V(r)$ and α, where α is the tuning parameter.

$$V^* = V(r), \qquad V(r) \geq E, \qquad 3.5$$

$$V^* = V(r) + \Delta V(r), \qquad V(r) < E, \qquad\qquad 3.6$$

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + E - V(r)} \qquad\qquad 3.7$$

Values of E and α are determined by running a short conventional MD. By decreasing the energy barrier, a system in a local minimum can overcome the barrier quicker and progress to the global minimum. Accelerated MD is still under development to work along with the currently existing force fields. All these methods enable better sampling but are not aimed at capturing the time evolution of the system. Therefore, we used conventional MD in this thesis.
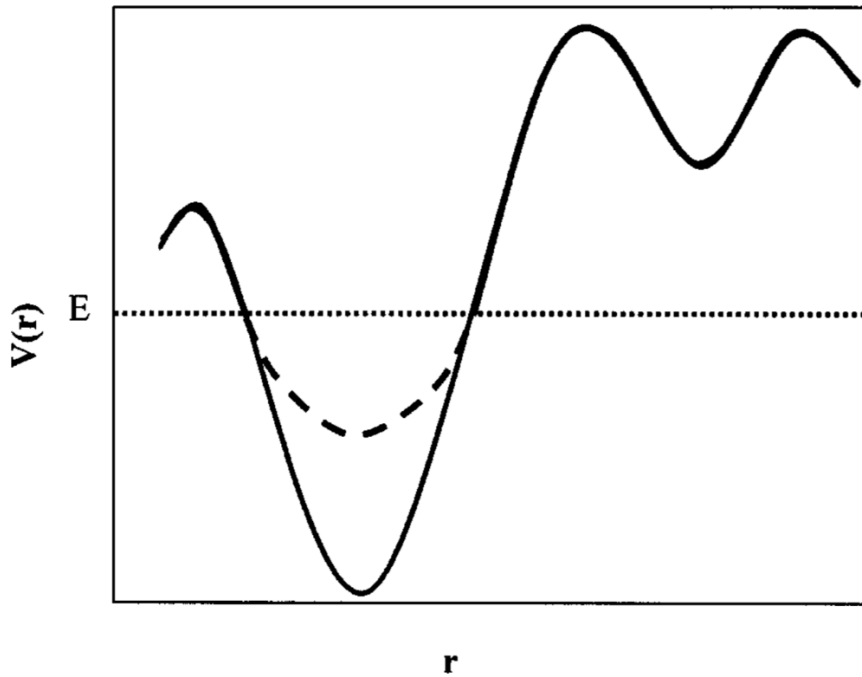

Figure 3-4 Biased potential in Accelerated MD: Schematic representation of normal potential (thick line) and a possible biased potential (dashed line). Image adapted from [154].

3.4 Simulations of glycosylated PrP

To study the effects of glycosylation on PrP, we chose the Hamster PrP molecule as the rates of susceptibility to prion disease is high in Hamster. The NMR structure of the protein is available at the Protein Data Bank under the ID "1b10" [54]. To study the effects of the size, the site of glycosylation and the composition of glycans, three different glycans with different structural complexity and chemical composition were chosen. First, a diacetylchitobiose, a very short glycan with two GlcNAc units and very little structural complexity. Next, a complex biantennary glycan, a relatively bulkier glycan with two branches and more subunits. Lastly, a complex biantennary glycan with sialic acid groups was chosen. These glycans were then attached, as described below, to a model of Hamster PrP (PDB 1b10 [54]) and micro-second long all-atom simulations were performed in explicit water for various glycoforms. Site specific effects were tested through simulations of singly glycosylated PrP with glycan at either N181 or N197. Dynamics of doubly glycosylated PrP was also studied and compared against singly glycosylated PrP. The difference in the size of the glycans allows us to study the effect of size at each site. Addition of sialic acid adds negatively charged units to the glycan chain, which allows testing the effects of presence SA of PrP.
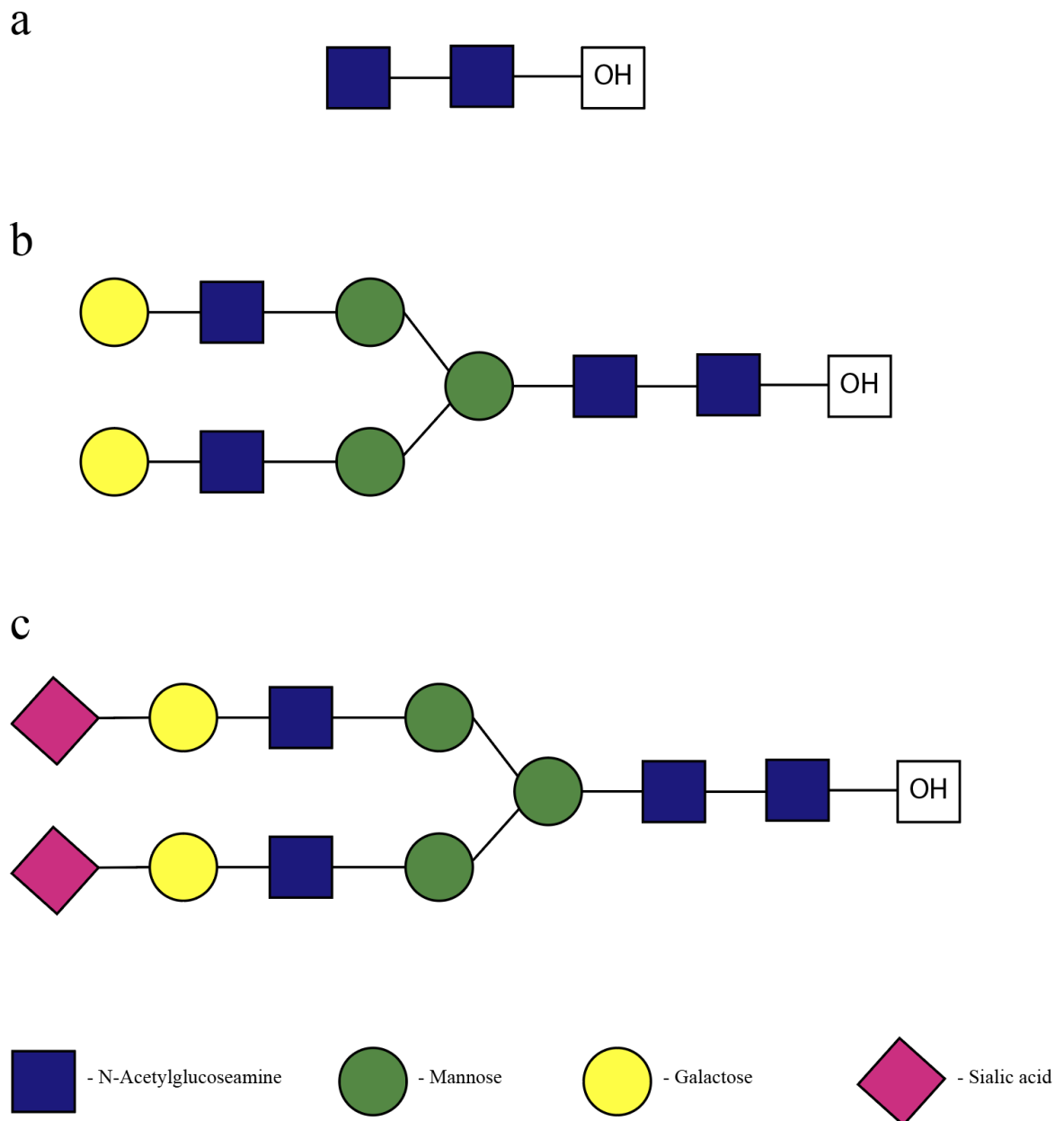
a

b

c

- N-Acetylglucoseamine   - Mannose   - Galactose   - Sialic acid

Figure 3-5 Block diagram of glycans used in the simulation: a) Diacetylchitobiose

b) A complex biantennary glycan c) A complex biantennary glycan with sialic acid.

All the simulations were performed using AMBER 18 molecular dynamics software and Compute Canada systems. Additional software used to study the structure and analyze the simulations are MOE (Molecular operating environment), and Python-3.7.

3.4.1 Preparation of the starting structure for the simulations

The PDB 1b10 has 25 submitted conformers [54]. When aligned and overlapped, these conformations appear nearly identical, and differences arise only in the flexible regions. Examining the structures in MOE for overlaps, it was noted that the difference in the root mean square deviation (RMSD) in the coordinates of not just the backbone atoms, but all atoms in the protein was less than 1.56 Å which clearly indicates great overlap among the structures for a 104 AA long protein. From the figure below, it can be noted that the difference between the conformers can be seen in the loop connecting helix 3 and sheet 2, in the first two residues in the N-terminus, and last two residues in the C-terminus of the protein. Since these regions are mostly unstructured, they are expected to be highly dynamic in the MD simulations compared to other regions of the protein. Therefore, any of the 25 structures would be a good starting point for the MD simulations. This is because the above-mentioned differences between the structures would have very little or no impact after the minimization and the heating steps of the MD. So, the first structure in the PDB file was selected. Using MOE, fixing missing hydrogen atoms, correcting protonation states of the AAs to pH 7, and other preliminary corrections were applied before it was used in the MD simulations.
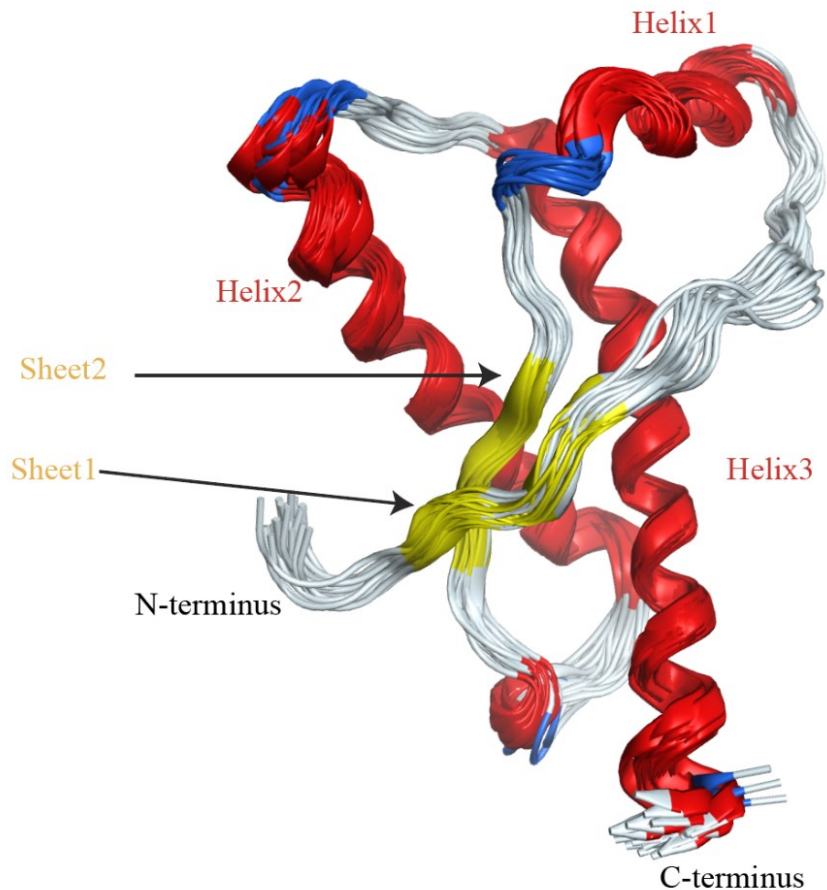
Figure 3-6 Hamster PrP 3D structure under the PDB ID-1b10: Overlap of all the 25 structures in the PDB entry.

The "LEaP" module of AMBER is used to read the PDB file of the starting structure, perform any modification to the starting structure, load force fields and libraries necessary for the simulation, and generate parameter files used in setting up the simulations in AMBER. The name "LEaP" is coined because it is used to link, edit and add derive parameters for the MD. AMBER offers a command line version and a graphical version of the LEaP module called "tleap" and "xleap" respectively.

Since Compute Canada systems were used to perform simulations, tleap was used extensively in the project.

A PDB file can be edited and modified by using any text editor. AMBER has very specific requirements for the input PDB files. The leap module recognizes AAs using three letter codes assigned to different AA within the PDB file. For every three-letter code, there is a corresponding template of atoms in the AMBER force field libraries. If there are missing atoms in the PDB across a three-letter code, AMBER will fill in the missing atoms using its libraries. Also, the three-letter code for AAs with any modifications is different from their natural state. For example, a cysteine (Cys) in its unmodified state is recognized using the code CYS and a cysteine involved in a disulfide bond is recognized by CYX, which has a different number of hydrogen atoms in its template. This is not true for other PDB readers and visualizers. Most PDB visualizers use the term "CONECT" followed by the atom numbers to read such linkages. If a PDB is not edited with the right codes according to the AMBER requirements, AMBER will not create the necessary linkage. Instead, it will fill in the missing atoms according to the template in the AMBER libraries which can lead to a system that is different from what is intended to be created. While the command "PDB4amber" in the LEaP module can be helpful in converting a PDB file to AMBER's readable form, we found editing them manually provided more control, especially when building the glycoforms.

3.4.2 Building glycoforms of PrP

The glycoforms of PrP were built manually. Firstly, the PDB file of the selected Hamster PrP structure was edited such that the protein was capped at both the N terminus and C terminus to neutralize the ends. Next, as mentioned earlier, the three-letter code for Cys residues involved in disulfide bridges were changed from "CYS" was changed to "CYX" and all the hydrogen atoms were removed from the residue. The appropriate number of hydrogens will be added by the leap module as it checks for missing atoms with the library and fills them automatically while reading the file. Along with the modification to Cys residues forming disulfide linkages, the Asn residues involved in the N-link was modified from "ASN" to "NLN" and hydrogen atoms were removed from the modified residue. In the case of doubly glycosylated variants, both N181 and N197 were modified appropriately. At the end of the capped protein chain in PDB, a "TER" card was used to indicate the end of the chain. Following that, the three-letter codes of the monosaccharides and atoms in the unit with their coordinates were listed one after the other with a "TER" card separating each monosaccharide. The AMBER specific glycan codes for various monosaccharides are available in the GLYCAM force fields libraries.

This edited PDB file was loaded into tleap, the command line version of the LEAP module, after importing the "leaprc.protein.ff14SB" and "leaprc.GLYCAM_06j-1" force fields, along with the TIP3P water box [155], [156]. These force fields are currently the recommended force fields by AMBER for glycoprotein systems. The bonds between the cysteines that form the di-sulfide bridge, bonds between asparagine and the first monosaccharide of the glycan chain,

and the bonds between the different monosaccharides were specified using the commands in tleap. A water filled periodic box with a buffer distance of 14 Å was constructed around the glycosylated protein, and Na+ ions were added to the system only to neutralize the residual charge. The parameter files necessary for the MD simulations were generated along with a PDB file of the starting structure. These steps were repeated for the three glycans, diacetylchitobiose, complex biantennary glycan and complex biantennary glycan with sialic acid. For each glycan, three glycoforms were prepared (mono glycosylation with glycan at N181, mono glycosylation with glycan at N197 and homo-diglycosylated variant with glycan at both N181 and N197).

AMBER also provides an online glycoprotein builder at www.glycam.org. The PDB file of a protein can be uploaded to the website, and the glycoform can be constructed using an interactive oligosaccharide builder or from a preexisting library of oligosaccharides or using a condensed GLYCAM notation recognized by the website. The server also provides the option to solvate the molecule and download the necessary parameter files. However, the website is not completely developed, and we noticed few limitations to using the web server. Firstly, the option to choose the desired force fields to generate the simulation parameters is absent. Next, even though it recognizes the site of glycosylation, changes to three letter code for the modification at the site of glycosylation is absent. Therefore, these modifications had to be performed manually after downloading the structure. The glycam server was useful in getting the relative spatial coordinates of the atoms in the glycan. It is for this reason, we used tleap to build the glycoprotein entirely while referring to the

structures made by the glycam server only for the rough initial coordinates of the atoms.
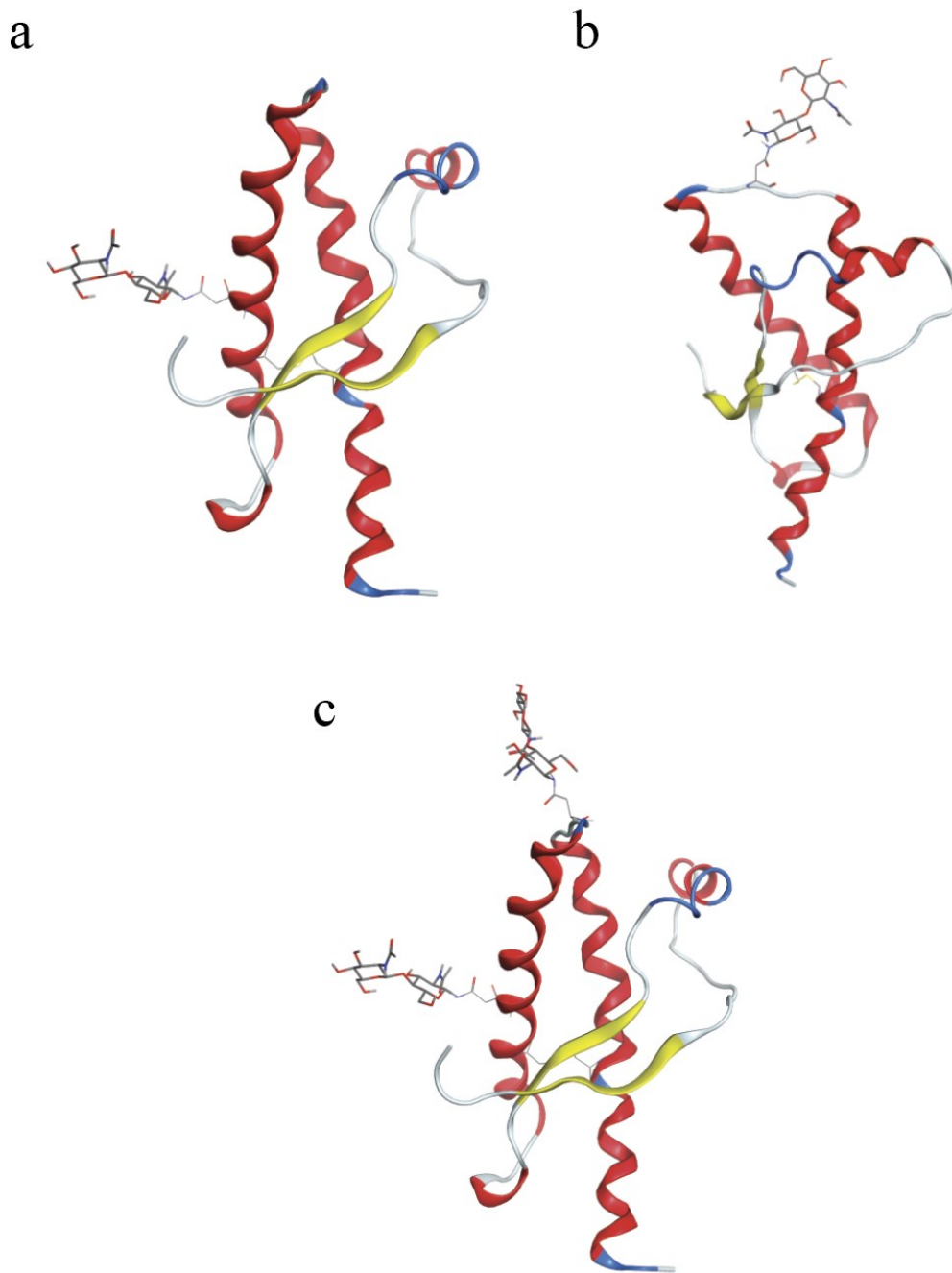
Figure 3-7 Simulation starting structures of PrP glycoforms with diacetylchitobiose: a) mono glycosylated PrP with glycan at N181 b) mono glycosylated PrP with glycan at N197 c) diglycosylated PrP.

Figure 3-8 Simulation starting structures of PrP glycoforms with complex biantennary glycan: a) Mono glycosylated PrP with glycan at N181 b) mono glycosylated PrP with glycan at N197 c) diglycosylated PrP.

Figure 3-9 Simulation starting structures of PrP glycoforms with complex biantennary glycan with sialic acid: a) Mono glycosylated PrP with glycan at N181 b) mono glycosylated PrP with glycan at N197 c) diglycosylated PrP.
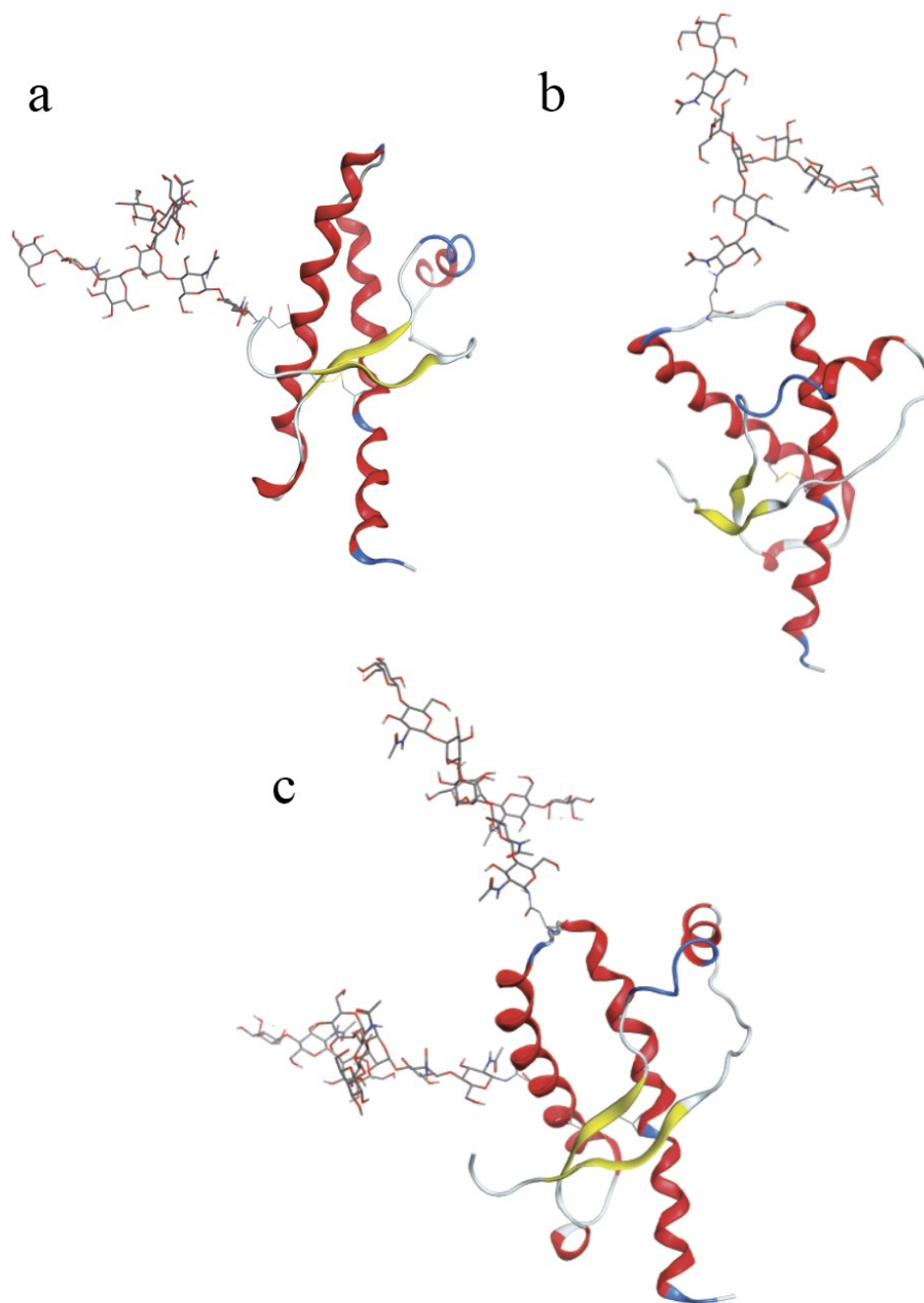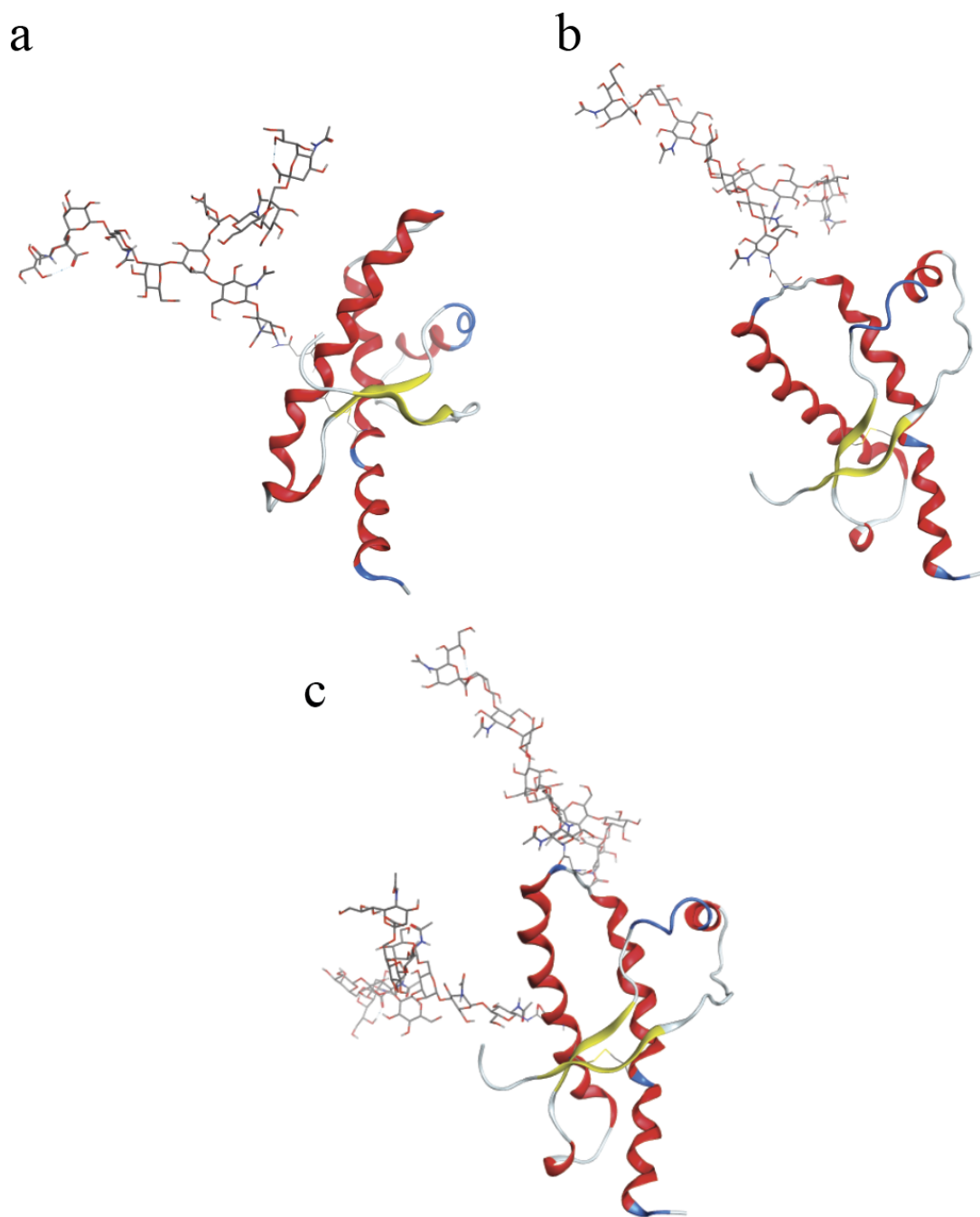
From the construction above it is clear that the size of the larger glycans used is comparable to the size of the protein. The solvated systems were subjected to energy minimization and then heated to 300 K gradually after which the systems were simulated for a total length of 1 microsecond at constant pressure. The different glycosylation variants used, and the number of replicas of each variant are listed in the table below. Simulations of the un-glycosylated variant were also performed to be used as the control system to the calculations. The simulations used GPUs available in the Compute Canada resources.

Analysis of the simulations was performed using CPPTRAJ and PYTRAJ tools within the AmberTools package [157]. CPPTRAJ is the main program in AMBER for processing trajectory files. Using commands specific to CPPTRAJ, various useful quantities can be extracted from the trajectory files. The list of actions commands are available in the CPPTRAJ manual (https://amber-md.github.io/cpptraj/CPPTRAJ.xhtml). PYTRAJ is a python package binding to the CPPTRAJ. It allows the user to utilize the functions of CPPTRAJ within a python ecosystem, thereby increasing the flexibility of data handling. Documentation for PYTRAJ is available currently at https://amber-md.github.io/pytraj/latest/index.html.

Table 3-1 Table of simulation stats: List of various glycoforms built, the number of independent simulations performed and their simulation time.

Glycan attached	Position of glycosylation	Total simulation time	Number of simulations
Unglycosylated	-	1 µs	3
Diacetylchitobiose	181	1 µs	3
Diacetylchitobiose	197	1 µs	3
Diacetylchitobiose	181 and 197	1 µs	3
Complex biantennary	181	1 µs	3
Complex biantennary	197	1 µs	3
Complex biantennary	181 and 197	1 µs	3
Complex biantennary with Sialic acid	181	1 µs	3
Complex biantennary with Sialic acid	197	1 µs	3
Complex biantennary with Sialic acid	181 and 197	1 µs	3

The various glycoforms built, the number of independent MD simulations performed, and the time of simulations are summarized in Table 3-1. The frames from the last 500 ns of the simulations were grouped into multiple clusters based on their structural similarity, which was quantified by the RMSD between the frames. This was repeated for all the simulations performed. The central structure in each cluster, called the "cluster center", is the representative structure for that cluster. The cluster centers of the top three clusters (three most populated clusters) of all the simulations were visualized and studied using MOE. Physical quantities extracted from the simulations are discussed below.

3.4.3 RMSD

The RMSD (root mean square deviations) of the Cα (Figure 1-1) in the protein backbone from the simulations were monitored to check for significant conformational changes. The value is computed by measuring deviations in atomic positions of selected atoms from each frame with respect to a reference frame. The reference frame used in this study was the NMR structure of Hamster PrP. RMSD of the protein structure in a frame with respect to the reference frame is defined as,

$$RMSD = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} \delta_i^2\right)} \qquad 3.8$$

where δ_i, is the distance between atom i^{th} atom in that frame and the reference frame. Usually, all the structures from the frames used in the calculations are

superimposed over the reference frame before computing RMSD to avoid deviations due to rotation and translation of the molecules. That way, large deviations in RMSD would indicate structural changes only. A plot of RMSD as a function of time (or frame number) would help understand the extent of changes to the structure through the course of the simulation.
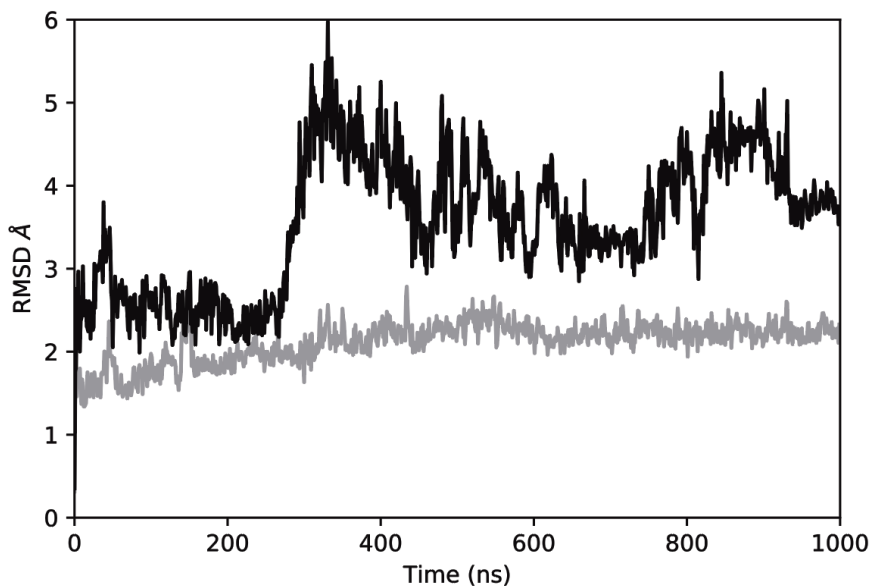
Figure 3-10 Example RMSD plot: Example RMSD plots of two systems, a very stable system with minimal structural changes (grey curve) and a system that's undergoing a lot of structural changes (black curve).

We used RMSD plots to recognize the stable segments of the simulations to carry out further calculations.

3.4.4 RMSF

The RMSF (root-mean-square-fluctuations) of the α carbons of all the residues were computed for a stable 100 ns towards the end of all the simulations and the

average RMSF for every glycoform was computed and plotted. RMSF is calculated by measuring fluctuations in the selected atoms across chosen frames. By choosing 100ns, towards the end of the simulation, in which no drastic conformational change occurred, we could understand the stability of different portions of the protein. RMSF plots were compared between different glycoforms to identify average effects due to glycan on the atomic fluctuations of the protein. The fluctuations in the un-glycosylated PrP was used to monitor any glycan induced effects. RMSF of jth atom is computed by,

$$RMSF_j = \sqrt{\frac{1}{n}(\sum_{i=1}^{n}(x_{ij}-< x_j >)^2)} \qquad 3.9$$

where, x_{ij} is the coordinate of jth atom in the ith frame of a simulation containing n frames.

3.4.5 Electrostatic interaction energy

Electrostatic interaction energy between the glycans and all the residues of the protein was calculated for the top three cluster centers and averaged over the three simulations for each glycoform. This was used to study the average glycan-protein interactions in the converged structures.

$$E_{elec} = k\frac{q_i q_j}{r_{ij}}(1 - \frac{r_{ij}^2}{r_{cut}^2})^2 \qquad 3.10$$

4. Results

From analyzing various frames from different parts of the simulation, it was observed that the protein backbone structure of the glycosylated variants did not change drastically with respect to the unglycosylated PrP. In a few simulations, changes were observed in helix 1 and helix 2. Secondary structure of these sites was affected. Most importantly, there were multiple conformations that each glycan could take on. This was understood from noticing different simulations of the same glycoform resulting in different stable converged conformations of glycan.

Unglycosylated hamster PrP did not undergo any structural changes, as expected. The top clusters from the three independent simulations of unglycosylated PrP, when aligned, show very good overlap (Figure 4-1). All of the cluster centers are similar to the conformations found in the NMR structure.
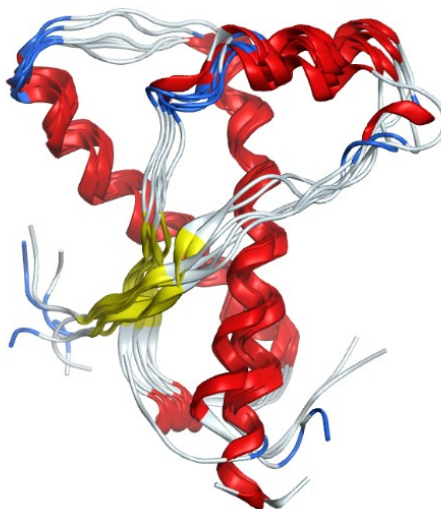

Figure 4-1 Cluster centers of unglycosylated PrP: Top 2 cluster centers of unglycosylated PrP from 3 independent simulations.

Glycoforms with Diacetylchitobiose:

Some cluster centers with different orientations of the glycan from all the simulations involving diacetylchitobiose are shown in figures below (Figure 4-2, Figure 4-5). Diacetylchitobiose being a small glycan, there are fewer sites that can form hydrogen bonds with the protein. From analyzing the cluster centers from the simulations, it could be noted that there is a large variation in the orientations of the glycan between clusters. This is true for all the three glycoforms involving diacetylchitobiose. The GlcNAc units in the glycan were observed to form hydrogen bonds mostly with immediate neighboring residues.

The sites of glycosylation limit the residues that can interact with the small sized glycan. When diacetylchitobiose is attached to N181, a site on helix 2 that faces away from the protein core. It is seen to predominantly interact with the only one or two residues at the local site of glycosylation, and with residues on helix 3. This can be seen in the plot of residue-wise average electrostatic interaction energy with the glycan (Figure 4-16). N197 is on a flexible loop, pointing away from the protein, and a similar behavior is seen despite the flexibility of the site. Diacetylchitobiose at N197 is seen to interact with the residues around the site of glycosylation and with helix 1. The hydrogen bonds formed by glycans at N197 reduce the atomic fluctuations of the flexible loop which in turn reduces the fluctuations of N-terminal region (Figure 4-3). Glycan at N181 does not seem to reduce the overall atomic fluctuations of the protein.

Of the three simulations of mono glycosylated PrP attached to diacetylchitobiose at N197, the architecture of the protein did not change in two

simulations, but the backbone of the helix 1 was noticed to be altered in the third simulation. Helix 1 is observed to be disrupted due to formation of hydrogen bonds between the glycan and Glu152. This modified architecture is found in all the top clusters of the third simulation suggesting that the glycan interaction to distort the protein at helix 1 is strong. In the doubly glycosylated variant, the glycans interact with protein independently, with very little interaction between the small sized glycan (image (k) of Figure 4-5). The interaction energy profile of the diacetylchitobiose at N181 and N197 in mono glycosylated PrP is very similar to that of diglycosylated variants (Figure 4-4).
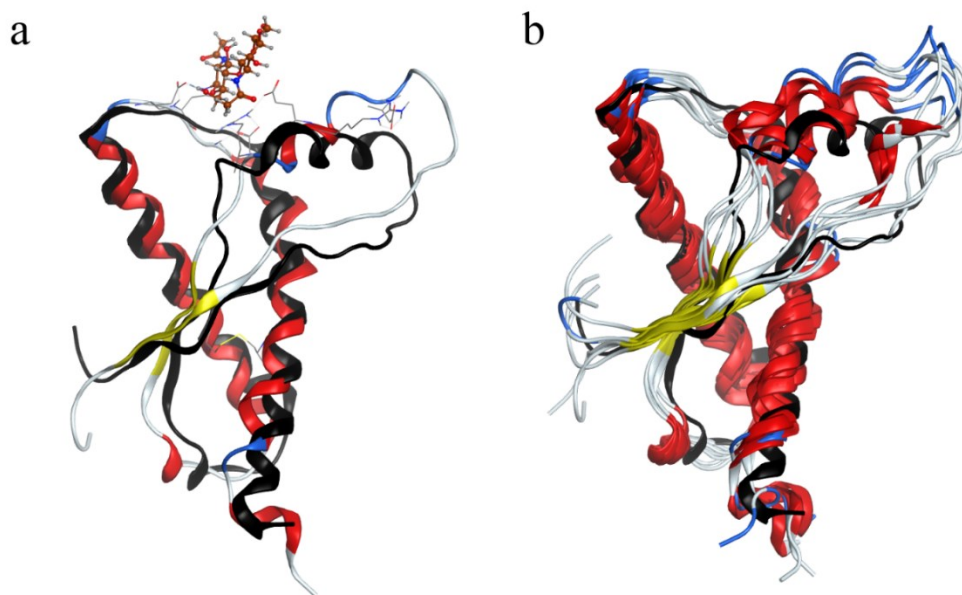

Figure 4-2 Distortion of helix 1 by diacetylchitobiose: a) Distortion of helix 1 due to glycan interaction seen in one of the three simulations of mono glycosylated PrP with diacetylchitobiose at N197. B) Overlap of the backbone of cluster centers of top 6 clusters from the same simulation. NMR structure of PrP is shown in black
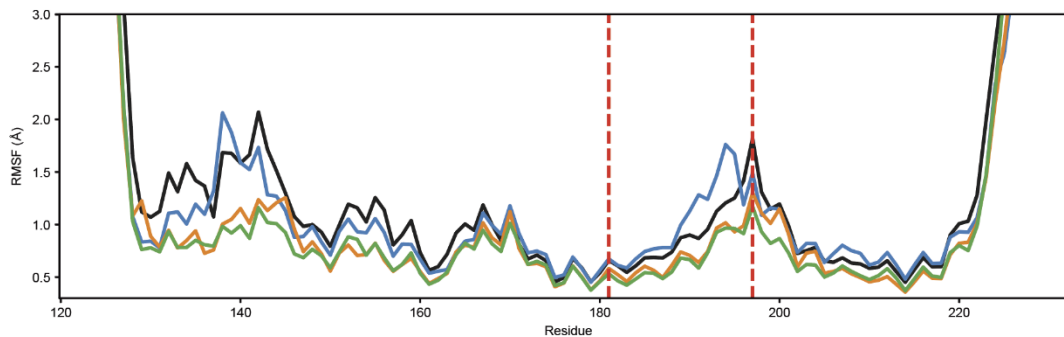
Figure 4-3 RMSF plots from simulations of PrP modified through addition of diacetylchitobiose: Average RMSF from simulations of mono glycosylated PrP with glycan at N181 represented by the blue curve, average RMSF from simulations of mono glycosylated PrP with glycan at N197 represented by the orange curve, and average RMSF from simulations of doubly glycosylated PrP represented by the green curve. Red dashed lines represent the site of glycosylation.
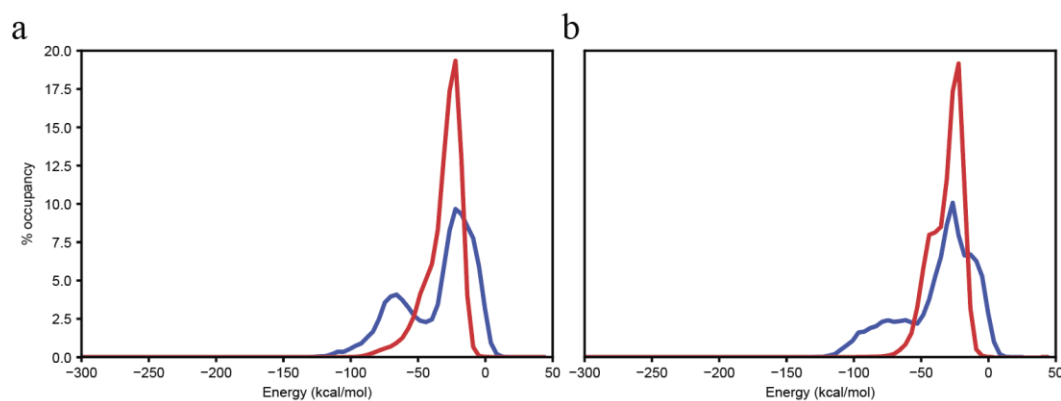
Figure 4-4 Interaction Energy Profile of diacetylchitobiose: Profile of electrostatic interaction energy between glycan and protein from simulations of a) mono glycosylated b) double glycosylated PrP. The curve in red represents glycan attached at N197 and the curve in blue represents glycan attached at N181.
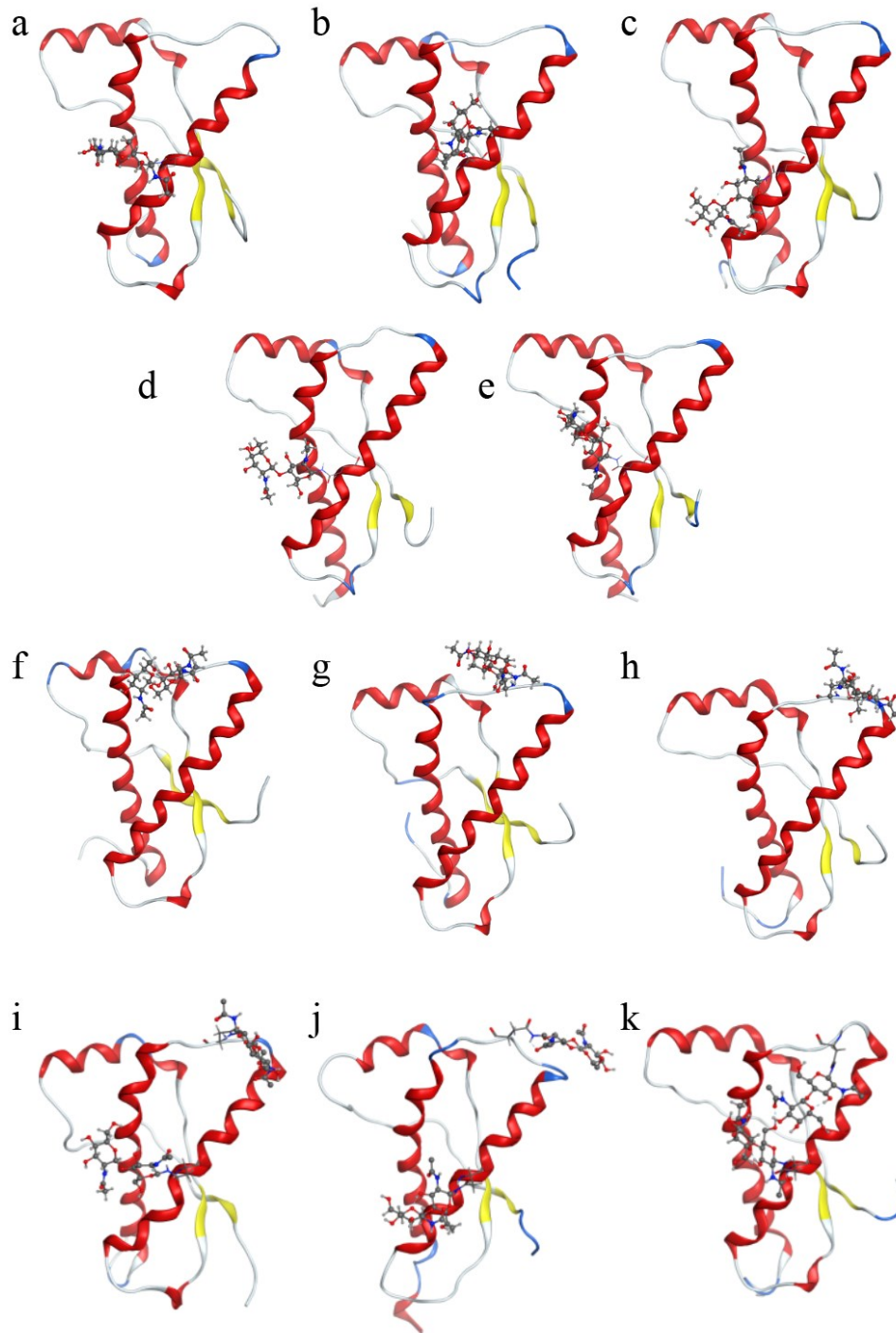
Figure 4-5 Cluster centers from the simulations of PrP attached to diacetylchitobiose: Cluster center apart from the ones in Figure 4-2 are shown here. (a-e) mono glycosylated PrP with glycan at N181, (f-h) mono glycosylated PrP with glycan at N197, (i-k) diglycosylated PrP with glycan at N181 and N197.

Glycoforms with Complex biantennary with no Sialic Acid:

Complex biantennary glycan with no sialic acid, being a much bigger and bulkier glycan compared to diacetylchitobiose, it is able to form a greater number of hydrogen bonds with the protein. The architecture of the protein did not undergo any change upon addition of the glycan, but some details of the secondary structure within the protein were affected in a few simulations. Some cluster centers with different orientations of glycan are shown in Figure 4-10 along with Figure 4-6 and Figure 4-7.

When linked to N181, the glycan can be seen to interact with local residues at the site of glycosylation. The branches of the glycan can also form hydrogen bonds with the helix 3 or with sheet 1 and sheet 2 depending on how the glycan converges. Since these sites are already stable, there is a very little noticeable decrease in the atomic fluctuations. However, in one of the three simulations of mono glycosylated PrP with complex biantennary glycan at N197, the region of helix 2 towards the N-terminus from the site of glycosylation dissociates to form a loop. Cluster centers of the top three clusters belonging to the above-mentioned simulation are shown below (Figure 4-6).
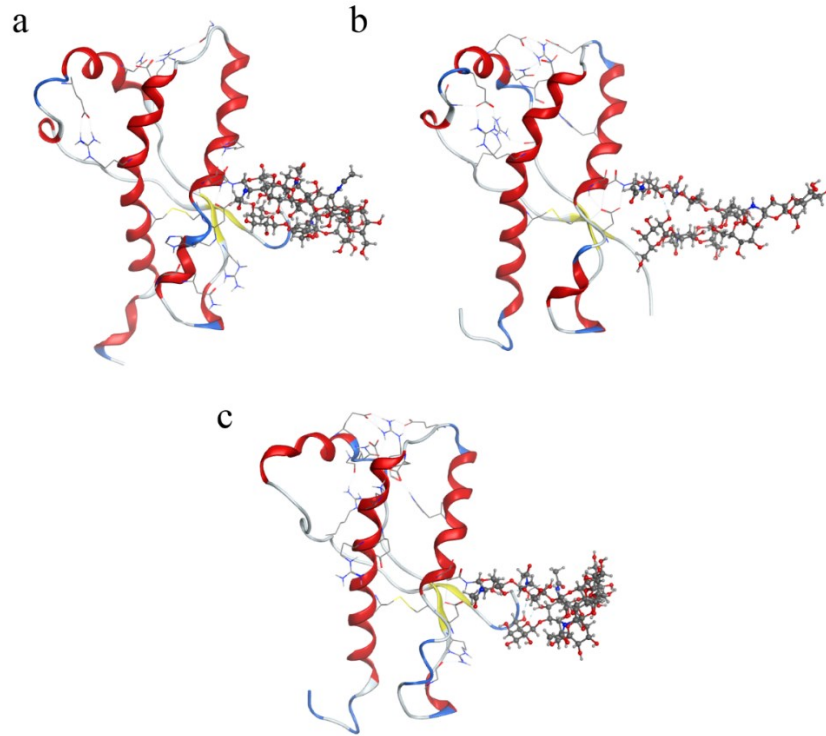
Figure 4-6 Helix 2 dissociated by complex biantennary glycan: Cluster centers of the top 3 clusters from the simulation of mono glycosylated PrP with glycan at N181 that showed dissociation of helix 2.

When linked to N197, the complex biantennary glycan is able to reduce the fluctuations of the protein drastically (Figure 4-8). This is because stabilizing hydrogen bonds are formed with the flexible regions. While the average atomistic fluctuations decrease, the secondary structure was affected in one simulation turning the beta sheet into loops. This is possibly due to the orientation of the glycan in these structures. One branch of the glycan forms hydrogen bonds with helix 1, while the rest of the glycan is seen to be stretching away from the protein. This pull on the helix 1 by the stretching of the glycan could be the reason for the dissociation of the beta sheet.
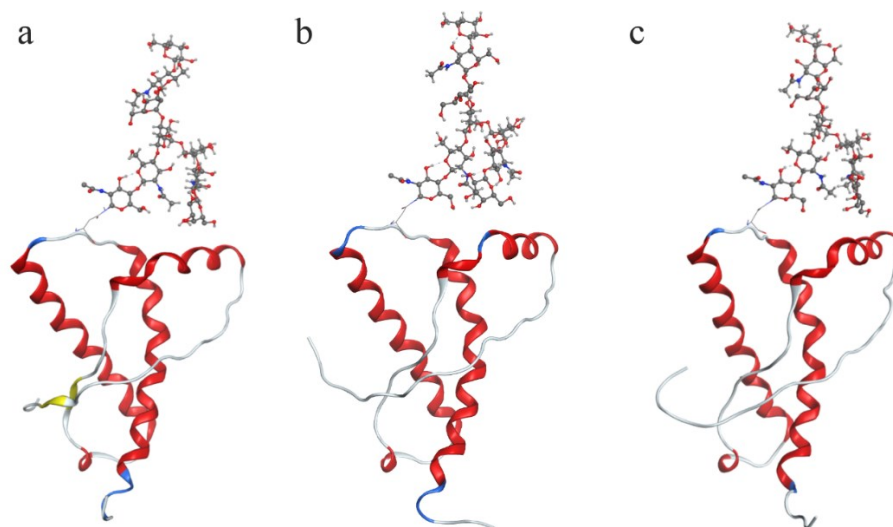
Figure 4-7 Reduction in beta sheet content due to complex biantennary glycan: Central structures of the top three clusters from a simulation of mono glycosylated PrP at N197 that showed a reduction in beta sheet structures.

The doubly glycosylated variant did not show any change in the structure of the protein. When doubly glycosylated, the glycans could most often interact with each other. This is due to the bulkiness of the glycans, which allowed them to stretch to form stable interactions. In some low populated clusters, they were also seen to converge with no hydrogen bonds between the glycans, converging into structures similar to the mono glycosylated variants. The strong interactions between the glycans changed the interaction energy profile of glycans when compared to the mono glycosylation variants (Figure 4-9). The modified pattern suggests that glycan-protein interaction for glycans at each site decreases in the presence of the doubly glycosylated state. The decrease in atomic fluctuations of doubly glycosylated variants is similar to mono glycosylation at N197 (Figure 4-8).
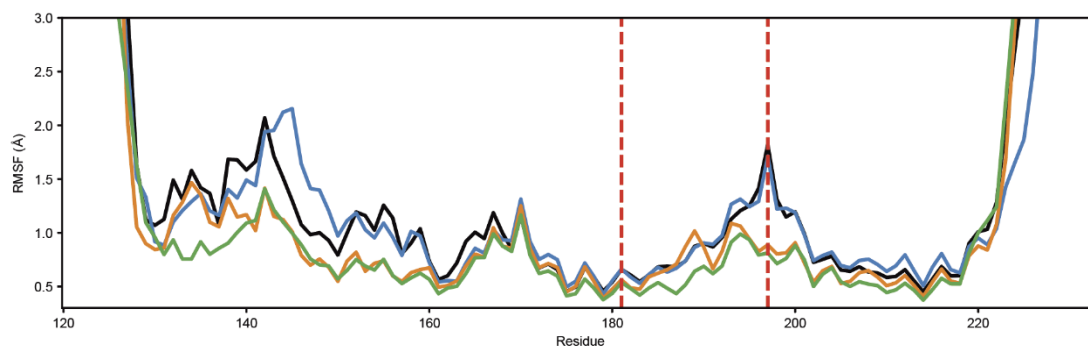
Figure 4-8 RMSF plots from simulations of PrP glycosylated with the addition of complex biantennary glycan: Average RMSF from simulations of mono glycosylated PrP with glycan at N181 represented by the blue curve, average RMSF from simulations of mono glycosylated PrP with glycan at N197 represented by the orange curve, and average RMSF from simulations of doubly glycosylated PrP represented by the green curve. Red dashed lines represent the site of glycosylation.
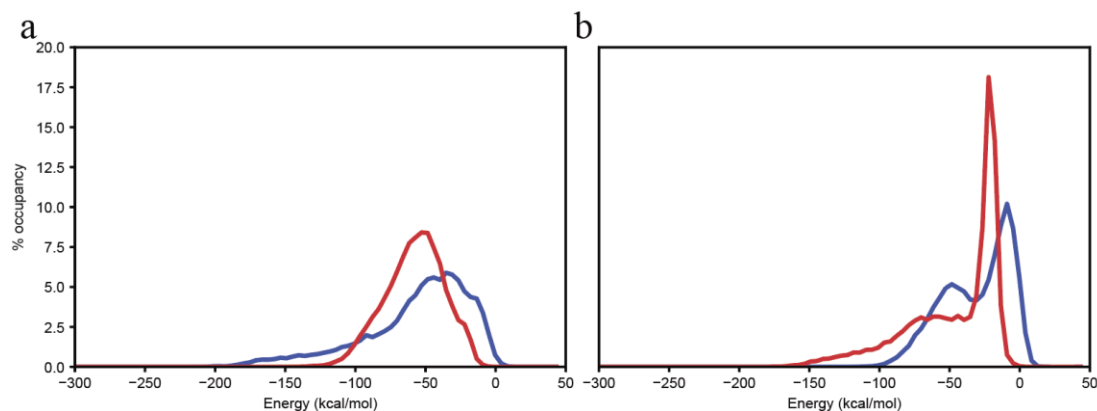
Figure 4-9 Interaction Energy Profile of complex biantennary glycan: Profile of electrostatic interaction energy between glycan and protein from simulations of a) mono glycosylated b) di-glycosylated PrP. The curve in red represents the energy profile of glycan attached at N197 and the blue curve represents the energy profile of glycan attached at N181.
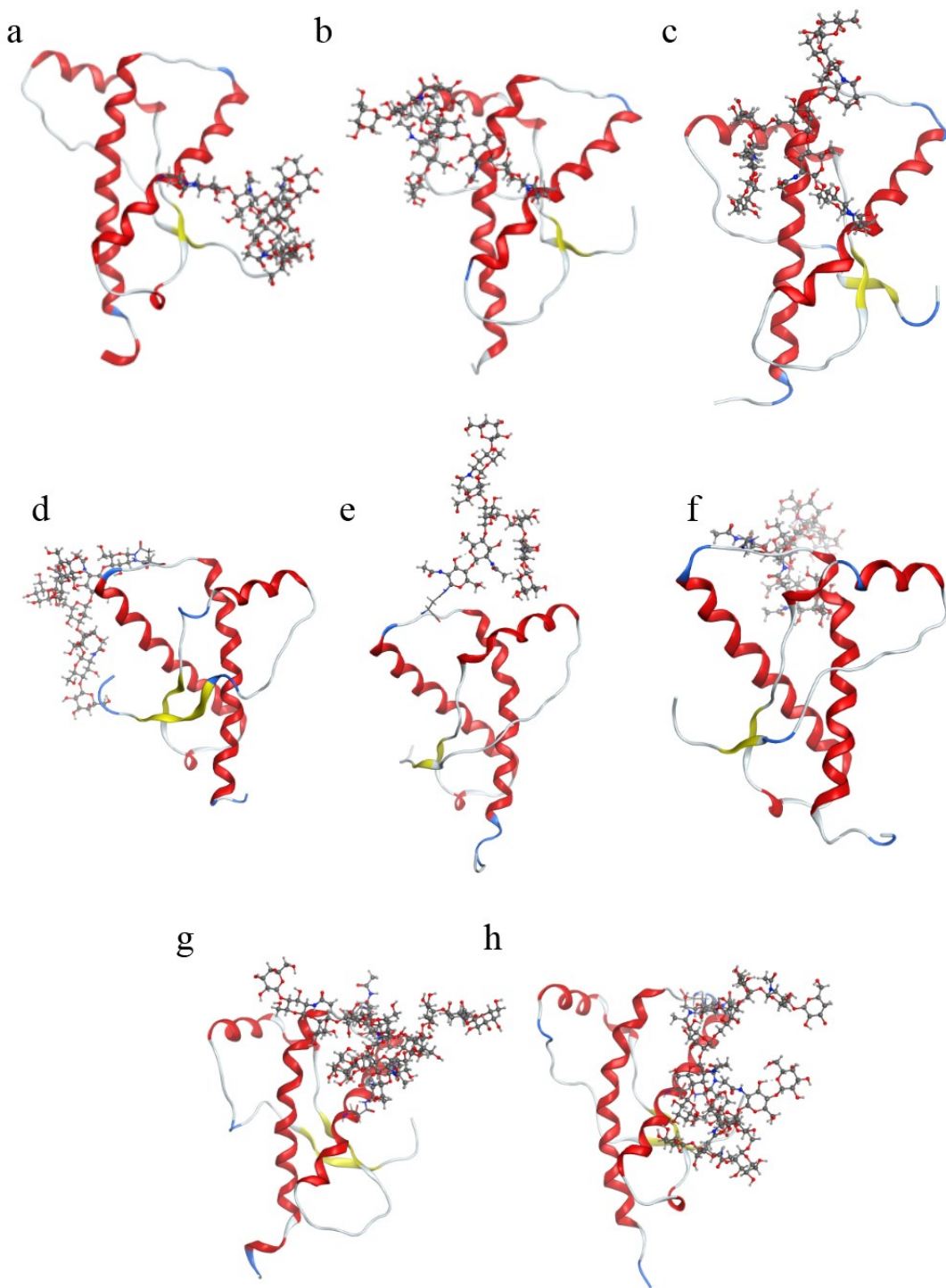
Figure 4-10 Cluster centers from the simulations of PrP attached to complex biantennary glycan: (a-c) mono glycosylated PrP at N181, (d-f) mono glycosylated PrP at N197, (g,h) diglycosylated PrP with glycan at N181 and N197.

Glycoforms with Complex biantennary glycan with Sialic acid:

The addition of sialic acid to the complex biantennary glycan does not change the length of the glycan drastically, yet it has a stronger impact on the protein. The backbone fluctuations decrease in all the three glycoforms (Figure 4-14). The average residue-wise interactions show increased magnitude of interaction energy suggesting an increase in the strength of the non-covalent bonds formed between glycan and the AA residues (Figure 4-17).

When linked to N181, the orientation of the glycans is similar to what was observed in the variant without the sialic acid, but with much stronger and stable interaction through both branches. Two of the three simulations of mono glycosylated PrP at N181 had no change in the 3D structure. However, in the third simulation, helix 1 was dissociated. In the top cluster, it can be noticed that the helix 1 turns into a loop (Figure 4-11), while cluster centers of second and third clusters show lesser dissociation.

Of the three simulations of mono glycosylated PrP with complex biantennary glycan with SA attached to N197, there were no structural changes to the protein in simulation 1. Simulation 2 had dissociated helix 2 (Figure 4-12) similar to what was seen in complex biantennary with no SA at N181. In simulation 3, helix 1 is found to be displaced from its original position (Figure 4-13).

More cluster centers that show glycan at various other orientations are shown in Figure 4-16.
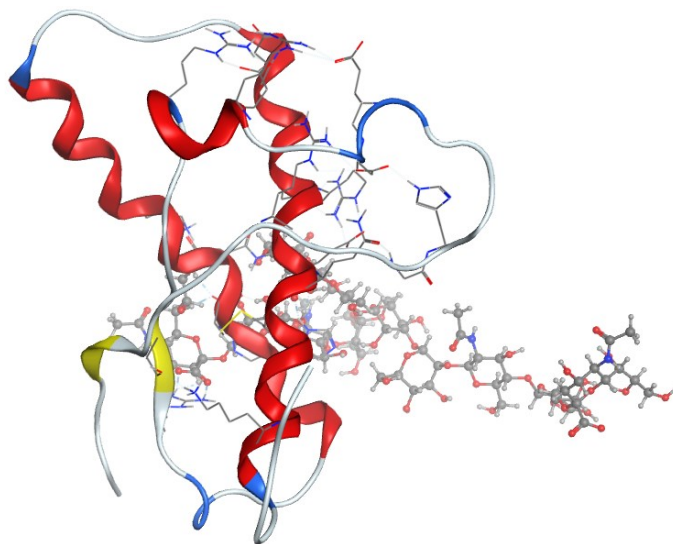
Figure 4-11 Helix 1 dissociated by complex biantennary glycan with sialic acid:
Cluster center of top cluster from a simulation of PrP attached to complex biantennary glycan with sialic acid, at N181, that showed dissociated helix 1.
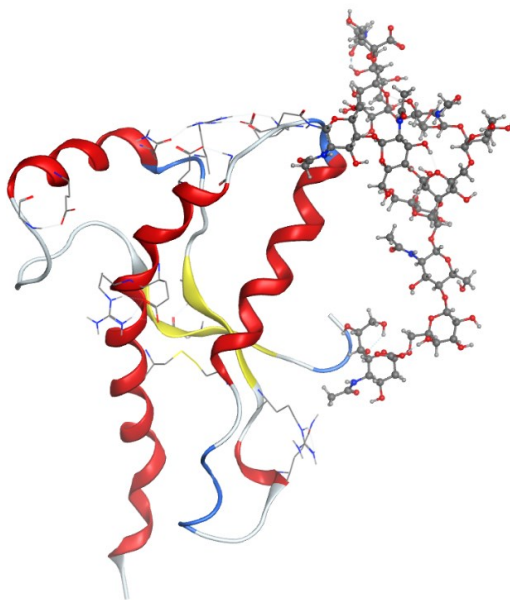
Figure 4-12 Helix 2 dissociated by complex biantennary glycan with sialic acid:
Cluster center of top cluster from a simulation of PrP attached to complex biantennary glycan with sialic acid, at N197, that showed dissociated helix 2
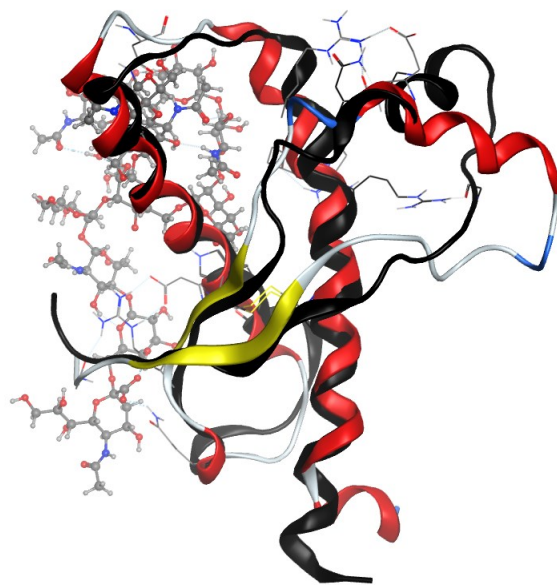
Figure 4-13 Helix 1 distorted by complex biantennary glycan with sialic acid:
Cluster center of top cluster from a simulation of PrP attached to complex biantennary glycan with sialic acid, at N197, that showed distorted helix 1. NMR structure of unglycosylated PrP shown in black.

In the diglycosylated models, stable interactions between the glycans at two sites were observed. But the interaction between protein and glycan was much stronger than what was observed with the other glycans (Figure 4-17). The overall 3D structure of the protein is unchanged. This is possibly due to interactions between the glycans reducing the disruptive interactions between the protein and the glycan.
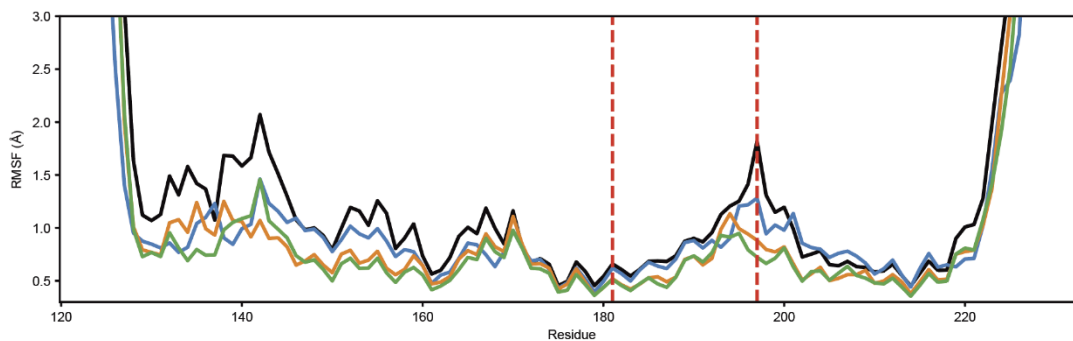
Figure 4-14 RMSF plots from simulations of PrP glycosylated with the addition of complex biantennary glycan with sialic acid: Average RMSF from simulations of mono glycosylated PrP with glycan at N181 represented by the blue curve, average RMSF from simulations of mono glycosylated PrP with glycan at N197 represented by the orange curve, and average RMSF from simulations of diglycosylated PrP represented by the green curve. Red dashed lines represent the site of glycosylation.
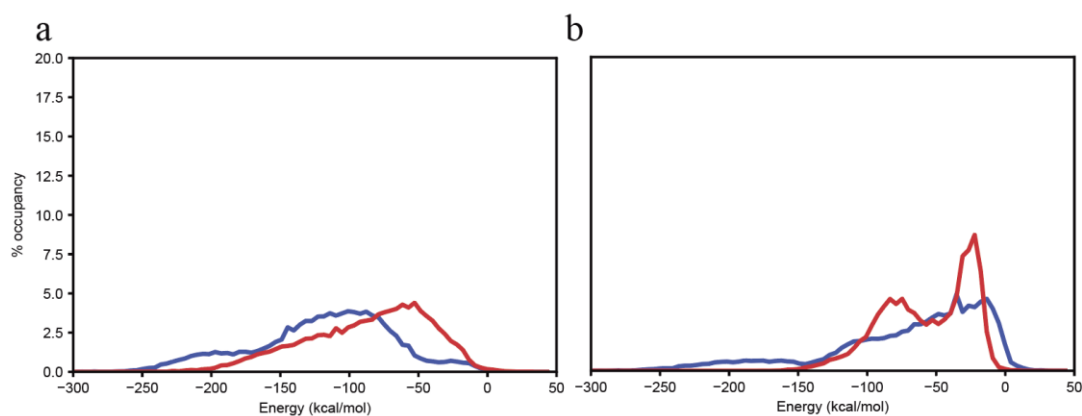
Figure 4-15 Interaction energy profile of complex biantennary glycan with sialic acid: Profile of electrostatic interaction energy between glycan and protein from simulations of a) mono glycosylated b) diglycosylated PrP. The curve in red represents glycan attached at N197 and the curve in blue represents glycan attached at N181.
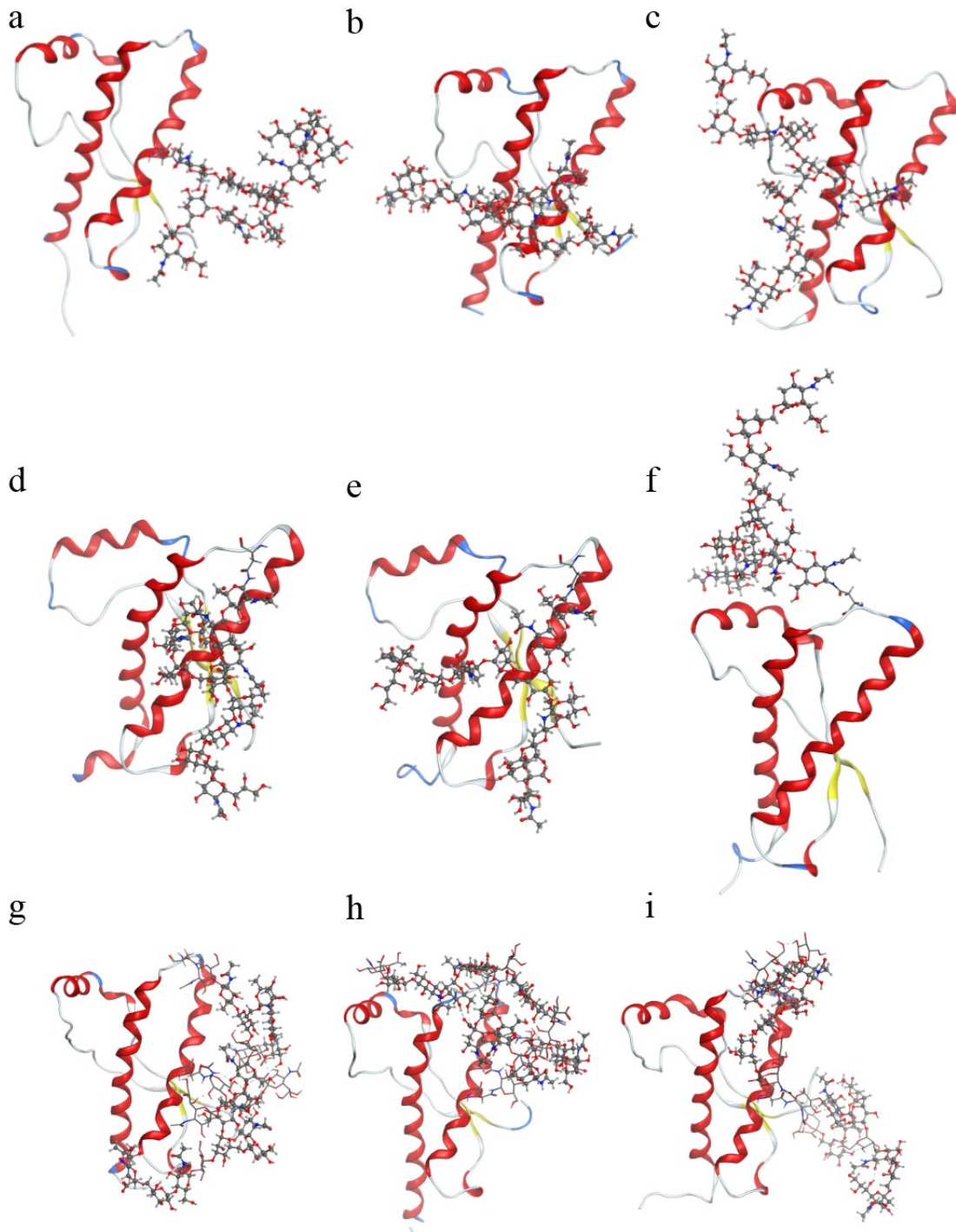
Figure 4-16 Cluster centers from simulations of PrP with complex biantennary glycan with sialic acid: (a-c) mono glycosylated PrP with glycan at N181, (d-f) mono glycosylated PrP with glycan at N197, (g-i) diglycosylated PrP with glycan at N181 and N197
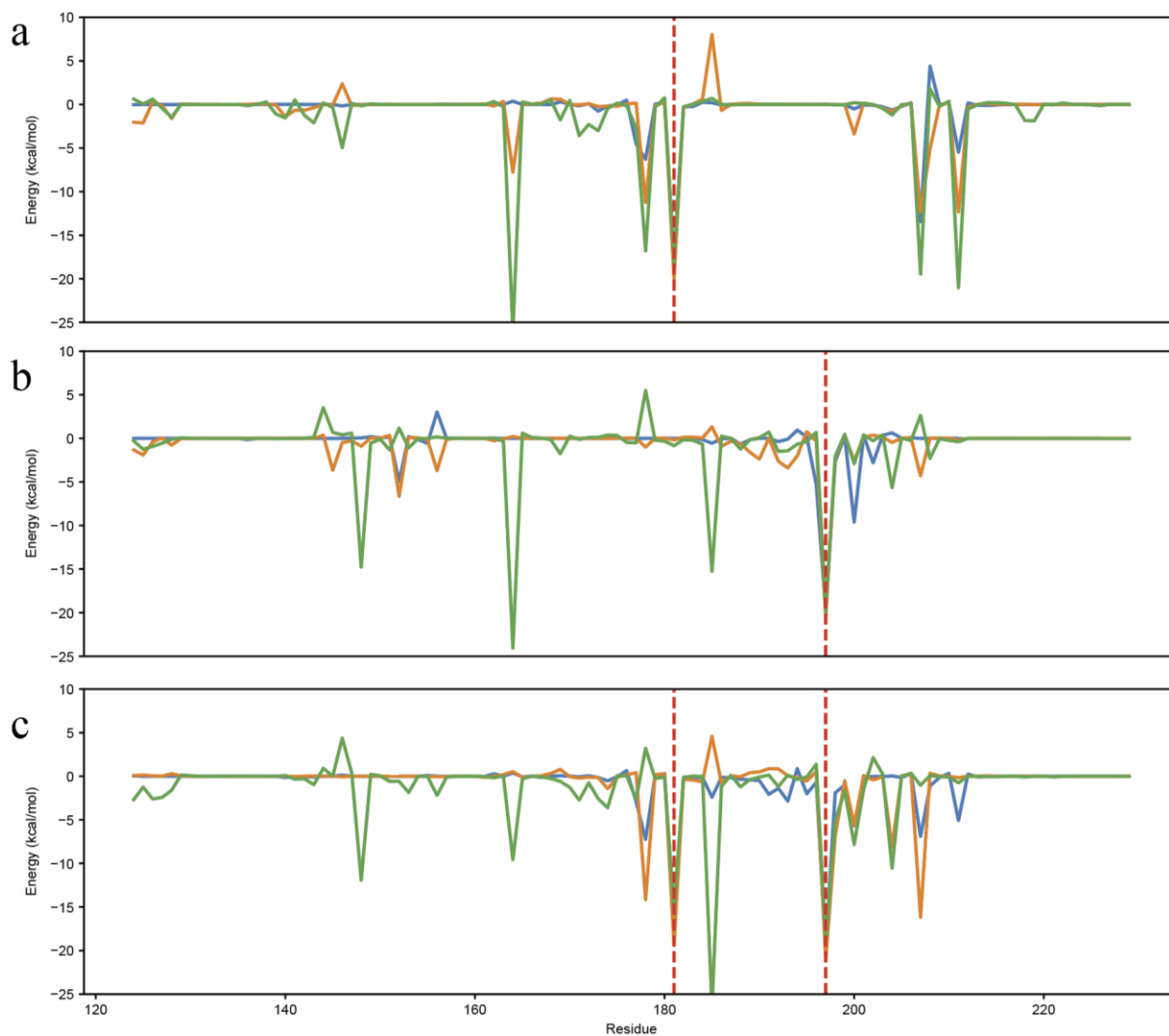
Figure 4-17 Residue-wise average electrostatic interaction energy: Average electrostatic interaction energy was computed for each residue in the cluster centers of the top 3 clusters of all the simulations. Average electrostatic interaction energy of PrP residues computed for mono glycosylated PrP with glycan at N181, mono glycosylated PrP with glycan at N197 and diglycosylated PrP are shown in graphs a), b) and c) respectively. In each graph, the blue curve represents glycoforms with diacetylchitobiose, the orange curve represents glycoforms with the complex biantennary glycan, the green curve represents glycoforms with the complex biantennary glycan with SA. Red dashed lines represent the site of glycosylation.

From the above observations, it is evident that glycans can have significant effects on the structure and stability of the PrP. Glycans are capable of modifying the secondary structure of some parts of PrP and can also displace the structured segments within the protein. Such modifications are observed even with the smallest of the glycans used. Previous studies of glycoproteins suggest that glycans in most cases stabilize the local structure of glycosylation, while in some cases they disrupt the local structure [14], [15], [29], [106], [110], [111]. The observations from the simulations in this study show that glycan may stabilize or disrupt the structure depending on their convergence, and, thereby, the resulting interaction. Multiple simulations of the same glycoform showed that there are different convergence possibilities, and each configuration impacts the protein differently. In the case of the complex biantennary glycan at N181, while two out of three simulations did not change the structure of the protein, one simulation resulted in partial dissociation of helix 1 near the site of glycosylation. Similarly, diacetylchitobiose at N197 displaces helix 1 and partially converts it to a loop in one simulation. Complex biantennary glycan with SA at N181 showed similar effects on helix 1 and complex biantennary with SA at N197 could also dissociate helix 2. It is clear that helix 1 and helix 2 can be affected by glycans at both sites of the mono glycosylated forms.

Such a reduction in alpha helix content in the protein structure due to glycans suggests that the internal non-covalent interactions are disturbed by the glycans. This means glycans can induce partial unfolding within the structure of PrP. This observation has not been reported before.

The stability of helix 1 in the unglycosylated PrP arises from the hydrogen bonds between Y149, R156 on helix 1 with various AA on the loop connecting helix 2 and helix 3. Diacetylchitobiose in one simulation formed additional hydrogen bonds with the helix 1 region, specifically with E152 forcing Y149 to be free due to steric effects. Diacetylchitobiose at N197 could also form hydrogen bonds with the AAs in loop between helix 3 and helix 2 which previously provided stability to helix 1. Such disruptive interaction of glycans with this region resulted in the loss of secondary structure in helix 1. This induced unfolding within the PrP has not been reported previously and could play a crucial role in its misfolding mechanism.

Bulkier the glycan, higher is the possibility in attaining a stable conformation through multiple interactions. The increased number of monosaccharide units naturally increase the number of possible sites that can form hydrogen bonds with the protein. The glycans used in the simulation, while they are of different sizes, they are much smaller than the glycans that are usually found in the strains of PrP [7]. For bulkier glycans, the orientations of glycans in diglycosylated PrP need not be the same as seen in mono glycosylated PrP. This is because glycans at one site could interact with glycan at another site on the protein when doubly glycosylated. This allows for more possible stable configurations for the glycans. The increase in size from diacetylchitobiose to complex biantennary glycan resulted in more sites of interaction, but with very little increase in the magnitude interaction energy. But the increase in the magnitude of interaction energy was much larger in the presence of SA groups, suggesting that the non-covalent interactions are stronger with the addition of SA. The presence of the negative charge is the reason for such behaviour.

5. Outlook

Glycans are dynamic molecules, with varied structure and composition. The fact that monosaccharides can link in any order, and the possibility of branching within glycans increases the complexity in the structure and composition of these sugar molecules. The length and flexibility of the long branches within a glycan can allow the glycan to interact with various segments in a protein. In glycosylated PrP, a glycan linked at a site of glycosylation, can take on multiple stable conformations, and the glycan can thus interact with different parts of the protein depending on its convergence and can impact the structure and stability accordingly. These newly formed interactions often change the non-covalent interactions within the protein. When glycans interact with flexible regions through hydrogen bonds, the atomic fluctuations of these regions reduce. But in a few cases, it was observed that disturbing the pre-existing non-bonded interactions can also change the nature of secondary structure within the region, for example, turning it into a loop. All the glycans used in the study showed this behaviour in at least one of the multiple simulations involving each glycan. It was interesting to note that doubly glycosylated PrP only showed increased stability and no dissociation of the secondary structures.

While all the glycans used in the simulations were found to interact with the neighboring residues at the site of glycosylation, bigger glycans with multiple branches and long arms could converge in such a way that they formed stable interactions through hydrogen bonds with parts of the protein far away from the site of the glycosylation as well. The resulting reduction in the atomic fluctuation of the

protein by glycans depends on the sites at which the glycan is attached as the site of glycosylation dictates the residues that can interact with the glycan.

PrP is almost always chemically modified by the addition of glycans. However, in vitro studies have not used these modifications extensively in their study. This study helps understand PrP in its more natural state as found in the cell and it emphasizes the impact of the presence of glycans. The effects of glycans on PrP certainly cannot be ignored. Site of glycosylation, size of the glycans, and the composition of glycans can all have specific effects on the PrP molecule.

This is the first study that explores the effects of the size, composition and site of glycosylation on the PrP in a systematic fashion. Even though glycans mostly stabilize the molecule, it is clear that structural changes in PrP can be brought about by either bigger glycans at either site or even a small glycan like diacetylchitobiose at N197. Structural changes, if any, were observed in the helix 1, helix 2, and in the beta strands within Hamster PrP. Glycan at N197 is most often found to form stabilizing interactions with the nearby helix 1 which is relatively a flexible part of the protein. These glycan-protein interactions stabilize the protein by reducing fluctuations of such flexible regions. The glycan at N181, while it forms interactions with the local site of glycosylation, the probability of it interacting with the more flexible helix 1 is not as high as in the case of a glycan at N197. While the strength of the interaction between protein and glycan can be increased by the size of glycan, large increases in strength of interaction can also be brought in by the presence of charged SA groups.

One of the crucial observations is the induced unfolding within the PrP due to glycan-protein interactions. In the context of PrP, this is an important effect that can presumably contribute to PrP misfolding. The unfolded protein with more flexible segments is more likely to be influenced by PrP^{Sc} in inducing misfolding.

This observation provides a motivation to further study dimerization simulations of glycosylated prion proteins with bulky glycans. Glycans have proven to form stabilizing hydrogen bonds with the protein and with other glycans. The dimerization simulation can shed lights into the initiation misfolding process itself. It is highly likely that at the initial stages of misfolding, the stability is provided by glycan interactions. It is also possible that a glycan from one PrP molecule could attract other PrP molecules through hydrogen bonds. If either of the PrP is misfolded, then glycans are likely to help in the templated seeded misfolding. Once the chemical techniques for specific glycosylation are developed, this speculation can be tested through aggregation studies of specific glycoforms of PrP. Studying the rate of aggregation for different glycoforms can hint at preferential effects if any.

Hamster PrP is known to have no intermediate structure along its folding pathway. This was uncovered in SMFS studies [40]. If site specific chemical glycosylation techniques are developed, it will provide a way to study force extension curves of glycosylated PrP. Intermediate states are held together by non-covalent interaction within the protein. It is highly likely that the presence of glycans can significantly modify the non-covalent interactions and has the potential to induce new intermediate states. This can have significant effects on the folding pathway.

This study, therefore, motivates to conduct several further studies to explore the glycan effects on PrP. These studies can be both experimental and computational in nature. Molecular dynamics of dimerization of glycosylated PrP can possibly provide insights into the misfolding mechanisms. Glycans with three, four, or more branches form the highest population among the glycans found in PrP^{Sc}, therefore, simulations of PrP with even more bulky glycans, with more branches can be tested. Effects on PrP due to glycans with fucose ring [158], a commonly found monosaccharide in mammals, attached to the first GlcNAc of the core of the N-glycan can be tested to understand effects of steric crowding closer to the site of glycosylation. Such steric crowding might play an important role when the glycan is attached to N197 due to its proximity to helix 1. Mannose rich glycans and hybrid glycans might impact PrP differently compared to complex glycans. MD can again be used to compare the effects of other types of glycans with the results of this study. With the development of chemical techniques to attach glycans at the desired site, several experimental studies can be carried out. To test the stability of specific glycoforms, thermal unfolding can be performed. Aggregation studies can hint at the preferential effects in misfolding of specific glycoforms (if any), and SMFS would provide insights into glycan's effects on the intermediate states, and on the folding pathway. Such analysis can be crucial to the complete understating of the mechanism of prion disease spread.

6. References

[1] D. W. Colby and S. B. Prusiner, "Prions," *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 1, pp. 1–22, Jan. 2011, doi: 10.1101/cshperspect.a006833.

[2] B. Caughey and G. S. Baron, "Prions and their partners in crime," *Nature*, vol. 443, no. 7113. Nature Publishing Group, pp. 803–810, Oct. 19, 2006, doi: 10.1038/nature05294.

[3] A. Aguzzi and A. M. Calella, "Prions: Protein aggregation and infectious diseases," *Physiological Reviews*, vol. 89, no. 4. Physiol Rev, pp. 1105–1152, Oct. 2009, doi: 10.1152/physrev.00006.2009.

[4] N. J. Cobb and W. K. Surewicz, "Prion diseases and their biochemical mechanisms," *Biochemistry*, vol. 48, no. 12, pp. 2574–2585, Mar. 2009, doi: 10.1021/bi900108v.

[5] A. Aguzzi, C. Sigurdson, and M. Heikenwaelder, "Molecular mechanisms of prion pathogenesis," *Annual Review of Pathology: Mechanisms of Disease*, vol. 3. Annu Rev Pathol, pp. 11–40, 2008, doi: 10.1146/annurev.pathmechdis.3.121806.154326.

[6] J. T. Jarrett and P. T. Lansbury, "Seeding 'one-dimensional crystallization' of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie?," *Cell*,

vol. 73, no. 6. Elsevier, pp. 1055–1058, Jun. 18, 1993, doi: 10.1016/0092-8674(93)90635-4.

[7]   P. M. Rudd, A. H. Merry, M. R. Wormald, and R. A. Dwek, "Glycosylation and prion protein," *Curr. Opin. Struct. Biol.*, vol. 12, no. 5, pp. 578–586, Oct. 2002, doi: 10.1016/S0959-440X(02)00377-9.

[8]   R. G. Spiro, "Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds," *Glycobiology*, vol. 12, no. 4. Oxford University Press, pp. 43–56, Apr. 01, 2002, doi: 10.1093/glycob/12.4.43R.

[9]   T. Endo, D. Groth, S. B. Prusiner, and A. Kobata, "Diversity of oligosaccharide structures linked to asparagines of the scrapie prion protein," *Biochemistry*, vol. 28, no. 21, pp. 8380–8388, Oct. 1989, doi: 10.1021/bi00447a017.

[10] N. G. Jayaprakash and A. Surolia, "Role of glycosylation in nucleating protein folding and stability," *Biochem. J.*, vol. 474, no. 14, pp. 2333–2347, Jul. 2017, doi: 10.1042/BCJ20170111.

[11] H. C. Joao and R. A. DWEK, "Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its individual glycoforms," *Eur. J. Biochem.*, vol. 218, no. 1, pp. 239–244, Jan. 1993, doi: 10.1111/j.1432-1033.1993.tb18370.x.

[12] C. Wang, M. Eufemi, C. Turano, and A. Giartosio, "Influence of the carbohydrate moiety on the stability of glycoproteins," *Biochemistry*, vol. 35, no. 23, pp. 7299–7307, Jan. 1996, doi: 10.1021/bi9517704.

[13] S. E. O'connor, B. Imperiali, "A molecular basis for glycosylation-induced

conformational switching," *Chem. Biol.*, vol. 5, pp. 427–437, 1998.

[14] D. Shental-Bechor and Y. Levy, "Effect of glycosylation on protein folding: A close look at thermodynamic stabilization," *Proc. Natl. Acad. Sci.*, vol. 105, no. 24, pp. 8256–8261, Jun. 2008, doi: 10.1073/pnas.0801340105.

[15] Y. Gavrilov, D. Shental-Bechor, H. M. Greenblatt, and Y. Levy, "Glycosylation May Reduce Protein Thermodynamic Stability by Inducing a Conformational Distortion," *J. Phys. Chem. Lett.*, vol. 6, no. 18, pp. 3572–3577, Sep. 2015, doi: 10.1021/acs.jpclett.5b01588.

[16] X. Y. Wang, C. G. Ji, and J. Z. H. Zhang, "Exploring the Molecular Mechanism of Stabilization of the Adhesion Domains of Human CD2 by N-Glycosylation," 2012, doi: 10.1021/jp304116d.

[17] R. Høiberg-Nielsen, C. C. Fuglsang, L. Arleth, and P. Westh, "Interrelationships of glycosylation and aggregation kinetics for Peniophora lycii phytase," *Biochemistry*, vol. 45, no. 15, pp. 5057–5066, Apr. 2006, doi: 10.1021/bi0522955.

[18] Y. Zhuo, J.-Y. Y. Yang, K. W. Moremen, and J. H. Prestegard, "Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell Adhesion Molecule 1 (CEACAM1)," *J. Biol. Chem.*, vol. 291, no. 38, pp. 20085–20095, Sep. 2016, doi: 10.1074/jbc.M116.740050.

[19] D. D. Banks, "The Effect of Glycosylation on the Folding Kinetics of Erythropoietin," *J. Mol. Biol.*, vol. 412, no. 3, pp. 536–550, Sep. 2011, doi: 10.1016/j.jmb.2011.07.061.

[20] P. M. Rudd *et al.*, "Glycosylation differences between the normal and pathogenic prion protein isoforms," *Proc. Natl. Acad. Sci.*, vol. 96, no. 23, pp. 13044–13049, Nov. 1999, doi: 10.1073/pnas.96.23.13044.

[21] N. J. Cobb, F. D. Sönnichsen, H. Mchaourab, and W. K. Surewicz, "Molecular architecture of human prion protein amyloid: A parallel, in-register β-structure," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 48, pp. 18946–18951, Nov. 2007, doi: 10.1073/pnas.0706522104.

[22] R. Tycko, R. Savtchenko, V. G. Ostapchenko, N. Makarava, and I. V. Baskakov, "The α-helical C-terminal domain of full-length recombinant PrP converts to an in-register parallel β-sheet structure in PrP fibrils: Evidence from solid state nuclear magnetic resonance," *Biochemistry*, vol. 49, no. 44, pp. 9488–9497, Nov. 2010, doi: 10.1021/bi1013134.

[23] K. W. Leffers, H. Wille, J. Stöhr, E. Junger, S. B. Prusiner, and D. Riesner, "Assembly of natural and recombinant prion protein into fibrils," *Biol. Chem.*, vol. 386, no. 6, pp. 569–580, 2005, doi: 10.1515/BC.2005.067.

[24] E. Vázquez-Fernández *et al.*, "The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy," *PLoS Pathog.*, vol. 12, no. 9, p. e1005835, Sep. 2016, doi: 10.1371/journal.ppat.1005835.

[25] C. Araman, R. E. Thompson, S. Wang, S. Hackl, R. J. Payne, and C. F. W. Becker, "Semisynthetic prion protein (PrP) variants carrying glycan mimics at position 181 and 197 do not form fibrils," *Chem. Sci.*, vol. 8, no. 9, pp. 6626–6632, 2017, doi: 10.1039/c7sc02719b.

[26] J. Zuegg and J. E. Gready, "Molecular dynamics simulation of human prion

protein including both N-linked oligosaccharides and the GPI anchor," *Glycobiology*, vol. 10, no. 10, pp. 959–974, Oct. 2000, doi: 10.1093/glycob/10.10.959.

[27]   M. L. Demarco and V. Daggett, "Characterization of cell-surface prion protein relative to its recombinant analogue: insights from molecular dynamics simulations of diglycosylated, membrane-bound human prion protein," *J. Neurochem.*, vol. 109, no. 1, pp. 60–73, Apr. 2009, doi: 10.1111/j.1471-4159.2009.05892.x.

[28]   L. Zhong and J. Xie, "Investigation of the effect of glycosylation on human prion protein by molecular dynamics," *J. Biomol. Struct. Dyn.*, vol. 26, no. 5, pp. 525–533, Apr. 2009, doi: 10.1080/07391102.2009.10507268.

[29]   M. M. Chen *et al.*, "Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 52, pp. 22528–22533, Dec. 2010, doi: 10.1073/pnas.1015356107.

[30]   D. P. Clark, N. J. Pazdernik, and M. R. McGehee, "Protein Synthesis," in *Molecular Biology*, Elsevier, 2019, pp. 397–444.

[31]   C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, no. 6968. Nature Publishing Group, pp. 884–890, Dec. 18, 2003, doi: 10.1038/nature02261.

[32]   M. M. Gromiha, "Proteins," in *Protein Bioinformatics*, Elsevier, 2010, pp. 1–27.

[33]   C. Levinthal, "How to fold graciously," *Mössbauer Spectrosc. Biol. Syst. Proc.*,

vol. 24, no. 41, pp. 22–24, 1969.

[34] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, "The Protein Folding Problem," *Annu. Rev. Biophys.*, vol. 37, no. 1, pp. 289–316, Jun. 2008, doi: 10.1146/annurev.biophys.37.092707.153558.

[35] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, "Navigating the folding routes," *Science*, vol. 267, no. 5204. Science, pp. 1619–1620, 1995, doi: 10.1126/science.7886447.

[36] K. A. Dill and H. S. Chan, "From Levinthal to Pathways to Funnels: The 'New View' of Protein Folding Kinetics," *Nat. Struct. Biol.*, vol. 4, no. 1, p. 10, 1997.

[37] L. Martínez, "Introducing the Levinthal's protein folding paradox and its solution," *J. Chem. Educ.*, vol. 91, no. 11, pp. 1918–1923, Nov. 2014, doi: 10.1021/ed300302h.

[38] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096. American Association for the Advancement of Science, pp. 223–230, Jul. 20, 1973, doi: 10.1126/science.181.4096.223.

[39] R. A. Goldbeck, Y. G. Thomas, E. Chen, R. M. Esquerra, and D. S. Kliger, "Multiple pathways on a protein-folding energy landscape: Kinetic evidence," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 6, pp. 2782–2787, Mar. 1999, doi: 10.1073/pnas.96.6.2782.

[40] H. Yu *et al.*, "Direct observation of multiple misfolding pathways in a single prion protein molecule," doi: 10.1073/pnas.1107736109/-/DCSupplemental.

[41] J. S. Valastyan and S. Lindquist, "Mechanisms of protein-folding diseases at a glance," *DMM Dis. Model. Mech.*, vol. 7, no. 1, pp. 9–14, Jan. 2014, doi:

10.1242/dmm.013474.

[42]   H. Saibil, "Chaperone machines for protein folding, unfolding and disaggregation," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 10. Nature Publishing Group, pp. 630–642, Sep. 12, 2013, doi: 10.1038/nrm3658.

[43]   B. Bukau and A. L. Horwich, "The Hsp70 and Hsp60 chaperone machines," *Cell*, vol. 92, no. 3. Cell Press, pp. 351–366, Feb. 06, 1998, doi: 10.1016/S0092-8674(00)80928-9.

[44]   F. U. Hartl and M. Hayer-Hartl, "Protein folding. Molecular chaperones in the cytosol: From nascent chain to folded protein," *Science*, vol. 295, no. 5561. Science, pp. 1852–1858, Mar. 08, 2002, doi: 10.1126/science.1068408.

[45]   A. L. Goldberg, "Protein degradation and protection against misfolded or damaged proteins," *Nature*, vol. 426, no. 6968. Nature, pp. 895–899, Dec. 18, 2003, doi: 10.1038/nature02263.

[46]   M. L. DeMarco and V. Daggett, "From conversion to aggregation: Protofibril formation of the prion protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 8, pp. 2293–2298, Feb. 2004, doi: 10.1073/pnas.0307178101.

[47]   M. T. Woodside and S. M. Block, "Reconstructing folding energy landscapes by single-molecule force spectroscopy," *Annu. Rev. Biophys.*, vol. 43, no. 1, pp. 19–39, Jun. 2014, doi: 10.1146/annurev-biophys-051013-022754.

[48]   M. A. Wulf, A. Senatore, and A. Aguzzi, "The biological function of the cellular prion protein: An update," *BMC Biology*, vol. 15, no. 1. BioMed Central Ltd., pp. 1–13, May 02, 2017, doi: 10.1186/s12915-017-0375-5.

[49]   G. A. Davies, A. R. Bryant, J. D. Reynolds, F. R. Jirik, and K. A. Sharkey, "Prion

diseases and gastrointestinal tract," *Canadian Journal of Gastroenterology*, vol. 20, no. 1. Pulsus Group Inc., pp. 18–24, 2006, doi: 10.1155/2006/184528.

[50] R. Linden, "The biological function of the prion protein: A cell surface scaffold of signaling modules," *Front. Mol. Neurosci.*, vol. 10, p. 77, Mar. 2017, doi: 10.3389/fnmol.2017.00077.

[51] N. Vassallo and J. W. Herms, "Cellular prion protein function in copper homeostasis and redox signalling at the synapse," *Journal of Neurochemistry*, vol. 86, no. 3. John Wiley & Sons, Ltd, pp. 538–544, Aug. 01, 2003, doi: 10.1046/j.1471-4159.2003.01882.x.

[52] H. Hara and S. Sakaguchi, "N-terminal regions of prion protein: Functions and roles in prion diseases," *International Journal of Molecular Sciences*, vol. 21, no. 17. MDPI AG, pp. 1–14, Sep. 01, 2020, doi: 10.3390/ijms21176233.

[53] B. Oesch *et al.*, "A cellular gene encodes scrapie PrP 27-30 protein," *Cell*, vol. 40, no. 4, pp. 735–746, 1985, doi: 10.1016/0092-8674(85)90333-2.

[54] T. L. James *et al.*, "Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 19, pp. 10086–10091, Sep. 1997, doi: 10.1073/pnas.94.19.10086.

[55] D. A. Lysek *et al.*, "Prion protein NMR structures of cats, dogs, pigs, and sheep," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 3, pp. 640–645, Jan. 2005, doi: 10.1073/pnas.0408937102.

[56] A. D. Gossert, S. Bonjour, D. A. Lysek, F. Fiorito, and K. Wüthrich, "Prion protein NMR structures of elk and of mouse/elk hybrids," *Proc. Natl. Acad. Sci.*

U. S. A., vol. 102, no. 3, pp. 646–650, Jan. 2005, doi: 10.1073/pnas.0409008102.

[57] G. Stubbs and J. Stöhr, "Structural biology of PrP Prions," *Cold Spring Harb. Perspect. Med.*, vol. 7, no. 6, Jun. 2017, doi: 10.1101/cshperspect.a024455.

[58] F. Chiti and C. M. Dobson, "Protein misfolding, functional amyloid, and human disease," *Annual Review of Biochemistry*, vol. 75. Annu Rev Biochem, pp. 333–366, 2006, doi: 10.1146/annurev.biochem.75.101304.123901.

[59] H. Wille and J. R. Requena, "The structure of PrPsc prions," *Pathogens*, vol. 7, no. 1, Mar. 2018, doi: 10.3390/pathogens7010020.

[60] C. Govaerts, H. Wille, S. B. Prusiner, and F. E. Cohen, "Evidence for assembly of prions with left-handed β-helices into trimers," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 22, pp. 8342–8347, Jun. 2004, doi: 10.1073/pnas.0402254101.

[61] C. Glynn *et al.*, "Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core," *Nat. Struct. Mol. Biol.*, vol. 27, no. 5, pp. 417–423, 2020, doi: 10.1038/s41594-020-0403-y.

[62] F. López García, R. Zahn, R. Riek, and K. Wüthrich, "NMR structure of the bovine prion protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 15, pp. 8334–8339, Jul. 2000, doi: 10.1073/pnas.97.15.8334.

[63] D. A. Lysek *et al.*, "Prion protein NMR structures of cats, dogs, pigs, and sheep," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 3, pp. 640–645, Jan. 2005, doi: 10.1073/pnas.0408937102.

[64] L. Calzolai, D. A. Lysek, D. R. Pérez, P. Güntert, and K. Wüthrich, "Prion protein NMR structures of chickens, turtles, and frogs," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 3, pp. 651–655, Jan. 2005, doi: 10.1073/pnas.0408939102.

[65] P. K. Baral, M. Swayampakula, A. Aguzzi, and M. N. G. James, "X-ray structural and molecular dynamical studies of the globular domains of cow, deer, elk and Syrian hamster prion proteins," *J. Struct. Biol.*, vol. 192, no. 1, pp. 37–47, 2015, doi: 10.1016/j.jsb.2015.08.014.

[66] M. Qasim Khan *et al.*, "Prion disease susceptibility is affected by β-structure folding propensity and local side-chain interactions in PrP," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 46, pp. 19808–19813, Nov. 2010, doi: 10.1073/pnas.1005267107.

[67] B. Christen, D. R. Pérez, S. Hornemann, and K. Wüthrich, "NMR Structure of the Bank Vole Prion Protein at 20 °C Contains a Structured Loop of Residues 165-171," *J. Mol. Biol.*, vol. 383, no. 2, pp. 306–312, Nov. 2008, doi: 10.1016/j.jmb.2008.08.045.

[68] Y. Zhang, W. Swietnicki, M. G. Zagorski, W. K. Surewicz, and F. D. Sönnichsen, "Solution structure of the E200K variant of human prion protein. Implications for the mechanism of pathogenesis in familial prion diseases," *J. Biol. Chem.*, vol. 275, no. 43, pp. 33650–33654, Oct. 2000, doi: 10.1074/jbc.C000483200.

[69] J. Collinge, K. C. L. Sidle, J. Meads, J. Ironside, and A. F. Hill, "Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD," *Nature*, vol. 383, no. 6602, pp. 685–690, Oct. 1996, doi: 10.1038/383685a0.

[70] J. Castilla, R. Morales, P. Saá, M. Barria, P. Gambetti, and C. Soto, "Cell-free propagation of prion strains," *EMBO J.*, vol. 27, no. 19, pp. 2557–2566, Oct. 2008, doi: 10.1038/emboj.2008.181.

[71] E. Katorcha, N. Makarava, R. Savtchenko, and I. V. Baskakov, "Sialylation of the prion protein glycans controls prion replication rate and glycoform ratio," *Sci. Rep.*, vol. 5, no. 1, pp. 1–11, Nov. 2015, doi: 10.1038/srep16912.

[72] J. Collinge and A. R. Clarke, "A general model of prion strains and their pathogenicity," *Science*, vol. 318, no. 5852. Science, pp. 930–936, Nov. 02, 2007, doi: 10.1126/science.1138718.

[73] T. W. Rademacher, R. B. Parekh, and R. A. Dwek, "Glycobiology," *Annual Review of Biochemistry*, vol. 57. Annu Rev Biochem, pp. 785–838, 1988, doi: 10.1146/annurev.bi.57.070188.004033.

[74] R. A. Somerville, "Host and transmissible spongiform encephalopathy agent strain control glycosylation of PrP," *J. Gen. Virol.*, vol. 80, no. 7, pp. 1865–1872, 1999, doi: 10.1099/0022-1317-80-7-1865.

[75] T. Kuczius, R. Koch, K. Keyvani, H. Karch, J. Grassi, and M. H. Groschup, "Regional and phenotype heterogeneity of cellular prion proteins in the human brain," *Eur. J. Neurosci.*, vol. 25, no. 9, pp. 2649–2655, Apr. 2007, doi: 10.1111/j.1460-9568.2007.05518.x.

[76] N. Mitra, N. Sharon, and A. Surolia, "Role of N-Linked Glycan in the Unfolding Pathway of Erythrina corallodendron Lectin," *Biochemistry*, vol. 42, no. 42, pp. 12208–12216, Oct. 2003, doi: 10.1021/bi035169e.

[77] K. A. Nishina *et al.*, "The stoichiometry of host PrPC glycoforms modulates the efficiency of PrPSc formation in vitro," *Biochemistry*, vol. 45, no. 47, pp. 14129–14139, Nov. 2006, doi: 10.1021/bi061526k.

[78] J. R. Piro *et al.*, "Prion Protein Glycosylation Is Not Required for Strain-Specific

Neurotropism," *J. Virol.*, vol. 83, no. 11, pp. 5321–5328, Jun. 2009, doi: 10.1128/jvi.02502-08.

[79]   T. Pan *et al.*, "Novel Differences between Two Human Prion Strains Revealed by Two-dimensional Gel Electrophoresis," *J. Biol. Chem.*, vol. 276, no. 40, pp. 37284–37288, Oct. 2001, doi: 10.1074/jbc.M107358200.

[80]   V. Beringue *et al.*, "Regional heterogeneity of cellular prion protein isoforms in the mouse brain," *Brain*, vol. 126, no. 9, pp. 2065–2073, Sep. 2003, doi: 10.1093/brain/awg205.

[81]   V. N. Uversky, "Posttranslational Modification," in *Brenner's Encyclopedia of Genetics: Second Edition*, Elsevier Inc., 2013, pp. 425–430.

[82]   R. Boyer, "Posttranslational modification of proteins: Expanding nature's inventory. Christopher T. Walsh, Roberts & Company Publishers, Greenwood Village, CO, 2005, 576 pp., ISBN 0-9747077-3-2, $98.00," *Biochem. Mol. Biol. Educ.*, vol. 34, no. 6, pp. 461–462, Nov. 2006, doi: 10.1002/bmb.2006.494034069996.

[83]   O. N. Jensen, "Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry," *Current Opinion in Chemical Biology*, vol. 8, no. 1. Elsevier Ltd, pp. 33–41, 2004, doi: 10.1016/j.cbpa.2003.12.009.

[84]   P. M. Rudd and R. A. Dwek, "Glycosylation: Heterogeneity and the 3D structure of proteins," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 32, no. 1. Informa Healthcare, pp. 1–100, 1997, doi: 10.3109/10409239709085144.

[85]   B. J. Arey and F. J. López, "Are circulating gonadotropin isoforms naturally

occurring biased agonists? Basic and therapeutic implications," *Rev. Endocr. Metab. Disord.*, vol. 12, no. 4, pp. 275–288, Dec. 2011, doi: 10.1007/s11154-011-9188-y.

[86] E. K. Culyba *et al.*, "Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns," *Science (80-.).*, vol. 331, no. 6017, pp. 571–575, Feb. 2011, doi: 10.1126/science.1198461.

[87] J. L. Price *et al.*, "N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability.," *Biopolymers*, vol. 98, no. 3, pp. 195–211, 2012, doi: 10.1002/bip.22030.

[88] S. KODAMA, M. TSUJIMOTO, N. TSURUOKA, T. SUGO, T. ENDO, and A. KOBATA, "Role of sugar chains in the in-vitro activity of recombinant human interleukin 5," *Eur. J. Biochem.*, vol. 211, no. 3, pp. 903–908, 1993, doi: 10.1111/j.1432-1033.1993.tb17624.x.

[89] T. Ishino *et al.*, "A protein engineering approach differentiates the functional importance of carbohydrate moieties of interleukin-5 receptor α," *Biochemistry*, vol. 50, no. 35, pp. 7546–7556, Sep. 2011, doi: 10.1021/bi2009135.

[90] E. S. Trombetta, "The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis," *Glycobiology*, vol. 13, no. 9. Oxford Academic, pp. 77R-91R, Sep. 01, 2003, doi: 10.1093/glycob/cwg075.

[91] E. Bieberich, "Synthesis, Processing, and Function of N-glycans in N-glycoproteins," in *Advances in neurobiology*, vol. 9, NIH Public Access, 2014, pp. 47–70.

[92] K. W. Moremen, M. Tiemeyer, and A. V. Nairn, "Vertebrate protein

glycosylation: Diversity, synthesis and function," *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 7, pp. 448–462, 2012, doi: 10.1038/nrm3383.

[93] S. A. Overview, A. Illiano, G. Pinto, C. Melchiorre, and A. Carpentieri, "Protein Glycosylation Investigated by Mass," pp. 1–22, 2020.

[94] P. Burda and M. Aebi, "The dolichol pathway of N-linked glycosylation," *Biochimica et Biophysica Acta - General Subjects*, vol. 1426, no. 2. Biochim Biophys Acta, pp. 239–257, Jan. 06, 1999, doi: 10.1016/S0304-4165(98)00127-5.

[95] J. Roth, "Protein N-glycosylation along the Secretory Pathway: Relationship to organelle topography and function, protein quality control, and cell interactions," *Chem. Rev.*, vol. 102, no. 2, pp. 285–303, Feb. 2002, doi: 10.1021/cr000423j.

[96] F. Schwarz and M. Aebi, "Mechanisms and principles of N-linked protein glycosylation," *Current Opinion in Structural Biology*, vol. 21, no. 5. Curr Opin Struct Biol, pp. 576–582, Oct. 2011, doi: 10.1016/j.sbi.2011.08.005.

[97] R. E. Dempski and B. Imperiali, "Oligosaccharyl transferase: Gatekeeper to the secretory pathway," *Current Opinion in Chemical Biology*, vol. 6, no. 6. Elsevier Ltd, pp. 844–850, Dec. 01, 2002, doi: 10.1016/S1367-5931(02)00390-3.

[98] L. Ellgaard and A. Helenius, "Quality control in the endoplasmic reticulum," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 3. Nature Publishing Group, pp. 181–191, Mar. 01, 2003, doi: 10.1038/nrm1052.

[99] L. Cao *et al.*, "Global site-specific analysis of glycoprotein N-glycan processing," *Nat. Protoc.*, vol. 13, no. 6, pp. 1196–1212, 2018, doi:

10.1038/nprot.2018.024.

[100] M. Nagae and Y. Yamaguchi, "Function and 3D Structure of the N-Glycans on Glycoproteins," *Int. J. Mol. Sci*, vol. 13, pp. 8398–8429, 2012, doi: 10.3390/ijms13078398.

[101] A. Varki *et al.*, *Essentials of glycobiology, third edition*. Cold Spring Harbor Laboratory Press, 2017.

[102] J. L. Price *et al.*, "Context-dependent effects of asparagine glycosylation on pin WW folding kinetics and thermodynamics," *J. Am. Chem. Soc.*, vol. 132, no. 43, pp. 15359–15367, Nov. 2010, doi: 10.1021/ja106896t.

[103] S. R. Hanson, E. K. Culyba, T.-L. T. L. Hsu, C.-H. C. H. Wong, J. W. Kelly, and E. T. Powers, "The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability," *Proc. Natl. Acad. Sci.*, vol. 106, no. 9, pp. 3131–3136, Mar. 2009, doi: 10.1073/pnas.0810318105.

[104] L. Q. Yang *et al.*, "Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms," *J. Biomol. Struct. Dyn.*, vol. 32, no. 3, pp. 372–393, Mar. 2014, doi: 10.1080/07391102.2013.770372.

[105] G. T. DeKoster and A. D. Robertson, "Thermodynamics of unfolding for kazal-type serine protease inhibitors: Entropic stabilization of ovomucoid first domain by glycosylation," *Biochemistry*, vol. 36, no. 8, pp. 2323–2331, Feb. 1997, doi: 10.1021/bi962580b.

[106] V. T. Pham, E. Ewing, H. Kaplan, C. Choma, and M. A. Hefford, "Glycation improves the thermostability of trypsin and chymotrypsin," *Biotechnol. Bioeng.*, vol. 101, no. 3, pp. 452–459, Oct. 2008, doi: 10.1002/bit.21919.

[107] S. Kaushik, D. Mohanty, and A. Surolia, "Role of glycosylation in structure and stability of Erythrina corallodendron lectin (EcorL): A molecular dynamics study," *Protein Sci.*, vol. 20, no. 3, pp. 465–481, Mar. 2011, doi: 10.1002/pro.578.

[108] C. F. Brewer, "Lectin cross-linking interactions with multivalent carbohydrates," in *Advances in Experimental Medicine and Biology*, 2001, vol. 491, pp. 17–25, doi: 10.1007/978-1-4615-1267-7_2.

[109] S. Sinha and A. Surolia, "Attributes of glycosylation in the establishment of the unfolding pathway of soybean agglutinin," *Biophys. J.*, vol. 92, no. 1, pp. 208–216, 2007, doi: 10.1529/biophysj.106.092668.

[110] S. Halder, A. Surolia, and C. Mukhopadhyay, "Impact of glycosylation on stability, structure and unfolding of soybean agglutinin (SBA): an insight from thermal perturbation molecular dynamics simulations," *Glycoconj. J.*, vol. 32, no. 6, pp. 371–384, Aug. 2015, doi: 10.1007/s10719-015-9601-y.

[111] S. Halder, A. Surolia, and C. Mukhopadhyay, "Dynamics simulation of soybean agglutinin (SBA) dimer reveals the impact of glycosylation on its enhanced structural stability," *Carbohydr. Res.*, vol. 428, pp. 8–17, Jun. 2016, doi: 10.1016/j.carres.2016.04.009.

[112] G. Kern, N. Schülke, R. Jaenicke, and F. X. Schmid, "Stability, quaternary structure, and folding of internal, external, and core-glycosylated invertase from yeast," *Protein Sci.*, vol. 1, no. 1, pp. 120–131, 1992, doi: 10.1002/pro.5560010112.

[113] J. Tian *et al.*, "Effect of Glycosylation on an Immunodominant Region in the

V1V2 Variable Domain of the HIV-1 Envelope gp120 Protein," *PLoS Comput. Biol.*, vol. 12, no. 10, p. e1005094, Oct. 2016, doi: 10.1371/journal.pcbi.1005094.

[114] K. A. Kulkarni *et al.*, "Effect of glycosylation on the structure of Erythrina corallodendron lectin," *Proteins Struct. Funct. Bioinforma.*, vol. 56, no. 4, pp. 821–827, May 2004, doi: 10.1002/prot.20168.

[115] V. . Srinivas, N. C. Singha, F. P. Schwarz, and A. Surolia, "Differential scanning calorimetric studies of the glycoprotein, winged bean acidic lectin, isolated from the seeds of *Psophocarpus tetrogonolobus*," *FEBS Lett.*, vol. 425, no. 1, pp. 57–60, Mar. 1998, doi: 10.1016/S0014-5793(98)00197-5.

[116] M. M. Prabu *et al.*, "Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution," *J. Mol. Biol.*, vol. 276, no. 4, pp. 787–796, Mar. 1998, doi: 10.1006/jmbi.1997.1568.

[117] V. R. Srinivas, K. Bachhawat-Sikder, S. Habib, S. E. Hasnain, and A. Surolia, "Expression of winged bean basic agglutinin in Spodoptera frugiperda insect cell expression system," *Biosci. Rep.*, vol. 21, no. 3, pp. 361–367, Jun. 2001, doi: 10.1023/A:1013294417847.

[118] G. N. Rogers$, G. Herrlerb, J. C. Paulson$ll, and H.-D. Klenkg, "THE JOURNAL OF BIOLOGICAL CHEMISTRY Influenza C Virus Uses 9-0-Acetyl-N-acetylneuraminic Acid as a High Affinity Receptor Determinant for Attachment to Cells*," 1966.

[119] B. Schultze and G. Herrler, "Recognition of N-acetyl-9-O-acetylneuraminic acid by bovine coronavirus and hemagglutinating encephalomyelitis virus," in

Advances in Experimental Medicine and Biology, 1994, vol. 342, pp. 299–304, doi: 10.1007/978-1-4615-2996-5_46.

[120] B. Schultze and G. Herrler, "Bovine coronavirus uses N-acetyl-9-O-acetylneuraminic acid as a receptor determinant to initiate the infection of cultured cells," *J. Gen. Virol.*, vol. 73, no. 4, pp. 901–906, 1992, doi: 10.1099/0022-1317-73-4-901.

[121] A. W. Barb, L. Meng, Z. Gao, R. W. Johnson, K. W. Moremen, and J. H. Prestegard, "NMR characterization of immunoglobulin G Fc glycan motion on enzymatic sialylation," *Biochemistry*, vol. 51, no. 22, pp. 4618–4626, Jun. 2012, doi: 10.1021/bi300319q.

[122] K.-T. Shade and R. Anthony, "Antibody Glycosylation and Inflammation," *Antibodies*, vol. 2, no. 4, pp. 392–414, Jun. 2013, doi: 10.3390/antib2030392.

[123] A. Varki, "Sialic acids in human health and disease," *Trends in Molecular Medicine*, vol. 14, no. 8. NIH Public Access, pp. 351–360, Aug. 2008, doi: 10.1016/j.molmed.2008.06.002.

[124] S. Carvalho *et al.*, "Preventing E-cadherin aberrant N-glycosylation at Asn-554 improves its critical function in gastric cancer," *Oncogene*, vol. 35, no. 13, pp. 1619–1631, Mar. 2016, doi: 10.1038/onc.2015.225.

[125] A. Guillot *et al.*, "Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans," *Sci. Rep.*, vol. 6, no. 1, p. 35666, Dec. 2016, doi: 10.1038/srep35666.

[126] I. V. Baskakov and E. Katorcha, "Multifaceted role of sialylation in prion diseases," *Frontiers in Neuroscience*, vol. 10, no. AUG. Frontiers Media S.A.,

p. 358, Aug. 08, 2016, doi: 10.3389/fnins.2016.00358.

[127] E. Cancellotti *et al.*, "Post-translational changes to PrP alter transmissible spongiform encephalopathy strain properties," *EMBO Journal*, vol. 32, no. 5. European Molecular Biology Organization, pp. 756–769, Mar. 06, 2013, doi: 10.1038/emboj.2013.6.

[128] F. K. Wiseman *et al.*, " The Glycosylation Status of PrP C Is a Key Factor in Determining Transmissible Spongiform Encephalopathy Transmission between Species ," *J. Virol.*, vol. 89, no. 9, pp. 4738–4747, May 2015, doi: 10.1128/jvi.02296-14.

[129] M. Aminpour, C. Montemagno, and J. A. Tuszynski, "An overview of molecular modeling for drug discovery with specific illustrative examples of applications," *Molecules*, vol. 24, no. 9, 2019, doi: 10.3390/molecules24091693.

[130] R. S. Katiyar and P. K. Jha, "Molecular simulations in drug delivery: Opportunities and challenges," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 8, no. 4, p. e1358, Jul. 2018, doi: 10.1002/wcms.1358.

[131] S. I. Omar, M. Zhao, R. V. Sekar, S. A. Moghadam, J. A. Tuszynski, and M. T. Woodside, "Modeling the structure of the frameshift stimulatory pseudoknot in SARS-CoV-2 reveals multiple possible conformers," doi: 10.1101/2020.06.08.141150.

[132] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. I. General method," *J. Chem. Phys.*, vol. 31, no. 2, pp. 459–466, Aug. 1959, doi: 10.1063/1.1730376.

[133] M. Karplus, J. A. McCammon, and B. R. Gelin, "Dynamics of folded proteins,"

Nature, vol. 267, no. June, pp. 585–590, 1977.

[134] S. A. Hollingsworth and R. O. Dror, "Molecular Dynamics Simulation for All," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018, doi: 10.1016/j.neuron.2018.08.011.

[135] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic Validation of Protein Force Fields against Experimental Data," *PLoS One*, vol. 7, no. 2, p. e32131, Feb. 2012, doi: 10.1371/journal.pone.0032131.

[136] D. E. Shaw *et al.*, "Anton, a special-purpose machine for molecular dynamics simulation," *Commun. ACM*, vol. 51, no. 7, pp. 91–97, Jul. 2008, doi: 10.1145/1364782.1364802.

[137] S. J. Marrink and D. P. Tieleman, "Perspective on the martini model," *Chem. Soc. Rev.*, vol. 42, no. 16, pp. 6801–6822, Jul. 2013, doi: 10.1039/c3cs60093a.

[138] T. Hou, J. Wang, Y. Li, and W. Wang, "Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations," *J. Chem. Inf. Model.*, vol. 51, no. 1, pp. 69–82, Jan. 2011, doi: 10.1021/ci100275a.

[139] S. Matysiak and C. Clementi, "Minimalist Protein Model as a Diagnostic Tool for Misfolding and Aggregation," *J. Mol. Biol.*, vol. 363, no. 1, pp. 297–308, Oct. 2006, doi: 10.1016/j.jmb.2006.07.088.

[140] A. V. Onufriev and D. A. Case, "Generalized Born Implicit Solvent Models for Biomolecules," *Annu. Rev. Biophys.*, vol. 48, no. 1, pp. 275–296, May 2019, doi: 10.1146/annurev-biophys-052118-115325.

[141] J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. Van Der Spoel, "Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents," *J. Chem. Theory Comput.*, vol. 13, no. 3, pp. 1034–1043, Mar. 2017, doi: 10.1021/acs.jctc.7b00169.

[142] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983, doi: 10.1063/1.445869.

[143] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-Grained Protein Models and Their Applications," 2016, doi: 10.1021/acs.chemrev.6b00163.

[144] C. Clementi, "Coarse-grained models of protein folding: toy models or predictive tools?," *Current Opinion in Structural Biology*, vol. 18, no. 1. Curr Opin Struct Biol, pp. 10–15, Feb. 2008, doi: 10.1016/j.sbi.2007.10.005.

[145] J. Hu, T. Chen, M. Wang, H. S. Chan, and Z. Zhang, "A critical comparison of coarse-grained structure-based approaches and atomic models of protein folding," *Phys. Chem. Chem. Phys.*, vol. 19, no. 21, pp. 13629–13639, May 2017, doi: 10.1039/c7cp01532a.

[146] R. Berraud-Pache, C. Garcia-Iriepa, and I. Navizet, "Modeling Chemical Reactions by QM/MM Calculations: The Case of the Tautomerization in Fireflies Bioluminescent Systems," *Front. Chem.*, vol. 6, no. APR, p. 116, Apr. 2018, doi: 10.3389/fchem.2018.00116.

[147] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," *Angewandte Chemie - International Edition*, vol. 48, no. 7. John Wiley & Sons,

Ltd, pp. 1198–1229, Feb. 02, 2009, doi: 10.1002/anie.200802019.

[148] J. Ilja Siepmann and D. Frenkel, "Conflgurational bias monte carlo: A new sampling scheme for flexible chains," *Mol. Phys.*, vol. 75, no. 1, pp. 59–70, 1992, doi: 10.1080/00268979200100061.

[149] M. N. Rosenbluth and A. W. Rosenbluth, "Monte carlo calculation of the average extension of molecular chains," *J. Chem. Phys.*, vol. 23, no. 2, pp. 356–359, Feb. 1955, doi: 10.1063/1.1741967.

[150] U. H. E. Hansmann, "Parallel Tempering Algorithm for Conformational Studies of Biological Molecules," *Chem. Phys. Lett.*, vol. 281, no. 1–3, pp. 140–150, Oct. 1997, doi: 10.1016/S0009-2614(97)01198-6.

[151] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Physical Chemistry Chemical Physics*, vol. 7, no. 23. The Royal Society of Chemistry, pp. 3910–3916, Dec. 07, 2005, doi: 10.1039/b509983h.

[152] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems," *Biochim Biophys Acta*, no. 5, pp. 872–877, 2015, doi: 10.1016/j.bbagen.2014.10.019.

[153] L. C. T. Pierce, R. Salomon-Ferrer, C. Augusto F. De Oliveira, J. A. McCammon, and R. C. Walker, "Routine access to millisecond time scale events with accelerated molecular dynamics," *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 2997–3002, Sep. 2012, doi: 10.1021/ct300284c.

[154] D. Hamelberg, J. Mongan, and J. A. McCammon, "Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules," *J.*

Chem. Phys., vol. 120, no. 24, pp. 11919–11929, Jun. 2004, doi: 10.1063/1.1755656.

[155] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3696–3713, Jul. 2015, doi: 10.1021/acs.jctc.5b00255.

[156] K. N. Kirschner *et al.*, "GLYCAM06: A generalizable biomolecular force field. carbohydrates," *J. Comput. Chem.*, vol. 29, no. 4, pp. 622–655, Mar. 2008, doi: 10.1002/jcc.20820.

[157] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, Jul. 2013, doi: 10.1021/ct400341p.

[158] D. J. Becker and J. B. Lowe, "Fucose: Biosynthesis and biological function in mammals," *Glycobiology*, vol. 13, no. 7. Oxford Academic, pp. 41R-53R, Jul. 01, 2003, doi: 10.1093/glycob/cwg054.