

# Molecular simulation of cellulose surface interactions

MOLECULAR INSIGHT INTO CELLULOSE NANOCRYSTALS  
AND THEIR INTERACTION WITH CELLULOSIC OLIGOMERS  
BY ALL-ATOM SIMULATION

By Naveen Kumar Vasudevan,

*A Thesis Submitted to the School of Graduate Studies in the Partial  
Fulfillment of the Requirements for the Degree Master of Applied Science*

McMaster University © Copyright by Naveen Kumar Vasudevan March  
23, 2018

McMaster University

Master of Applied Science (2018)

Hamilton, Ontario (Department of Chemical Engineering)

TITLE: Molecular Insight into Cellulose Nanocrystals and their Interaction with Cellulosic Oligomers by All-Atom Simulation

AUTHOR: Naveen Kumar Vasudevan (McMaster University)

SUPERVISOR: Dr. Li Xi

NUMBER OF PAGES: xix, 135

# Abstract

Cellulose nanocrystal (CNC) has found application in a variety of novel products due to its spectrum of properties. Notably, the CNC-polymer systems have seen numerous applications in special materials like Pickering emulsions, foams and gels etc. CNC interacts with different polymers to a different extent. These interactions include molecular level and bulk interactions. Subsequently, they modify the interfacial properties. Though vibrant, the CNC-polymer molecular interaction is still unclear. We took this void in our understanding as our motivation to explore these interactions. In this work, we tried to understand why CNC interacts differently with different polymers and what drives the adsorption of polymer on CNC. Our work can also help us to understand the configurations and origins of CNC-polymer system properties. The broad range of length and time scales covered by this physical process requires a multiscale simulation approach. In this thesis, we start with the all-atom molecular simulation and focus on the specific energetic interactions between CNC surfaces and unrealistically short polymer chains. In future work, we will build on this model and develop a multiscale modeling approach for capturing the full scope of CNC-polymer interaction, including the configuration and dynamics of realistic long polymer chains around CNCs. We propose that there are two driving forces for adsorption based on the free energy difference values obtained via PMF (potential of mean force) calculations done on eight systems with different physical components. Overall, we conclude that the balance between polymer's ability to form hydrogen bonds with the surface and their interactions with the bulk solvent control the adsorption and desorption phenomenon. A larger coarse-grained model developed from our simulations will help to understand

these systems better. This presented work deals with the specific energy interactions and information which we will need for the systematic coarse-graining of these systems.

# *Acknowledgements*

Words cannot express my gratitude to my supervisor Dr. Li Xi for the trust he put in me. His guidance, patience, insight and impeccable interest on me and others in The Xi Research Group is what made this work possible. I would like to thank Dr. Emily D. Cranston and her students whose work served as the backbone of my study.

I would also like to acknowledge the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada for providing the computational resources used for this work.

I would like to express my heartfelt gratitude to my good friends here at The Xi Research group and McMaster University, especially to Dongyang Li and Lingli Liu for the countless hours of fruitful discussions and support.

Personally, I wish to extend my thanks to all the technical and non-technical staff members of The Department of Chemical Engineering for their benevolence and selfless work in making all of our lives a lot easier.

And above all, I wish to express my heartfelt gratitude and thanks to my family here and back home. It is my parent's trust, love and support which encourages me to push my limits and achieve great heights.

# Contents

<b>Abstract</b>	<b>iii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>Declaration of Authorship</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.1.1 Cellulose . . . . .	2
1.1.1.1 Occurrence . . . . .	3
1.1.1.2 Structure . . . . .	3
1.1.1.3 Morphology . . . . .	5
1.1.2 Cellulose Nanocrystals . . . . .	7
1.1.2.1 Cellulose-I <sub>α</sub> . . . . .	10
1.1.2.2 Cellulose-I <sub>β</sub> . . . . .	11
1.1.2.3 Surfaces (100) and (010) . . . . .	13
1.1.3 Hydrogen bonding . . . . .	14
1.1.4 Sulphated surface . . . . .	15
1.2 Motivation . . . . .	16
1.2.1 Physical understanding . . . . .	18
1.2.1.1 Chain length . . . . .	19
1.3 Objective . . . . .	20
1.4 Literature review . . . . .	20
1.5 Thesis outline . . . . .	25
<b>2 Methodology</b>	<b>27</b>
2.1 Molecular Dynamics . . . . .	27

2.1.1	Theory . . . . .	28
2.1.1.1	Newton’s equation of motion . . . . .	29
2.1.2	Scope of MD . . . . .	30
2.1.3	Periodic boundary condition (PBC) . . . . .	31
2.1.4	Thermodynamic ensembles . . . . .	32
2.1.5	Forcefields . . . . .	33
2.1.6	Limitations of MD . . . . .	35
2.2	Molecular Models . . . . .	35
2.2.1	Physical systems . . . . .	36
2.2.1.1	(100) surface systems . . . . .	36
2.2.1.2	(010) surface systems . . . . .	38
2.2.1.3	Surface modification . . . . .	39
2.2.2	Water model - TIP3P . . . . .	40
2.3	Forcefield - GLYCAM06j . . . . .	41
2.4	Simulation . . . . .	44
2.4.1	Pre-Processing . . . . .	44
2.4.1.1	Cellulose-Builder . . . . .	44
2.4.1.2	Glycam-web . . . . .	45
2.4.1.3	Packmol . . . . .	46
2.4.1.4	Avogadro . . . . .	47
2.4.1.5	Assisted Model Building with Energy Refinement (AMBER) . . . . .	48
2.4.2	Main Simulation . . . . .	50
2.4.2.1	Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) . . . . .	50
2.4.2.2	Simulation Strategy . . . . .	50
2.4.2.3	Free energy calculation . . . . .	53
2.4.2.4	Overview . . . . .	54
2.4.2.5	Umbrella sampling . . . . .	54
2.4.2.6	Steered molecular dynamics (SMD) . . . . .	56
2.4.2.7	collective variable (COLVAR) . . . . .	58
2.4.2.8	weighted histogram analysis method (WHAM) . . . . .	60
2.4.2.9	Histograms . . . . .	61
2.4.3	Post-Processing . . . . .	62
<b>3</b>	<b>Results and Discussion</b> . . . . .	<b>63</b>
3.1	Model Validation . . . . .	63
3.1.1	Crystal structure . . . . .	64
3.1.2	Forcefield comparison . . . . .	64
3.1.3	Water model . . . . .	66
3.2	Direct MD results . . . . .	68
3.2.1	Adsorption and desorption . . . . .	68

3.2.2	Water molecules . . . . .	70
3.2.2.1	Water density profile . . . . .	72
3.2.3	Oligomer diffusivity . . . . .	74
3.2.3.1	Oligomer movement along the (100) and (010) surfaces . . . . .	74
3.2.4	Hydrogen bond - Oligomer and water . . . . .	79
3.3	Free energy calculation . . . . .	82
3.3.1	PMF profile . . . . .	82
3.3.1.1	Zones in the PMF profile . . . . .	83
3.3.2	Configurations along the PMF curve . . . . .	86
3.3.3	Overall comparison . . . . .	88
3.3.3.1	Salt effects . . . . .	91
3.3.3.2	Bulk effects . . . . .	94
3.4	Other analyses . . . . .	97
3.4.1	Oligomer Properties . . . . .	97
3.4.2	Intramolecular Hydrogen bond within the oligomer . . . . .	100
3.5	Summary . . . . .	106
<b>4</b>	<b>Conclusions</b>	<b>112</b>
<b>A</b>	<b>Post-processing</b>	<b>117</b>
MDAnalysis	. . . . .	117
MDTraj	. . . . .	118
CPPTRAJ	. . . . .	118
VMD	. . . . .	118
<b>B</b>	<b>Input files</b>	<b>119</b>
A1	Cellulose-builder . . . . .	119
Input	. . . . .	119
Command	. . . . .	119
A2	Packmol . . . . .	120
Input	. . . . .	120
Command	. . . . .	121
A3	AMBER . . . . .	121
tleap	. . . . .	121
CPPTRAJ	. . . . .	121
<b>C</b>	<b>Calculations</b>	<b>123</b>
A1	Water molecules count . . . . .	123
A2	Salt ions count . . . . .	124
	<b>Bibliography</b>	<b>124</b>

# List of Figures

1.1	Cellobiose, a disaccharide with the formula $C_{12}H_{22}O_{11}$ . The oxygen, carbon and hydrogen atoms are represented in red, grey and white colours respectively and this colour code will be followed from here on. . . .	3
1.2	a) Cellulose chain (linear) b) Starch chain (non-linear). The orientation of the anomeric -OH group which results in structural differences is highlighted with red circles. . . . .	4
1.3	Classification-carbohydrates from sugars to crystalline cellulose surfaces simulated. Materials of interest in this thesis are highlighted in 'red'. . . . .	6
1.4	Crystalline and Amorphous zones in a cellulose microfibril. . . . .	8
1.5	Unit cell parameters and representation of cellulose- $I$ - $\alpha$ and $\beta$ crystals. . . .	9
1.6	Cellulose- $I_\alpha$ along yz-plane(left), xz-plane(middle), xy-plane(right). The x, y, z-axes are represented in red, green and blue respectively. . . . .	11
1.7	Cellulose- $I_\beta$ along yz-plane(left), xz-plane(middle), xy-plane(right). The x, y, z-axes are represented in red, green and blue respectively. . . . .	12
1.8	Cellulose- $I_\beta$ unit cell and lattice parameters. . . . .	12
1.9	Interchain hydrogen bonding in cellulose- $I_\beta$ . The x, y, z-axes are represented in red, green and blue respectively. . . . .	14
1.10	Interchain and Intrachain hydrogen bonding in two origin chains of Cellulose- $I_\beta$ . The x, y, z-axes are represented in red, green and blue respectively. . . . .	14
1.11	Sulphate groups at the C6 position. . . . .	15
2.1	2D simulation box showing PBC. The red particle, which is seen exiting through one direction can be seen reentering through the other face along the same direction. . . . .	32
2.2	Bonds, Angles, Dihedrals and Improper dihedrals. . . . .	33
2.3	MC over the (100) surface. The x, y, z-axes are represented in red, blue, green respectively. The water molecules are not shown for clarity. . . . .	38
2.4	MC over the (010) surface. The x, y, z-axes are represented in red, blue, green respectively. The water molecules are not shown for clarity. . . . .	39

2.5	Sulphate groups on (010) surface. . . . .	40
2.6	log(MSD) vs log(time). The plot shows slope along the log-log curve used to calculate the self-diffusion coefficient value of water. . . . .	42
2.7	Water density profile in the simulation box for a well equilibrated 286mc system. The blue dots are the COM of water molecules along the periodic box form one snapshot of the periodic box. . . . .	43
2.8	Residue, linkage and substitution on the MC oligomer chain. Each bead represents one glucose unit, b represents the beta linkage between the units. '6Me' denotes methyl group substitution at the C6 carbon's -OH group, similarly for '2Me'. . . . .	46
2.9	Residue, linkage and substitution on the pentaose oligomer chain. Each bead represents one glucose unit, b represents the beta linkage between the units. . . . .	46
2.10	A completely packed 286pe system with all three components (CNC, pentaose, water). The oxygen, carbon and hydrogen atoms are represented by red, green and white spheres respectively. . . . .	47
2.11	Umbrella Sampling technique. . . . .	57
2.12	Window-wise histogram for four systems. . . . .	61
3.1	CNC crystal structure comparison - initial and after equilibration runs. The unitcell along the $yz$ -plane is marked in yellow, the x,y and z-directions are marked in red, green and blue respectively. . . . .	65
3.2	Potential energy of CNC in vacuum as a function of time for three different forcefields. . . . .	66
3.3	TIP3P water model observations. . . . .	69
3.4	Component COMs of each system . . . . .	71
3.5	MSD and Self-diffusion coefficient of water. . . . .	73
3.6	Water number density ( $\rho_N$ ) . . . . .	75
3.7	MSD of oligomers. . . . .	76
3.8	Oligomer movement along the (100) surface. . . . .	77
3.9	Oligomer movement along the (010) surface. . . . .	78
3.10	Hydrogen bond between Oligomer and water - I. The number of unique hydrogen bonds between oligomer and water in the (100)pes system as a function of time, in terms of frames (1 frame = 0.002 ns) is shown, the number of hydrogen bonds in each unique type per frame is denoted along the right-hand side y-axis . . . . .	80
3.11	Hydrogen bond between Oligomer and water - II. The number of unique hydrogen bonds between oligomer and water in the (010)mc system as a function of time, in terms of frames (1 frame = 0.002 ns) is shown, the number of hydrogen bonds in each unique type per frame is denoted along the right-hand side y-axis . . . . .	81

3.12	PMF as the function of $\xi$ for the (100)pes system with 25 windows. The curve has three different zones namely, the favourable zone, marked by 1; the slopes, marked by 2, 4, 6 and the plateaus, marked by 3, 5, 7. Positions of interest where the oligomer configurations were captured are marked in red. . . . .	84
3.13	Configurations of pentaose oligomer at different positions along reaction coordinate ( $\xi$ ). The x, y and z-directions are given in red, green and blue colour respectively. . . . .	89
3.14	PMF curves of the systems as the function of $\xi$ . Two major groups belonging to two different surfaces can be seen. . . . .	92
3.15	The position of ions on (100) surface, the Sodium ions are given in blue and the Chloride ions are given in red. . . . .	95
3.16	The position of ions on (010) surface, the Sodium ions are given in blue and the Chloride ions are given in red. . . . .	96
3.17	End-to-end distance of the oligomer chains. . . . .	101
3.18	RMSD values of the oligomer chains. . . . .	102
3.19	$R_g$ of the oligomer chains. . . . .	103
3.20	SASA of the oligomer chains. . . . .	104
3.21	First and second water shells. . . . .	105
3.22	Intramolecular Hydrogen bond within the oligomer in windows 0-11 of the (100)pes. The plots show the number of intramolecular hydrogen bonds found in the oligomer as the function of time. . . . .	107
3.23	Intramolecular Hydrogen bond within the oligomer in windows 13-24 of the (100)pes system. The plots show the number of intramolecular hydrogen bonds found in the oligomer as the function of time. . . . .	108

# List of Tables

1.1	Unit cell parameters for Cellulose - I - $\alpha$ , $\beta$ . . . . .	9
2.1	Thermodynamic Ensembles and their constant parameters. . . . .	33
2.2	Simulated system nomenclatures and specifications. . . . .	37
2.3	Residue information for all the oligomer structures. . . . .	49
2.4	Initial configuration set up. . . . .	51
2.5	Equilibration steps followed for the systems. . . . .	52
2.6	Colvar module sampling specification used for all systems. . . . .	59

# Acronyms

## Abbreviations

**2D** two dimensional

**3D** three dimensional

**AMBER** Assisted Model Building with Energy Refinement

**CHARMM** Chemistry at Harvard Macromolecular Mechanics

**CLSM** confocal laser scanning microscopy

**CMC** carboxymethyl cellulose

**CNC** cellulose nanocrystals

**CNF** cellulose nanofibrils

**CO** cellulosic oligomers

**COD** Crystallography Open Database

**COM** centre of mass

**CP** cellulosic polymers

**COLVAR** collective variable

**DEX** dextran

**DOP** degree Of polymerization

**DS** degree of substitution

**ELBA** electrostatics-based

**GAFF** general AMBER force field

**GROMACS** GRONingen MACHine for Chemical Simulations

**GROMOS** GRONingen Molecular Simulation

**H<sub>2</sub>SO<sub>4</sub>** sulphuric acid

**HCl** hydrochloric acid

**HEC** hydroxyethyl cellulose

**HLB** hydrophilic lipophilic balance

**LBG** locust bean gum

**LAMMPS** Large-scale Atomic/Molecular Massively Parallel Simulator

**HPG** hydroxypropyl ugar

**LCPO** linear combinations of pairwise overlaps

**LCST** lower critical solution temperature

**LDOP** level-off degree of polymerization

**MC** methyl cellulose

**MD** molecular dynamics

**MEP** molecular electrostatic potential

**MSD** mean-squared displacement

**MS** molar substitution

$\mu$ **VT** Grand canonical ensemble

**NaCl** sodium chloride

**NMR** Nuclear magnetic resonance spectroscopy

**NPT** Isothermal-isobaric ensemble

**NVE** Microcanonical ensemble

**NVT** Canonical ensemble

**PBC** periodic boundary condition

**PDADMAC** polydiallyldimethylammonium chloride

**\*.PDB** protein database file

**PMF** potential of mean force

**\*.PSF** protein structure file

**QCM** quartz crystal microbalance

**RDF** radial distribution function

**RESP** restrained electrostatic potential

**RMSD** root-mean-square deviation

**SASA** surface accessible surface area

**SD** steepest descent algorithm

**SEM** scanning electron microscopy

**SMD** steered molecular dynamics

**SPR** surface plasmon resonance

**S-CNC** sulphated cellulose nanocrystal

**TIP3P** Three-site transferrable intermolecular potential

**VMD** Visual Molecular Dynamics

**WHAM** weighted histogram analysis method

**\*.XYZ** XYZ coordinate file

## Symbols

**$a$**  unit cell lattice axis length along x

$\alpha$  the angle between adjacent unit cell axes  $b$  and  $c$

$b$  unit cell lattice axis length along  $y$

$\beta$  the angle between adjacent unit cell axes  $b$  and  $a$

$c$  unit cell lattice axis length along  $z$

$\gamma$  the angle between adjacent unit cell axes  $a$  and  $c$

$k_B$  Boltzmann constant ( $1.38064852 \times 10^{-23} \text{ m}^2 \text{ Kg s}^{-2} \text{ K}^{-1}$ )

$\Delta G$  Gibbs-Free energy difference

**fs** femtoseconds

$l_p$  persistence length

$\mu\text{s}$  microseconds

$N_A$  Avogadro's number ( $6.022 \times 10^{23} \text{ mol}^{-1}$ )

**ns** nanoseconds

$h$  Planck's constant ( $6.62607004 \times 10^{-34} \text{ m}^2 \text{ kg/s}$ )

$\mathbf{R}_g$  radius of gyration

$\xi$  reaction coordinate

# Declaration of Authorship

I, Naveen Kumar Vasudevan, declare that this thesis titled, “Molecular Insight into Cellulose Nanocrystals and their Interaction with Cellulosic Oligomers by All-Atom Simulation” and the work presented in it are my own.

*To everyone who believes in me...*

# Chapter 1

## Introduction

Earth has an abundance of materials which vary in their characteristics, availability, and versatility when it comes to application. The development of special materials for specific use has been of keen interest to researchers. Biomaterials like large macromolecules and biopolymers are carefully engineered to cater targeted needs due to the novel characteristics they exhibit. Carbohydrates are one of the most sought-after macromolecules due to their close integration with both plant and animal life forms.

Cellulose, the most bountiful carbohydrate<sup>[35]</sup>, has immense significance in the development of bio-based materials (gels<sup>[24]</sup>, colloids<sup>[26]</sup>, emulsion<sup>[25]</sup> etc.), biomimetic systems<sup>[67]</sup> and applications such as polymer matrix reinforcement<sup>[40]</sup> etc. Cellulose as a material is one of the most extensively studied biopolymers. Much research has been conducted on cellulose to develop it into alternative products<sup>[40]</sup>, owing to its biodegradable, crystalline, thermal, mechanical and interfacial properties.

To understand any of these mentioned properties, a firm grasp of the molecular level

attributes and interactions is the key. And molecular dynamics (MD) is one of the methods which will enable us to have insight into the molecular, even atomic level interactions like hydrogen bonding, adsorption etc. This thesis addresses the same with the focus towards the interaction of cellulose nanocrystals (CNC) with cellulosic oligomers (CO) using all-atom MD method. We will see why and how different surface morphology affects CNC interaction with other polymeric and bulk systems.

In the following sections, we will discuss the occurrence, structure and chemical nature of cellulose. Within the scope of this thesis, we will discuss the interaction of the two crystalline surfaces (100) and (010) of cellulose- $I_{\beta}$  - the crystal form of CNC with cellulosic polymers (CP)s, Cellopentaose and methyl cellulose (MC).

## 1.1 Background

### 1.1.1 Cellulose

Cellulose is a carbohydrate with a general formula of  $(C_6H_{10}O_5)_n$ , formed by the union of hundreds of D-glucose units linked by (1 $\rightarrow$ 4) glycosidic bonds<sup>[21,64,65]</sup>. Broadly, cellulose belongs to carbohydrates and is one of the polymers in the family of polysaccharides like glycogen and starch, which are formed by the union of glucose units. Cellulose is formed by the  $\beta$ -glucose monomer units, instead of the  $\alpha$ -glucose units, which make starch and glycogen. The cellulose composition was first deduced in 1842 by Payen. The structure of a cellobiose unit (a disaccharide) is shown in fig. 1.1, the bonds and structure resulting

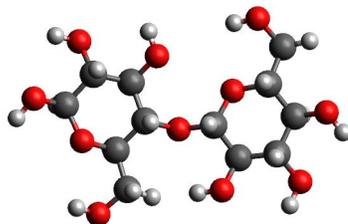


FIGURE 1.1: Cellobiose, a disaccharide with the formula  $C_{12}H_{22}O_{11}$ . The oxygen, carbon and hydrogen atoms are represented in red, grey and white colours respectively and this colour code will be followed from here on.

from the union of two glucose monomer units can be clearly seen. Cellulose is usually hundreds of units in length and occurs in various forms.

#### 1.1.1.1 Occurrence

Cellulose can be derived from various sources by a variety of techniques. Commonly it is extracted from wood, cotton, algae or tunicates. Cellulose obtained from bacteria, known as bacterial cellulose<sup>[9,59,60]</sup>, is also commonly seen. Their properties like molecular structure and physical state, which subsequently alter their chemical interactions, vary by a great extent based on their sources. Cellulose has shown interesting features due to its occurrence in various structural morphologies. The next section discusses its structure at a molecular level.

#### 1.1.1.2 Structure

Like it was mentioned in the previous section, cellulose's physical structure is influenced significantly by its origin. Cellulose, in general, occurs in various forms. Some of those forms are natural while other forms are all synthetically obtained by chemical and/or

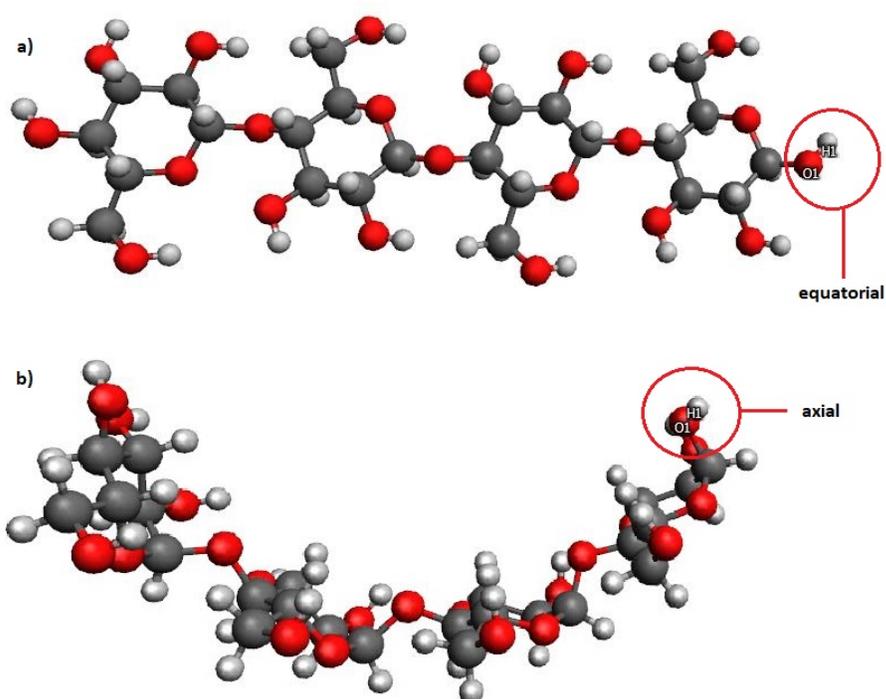


FIGURE 1.2: a) Cellulose chain (linear) b) Starch chain (non-linear). The orientation of the anomeric -OH group which results in structural differences is highlighted with red circles.

mechanical treatments of naturally occurring forms. To appreciate the structural differences and their influence over cellulose's interaction, one must look at its buildup starting from the molecular level. Starch, glycogen, and cellulose are made from the same monomer unit, glucose, but they vary by the association, type of the glucose units and the orientation of C<sub>1</sub> carbon's -OH group. Starch is formed by the glucose- $\alpha$  monomers while cellulose is formed from glucose- $\beta$  monomer units. The orientation of -OH groups influences the chain propagation (equatorial- $\beta$  form and axial- $\alpha$  form). It is this feature of cellulose which gives its characteristic alternately twisting glucose units bound by a glycosidic bond, resulting in an unbranched structure.

This unbranched structure results in a lot of interesting morphologies which are discussed in subsequent sections. Figure 1.2 shows short oligomer chains of both cellulose and starch. It can be seen that cellulose chain is linear while the starch chain is non-linear and this is due to the orientation of the hydroxyl group on the reducing end, which is the -OH group attached to the anomeric carbon atom. In cellulose the reducing end's -OH group is equatorial and in starch it is axial. This gives rise to the changes in their chain propagation and results in structural differences.

### 1.1.1.3 Morphology

It is evident that cellulose is significantly different from starch and other homopolysaccharides, both, physically and chemically (interactions). As mentioned earlier the unbranched structure of cellulose has influence over its macro-structures. This unbranched structure of cellulose gives rise to other macrostructures like fibres and cellulose nanocrystals (CNC). Their characteristics and morphology vary at the

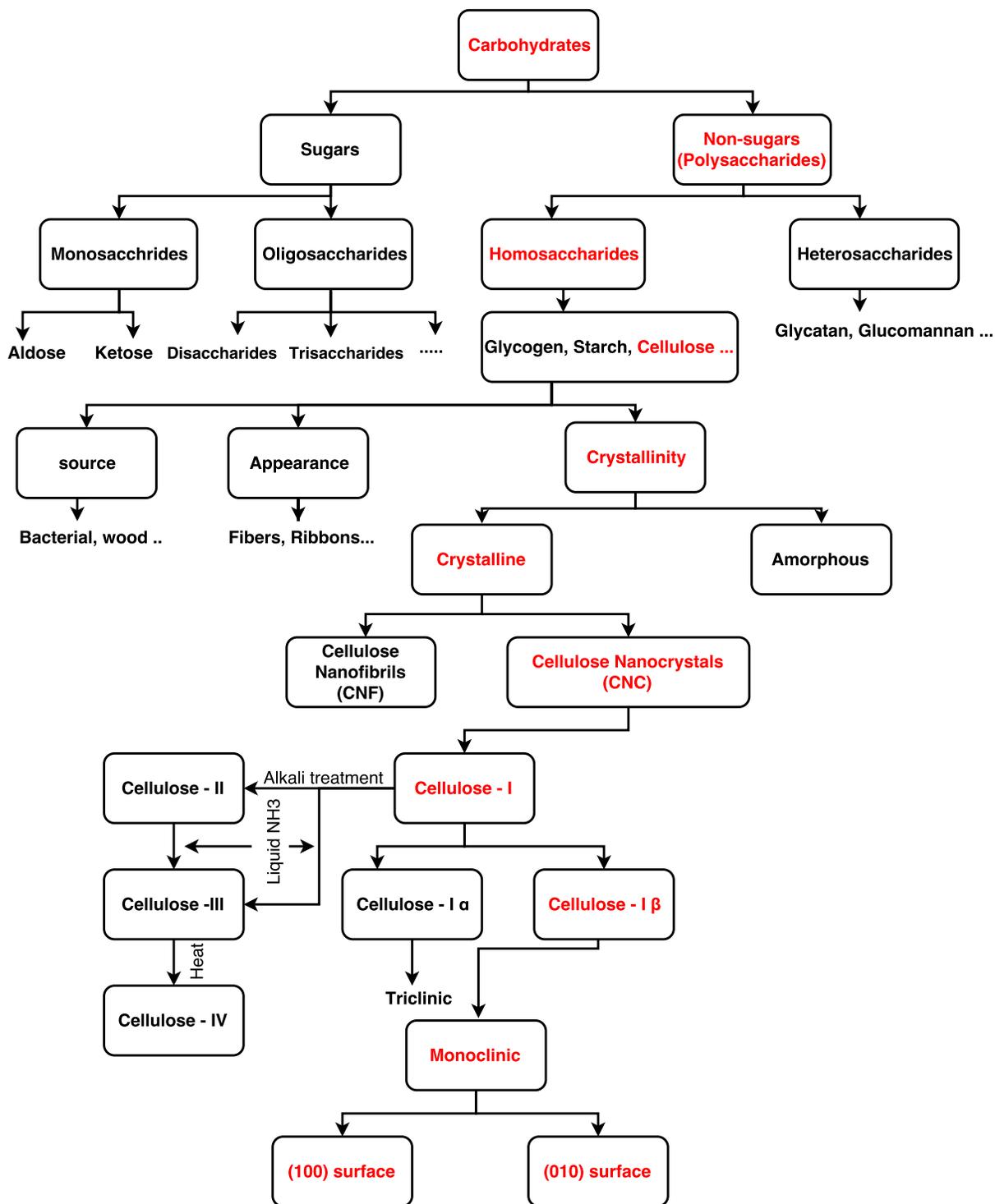


FIGURE 1.3: Classification-carbohydrates from sugars to crystalline cellulose surfaces simulated. Materials of interest in this thesis are highlighted in 'red'.

nano-level. Based on the source, the dimensions of the macrostructure varies (100 - 1000 units) and different treatments used to extract cellulosic matter from the source results in different morphologies. Figure 1.3 shows a detailed classification of carbohydrates and the systems of interest are highlighted in red. Cellulose can occur in the form of fibers<sup>[12]</sup>, ribbons or sheets but the most interesting physical occurrence of the cellulose is in its nano-state.

Nanocellulose refers to the cellulose structure with at least one of its dimensions in the nanometer range. Cellulose nanostructures became materials of interest in 1983 after Turbak et al. synthesized and coined the term microfibrillated cellulose. Cellulose as a nanomaterial has seen significant development and focus in recent years. Cellulose at a nano level can be classified into cellulose nanofibrils (CNF) and CNCs. Though CNF is of great interest in reinforced materials<sup>[40,74]</sup>, our focus is primarily on CNCs.

Although cellulose nanocrystals<sup>[56]</sup> were reported as early as 1949, it is quite recently that CNC has become a product of commercial interest<sup>[62,72]</sup>. This is due to the unique and versatile features of CNC and the novel products like grafted polymers and composite films etc. developed from it<sup>[24-26,53]</sup>. CNC has seen applications in a variety of fields like bio-imaging, ion-exchange applications etc.<sup>[10,11,19,53]</sup>

### 1.1.2 Cellulose Nanocrystals

One of the main reason for the difference in starch and cellulose structures is the linkage. Starch has the  $\alpha$  linkage while cellulose has the  $\beta$  linkage. In case of  $\beta$  linkage, each

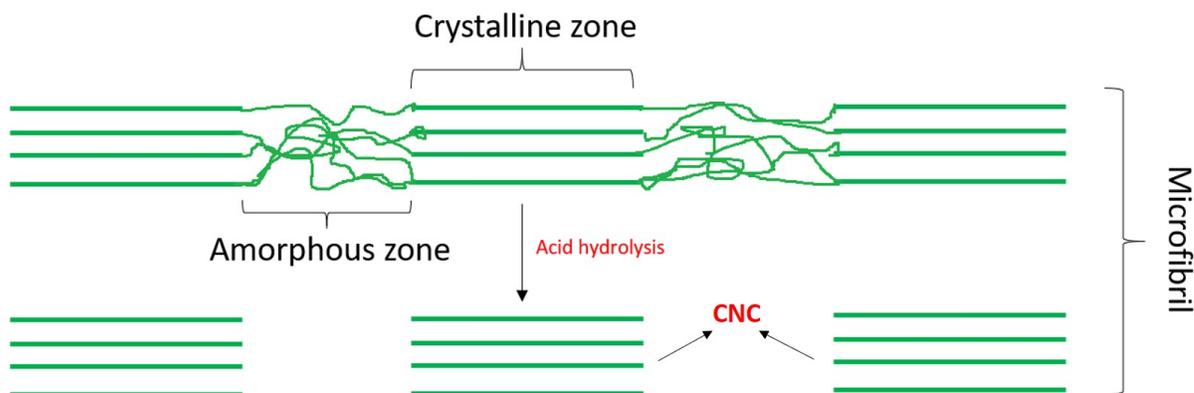


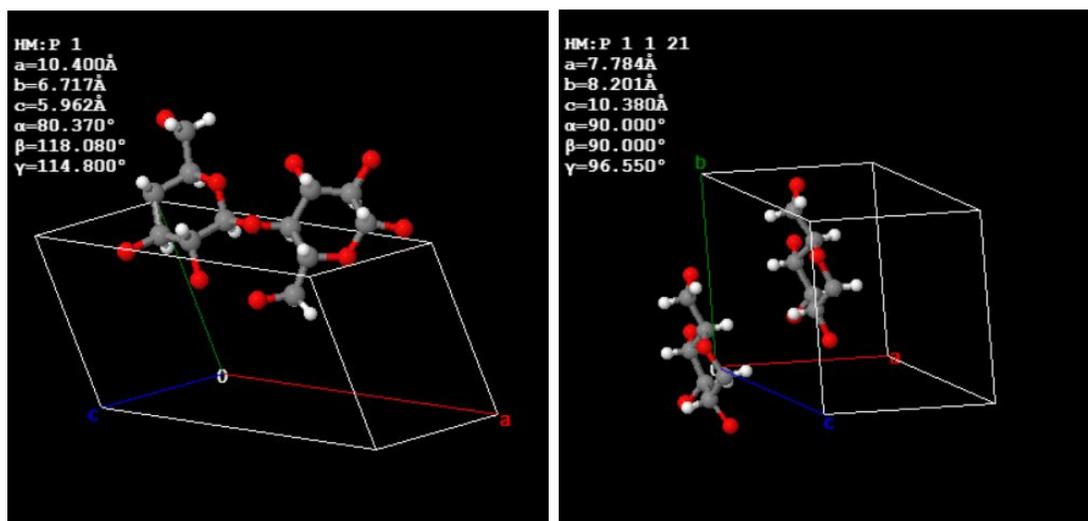
FIGURE 1.4: Crystalline and Amorphous zones in a cellulose microfibril.

glucose unit is rotated by  $180^\circ$  along the axis of the polymer backbone chain relative to the previous unit giving what we call the cellobiosyl repeat. This property enables cellulose to exhibit different morphologies and one of the most interesting structures resulting from it are CNCs. Cellulose microfibril has both crystalline and amorphous domains. CNC is extracted from the crystalline regions of a cellulose microfibril. CNC is usually extracted using acid hydrolysis. The amorphous and disordered segments of the fibril are more susceptible to acid attack. This acid-induced hydrolysis decreases the degree of polymerization (DOP) rapidly to a level called the level-off degree of polymerization (LDOP).

The difference in hydrolysis at different zones can be attributed to the difference in kinetics at the amorphous and crystalline domains. The CNC is obtained in a needle-like structure. These needle-like structures might vary in size based on the source or the extraction method. Figure 1.4 shows the different zones within a microfibril. Like it was shown in fig. 1.3, cellulose occurs in both natural and synthetic crystalline structures. It naturally occurs in two crystalline forms, cellulose- $I_\alpha$  and cellulose- $I_\beta$ <sup>[70]</sup>.

TABLE 1.1: Unit cell parameters for Cellulose - I -  $\alpha$ ,  $\beta$ .

Polymorph	$I - \alpha$	$I - \beta$
Type	Triclinic P1	Monoclinic P2 <sub>1</sub>
a (Å)	6.717	7.784
b (Å)	5.962	8.201
c (Å)	10.400	10.38
$\alpha$	118.08°	90°
$\beta$	114.80°	90°
$\gamma$	80.37°	96.5°
Chain(s)	1	2

(a)  $I_{\alpha}$ .(b)  $I_{\beta}$ .FIGURE 1.5: Unit cell parameters and representation of cellulose- $I$ - $\alpha$  and  $\beta$  crystals.

The distinguishing feature between the two crystal structures is their crystalline nature and space-group. Cellulose-I $_{\alpha}$  has triclinic structure while cellulose-I $_{\beta}$  has a monoclinic structure. These structures are discussed further in the next subsection. Table 1.1 gives the unit cell parameters, unit cell lattice axis length along x ( $a$ ), the unit cell lattice axis length along y ( $b$ ), unit cell lattice axis length along z ( $c$ ), the angle between adjacent unit cell axes  $b$  and  $c$  ( $\alpha$ ), the angle between adjacent unit cell axes  $b$  and  $a$  ( $\beta$ ) and the angle between adjacent unit cell axes  $a$  and  $c$  ( $\gamma$ ), for both  $\alpha$  and  $\beta$  crystalline structures of cellulose. The  $\alpha$  form of cellulose can be obtained from algae and bacteria, while  $\beta$  form from other plants such as tunicins etc. These two polymorphs together constitute the cellulose-I category. Figure 1.5 shows both the unit cell parameters and representation for cellulose-I- $\alpha$  and  $\beta$  crystal forms. These representations were made with the coordinate information obtained from Crystallography Open Database (COD)<sup>[50,51]</sup>.

#### 1.1.2.1 Cellulose-I $_{\alpha}$

Cellulose-I $_{\alpha}$  is one of the naturally occurring crystalline forms of cellulose and is mainly sourced from the cell wall of algae or bacteria. Cellulose-I $_{\alpha}$  has a triclinic one chain crystalline structure and it belongs to the P1 space group<sup>[51]</sup>. Space groups are also known as crystallographic groups and they represent the symmetry of a crystal. There are two types of space groups for triclinic crystals, namely P1 and P-1. All other space groups possess the symmetry elements of the P1 space group. P1 space group does not have any rotation axes except for the identity axes. For space group P1, there is only one symmetry equivalent position within the unit cell.

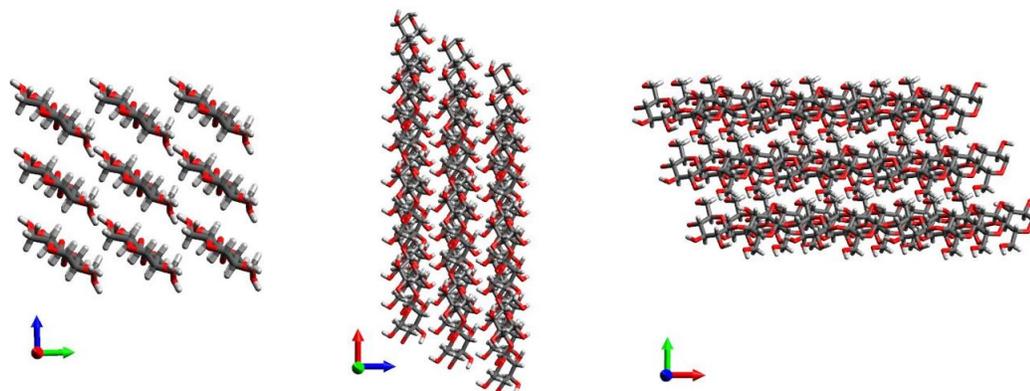


FIGURE 1.6: Cellulose- $I_\alpha$  along  $yz$ -plane(left),  $xz$ -plane(middle),  $xy$ -plane(right). The  $x$ ,  $y$ ,  $z$ -axes are represented in red, green and blue respectively.

Table 1.1 gives the lattice parameters for its unit cell and fig. 1.6 shows a simple cellulose- $I_\alpha$  crystal with three unit cells in all three lattice axes. It has a progressive parallel shear along the chain axis. The cellobiose exhibit a diagonal shift to the same extent as that of a cellulose- $I_\beta$ , which results in each dimer unit being aligned at a particular shear angle. This value differs depending on which crystallographic face is being viewed. The adjacent chains are held together by hydrogen bonds.

### 1.1.2.2 Cellulose- $I_\beta$

Cellulose- $I_\beta$  similar to  $I_\alpha$  is one of the naturally occurring crystalline forms of cellulose. It is more significant and of greater interest, as it can be obtained from various larger floras. It has a monoclinic crystal structure and it belongs to the  $P2_1$  space group<sup>[50]</sup>. It is the most frequently occurring space group. This space group has three types of symmetry element: two-one screw axes, glide planes, and points of inversion. cellulose- $I_\beta$  has a shift along the  $c$ -direction like the  $\alpha$  form but the key difference is that it does not

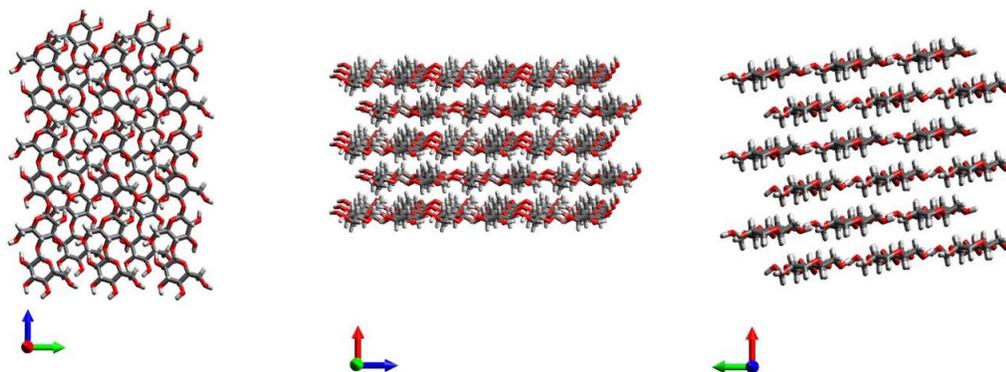


FIGURE 1.7: Cellulose- $I_{\beta}$  along yz-plane(left), xz-plane(middle), xy-plane(right). The x, y, z-axes are represented in red, green and blue respectively.

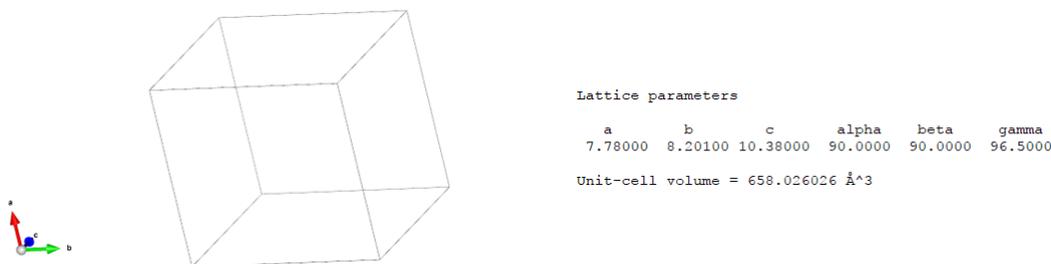


FIGURE 1.8: Cellulose- $I_{\beta}$  unit cell and lattice parameters.

have diagonal shear but alternates between adjacent sheets. The cellobiose along the c-directions has a diagonal shift of 0.26 nm. It has an alternate origin and centre chain sheets held together by hydrogen bond.

Cellulose- $I_{\beta}$  crystal is made up of two different sheets, they differ based on the cellobiosyl units which make up the unit cell. The origin units make up the flat sheet and the center units make up the slightly slanted sheets. Different hydrogen bonding network has been observed in these sheets. The different hydrogen bond patterns were studied by [Nishiyama et al.](#), using synchrotron X-ray and neutron diffraction<sup>[50]</sup>. Their

work provided a clear overview of the complexity and intricacy of hydrogen bonding interactions within cellulose crystal structures. Hydrogen bonds are attributed to some of the special properties and characteristics exhibited by cellulose and its different structural morphologies like fibre twists etc.<sup>[31]</sup>

Figure 1.7 shows the different axial views of a simple structure with three unit cell lengths in all crystallographic axes of cellulose-I<sub>β</sub> and fig. 1.8 shows the lattice parameters and unit cell of one cellulose-I<sub>β</sub> unit. Cellulose-I<sub>β</sub> is significant among the other polymorphs due to its easy availability since it occurs naturally and does not need any chemical alterations except for acid assisted hydrolysis, which yields CNC from microfibrils. Hydrogen bonding within cellulose crystals and chain strongly influence their characteristics such as insolubility in polar solvents. Within the scope of this thesis, all interactions have been studied for the different surfaces of cellulose-I<sub>β</sub>.

### 1.1.2.3 Surfaces (100) and (010)

In the above sections, the different crystal forms of cellulose were discussed. Each of these crystalline forms consists of different crystallographic planes, which vary in their properties. In case of cellulose-I<sub>β</sub>, (100) and (010) are the surfaces of interest. The monoclinic unit cell structure has three major planes along which the molecular configurations vary. The other three planes are considered to be mirror images. The crystal configuration along the z-direction is not considered usually as it is taken as the crystal growth direction. This will be discussed in detail in chapter 2 along with other model attributes. Two planes, one with an intercept in the x-axis, (100) and one with an intercept in the y-axis, (010) are considered for simulation studies in the presented work.

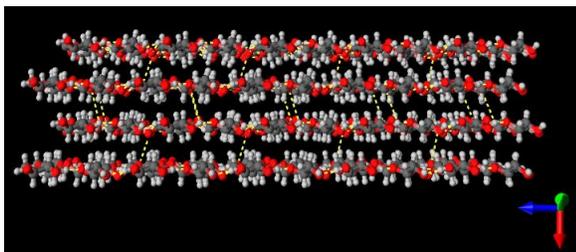


FIGURE 1.9: Interchain hydrogen bonding in cellulose-I<sub>β</sub>. The x, y, z-axes are represented in red, green and blue respectively.

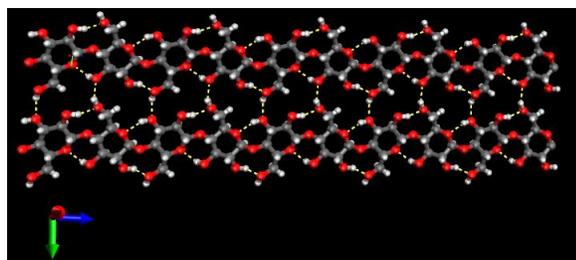


FIGURE 1.10: Interchain and Intrachain hydrogen bonding in two origin chains of Cellulose-I<sub>β</sub>. The x, y, z-axes are represented in red, green and blue respectively.

### 1.1.3 Hydrogen bonding

Hydrogen bonding has strong implications on the behaviours of cellulose interactions with solvents, ionic<sup>[66,78]</sup> or polar<sup>[32]</sup>, and polymers<sup>[37]</sup>. The interaction varies at various sites along the cellulose surface, depending on the orientation of -OH groups, but, it strongly influences the cellulose interactions with solutes and solvents. Cellulose-I has strong interchain and intrachain hydrogen bonds. Its complex hydrogen bonding network has been studied earlier using neutron fiber diffraction<sup>[50,51]</sup>. Some sample hydrogen bonding network configurations observed in Cellulose-I<sub>β</sub> are shown in figs. 1.9 and 1.10.

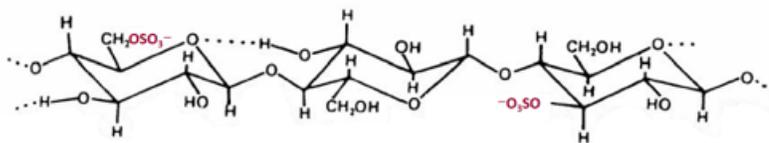


FIGURE 1.11: Sulphate groups at the C6 position.

### 1.1.4 Sulphated surface

CNC is prepared by treating the cellulose source with mineral acids like hydrochloric acid (HCl), sulphuric acid (H<sub>2</sub>SO<sub>4</sub>) etc. Depending on the mineral acid used, the surface functionality varies and the surface charge is altered subsequently, but the dimension and actual morphology of the synthesized CNC depend on various factors like source, acid hydrolysis conditions and the crystallinity of the source itself. Within the scope of this thesis, the simulated systems consist of sulphated CNC which are experimentally produced by hydrolysis using H<sub>2</sub>SO<sub>4</sub>.

The surface of crystals produced by this method has sulphate half ester groups. Figure 1.11 shows the sulphate groups bonded to the C6 positions along a cellulose chain. The surface group concentration and the ratio of sulphate groups to cellulose rings are discussed under the design aspect in subsection 2.2.1.3. The C6-OH (hydroxyl groups) are easily accessible along the (010) plane of cellulose-I<sub>β</sub>. Hence, this surface is sulphated. Sulphated CNC shows distinct properties in solutions and emulsion such as emulsion stabilization owing to their surface charges<sup>[24-26]</sup>.

## 1.2 Motivation

CNC has great potential to be used as a green alternative material to steel or carbon nanotubes due to its strong mechanical<sup>[75]</sup> properties and as stabilizer due to its physical interactive properties<sup>[24-26]</sup>. Cranston and Gray characterized some interesting morphological and optical properties of polyelectrolyte multilayers by incorporating nanocrystalline cellulose<sup>[7]</sup> in 2006. A spectrum of unique properties has been discovered since then. These properties have had some notable application in a variety of systems like gels<sup>[24]</sup>, colloids<sup>[26]</sup>, emulsion<sup>[25]</sup> and films<sup>[6,8]</sup> etc. Naturally, a system with such versatility will be studied intensively to explore newer properties in order to be applied in the production of novel materials.

CNC along with polymers, especially cellulosic polymers has given rise to many interesting observations<sup>[23]</sup>. Notably many interfacial properties have become topics of interest in the past decade. Sulphated CNCs form stable interfacial systems with CP like hydroxyethyl cellulose (HEC) and MC. These physical occurrences have become a subject of interest in recent days<sup>[3,76,77]</sup>. Hu et al. have studied the gelation capability of CNC and other polymeric materials<sup>[24]</sup> like HEC, MC, locust bean gum (LBG), hydroxypropyl ugar (HPG) and dextran (DEX). In most cases, they found good interactions between the pairs and also that the performance of CNC-polymer systems was better than that of CNC or polymer individually. The observed properties were attributed to the CNC-polymer interaction present in these systems. Another series of notable systems developed due to CNC-polymer interactions were re-dispersible Pickering emulsion and stable oil-water emulsion etc.<sup>[25,26]</sup>

In their cases, the main observation was the ability of CNC-polymer pairs especially CPs in stabilizing emulsion droplets of various sizes. Again enhanced stabilization was observed in case of the pairs compared to the cases of individual components. They quantified the droplet size using scanning electron microscopy (SEM) and confirmed the stability by measuring the mean droplet diameter, the pendant drop experiments were carried out by them to assess the adsorption of polymers and CNCs via interfacial tension measurements. In all these experiments they found proofs for interaction between polymers and the system stability. Similar proofs were provided in case of the gel systems.

One underlying factor, which is the foundation for all these observations was not explained clearly and that is the actual molecular interactions between CNC and the polymers. The key to understanding and applying any system's properties lie on our understanding of that system's molecular interactive property and that has been not yet been completely unravelled in case of CNC-polymer systems. This gap in our understanding when it comes to CNC-polymer systems is our main motivation for studying these systems. Though it is possible to detect and quantify these interactions, it is difficult to observe and pinpoint specific molecular interactions via experimental methods. It can be done systematically using simulation methods like molecular dynamics. Hence the significant void in our understanding when it comes to the molecular interactions within these systems is the main motivation behind us taking up this challenging topic.

### 1.2.1 Physical understanding

We expressed the reason behind our interest in exploring CNC-polymer systems in the previous section. In this section, we will give a physical overview of what has been done in this study and the reasons behind it. Our motive is to understand the CNC-polymer interactions in different conditions such as different crystal surfaces, presence or absence of salts etc. So we chose two oligomers, 1. methyl cellulose (MC) and 2. cellopentaose, which is a short cellulose oligomer chain with five glucose units and simulated them over two crystal surfaces ((100) and (010)) of CNC. The reason being, this way we can estimate the extent of CNC-MC interactions with respect to CNC-cellulose chain interactions.

The surface morphologies are different for both (100) and (010) surfaces. The (010) surface carry the sulphate half ester groups ( $\text{OSO}_3^-$ )<sup>[57]</sup> due to the acid hydrolysis, this was discussed earlier in subsection 1.1.4. In addition to this, the (100) surface is hydrophobic, while (010) is hydrophilic. Hence it is necessary to study these combinations in parallel. Also, these combinations were simulated under two different scenarios, with and without the presence of salt. It was experimentally observed that the presence or absence of salt resulted in different degrees of CNC-polymer interactions so the second set of combinations with salt to observe its effects on the interactions were modelled and simulated. Hence in total eight systems were modelled and simulated in this study.

The physical adsorption of the polymer chains on to the surface of CNC can be attributed to many physical interactions such as the formation of new hydrogen bonds

between the substrate and polymer, hydrophilic or hydrophobic interactions between bulk and the crystal etc. The system is studied in presence of a polar solvent, water. Hence the hydrogen bonding interaction may include the solvent-solute-solvent bridging effects. So a series of analysis has been carried out to pinpoint the contribution of different interactions observed in these systems.

#### **1.2.1.1 Chain length**

One key factor to note is the chain length of the oligomers used in this study. The length of the oligomers is 5 units in both the cases of MC and pentaose. There are two major reasons for this, 1. The oligomer chain if longer than the substrate i.e. CNC will interact with its own periodic image and will result in unrealistic dynamics and 2. The main focus is to understand the molecular interaction between the oligomer and CNC hence a smaller chain would be easier to simulate and follow along to pinpoint individual interactions. Another key reason is the computation cost, simulating a long oligomer chain would be very costly and in most cases will not provide much useful information. Hence a decision to simulate a smaller chain which can be a useful tool in setting up a larger system was made. The observations and data obtained from these simulations can be used to model complex and larger CNC-polymer systems. To replicate larger scale interactions, the potentials and other information have to be obtained by extensive runs and analysis of smaller scale systems. The model and system setup are discussed in [chapter 2](#) and it will shed more light on this aspect of the model development.

### 1.3 Objective

The overall objective of this work is to determine the actual surface interactions occurring between the substrate, sulphated cellulose nanocrystal (S-CNC) and solutes, the oligomers. Also, we wish to establish whether other physical interaction such as hydrophilicity and/or hydrophobicity, the presence of salt etc. have any impact on the adsorption. The interactions between two oligomers on two different surfaces of the sulphated CNC were studied. One of the main focuses is to discuss to what degree the MC adsorbs on to the CNC surface relative to a short cellulose chain and to determine the polymer specific interactions exhibited in similar conditions. The objectives are summarized here.

1. To determine the extent of interactions between different oligomers and different surfaces of CNC under different conditions.
2. To determine the driving force for adsorption.

### 1.4 Literature review

Polymer-polymer interaction has always been a topic of interest even since the 1950s, cellulose is one of the most widely used and studied bio-polymer owing to its mechanical<sup>[40]</sup>, chemical and interfacial properties<sup>[24-26]</sup>. The adsorption of polymer on the surface of CNC or CNF causes interesting changes to their properties in bulk systems like water binding capacity and aggregation. Ahola et al. have studied the effects of polymer adsorption on cellulose nanofibrils and they have given some

interesting observation as to how these properties vary from polymer to polymer using quartz crystal microbalance (QCM) with dissipation, confocal laser scanning microscopy (CLSM) and surface plasmon resonance (SPR)<sup>[1]</sup>. They studied the interaction of cellulosic fibril with three polymers namely, polydiallyldimethylammonium chloride (PDADMAC), carboxymethyl cellulose (CMC) and xyloglucan. They concluded based on their observations that at the interfacial level, the type of polymer adsorbed affects the fibril properties and its water content. They also observed its impact on macroscopic properties such as suspension aggregation. Their study presented the major changes observed in the properties of the polymer-polymer pair in bulk but did not give any insight into the actual polymer interactions.

The CNC adsorbed with polymer exhibits a wide range of properties from foam formation<sup>[27]</sup> to emulsion stabilization<sup>[25]</sup> etc. [Hu et al.](#) studied the capacity of CNC-MC pair in forming stable aqueous foams. They observed that CNC and MC are together better stabilizers than CNC or MC alone. The CNC decreases the air bubble size and subsequently the initial foam densities. Even with a low concentration of about 2 and 0.5 weight % for CNC and MC respectively, low-density foams were observed by them. They have proposed that the CNC adsorbs on to the large coils of MC, which was eventually attributed to the foam stabilization properties of MC. Even at a high temperature of 70°C, they have observed stable foams for 6 hours. Same as the previously discussed work they have reported some interesting interactive properties and observation but there is no account of the actual molecular or morphological interaction. They have suggested a set of possible schemes of the CNC and MC bindings but they were not able to test them directly.

Hence, it necessitates the investigation carried out in this work. The presented study could be very pivotal in determining the CNC-MC interactions. In another study by [Hu et al.](#) detailing the emulsion and emulsion gel stabilization properties of water-soluble cellulosic polymers and CNC, they have observed enhanced performance as in the case of CNC-MC in foam stabilization<sup>[27]</sup>. In this work, they studied the CNC-polymer effect on emulsion stabilization for three cases CNC-HEC, CNC-DEX and CNC-MC using CLSM and QCM. Similar to a previous work<sup>[27]</sup> they observed enhanced stabilization properties than exhibited by CNC or MC individually, except for CNC-DEX case as they did not observe stable gel formation in its case even at high concentration.

They found CNC to be a good emulsion stabilizer as it stabilized the oil-water emulsion without any visible creaming or separation. They have observed and reported a wide set of results which includes droplet size of the emulsion, oscillatory rheological measurements, emulsion stability, interfacial tension etc. The molecular interaction part still remains unclear. This lack of insight into the fundamental interactions which result in such a spectrum of CNC-polymer properties has not been addressed yet.

The surface interaction can be reasoned by physical interactions like hydrogen bonding between solute and substrate. In case of carbohydrates like cellulose and cellulosic polymers, this interaction is significant as cellulose structure be it fibril or crystalline does exhibit an excess of it and this has been accounted for its structural properties and stable morphologies<sup>[50,51,70]</sup>. The surface free energy of desorption will be a good estimate of the degree of affinity for each polymer towards CNC, relative to a cellulose chain. These measurements can be estimated using MD analysis. There are some well-established parameter sets for carbohydrates<sup>[34]</sup> and this makes these calculations

highly reliable. [Muthukumar and Khare](#) have done interesting works on dynamics of cello-oligosaccharides<sup>[54]</sup> and the free energy of desorption from CNC surfaces<sup>[49]</sup>. They have conducted a study on the free energy of desorption for a celohexaose (cellulose chain with six glucose units) from two cellulose-I<sub>β</sub> surfaces namely (100) and (110).

They calculated the free energy of desorption of celohexaose, for three different cases of  $\xi$ , on two different cellulose-I<sub>β</sub> surfaces, (100) and (110).

- **case i:**  $\xi$  normal distance between the centre of mass (COM) of the first layer of CNC and COM of the first glucose residue/ring.
- **case ii:**  $\xi$  normal distance between the COM of the first layer of CNC and the COM of last glucose residue/ring.
- **case iii:**  $\xi$  normal distance between the COM of the first layer of CNC and the COM of the whole glucose chain.

Similar works presenting free energy of decrystallization of cellulose-I<sub>β</sub><sup>[52]</sup>, nanoscale decrystallization thermodynamics & examination of  $\alpha$ -Chitin structures<sup>[2]</sup> have been conducted. [Muthukumar and Khare](#) had similar results which are consistent with previously conducted free energy studies and their PMF profiles<sup>[2,52]</sup>. [Beckham and Crowley](#) have obtained PMF curves for cases similar to that of the ones tested by [Muthukumar and Khare](#). The different cases were either edge of the chain or the whole chain being considered for PMF calculation. [Muthukumar and Khare](#) have cited the two types of hydrogen bonds, conventional O-H-O<sup>[52]</sup> and non-conventional C-H-O<sup>[17]</sup> for the molecular interaction between the oligomer and crystal structure which corresponds to different sections of the PMF profile.

They have detailed the different zones of PMF curve with different intermolecular and intramolecular hydrogen bonds. Similar PMF curves, observations and reasonings were made by them for the different crystal surfaces. The slight differences in the PMF values between their observations and previously existing studies<sup>[52]</sup> was attributed to the difference in forcefields, COM position and residue units etc. Their work gives a detailed view on the PMF values resulting from hydrogen bonding interactions expected in an oligomer chain during the desorption process on two different surfaces.

They have conducted the free energy calculation using Canonical ensemble (NVT), while most of the dynamics were carried out with Isothermal-isobaric ensemble (NPT) ensemble. The theoretical background of ensembles will be discussed in detail in chapter 2. They have removed other layers of the CNC except for the top layer involved in the PMF calculation for the sake of computational efficiency, but these layers might fall within the interaction range for higher non-bonded interaction cut-off values. It is unclear whether the solvent effect has any implications on the polymer's hydrogen bonding network and this has not be accounted for in their study.

Also, they have used different force constant for different cases which influence the biasing potential and might result in slight differences in the PMF curve. Fixed force constant value will be better to compare multiple systems with similar components such as the ones in the presented study. They have sampled the ensemble for 1 nanoseconds (ns), an extended period of sampling will also have a strong impact on the PMF curve. The longer the system gets sampled, the better the PMF value corresponds to the actual occurrence. Their study has shown and detailed the various interactions during

desorption of short cellulose oligomer and CNC, but it cannot be extended directly to the CNC and oligomer systems in this presented work.

We are in need of a model or simulation study which includes bulk system interactions like solvent bridging effect or impact of salt concentration. Also, a common frame of reference in terms of applied biasing potential is important. The applied bias is related to the force constant used in the PMF calculations. The NPT ensemble will be a better fit and more relatable to describe the system as the presented work roots its motivation on actual experimental studies<sup>[25,26]</sup>. Overall [Muthukumar and Khare](#)'s study deals with the desorption of a short cellulose oligomer chain from two different surfaces of CNC. The PMF curves and free energy differences were calculated for the same components and initial configurations. The presented work compares different oligomer's interaction on different surfaces of CNC. The presented work addresses the above-mentioned requirements to bridge our gap in understanding the actual bulk or molecular interactions involved in reasoning the properties exhibited by CNC-polymer pairs.

## 1.5 Thesis outline

- **Chapter 1:** This chapter discusses the backgrounds of the simulated physical systems. It gives details about cellulose's chemical nature, physical occurrences, structural morphologies and their properties in general. The crystalline structure, their importance and physical properties were reviewed in the initial sections of this chapter. In section [1.2](#) and its subsections, the motivation for this study and an overview of the planned study was discussed in detail. Currently existing work and

studies which have been carried out and the significance of the presented work in comparison to those were pointed out towards the end of this chapter.

- **Chapter 2:** This chapter discusses the methods, software and techniques used to study the system interactions. The model developed for the MD analysis, the various constraints and references followed for these model will be discussed in this chapter. The process steps used in setup, simulation and validation are explained in detail in this chapter. The justification for the choice of forcefield used for the model and the model evaluation will be discussed in general. Also, in this chapter, the various simulation conditions and strategies used to conduct the dynamic study will be discussed. Various analysis methods used, their background, inter-relation and their significance to the studied properties will be discussed in this chapter.
- **Chapter 3:** This chapter includes all the results obtained from the simulation studies carried out on the various systems mentioned in the introduction and model sections of chapters 1 and 2. The major results include system interactions in bulk and at short range, molecular diffusion etc. The free energy of desorption for different oligomer units on different surfaces will be discussed in this chapter. Other properties like the radius of gyration, solvent accessible surface area and interaction energies etc. will be compared in parallel for different systems.
- **Chapter 4:** This chapter summarizes the major findings and contribution of this work.

# Chapter 2

## Methodology

### 2.1 Molecular Dynamics

The development of computers triggered the widespread use of numerical methods. Computers were and are still being used to solve complex and huge problems, which were and are nearly impossible to approach by conventional and analytical methods. One of such problems is the modelling of chemical and physical systems. MD is a method used to simulate the movements of atoms and molecules by following a set of constraints and utilizing computational techniques.

MD had its roots in the computer simulation work carried out by Alder and Wainwright in 1957 to study the phase transition in hard sphere collision, these days it has become a standard method in Computational Chemistry and Chemical Engineering. A system can be modelled by observing the interaction of its subunits, atoms/molecules for a specific period of time ranging from femtoseconds (fs) and sometimes to microseconds ( $\mu\text{s}$ ). Based on the dynamics, the system properties can be predicted or evaluated. MD is a

very effective tool to model large-scale intricate biomolecular systems via all-atom or coarse-grained simulation methods. The principle which serves as the backbone for MD is very straightforward. A given system will have a number of interacting particles, they could be atoms, molecules or beads etc. held inside an imaginary two dimensional (2D) plane or three dimensional (3D) box.

These interactions are governed and constrained by a set of rules which are described in the form of potential functions called 'Forcefields' in general. These interactions are carried out over a span of time until the system stabilizes and gives statistically meaningful information. From the simulation data, the trajectory of each individual entity in the simulation box is obtained and it is used to make a time evolution of the system, which is used to extract system information.

### **2.1.1 Theory**

MD method is a simulation technique used to find the evolution of a given system, some rapid movements like the movement of electrons are averaged out as it is instantaneous compared to the relatively slow movements of atomic nuclei and atoms (Born-Oppenheimer approximation). The Schrödinger equation,

$$\frac{\hbar^2}{2m} \nabla^2 \Psi + V(\mathbf{r})\Psi = -i\hbar \frac{\partial \Psi}{\partial t} \quad (2.1)$$

describes the changes a system undergoes by inculcating quantum effects. It can be segregated into a time-dependent equation for nuclei and the time-independent equation for electrons. But in the case of classical MD simulation methods, the interactions are

considered purely classical and are governed by classical mechanics. Hence, the quantum part of the nuclei equation too is neglected. The nuclei propagate in the given space following the potential provided by the forcefield in accordance to Issac Newton's equations of motion. This observation is elevated on longer time and length scales to make a meaningful observation of the system.

### 2.1.1.1 Newton's equation of motion

For a given system with N particles, the coordinates of the N individual particles form the potential function,

$$U(r_1, r_2, \dots, r_n) \quad (2.2)$$

since the system follows classical mechanics, the particle dynamics can be solved numerically using Newton's equation of motion,

$$\vec{F} = m\vec{a} \quad (2.3)$$

which in its differential form is,

$$\vec{F}_i = -\frac{\partial U}{\partial r_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2.4)$$

where the acceleration is given by,

$$\vec{a}_i = \frac{d^2 r_i}{dt^2} = \frac{d\vec{v}_i}{dt} \quad (2.5)$$

The  $\vec{F}$  is the force on a particle of mass  $m$ , and  $\vec{a}$  is the particle's acceleration,  $t$  is the system time,  $r_i$  is the coordinate of the atom  $i$  and  $\vec{v}_i$  is the velocity of the atom  $i$ . The force  $\vec{F}$  is the gradient of the potential function. The system evolution is captured by solving these equations for a predetermined time step, which gives the system trajectory.

### 2.1.2 Scope of MD

One of the practical limitations placed on MD method from extending it to a very large system is the availability of computational resources, all-atom simulation even with the incorporation of impressive simulation strategies can best reach  $\mu$ s time scale. Hence the MD simulation method has a practical restriction on the time and length scale of the problem. Usually, simulation boxes of a few  $100 \text{ nm}^3$  in size for a few ns in time are simulated via all-atom simulation methods.

There are other methods like united-atoms, where in general smaller atoms like aliphatic hydrogen atoms in an aliphatic compound or aromatic ring are grouped together with their respective carbon(C) atoms. The forcefield like GRONingen Molecular Simulation (GROMOS) used in GRONingen MACHine for Chemical Simulations (GROMACS)<sup>[22,69]</sup> is one good example of such an approach.

There are coarse-graining techniques, where the larger units like a whole aromatic ring are considered to be interacting beads connected by a spring. Using this method larger time and length scales can be simulated. But, each method has its own benefits and

drawbacks. To study system specific or in some biomolecular cases where one needs to study site-specific interaction, all-atom simulation methods come in handy.

### **2.1.3 Periodic boundary condition (PBC)**

It was mentioned in the previous section that the size of the simulated systems in MD are usually very small when compared to the macroscopic systems. The size ratio will be different than in reality and this would result in the MD simulation ending up with insignificant or artificial observations. To overcome this challenge, periodic boundary condition (PBC) is applied to the simulation box where the simulation box is surrounded by the periodic images of itself. These are identical images of the central/principle box.

The movements are identical in these boxes, in the sense, any particle exiting the central box from one end immediately reenters the box through the other end and this can be extended to all three dimensions. This provides an infinitely long simulation box. Figure 2.1 shows a 2D simulation box exhibiting PBC. The red particle seen exiting from the top boundary reenters the bottom boundary along the same directional vector.

Now we can extend this to all three dimensions where a particle exiting the simulation box through one of its sides will reenter the simulation box through the side which is perpendicular to the exit face.

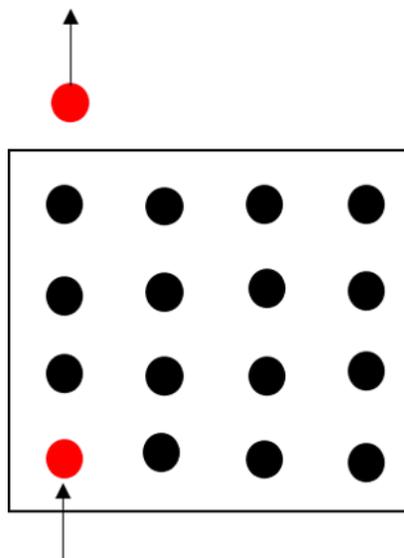


FIGURE 2.1: 2D simulation box showing PBC. The red particle, which is seen exiting through one direction can be seen reentering through the other face along the same direction.

### 2.1.4 Thermodynamic ensembles

Ensembles or statistical ensembles are the probability distribution of the state of a system. It is also used to describe the small set of states sampled from a full set of available states for a given system. Thermodynamic ensembles are a specific case of statistical ensembles, where they are in statistical equilibrium and are used to predict thermodynamic properties of a system.

The most common and principle thermodynamic ensembles include NVT, NPT, Microcanonical ensemble (NVE) and Grand canonical ensemble ( $\mu$ VT). Based on the ensemble, some of the system parameters like the number of particles, volume, pressure and temperatures are kept constant while others are allowed to vary. Table 2.1 gives the list of ensembles described by Gibbs and their corresponding constant parameters.

TABLE 2.1: Thermodynamic Ensembles and their constant parameters.

Ensemble	Representation	Constant parameters
Isothermal-isobaric	NPT	number of particles N, pressure P and temperature T
Canonical	NVT	number of particles N, volume V and temperature T
microcanonical	NVE	number of particles N, volume V and energy E
Grand canonical	$\mu$ VT	chemical potential $\mu$ , volume V and temperature T

### 2.1.5 Forcefields

The forcefields are the backbone of any MD simulation method. In order to obtain reliable results from MD simulations, the forcefield used must be of a good quality. The development of forcefields is a labour intensive job, involving multiple layers of verification and modelling until it reflects the desired and observed system properties. No forcefield is complete, each forcefield is designed to reflect a specific factor of the system. Hence the choice of parameter is a crucial part of any MD simulation techniques.

Parameter sets of the same particles vary from forcefield to forcefield. There is always a

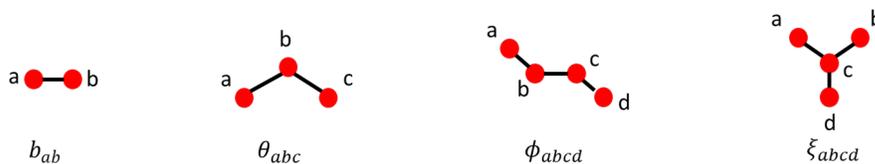


FIGURE 2.2: Bonds, Angles, Dihedrals and Improper dihedrals.

balance between accuracy and interactions. Neglecting interactions may result in a reduction in accuracy. Hence, care is always taken while parameterizing forcefields. Though interactions can be segregated into various types, in general, forcefield

interactions can be reduced down to bonded and non-bonded interactions. Figure 2.2 shows the set of bonded interactions observed in a common system. The simulation facilitates particle interaction based on the provided potential, from which subsequently the forces, velocity and trajectory for each individual constituent are determined.

The forcefield potentials as mentioned earlier are a function of positions of the particles and can be represented as

$$U(r) = U^{\text{bonded}}(r) + U^{\text{non-bonded}}(r). \quad (2.6)$$

Where  $r$  is the position of each individual particle and  $U$  is the potential function. The bonded interactions are restricted only to atoms which are bounded by actual bonds.

As the number of particles increases, their interactions vary from collinear, coplanar to inter-planar via bonds, angles and dihedrals respectively. On the other hand, the non-bonded interactions are for all the atoms in the system. Though we restrict the extent of interaction with a distant cut-off, in theory, all particles involve in non-bonded interactions and they can be further segregated into Leonard-Jones and electrostatic/Coulombic interactions. The above-mentioned interactions and the corresponding potentials are represented as follows,

$$U^{\text{bonded}}(r) = U^{\text{bonds}}(r) + U^{\text{angles}}(r) + U^{\text{dihedrals}}(r) + U^{\text{improper-dihedrals}}(r) \quad (2.7)$$

$$U^{\text{non-bonded}}(r) = U^{\text{LJ}}(r) + U^{\text{coulomb}}(r). \quad (2.8)$$

### **2.1.6 Limitations of MD**

It would have been evident from the above discussion that though MD is a very flexible and highly customizable technique, it does have some limitations which prevent its potential from being extended to a whole variety of systems. The first and foremost one is the limited time and length scales, very large systems cannot be simulated for an extended period of time. Lack of generalization for forcefields prevents them from being utilized for a range of systems.

Most forcefields are modelled bearing a specific system and/or a set of specific properties. In addition, parametrization of forcefields is a tedious process. But, fortunately, standardized force fields can be found for most organic and biological molecules as well as a wide variety of synthetic polymers. A well-validated forcefield coupled with a robust simulation software enables one to model complex system interactions and observations which are difficult to observe or explain on macroscale with ease.

## **2.2 Molecular Models**

This section is dedicated to explaining the various physical systems, modelling techniques, routines, software and custom-codes used to set up the initial model for MD simulations in this study, each software is explained at the first occurrence along the process flow.

## 2.2.1 Physical systems

In this section, the various physical systems modelled in this work will be discussed. The CNC-oligomer interactions are the primary focus of this study. In order to study the relative interaction of different oligomers with the different surfaces with and without the presence of salt sodium chloride (NaCl), a total of eight systems were simulated. The simulated CNC was of type  $I_\beta$ , and the two different oligomers were methyl cellulose and a short cellulose chain (pentaose), both with 5 monomer units.

They were simulated on two different surfaces of the crystal namely (100) and (010). The water model used for all the systems was the Three-site transferrable intermolecular potential (TIP3P) model. [Jorgensen et al.](#) have made a detailed comparison between different water model. The justification for the selection of this particular model is provided in subsection 2.2.2 and section 2.3. Table 2.2 lists the nomenclature which will be used in the rest of the thesis.

### 2.2.1.1 (100) surface systems

In these systems, the interaction of oligomers with the (100) surface of the CNC was studied and salt, if used, were maintained at a concentration of 50 mM. The crystal structure was standard for similar surfaces, i.e. in terms of unit cell count and periodicity. This was done to maintain consistency in observation and extent of interactions. The simulated CNC had unit cell lengths of (2, 8, 6) in ( $x$ ,  $y$ ,  $z$ ) directions respectively. No specific surface modifications were made to this surface. The difference

TABLE 2.2: Simulated system nomenclatures and specifications.

System	Crystal surface	Crystal unit cells (x, y, z)	Oligomer	Salt
286mc	(100)	2, 8, 6	MC	no
286pe	(100)	2, 8, 6	Pentaose	no
1026mc	(010)	10, 2, 6	MC	no
1026pe	(010)	10, 2, 6	Pentaose	no
286mcs	(100)	2, 8, 6	MC	yes
286pes	(100)	2, 8, 6	Pentaose	yes
1026mcs	(010)	10, 2, 6	MC	yes
1026pes	(010)	10, 2, 6	Pentaose	yes

in surface functionality was discussed earlier in subsection 1.1.4.

The CNC surface bears functional groups depending on the type of acid hydrolysis carried out to separate them. The (100) surface does not have any surface C6-OH groups, which are readily available to interact with  $\text{H}_2\text{SO}_4$  and form sulphate half ester groups. Hence, this surface was left unaltered. The normal direction was along  $x$  and the surface plane was the  $YZ$  plane. Figure 2.3 shows an MC oligomer chain placed near the (100) plane. The CNC can be seen with 4 layers i.e. two unit cells in thickness along the  $x$ -direction. The oligomer placed about 1.5 nm away from the surface so that it just falls within the interaction zone of CNC.

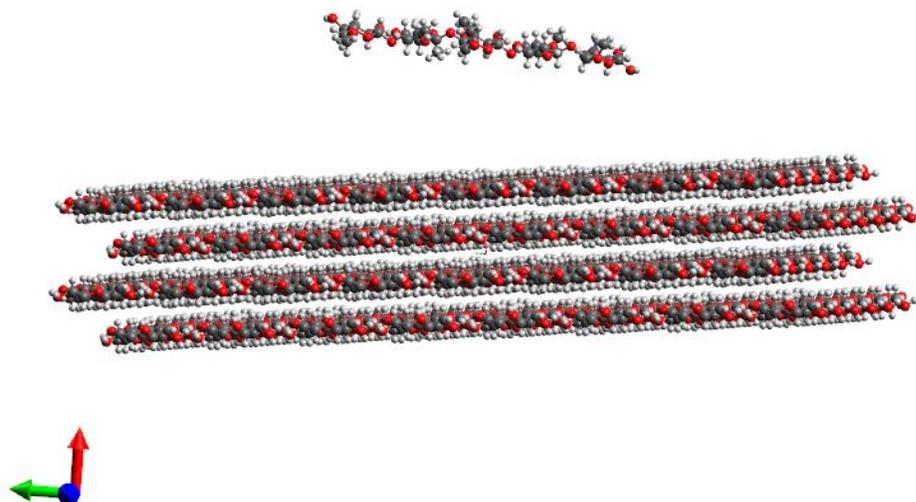


FIGURE 2.3: MC over the (100) surface. The  $x$ ,  $y$ ,  $z$ -axes are represented in red, blue, green respectively. The water molecules are not shown for clarity.

### 2.2.1.2 (010) surface systems

In these systems, the interaction of oligomers with the (010) surface of the CNC was studied and salt, if used, were maintained at a concentration of 50 mM. The crystal structure was standard for similar surfaces, i.e. in terms of unit cell count and periodicity to maintain consistency in observation and extent of interactions. The simulated CNC has unit cell lengths of (10, 2, 6) in ( $x$ ,  $y$ ,  $z$ ) directions respectively. The normal direction was along  $y$  and the surface plane was the  $XZ$  plane. Figure 2.4 shows an MC oligomer chain placed near the (010) plane.

It can be noted that the unit cell count along the normal direction was maintained at a constant value of '2' for both the surfaces. This was to make sure that the extent of

interaction and simulation box size was similar in both the cases. Because the water molecule count depends on the box size as the mass density of water was used to calculate the number of water molecules. The unit cell count on the other two directions was adjusted accordingly to give a stable crystalline structure. The oligomer is closer to the surface as this particular snapshot was captured from a dynamic run.

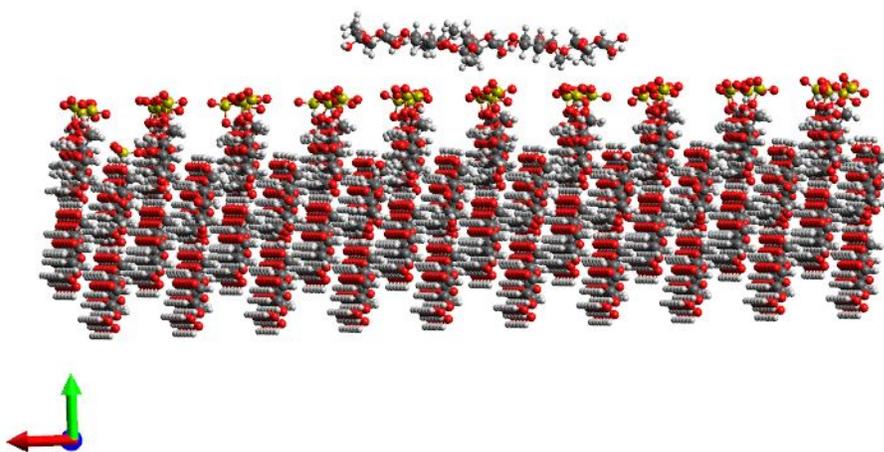


FIGURE 2.4: MC over the (010) surface. The x, y, z-axes are represented in red, blue, green respectively. The water molecules are not shown for clarity.

### 2.2.1.3 Surface modification

It was briefly mentioned in subsection 1.1.4 that because of the acid hydrolysis process the CNC surface is modified with sulphate half ester groups. The experimental system based on which this work has been carried out used CNC processed using  $\text{H}_2\text{SO}_4$ . Hence they bear surface sulphate groups and are sulphated-CNCs. The surface sulphate substitution is mainly observed on the (010) surface. The acid attack takes place in the amorphous

region of the microfibril and the resultant CNC will have sulphate half ester groups ( $\text{OSO}_3^-$ ). These groups are available at the C6 position since they are readily accessible perpendicular to the contact plane of CNC. Figure 2.5 shows a segment of the (010) surface which is grafted with sulphate groups along the +y direction. The surface charge

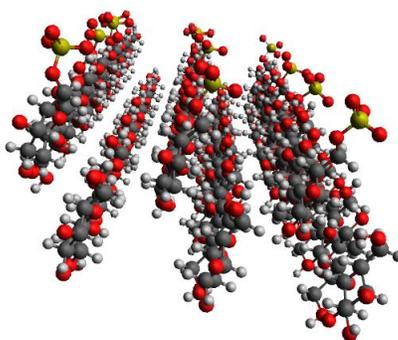


FIGURE 2.5: Sulphate groups on (010) surface.

is determined by the half ester sulphate groups to surface rings ratio, which according to the observations made by Reid et al. is one sulphate group for every two to three glucose rings. The lower bound of the surface sulphate group concentration, i.e., one sulphate group per three rings, was chosen for modelling efficiency. The (010) and its periodic image surface (020) had a total of 120 accessible surface C6-OH groups and the crystal was modified to have a 1:3 ratio of sulphate to glucose rings. Hence a total of 40 sulphate half ester groups ( $\text{OSO}_3^-$ ) were grafted at different sites to finalize the (010) surface.

### 2.2.2 Water model - TIP3P

It was mentioned earlier that the water model chosen for the systems is the TIP3P model. Two of the main reasons for the selection of this model over other models are (1) the forcefield chosen for cellulose and methyl cellulose in this study is highly compatible

with the potentials of this water model. The forcefield will be discussed in details in section 2.3 and (2) the TIP3P partial charge distribution and potential usually provide a well-equilibrated and uniformly dense system.

The second reason is quite significant, as when simulating liquids caution must be maintained to make sure the periodic images do not overlap as the non-bonded interactions will make the system unstable. This choice is further discussed in section 2.3. Figure 2.6 shows that the mean-squared displacement (MSD) of the water molecules as the function of time. The self-diffusion coefficient is derivative of the MSD curve.

The self-diffusion coefficient of the TIP3P model used with the systems is consistent with the observed values of  $4.6 \times 10^{-5} \text{ cm}^2/\text{s}$ <sup>[42]</sup>. As in the case of most water models, TIP3P is also higher than the observed experimental value of  $2.3 \times 10^{-5} \text{ cm}^2/\text{s}$ <sup>[48]</sup>. Figure 2.7 shows the uniform density distribution observed in a well-equilibrated 286mc system's dynamic run. The region between  $-10 \leq x \leq 10$  is where the CNC crystal was located. There are no signs of locally dense zones, which is a good indication of the compatibility of the water model and the rest of the system.

## 2.3 Forcefield - GLYCAM06j

It was mentioned in fig. 2.2 that the forcefields are the backbone for MD simulations and that the choice of forcefield is critical to obtain accurate results. Also, that the forcefields are developed for specific systems and to replicate specific interactions. The

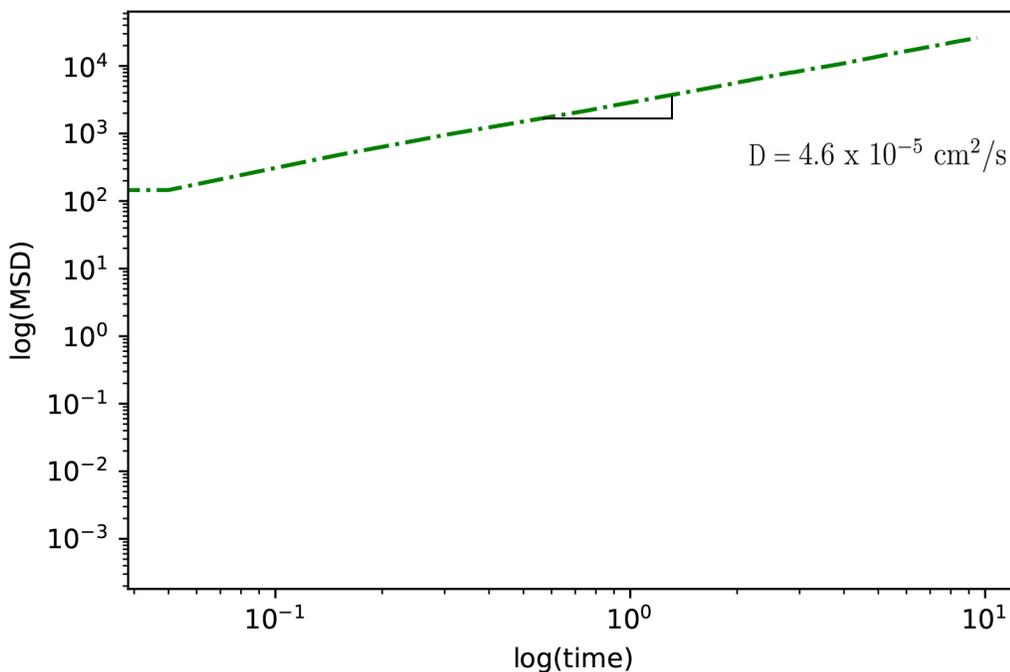


FIGURE 2.6:  $\log(\text{MSD})$  vs  $\log(\text{time})$ . The plot shows slope along the log-log curve used to calculate the self-diffusion coefficient value of water.

GLYCAM\_06j-1<sup>[34]</sup> was chosen as the forcefield to simulate the systems in this study. GLYCAM06 was developed by [Kirschner et al.](#) specifically for the carbohydrates and it is highly consistent with already existing AMBER forcefields<sup>[71]</sup>. [Kirschner and Woods](#) found that the solvents interactions influence the carbohydrates conformations, which emphasize the use of consistent water model. The GLYCAM06 forcefield was developed with water potential from the TIP3P model.

This is another reason for choosing the TIP3P model over other existing water models. The inter-glycosidic linkage conformations are very difficult to be determined from spectroscopic studies like the Nuclear magnetic resonance spectroscopy (NMR). GLYCAM06 was observed to provide good dihedral and glycosidic linkage data by [Foley](#)

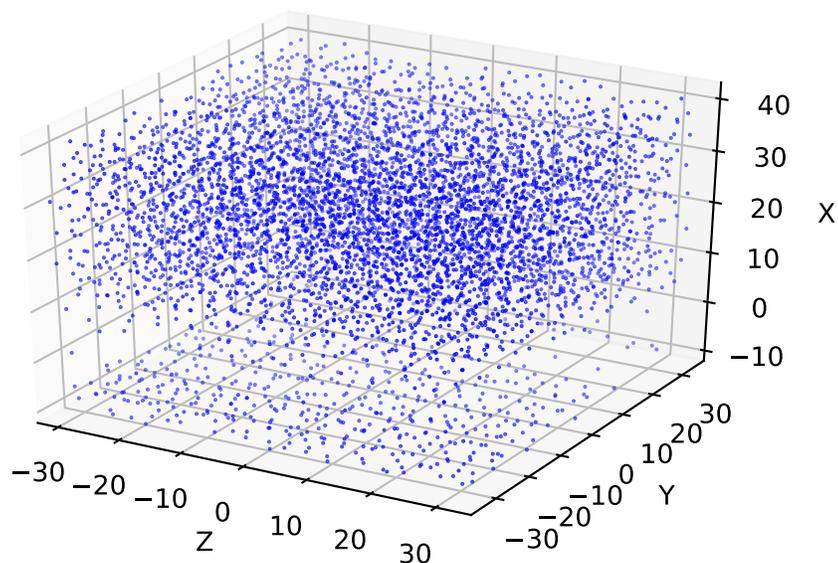


FIGURE 2.7: Water density profile in the simulation box for a well equilibrated 286mc system. The blue dots are the COM of water molecules along the periodic box form one snapshot of the periodic box.

*et al.* One more interesting feature of GLYCAM06 is that it is highly flexible and portable between simulation software.

It is highly compatible with existing simulation software like GROMACS<sup>[22,69]</sup>, Assisted Model Building with Energy Refinement (AMBER)<sup>[5]</sup> and Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)<sup>[55]</sup>. The features like relative functionality, consistency with observations, portability and flexibility of GLYCAM06 made it a natural choice for the preferable forcefield for the considered systems.

## **2.4 Simulation**

### **2.4.1 Pre-Processing**

#### **2.4.1.1 Cellulose-Builder**

The model building steps were same for all the systems except for the (010) surface systems where an extra surface modification step was included. The initial structure or layout for the CNC was made using the Cellulose-Builder tool<sup>[15]</sup>. The Cellulose-builder tool developed by [Gomes and Skaf](#) enables the user to define the crystal type, unit cell counts along each crystallographic direction and the periodic conditions.

For these systems, all surfaces were considered to periodic, which forms an infinitely long crystal structure. All CNC configuration obtained from cellulose-builder were periodic and we designed the system in a way such that the structural integrity holds during the MD simulation. The CNC structure is layered. If simulated with less number of layers the CNC structure will break apart during MD simulations.

Hence a series of crystal sizes were tested and an optimal unit cell count was determined and it is listed in table [2.2](#). Structure files generated by the cellulose builder include atom coordinates in the protein database file (\*.PDB) or XYZ coordinate file (\*.XYZ) format and atomic connectivity information in the form of protein structure file (\*.PSF). These structures files were used to build the model further.

### **2.4.1.2 Glycam-web**

The oligomer structures were made using the online tool, Glycam-web<sup>[73]</sup>. Both the oligomer chains were made of 5 monomer units each. The pentaose structure had no substitutions along the chain length, it was a linear oligomer chain. On another hand, the MC chain has substitutions at different sites, in general, the MC oligomer chain will have a minimum of a few 100 to 1000 monomer units and along the chain. There are hydrophilic and hydrophobic zones based on the degree of substitution (DS). The MC used in the actual experiments were observed to have lower molar substitution (MS) values than the DS.

This shows that there were no side chains from the point of substitution. Along the cellulose chain, each glucose monomer has three possible sites for substitution. They are, the -OH groups attached to C-2,3,6 positioned carbon atoms. The other carbon atoms form the glycosidic bonds or do not have hydroxyl groups to involve in substitution. Substitutions were made at the three available substitution sites along the oligomer chain.

This online tool was used to have consistent residue information, the uses of which will be discussed in subsection 2.4.1.5. Figures 2.8 and 2.9 show the residue information, chain substitution sites and glycosidic linkage sites for both MC and pentaose. The chains are linear and cellulosic in nature hence they were modelled with 1-4 glycosidic linkages.

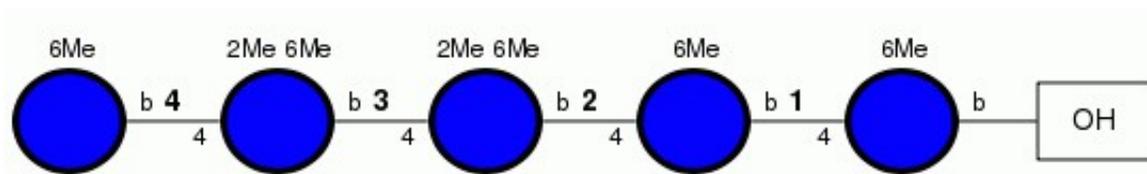


FIGURE 2.8: Residue, linkage and substitution on the MC oligomer chain. Each bead represents one glucose unit, b represents the beta linkage between the units. '6Me' denotes methyl group substitution at the C6 carbon's -OH group, similarly for '2Me'.

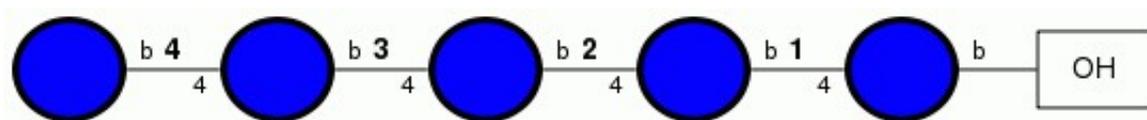


FIGURE 2.9: Residue, linkage and substitution on the pentose oligomer chain. Each bead represents one glucose unit, b represents the beta linkage between the units.

### 2.4.1.3 Packmol

Packmol is a software tool used to develop the initial configuration for MD simulations<sup>[45]</sup>. Packmol enables the user to utilize the whole geometry of the system<sup>[44]</sup>. The molecules need to be packed in a simulation box with a reasonable density and no overlaps before they could be simulated. Packmol is effective in achieving this operation. In case of the systems in this study, the molecules were placed near the center of the planned simulation box, where the oligomer chain will rest in the bulk water at the initiation of the simulation.

The CNC structures with oligomers are shown in figs. 2.3 and 2.4. The structure with CNC and an oligomer chain was packed with water molecules made with TIP3P specifications until the separation distance criteria set to avoid overlap was met. The resulting structure from this step is used as the initial configuration for the MD

simulation. Figure 2.10 shows a complete simulation box packed with all the components.

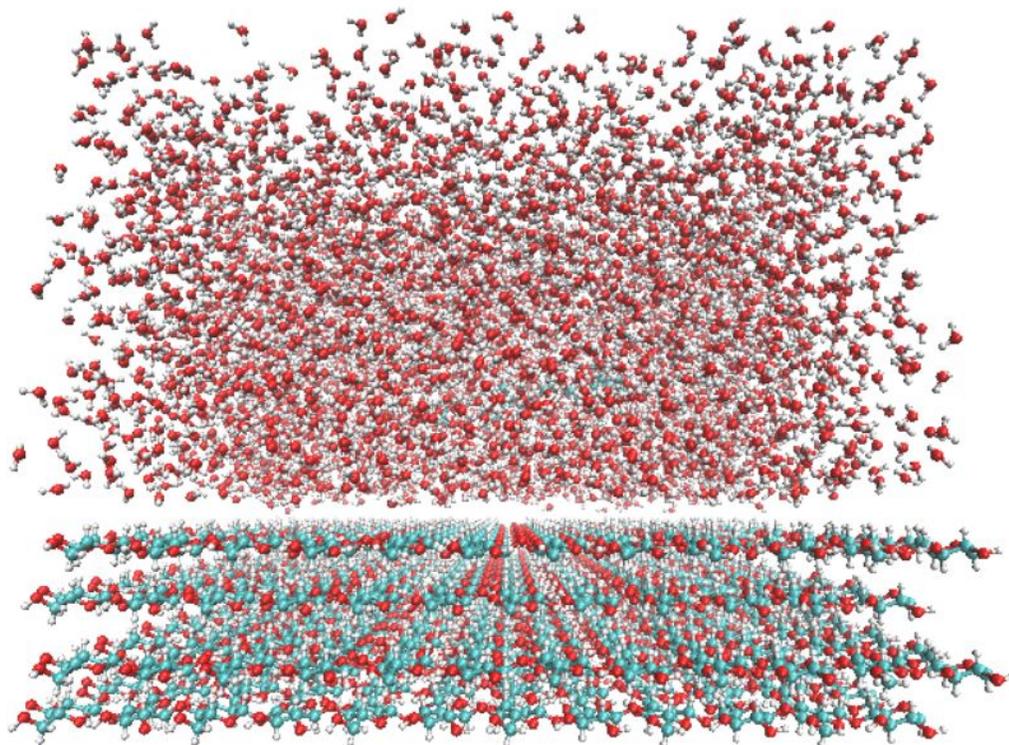


FIGURE 2.10: A completely packed 286pe system with all three components (CNC, pentose, water). The oxygen, carbon and hydrogen atoms are represented by red, green and white spheres respectively.

#### 2.4.1.4 Avogadro

Avogadro is an open visualization and molecule editing tool<sup>[20]</sup>, in this work it was used mainly for three major purposes,

- To graft the surface sulphate groups on the (010) surface.
- To check for any defects in the structure of the system at each step.

- To convert the structure files between formats required by different software tools.

The structure files from Avogadro and Packmol do not contain the residue information required for parameterizing molecules in the systems. The parameterization operation and its uses are discussed in the next section.

#### **2.4.1.5 AMBER**

AMBER is the next tool used in the process flow, AMBER<sup>[5]</sup> and Ambertools available with AMBER16 was used to process the structure files of each system. AMBER is a collection of programs for MD simulation, it has numerous tools for pre/post-processing simulation data. Before processing the structure file in AMBER, two other information namely, 'atom type' and 'partial charges' had to be appended to the structure file based on the forcefield and residue information. Table 2.3 shows all the residues used in this study and their corresponding residue information.

These information are crucial as the forcefield parameters are assigned to each atom in the system based on these parameters as identification. Care was taken to make the system modular, as a substitution at a site on the glucose ring will change the overall charge of the ring. As per the instructions provided by the GLYCAM team, the modular charge corrections were made for respective residues with every substitution.

In GLYCAM06, the partial charges calculations for condensed phase simulation, which are appropriate for TIP3P water simulations were made with HF/6-31G\*//HF/6-31G\* molecular electrostatic potential (MEP) method. The restrained electrostatic potential

(RESP) charge fitting methodology was employed to derive the final partial charges. The residue names given by three-letter nomenclature for the cellulosic units is broken down as follows, '4GB' - '4' stands for the number of valence positions; 'G' denotes the glucopyranose ring and 'B' denotes beta configuration.

TABLE 2.3: Residue information for all the oligomer structures.

S.No.	Residue	Structures containing the residue	Number of atoms	Net charge (C)
1	4GB	CNC (100), MC, pentaose	21	0
2	UGB	CNC (010), MC	20	-0.194
3	ROH	MC, pentaose	2	-0.194
4	QGB	MC	19	-0.388
5	RGB	MC	19	-0.388
6	6GB	MC	21	0
7	MEX	MC	4	0.233
8	SO3	CNC (010) sulphate group	4	-0.837
9	WAT	water	3	0

The assignment of partial charges and atom type parameters was done using a custom code. Having processed the structure file with the information required by the forcefield, the structure file was loaded into AMBER and then by loading suitable forcefield data files the structure file was converted into two AMBER compatible files, \*.top (topology file) and \*.crd (coordinate file). These two files were then converted into a data file which was used for the further processes using an available LAMMPS tool, 'amber2lammps.py'.

## **2.4.2 Main Simulation**

### **2.4.2.1 LAMMPS**

All MD simulations in this study were carried out using LAMMPS<sup>[55]</sup>. LAMMPS is a classical MD code and it stands for Large-scale Atomic/Molecular Massively Parallel Simulator. LAMMPS is highly efficient and it can handle millions to billions of particles. Unlike other simulation packages like GROMACS or AMBER, which are mostly constrained for bio-systems and few polymers, LAMMPS is versatile and can be used to simulate a wide variety of systems such as metallic, coarse-grained, bio-molecular and polymeric ones.

Although LAMMPS is powerful and robust, it cannot do pre/post-processing on its own, hence it requires the user to develop custom codes or use available tools like Visual Molecular Dynamics (VMD)<sup>[28]</sup>, MDAnalysis<sup>[16,47]</sup>, MDTraj<sup>[46]</sup> and topotools<sup>[36]</sup> from VMD etc. for post-processing and the previously mentioned tools and codes in subsection 2.4.1 for pre-processing. One significant factor about LAMMPS is that the system simulated using it is highly customizable and the user can control all factors of the simulation ranging from on the fly particle addition, deletion and once can constrain particles to observe specific interactions etc.

### **2.4.2.2 Simulation Strategy**

Irrespective of the CNC surface, the same set of simulation strategy was used to simulate all eight systems. Both the surface simulations followed the same initial configuration,

setups and minimization strategy, but changes were made in the dynamics setting slightly to accommodate system stability. Table 2.4 shows the initial configuration followed for the systems.

TABLE 2.4: Initial configuration set up.

<b>Configurations</b>	<b>I</b>
Systems	286mc(s), 286pe(s), 1026mc(s), 1026pe(s)
Bond, Angle, Dihedral	Harmonic
Pairwise interaction cutoff	12 Å
Long-range interaction solver	Ewald
time step (dt)	1

In both the cases, the initial setup was the same, bonds, angles and dihedrals were treated as harmonic potentials. The pairwise interaction cut off was set to 12 Å. We have tested longer cutoffs and found that no significant difference was observed as we increase the cutoff to 16Å. Both the non-bonded interactions LJ/Coulombic were set to the same cut-off. The Ewald long-range interaction solver was used with a tolerance of  $10^{-6}$  Kcal/mole-Angstrom in force, to handle the long-range interactions of atoms beyond the short-range interaction cutoff. The initial timestep which was used to equilibrate the system was set at 1 femtosecond (fs).

After the initial setup, the system was equilibrated using a series of minimization and NVT ensemble runs to reduce the potential energy of the system and bring it to a constant value before running the production dynamics. This step is carried out to make sure that the system remains stable when velocity is assigned gradually to initiate

dynamics. The operation of minimization is nothing but the step by step reduction of potential energy of the system to a constant value.

TABLE 2.5: Equilibration steps followed for the systems.

Step	Operation/ensemble	type/temperature	time steps
I	minimization	SD	100
II	NVT	150 K	5000
III	minimization	SD	500
IV	NVT	200 K	5000
V	minimization	SD	500
VI	NVT	300 K	5000
VII	minimization	SD	5000

All minimization and operations were carried out at 1 fs time step. The steepest descent algorithm (SD) at a tolerance of  $10^{-8}$  Kcal/mole-Angstrom for force and  $10^{-8}$  Kcal for energy was used to minimize the systems. The intermediate NVT ensemble runs to equilibrate the system were carried out with a gradual increase in temperature until the desired production run temperature was achieved. The production run temperature was 298 K. Table 2.5 lists all the steps carried out in the equilibration routine.

Followed by the equilibration step, the production run was carried out. Two schemes were followed for the two different surfaces. Only the production run's damping parameter had differences between the two simulation schemes. The production run was carried out with a time step of 2 fs and the SHAKE algorithm invented for schemes like standard Verlet time-stepping was used to constraint the covalent bonds between O, H atoms in water, oligomers and C, H atoms in the oligomers and CNC. This enables the

usage of a larger time step. The NPT ensemble was used to run all production and dynamic runs in all the systems. All simulations were carried out at a standard temperature of 298 K and pressure of 1 atm. The key differences in the two simulation schemes were the damping parameter used to regulate the pressure coupling and temperature in the simulation box, pressure damping parameter 'pdamp' was set at 50 for the (100) systems and 250 for the case of (010) systems and the temperature damping parameter 'tdamp' were set at 10 and 25 for (100) and (010) systems respectively.

The damping parameters carry the time unit. Temperature and pressure fluctuate during the production run and these damping parameters define the time LAMMPS can use to adjust this fluctuation. These values need to be optimal and were determined using trial and error approach. If the values are set too high, the setup parameters fluctuate heavily and if the values are set too low, the setup parameters will not have the necessary time required to adjust the fluctuation and will eventually fail in both the cases. Along with production run few other specific modules in LAMMPS were used to run series of analyses on the fly. These analyses and their theories are discussed in the following sections. A list of input commands and files used with some of the preprocessing software is shown in [Appendix B](#).

#### **2.4.2.3 Free energy calculation**

The major simulation which was carried out during the production run was the free energy difference calculation. The free energy difference calculation requires good sampling of the system over a long period of time at the order of a few nanoseconds. To calculate free

energy two parameters has to be decided, what are the  $\xi$  along which free energy is to be calculated and the macrostates between which the free energy difference is defined. In our case, the reaction coordinate is the direction perpendicular to the adsorption surface and the macrostates are the adsorbed and desorbed states of oligomers with respect to the CNC surface. Since all the simulation were carried out in the NPT ensemble the calculated free-energy difference is the Gibbs-Free energy difference ( $\Delta G$ ).

#### **2.4.2.4 Overview**

Free energy calculations in this work were carried out using the collective variable (COLVAR), steered molecular dynamics (SMD) modules in LAMMPS and the weighted histogram analysis method (WHAM) code. Free energy differences were expressed in terms of the potential of mean force (PMF). This value was calculated using a technique called umbrella sampling, this technique was carried out using the COLVAR and SMD modules in LAMMPS. The values obtained from umbrella sampling were used to calculate the PMF value via the WHAM code. A detailed explanation and the relationship between each of these steps and tools and their underlying theory is explained in the following sections.

#### **2.4.2.5 Umbrella sampling**

There are numerous techniques to calculate free energy in MD simulation. Some of them are free energy perturbation, thermodynamic integration, umbrella sampling and histogram methods etc. The technique used to calculate the free energy in this study is the umbrella sampling. This technique is used to improve the sampling of a system i.e.

capture of different configurations of the same system. The free energy changes along the reaction coordinate, with global and local minima showing the most favourable zones along it. The free energy surface (one dimensional in this case) will have barriers which need to be crossed in order to sample all possible configuration of a given system. In order to do that a restraint to force is added. This value (5-10 Kcal/mol) was chosen based on the values initially used to study the forcefields compatibility and some previous studies<sup>[14,34,49]</sup>.

After sampling, this bias needs to be removed and to do so multiple overlapping restraints in the form of umbrellas need to be added along the reaction coordinate, thus giving the name umbrella sampling. A reaction coordinate ( $\xi$ ) can be a distance, periodic angle or atom positions etc. Finally, the bias in the final value is removed, provided the bias is a function of the reaction coordinate. Umbrella sampling avoids the sampling problem by altering the Hamiltonian so that the unfavourable states are sampled sufficiently. The free energy can be related to the partition function ( $Q_{NPT}$ ) of the ensemble as follows,

$$G_{NPT} = -k_B T \ln(Q_{NPT}). \quad (2.9)$$

The partition function  $Q_{NPT}$  is given by,

$$Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_o} \int dV \int dr dp e^{-(E+PV)/k_B T}. \quad (2.10)$$

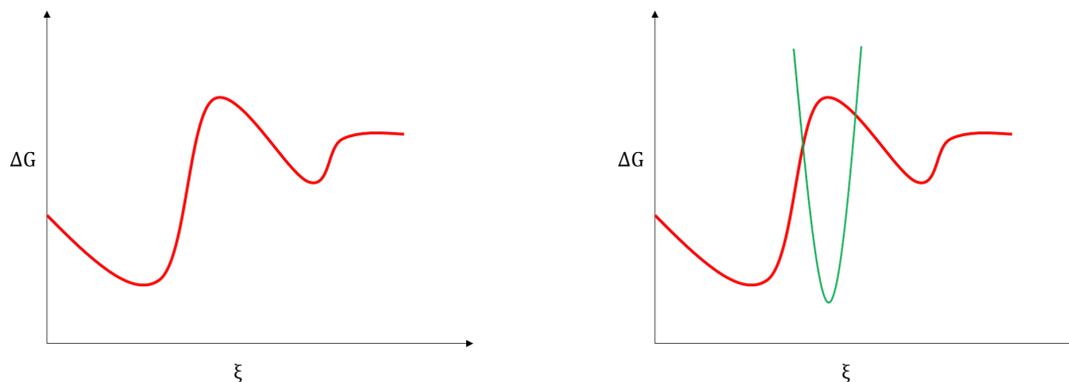
Where  $V_o$  is the constant volume,  $N$  is the system size, Planck's constant ( $6.62607004 \times 10^{-34} \text{ m}^2 \text{ kg/s}$ ) ( $h$ ),  $r$  is the position,  $p$  is the momentum,  $P$  is the pressure,  $k_B$  is the Boltzmann constant,  $V$  is the volume,  $E$  is the Hamiltonian and  $T$  is the temperature.

Figure 2.11 shows a simple schematic of the umbrella sampling technique. Figure 2.11(a) shows a simple one-dimensional free energy difference curve as a function of the reaction coordinate ( $\xi$ ) and it can be seen that there are a series of elevation and depression along the curve. These denote different configurations bore by the system's component and they all possess different energies. In fig. 2.11(b) a small restraint as a bias is added along the curve to help the system cross this energy difference and this becomes a window with a width in the same units as that of the reaction coordinate, Å for our systems. Figure 2.11(c) shows multiple biases being added along the curve to overcome the energy differences along the curve, this is done to make sure that all possible configurations with all possible energy differences are captured and finally fig. 2.11(d) shows the net results, the red curve denotes the actual PMF/free energy curve and the green curve denotes the sequential bias given by us via umbrella sampling. These biases negate the energy differences between the subsequent windows and making the curve flat with no energy difference. This enables perfect sampling of the given system. Since the bias is a function of the reaction coordinate, they can be removed during the final PMF calculation to get the actual PMF curve.

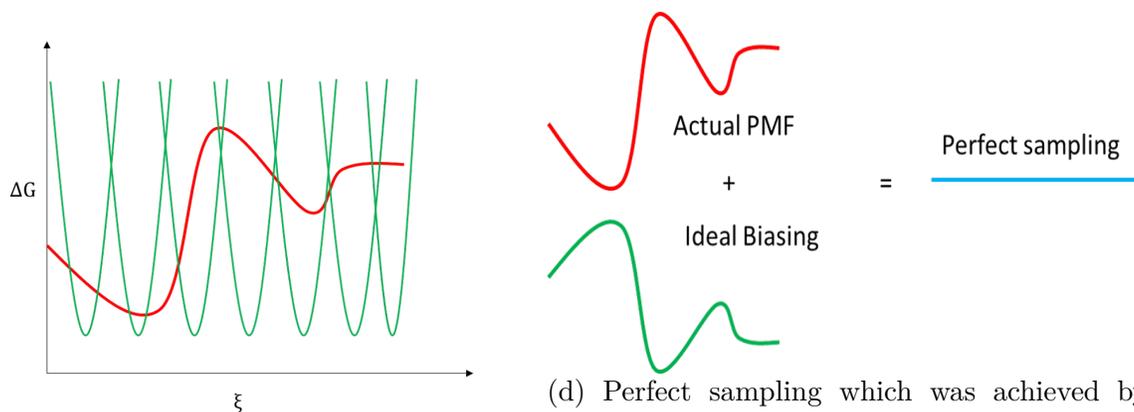
#### 2.4.2.6 Steered molecular dynamics (SMD)

Steered Molecular Dynamics (SMD) is done to steer the system or its component in a specific direction or pathway. The system is restrained to act in a specific way to observe the occurrences of interest. SMD in LAMMPS can be carried out in fixed force or fixed velocity modes. In our case, SMD was done to get the initial configuration for each window, which was described in the previous section. The system under study

FIGURE 2.11: Umbrella Sampling technique.



- (a) True PMF curve with free energy as the function of reaction coordinate is given, this represents the actual free energy difference between the macrostates.
- (b) The red curve denotes the actual PMF curve and the restraint to cross the energy difference at one point along the reaction coordinate is represented by the green umbrella curve.



- (c) Series of overlapping windows of restraints along the reaction coordinate to overcome energy differences is shown.
- (d) Perfect sampling which was achieved by adding bias to the actual PMF curve is shown by the blue curve. The red and green curves represent the actual PMF curve and the ideal biasing provided by us.

needs to be sampled in all zones along the reaction coordinate ( $\xi$ ). Since the calculation requires perfect sampling along the coordinate, for longer coordinates performing the sampling at a stretch could be tedious and almost impossible. Hence the reaction coordinate was split into different segments with the same width  $0.5 \text{ \AA}$  and each segment is called a 'window'.

The SMD was carried out as a single simulation. The SMD routine used in this study is as follows, an initial configuration which was well equilibrated was chosen for each system and then a directional vector, force constant and velocity was chosen for each system, and they were, the normal direction to the surface, 5 Kcal/mol and 0.0001 Å/ns respectively. The force constant is discussed in detail in subsection 2.4.2.7. The same values were used irrespective of the system. The oligomer unit then follows the direction vector and moves along the same direction gradually away from the surface with the defined velocity. Once it reaches the desired reaction coordinate distance (final window), the direction was changed and the oligomer was moved towards the surface with the same velocity by using the same force constant. The system configurations were captured along the coordinate which was then used as the initial configuration for each individual window. Each of these windows was simulated individually over a fixed period of time. This was done to reduce the overall simulation time and to make sure perfect sampling was achieved.

#### **2.4.2.7 COLVAR**

COLVAR stands for collective variable. In MD simulations, it is useful to reduce the degrees of freedom to fewer parameters, whose statistical distribution can be analyzed individually. These parameters are referred to as order parameters, collective variables etc. The COLVAR module<sup>[13]</sup> in LAMMPS was used to record the collective variable along the reaction coordinate, in this case, the reaction coordinate direction was the direction perpendicular to the adsorption surface and the recorded collective variable was the distance between the COMs of the first layer of CNC and the oligomer. COLVAR

module interfaces with LAMMPS. It puts an imaginary spring between the two groups of atoms, this spring has a spring constant value which is equivalent to the set force constant. The force/spring constant ( $k$ ) is related to the biasing potential  $V$  by,

$$V = \frac{1}{2}(x - x_o)^2. \quad (2.11)$$

Where  $x$  and  $x_o$  are the colvar values of the current and initial states respectively. Like it was mentioned in the previous section, SMD is done and initial configuration for each window is recorded from that run. These initial configurations were set up as each individual MD simulation to record the colvar values for each window. The temperature was maintained to the same value as that of the production NPT ensemble run. As the simulation proceeded the COLVAR records the collective variable value for every 1000 time steps. For the systems in this study, there were a total of 25 windows, all of them were simulated and sampled for 4 ns.

The oligomers were restrained at a specified value( $x_o$ ) for each windows using a biasing potential/force constant/spring constant of 8 Kcal/mol. This value was maintained for each window in all simulations for the sake of consistency. Table 2.6 shows the windows and COLVAR set up used for all the systems. The biasing potential value does not have to be constant for each window even within the system, as the bias for each window will be explicitly mentioned and used in PMF calculation.

TABLE 2.6: Colvar module sampling specification used for all systems.

Systems	Force constant (Kcal/mol)	Windows	Sampling time (ns)
All	8	25	4

### 2.4.2.8 WHAM

The WHAM method<sup>[39,61]</sup> is a well-known method to calculate the PMF. The values from the recorded collective variable along the reaction coordinate for every window from every system was used to compute the PMF values. The WHAM code created by [Grossfield](#) was used for the purpose. The WHAM code gives the value of PMF along with the errors for a given temperature. The PMF value is equivalent to the free energy difference in the observed reference frames. The PMF value is basically the free energy profile along the reaction coordinates with all the other degrees of freedom being averaged out. For example, given a separation distance ( $r$ ) between groups, the distribution of  $r$  can be given by radial distribution function  $g(r)$  and the corresponding relationship with PMF is given as,

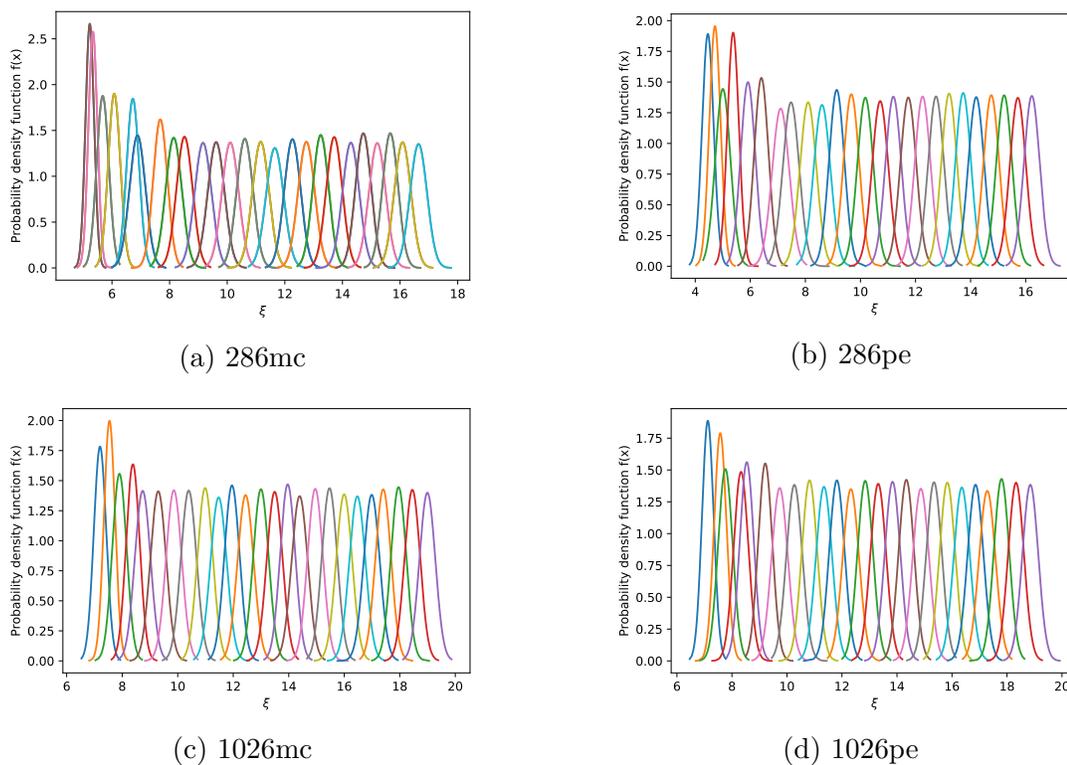
$$F(r) = -k_B T \ln g(r). \quad (2.12)$$

Where, Boltzmann constant ( $1.38064852 \times 10^{-23} \text{ m}^2 \text{ Kg s}^{-2} \text{ K}^{-1}$ ) ( $k_B$ ),  $T$  is the temperature. Since the relationship is logarithmic small changes in free energy will correspond to drastic changes in values of the computed function's value, PMF and this is the main reason for the need for extensive sampling. The umbrella potential/biasing potential ( $U'(r)$ ) is given by eq. (2.11). This value is the bias given to the system in order for it to cross the energy difference and sample every configuration. The biased probability ( $P'(r)$ ), which is the calculated biased PMF value from each window's COLVAR simulation is given in the same form as eq. (2.12). Having these values the unbiased free energy is calculated as,

$$G(r) = -k_B T \ln P'(r) - U'(r) + F. \quad (2.13)$$

where  $F$  is dependant on  $U'(r)$  and  $P'(r)$  is estimated from each simulation. Finally after

FIGURE 2.12: Window-wise histogram for four systems.



removing the bias, bootstrapping was carried out to determine the possible errors in the PMF values and the PMF values were obtained as a function of the reaction coordinates. These values were used for further analysis.

### 2.4.2.9 Histograms

It was mentioned numerous time in the above sections that the system needs to be sampled well along the reaction coordinate to get reliable PMF values. Histograms of the measured values (collective variables) in each window for a given system will be a good measure

to check sampling. Figure 2.12 shows the histograms for each window for four of the eight systems. The histogram plot is shown with the probability density function  $f(x)$  as a function of the reaction coordinate. The probability of the collective variable value is recorded for each window in every system. It can be seen that the windows have good overlap and cover the whole  $\xi$ . This is a good indication that the whole reaction coordinate was sampled well.

### 2.4.3 Post-Processing

The data from LAMMPS simulations were used to process and extract information on the system. Like it was mentioned in subsection 2.4.2.1, except for the modules present in LAMMPS which can be used to perform analysis on the fly, most post-processing is carried out using different codes or tools available from other packages and software tools. Some of the tools which were used include MDAnalysis<sup>[16,47]</sup>, MDTraj<sup>[46]</sup>, VMD<sup>[28]</sup>, Topotools<sup>[36]</sup>, CPPTRAJ<sup>[58]</sup>. Their usage specifications are listed in Appendix A. Most of the analysis tools carry out the operation through a geometric approach, i.e. the coordinate values of each atom or particle in the system are used to extract information. For example, to perform hydrogen bond analysis, the geometric criteria like separation distance and the angle between donor-acceptor pairs are used.

# Chapter 3

## Results and Discussion

The results chapter has been classified into three sections, (1) Model Validation, (2) Direct MD results and (3) Free energy calculations and related observations. All the discussion for each result is done as and when it is introduced.

### 3.1 Model Validation

It is necessary to validate the model before conducting any simulation study, it is true especially if the model involves specific structural properties. Like in our case where we deal with crystalline structures. So in order to check the models, we conducted a series of simulation runs to assess the integrity of simulated components along the simulation timeline. We checked factors like the crystal structure and water model performances in our systems and the results from these runs are presented in this section.

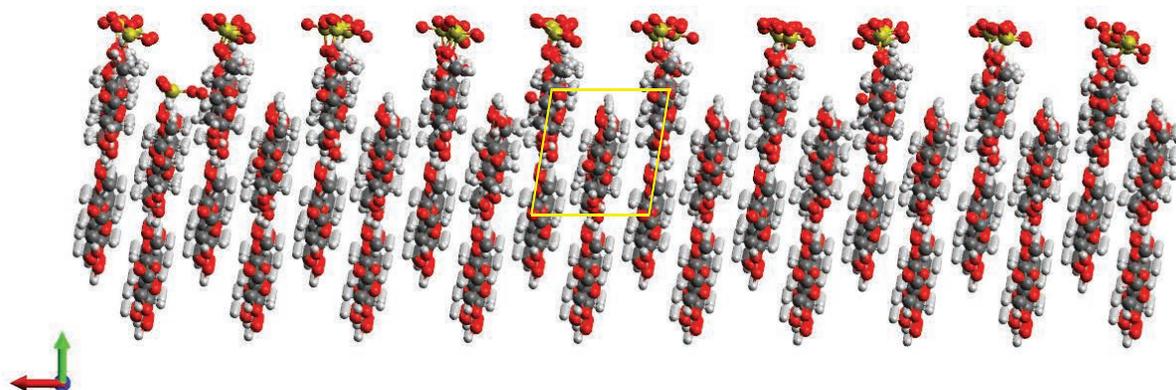
### 3.1.1 Crystal structure

As a part of the model validation, we checked the differences in crystal structure integrity and stability before and after simulation runs i.e. the configuration at the initial stage and after a series of equilibration steps. Figure 3.1 shows the configurations observed in the CNC structure used to study the (010) surface systems. The unit cell along the  $XY$  plane is marked in yellow in both the cases. It can be seen that the crystal structure hold well even after going through a series of equilibration and production MD runs. This was same in case of the other surface ((100) surface) too.

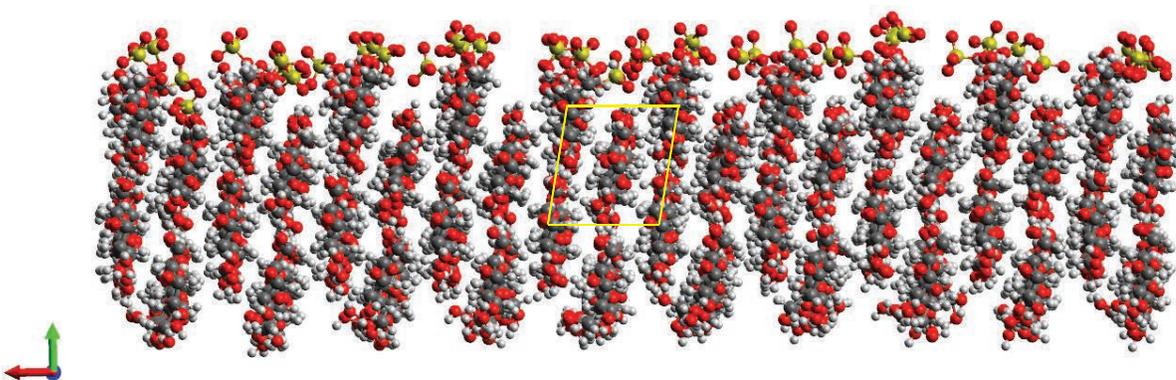
### 3.1.2 Forcefield comparison

A detailed account on forcefields and their importance in MD simulations was given in chapter 2. As a part of the model validation, we initially conducted small scale MD simulations to check for the feasibility of different forcefields such as Chemistry at Harvard Macromolecular Mechanics (CHARMM), general AMBER force field (GAFF) and GLYCAM. The initial which we checked was the potential energy of the CNC in vacuum.

This is done to see the convergence of each forcefield when simulated with NVT ensemble. The results of which are shown in fig. 3.2. The potential energy of CNC as a function of time is shown and almost constant value is observed in all the cases. Hence all of these forcfields as far as general properties go are highly compatible with LAMMPS. We chose GLYCAM due to the fact that it was developed specifically for



(a) Before.



(b) After

FIGURE 3.1: CNC crystal structure comparison - initial and after equilibration runs. The unitcell along the  $yz$ -plane is marked in yellow, the  $x,y$  and  $z$ -directions are marked in red, green and blue respectively.

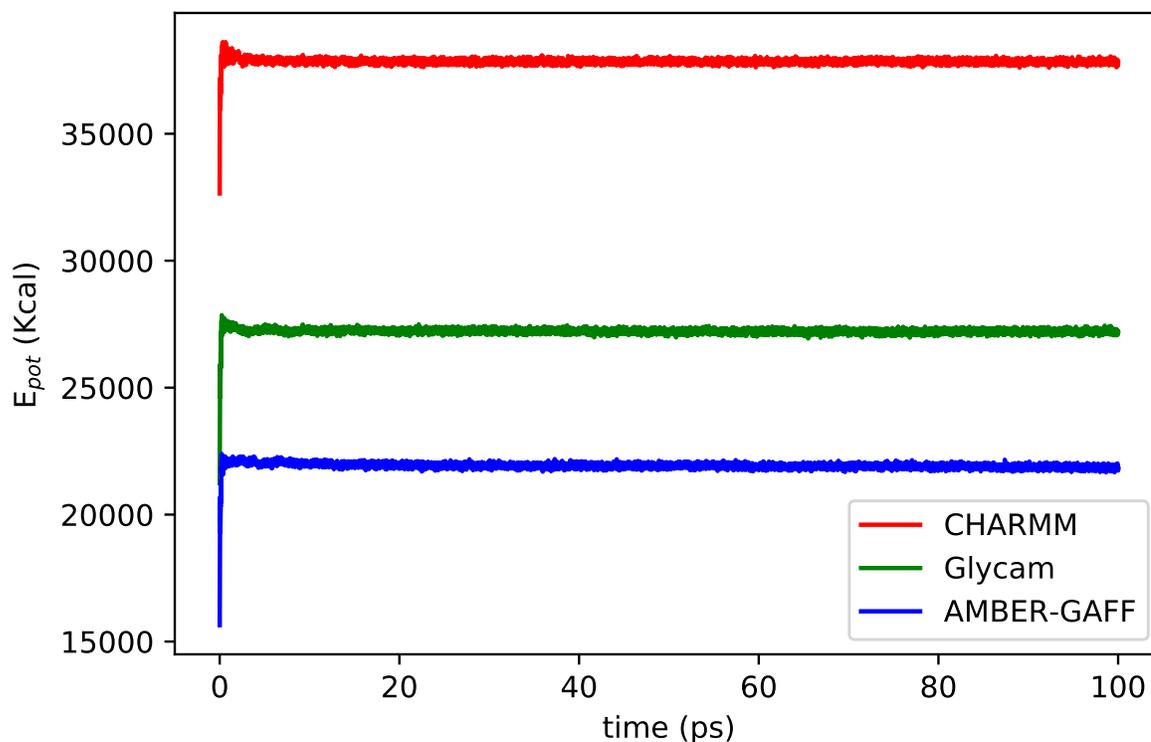


FIGURE 3.2: Potential energy of CNC in vacuum as a function of time for three different forcefields.

carbohydrates and also due previously existing works which show its potential in accurately simulation carbohydrates like cellulose.

### 3.1.3 Water model

The next model component we assessed is the water model. Water models can be tricky in terms of selection and simulation, the user needs to pay utmost care to the water model specifications and the possible impacts it might have on a system simulated using it. The water model used in this study is the TIP3P model. The first and foremost reason for this choice is that the GLYCAM06 forcefield parameters were developed with

respect to this water model.

Without some major test and revision of forcefield parameters, they cannot be used with a different water model. A general overview of pros and cons of TIP3P water model and results from pure bulk water simulations are presented in this section to correlate them with our choice. The TIP3P water model has very high diffusivity value compared to the observed experimental values but this is the case in most of the water models.

The electrostatics-based (ELBA) model predicts diffusivity values which are closer to the experimental ones. But, The TIP3P model from a geometric perspective gives closer resemblance to the actual water molecules. The angle and bond distances observed in water by experiments ( $104.45^\circ$ ,  $0.9584 \text{ \AA}$ ) are closer to the TIP3P water parameters of  $0.9572 \text{ \AA}$  and  $104.52^\circ$ <sup>[29]</sup>. Also, the TIP3P model shows better performance in the calculation of thermodynamic properties such as specific heats<sup>[41]</sup>.

In order to test the TIP3P model, we made some series of observations such as radial distribution function (RDF) of oxygen-oxygen pairs, hydrogen bond counts, the distribution of hydrogen bond separation distance and angle etc. The results from these observations are shown in fig. 3.3. It is also to be noted that the average intermolecular potential energy ( $E_{pot}$ ) calculated from our simulation,  $-9.9254 \text{ Kcal/mol}$  is consistent with the values observed by Mark and Nilsson. Figure 3.3(a) shows the RDF of oxygen-oxygen pairs, it can be seen that the first peak is around  $2.7 \text{ \AA}$  as expected and is consistent with previously observed ones<sup>[43]</sup>.

One interesting factor to note here is that the average separation distance for hydrogen

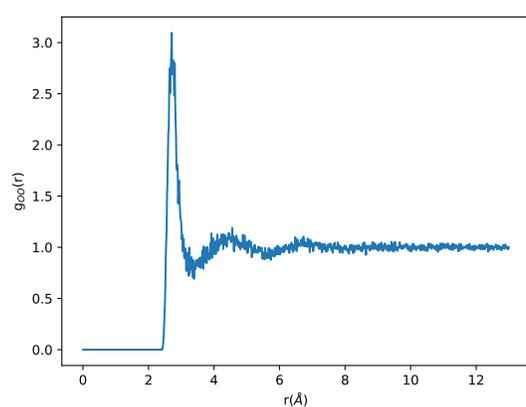
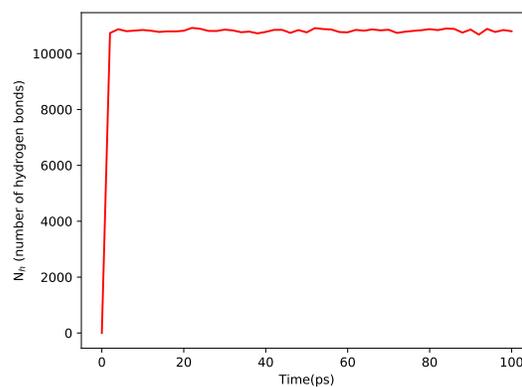
bonds found in fig. 3.3(d) is closer to the first peak of seen in the RDF curve, which denotes the presence of uniform hydrogen bond network within the bulk water system modelled using TIP3P. Also, this is complemented by the average number of hydrogen bonds identified, which is uniform. The average number of hydrogen bonds is shown as the function of time in fig. 3.3(b). Finally, in fig. 3.3(c) the average hydrogen bonding angle is shown.

On evaluating all these results, it is evident that the TIP3P model is acceptable and apt for the systems. There are other 4-point models with Ewald corrections, which are even more promising but those models are computationally costly. The TIP3P model has a good balance in terms of accuracy and cost. All these factors support our choice to select the TIP3P water model over other models.

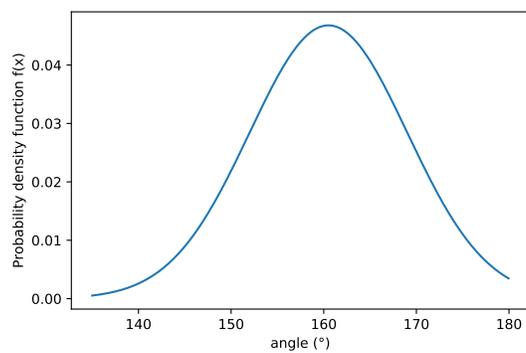
## **3.2 Direct MD results**

### **3.2.1 Adsorption and desorption**

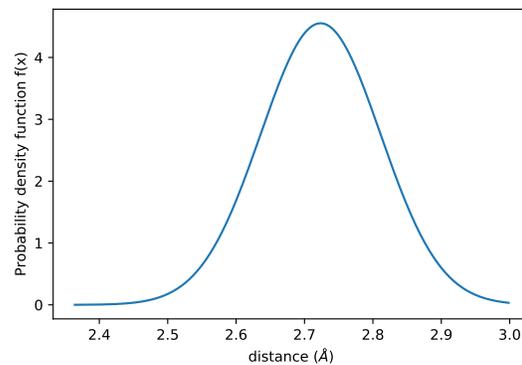
Since our primary interest is on the CNC-polymer interaction in the systems, it is natural that we observe the oligomer dynamics and interactions with the CNC surfaces for an extended period of time. We conducted a series of MD runs to observe these interactions and recorded parameters such as the COM of the oligomer along different axis. In fig. 3.4 the COM values of the different components of each system as a function of time is shown. All of these observations are from direct MD simulations carried out for 8 to 10 ns. The COM values serve as a measure of the position of each component along the normal axis,

(a) RDF of water showing the  $g_{OO}$  curve

(b) Number of hydrogen bonds in pure bulk water



(c) the water hydrogen bond angle distribution



(d) the water hydrogen bond separation distance distribution

FIGURE 3.3: TIP3P water model observations.

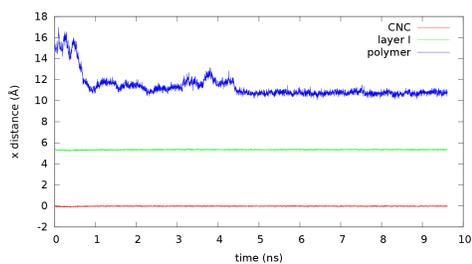
$x$  in case of (100) systems and  $y$  in case of (010) systems. It can be noted that the COM values of the oligomer show slight fluctuations and both adsorbed and desorbed states are seen along the length of the simulations. There are some cases where the changes in COM values are very low relative to other stages, like the (100)pe system shown in fig. 3.4(c). These might be because of the partially desorbed states being unstable compared to the adsorbed/desorbed states.

### 3.2.2 Water molecules

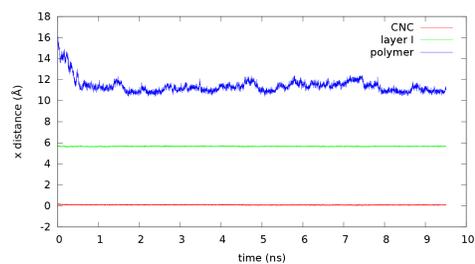
A good justification for the use of TIP3P water model was given already in subsections 2.2.2 and 3.1.3 and section 2.3. In this section, we discuss the observed properties of the water molecules in direct MD simulations, fig. 3.5 shows the MSD and self-diffusion coefficients of the water molecules observed in different systems. It was mentioned in subsection 3.1.3 that the diffusivity of TIP3P water model is high compared to the experimental value. Figure 3.5(b) shows the self-diffusivity coefficient values observed in each system alongside the values observed in pure water simulations and experiments.

The values are lower in presence of solutes, especially salts. As expected the diffusivity is slightly lower with the presence of solutes and this can also be attributed to the configuration of the systems. In general, all systems dealt in this study are heterogeneous in nature and do not have rotational symmetry in all three coordinates. The MSD values of water molecules for each system after series of equilibration steps were recorded well away from the super and sub-diffusion zones. The MSD values were

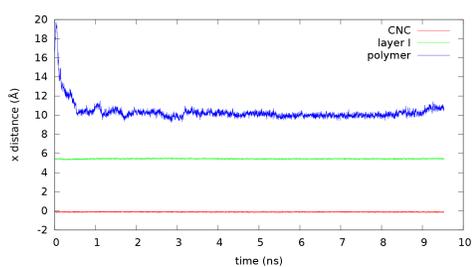
FIGURE 3.4: Component COMs of each system



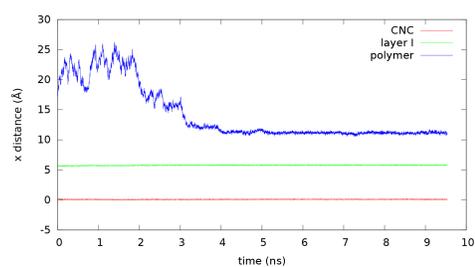
(a) (100)mc



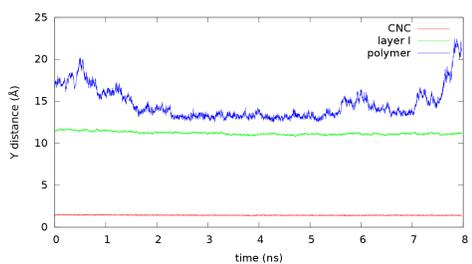
(b) (100)mcs



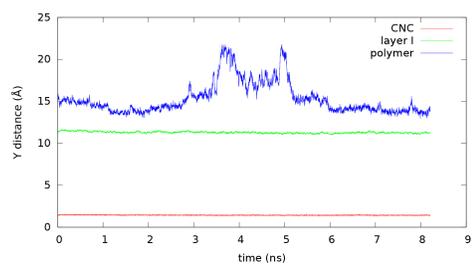
(c) (100)pe



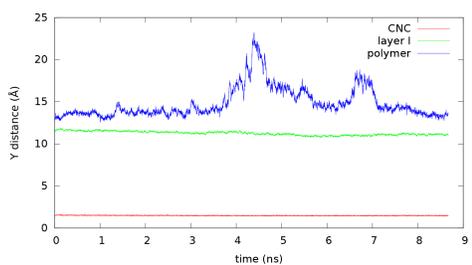
(d) (100)pes



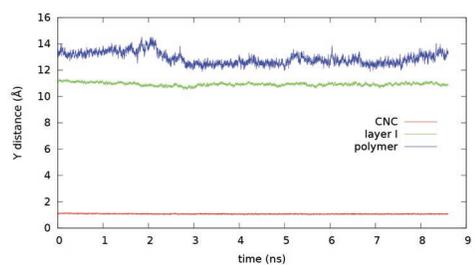
(e) (010)mc



(f) (010)mcs



(g) (010)pe



(h) (010)pes

found to be in a linear relationship with the time given by,

$$\langle r^2 \rangle \propto D \tau. \quad (3.1)$$

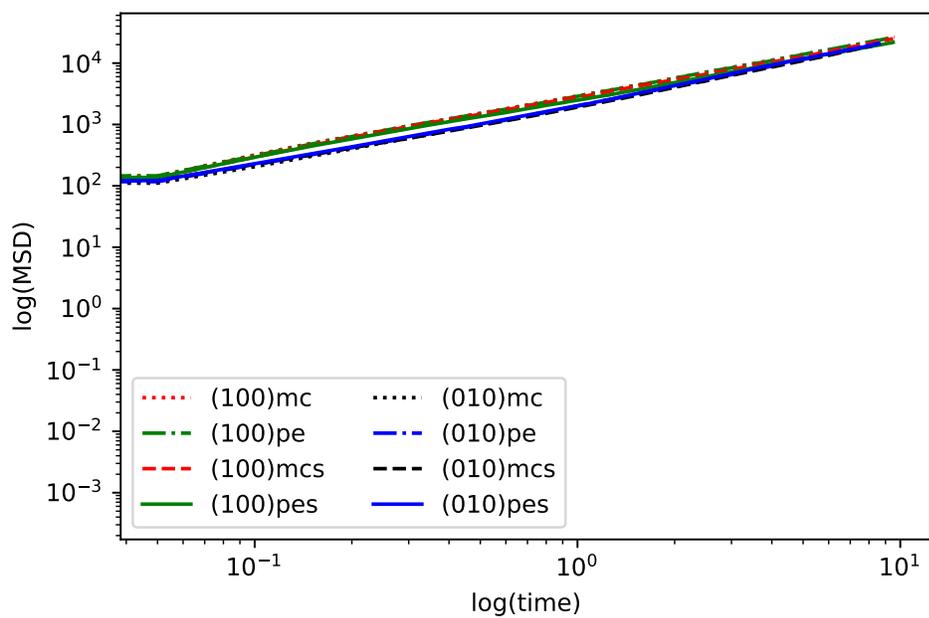
Where  $r$  is the displacement,  $D$  is the self-diffusion coefficient and  $\tau$  is time. These results are summarized in fig. 3.5(b).

### 3.2.2.1 Water density profile

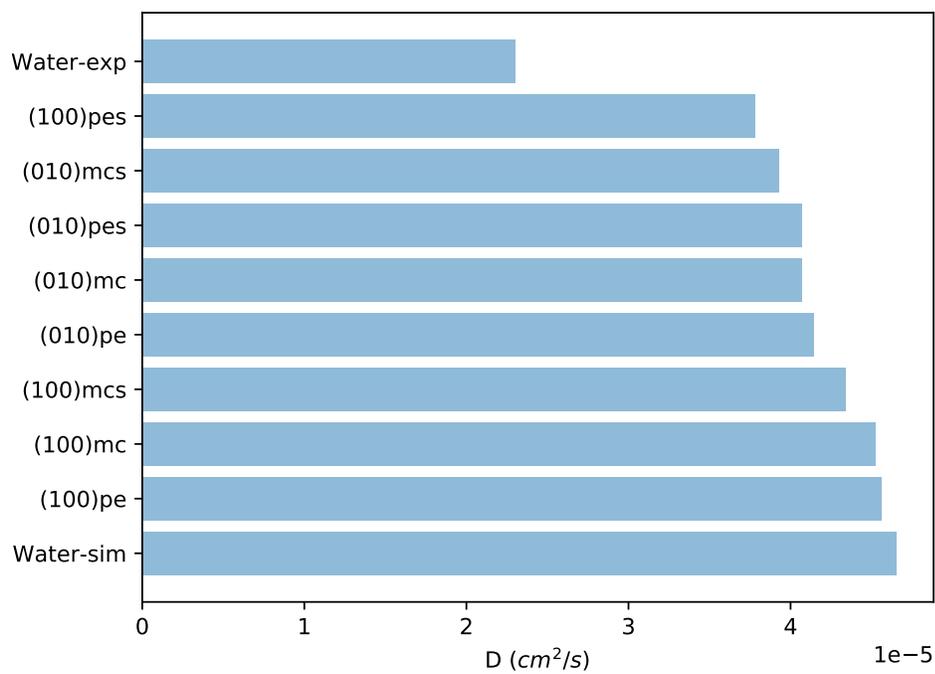
Another water property which was observed from direct MD simulations was the number density ( $\rho_N$ ) along the reaction coordinate. Figure 3.6 shows the water molecules  $\rho_N$  as a function of distance along the normal axis. The values were calculated with a width of  $0.5\text{\AA}$ . The region  $-10 \leq x \leq 10$  correspond to the CNC in both the cases of (100) and (010) systems. In case of the (010) surface, the sulphate groups extend from the surface and mark the margin of the CNC surface at about  $12\text{\AA}$  along the normal direction. In fig. 3.6(b), it can be seen that there is a slight fluctuation in the  $\rho_N$  within the region of  $10 \leq x \leq 12$ , this is due to the nonuniform presence of water molecules near the surface due to the sulphate groups. The rest of the profiles were uniform along the simulation box, except for the case of (010)pe and (100)mcs. In fig. 3.6(a), initially small peaks can be seen, this is due to the presence of water molecules closer to the CNC zone.

The CNC crystal (100) surface was slightly oriented in that plane due to the independent angle ( $\gamma = 96.5^\circ$ ). This has been the case even from the initially developed model, this is something which we wish to correct in the future models. The other

FIGURE 3.5: MSD and Self-diffusion coefficient of water.



(a) MSD.



(b) Self diffusion coefficient.

changes in the density profile within the bulk region can be attributed to the water's interactions with oligomer and salt ions.

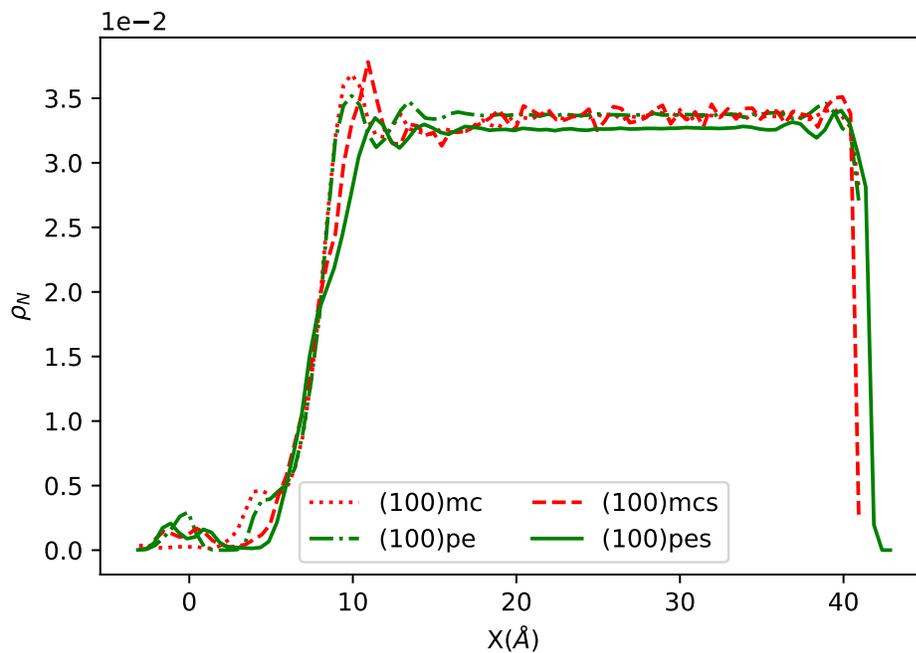
### 3.2.3 Oligomer diffusivity

The oligomer diffusivity in terms of MSD is shown in fig. 3.7. The zones on the MSD curve which are closer to being linear are where the oligomers exhibit almost constant diffusivity i.e. it corresponds to the period where the oligomer was in the bulk, where it had more freedom to move. Along the curve, the regions which show lower values correspond to the period where the oligomer was adsorbed on to the CNC surface. The subsequent increase and decrease in MSD values correspond to the adsorption and desorption experienced by the oligomer. This is can be compared with the COM values of each oligomer as a function of time given in fig. 3.4. It can be seen that these two curves correlate well.

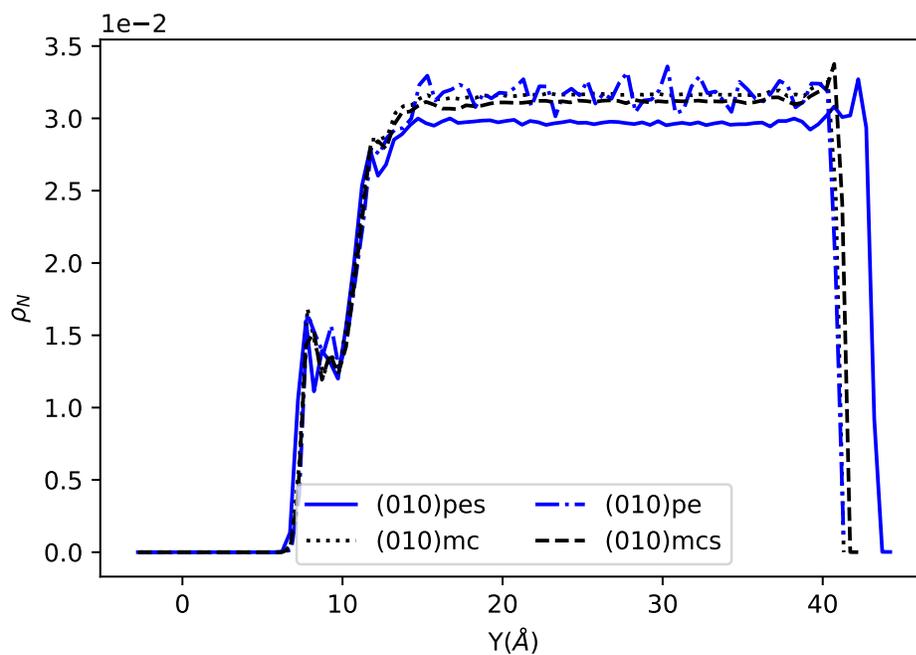
#### 3.2.3.1 Oligomer movement along the (100) and (010) surfaces

Though the oligomers have lesser freedom to move when adsorbed to the surface, it is usually restricted along one direction i.e. it is harder for the oligomers to mover along the normal direction compared to other two directions. Hence we can expect movements along the adsorbed surface. To understand the surface movements, the positions of the oligomer COMs on the normal surfaces ( $yz$  for (100) systems and  $xz$  for (010) systems) as a function of time were plotted and the results of which are shown in figs. 3.8 and 3.9.

The planar area covered by oligomer COMs over a span of  $\approx 7$  (ns) varies within a range

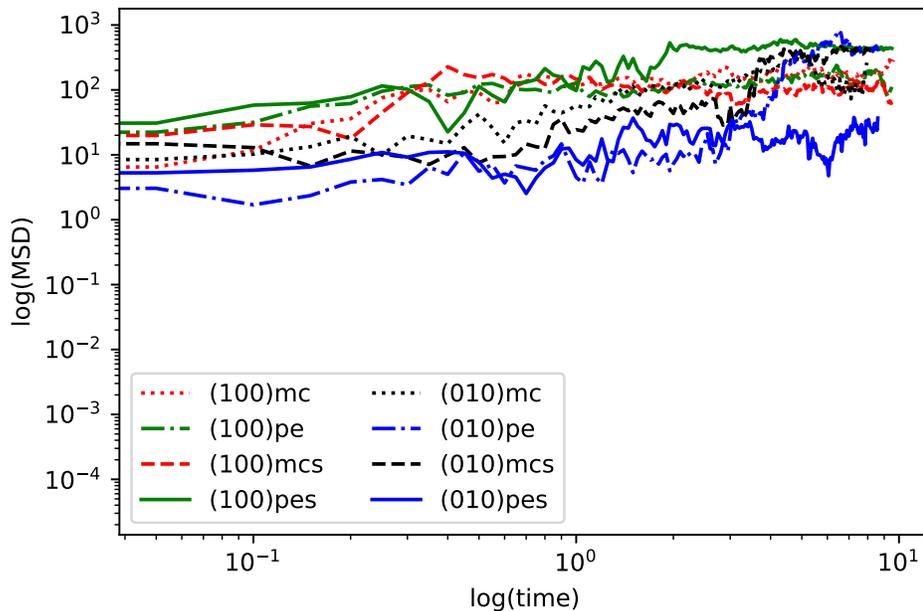
FIGURE 3.6: Water number density ( $\rho_N$ )

(a) (100) systems.



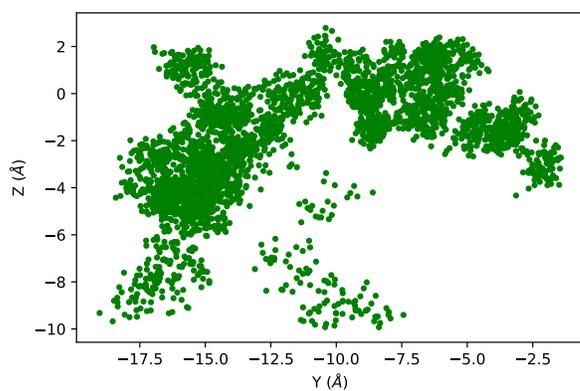
(b) (010) systems.

FIGURE 3.7: MSD of oligomers.

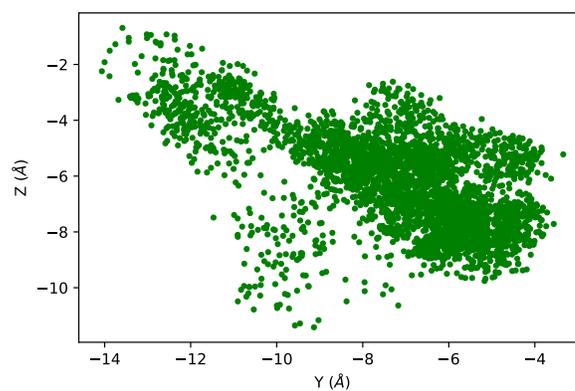


of 50 to 500 Å<sup>2</sup>. This is smaller compared to the overall surface area. Hence we can conclude that, though the oligomers move along the surface (slide), it is confined to a smaller region on the surface and it is typically in the neighbourhood of the adsorption site. There were no considerable differences between the surfaces in terms of overall movement but it should be noted that in the (010) surface the movement along the crystal's length axis ( $z$ ) was lower than the movement along the  $y$ -axis. This was not the case on the (100) surface. The movement was almost uniform along both surface axes, this could be because of the (100) surface being uniform.

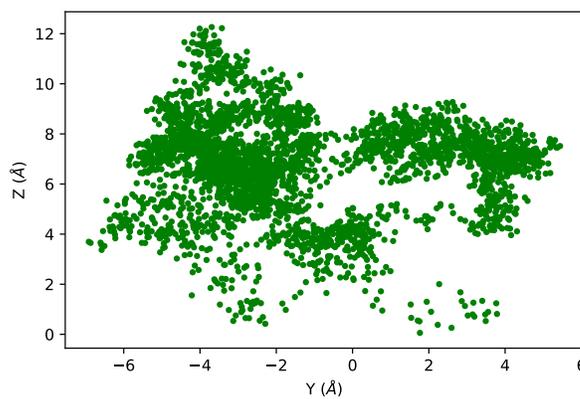
FIGURE 3.8: Oligomer movement along the (100) surface.



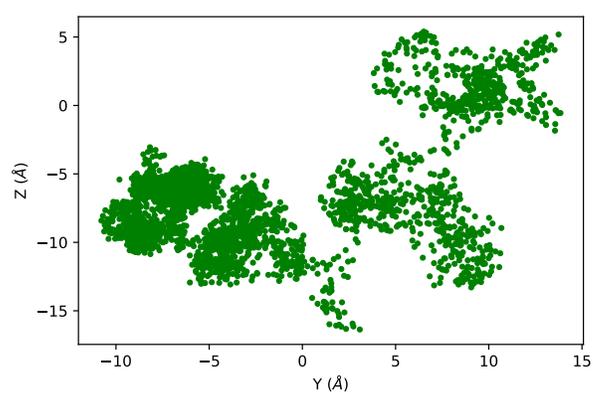
(a) (100)mc



(b) (100)mcs

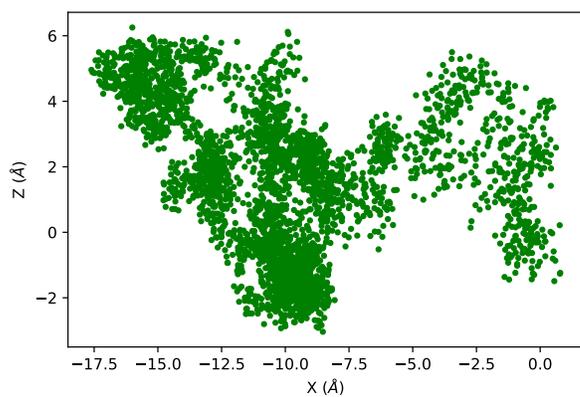


(c) (100)pe

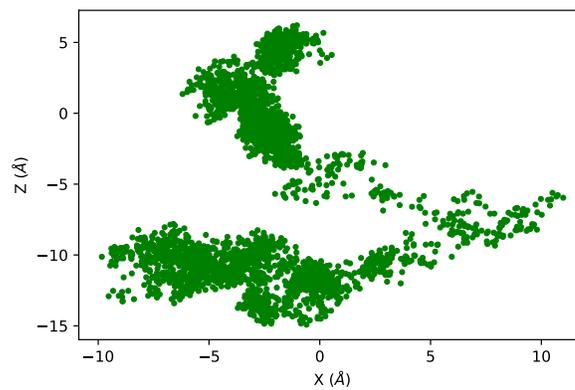


(d) (100)pes

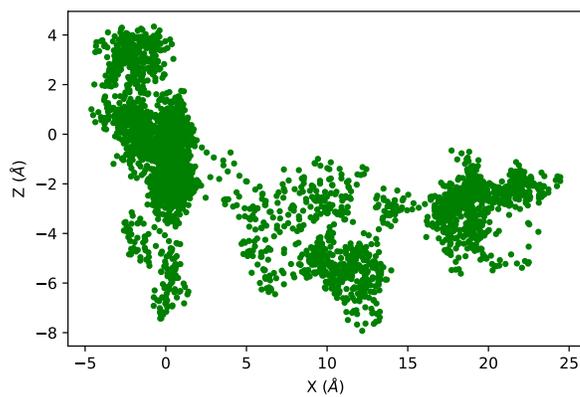
FIGURE 3.9: Oligomer movement along the (010) surface.



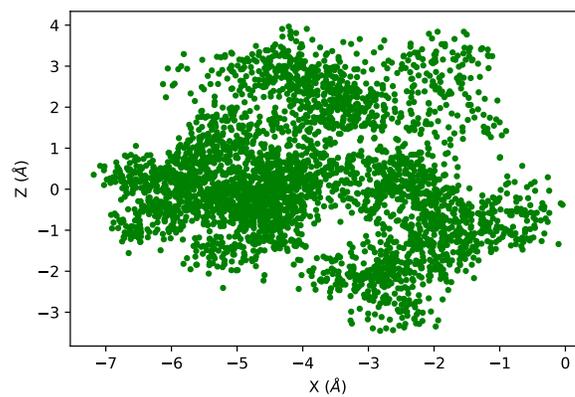
(a) (010)mc



(b) (010)mcs



(c) (010)pe



(d) (010)pes

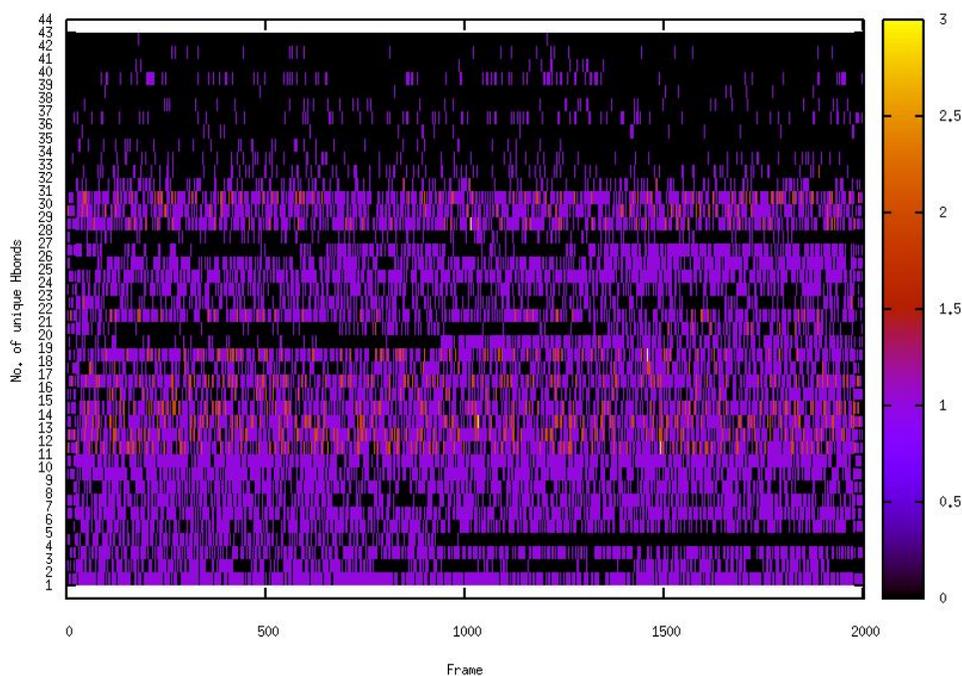
### **3.2.4 Hydrogen bond - Oligomer and water**

The hydrogen bond between oligomers and water is another interesting observation to make as this is an indicator of the oligomer's bulk interactions. Since most of the observations are quite intuitive, a mix of systems was considered to represent the oligomer-water interaction.

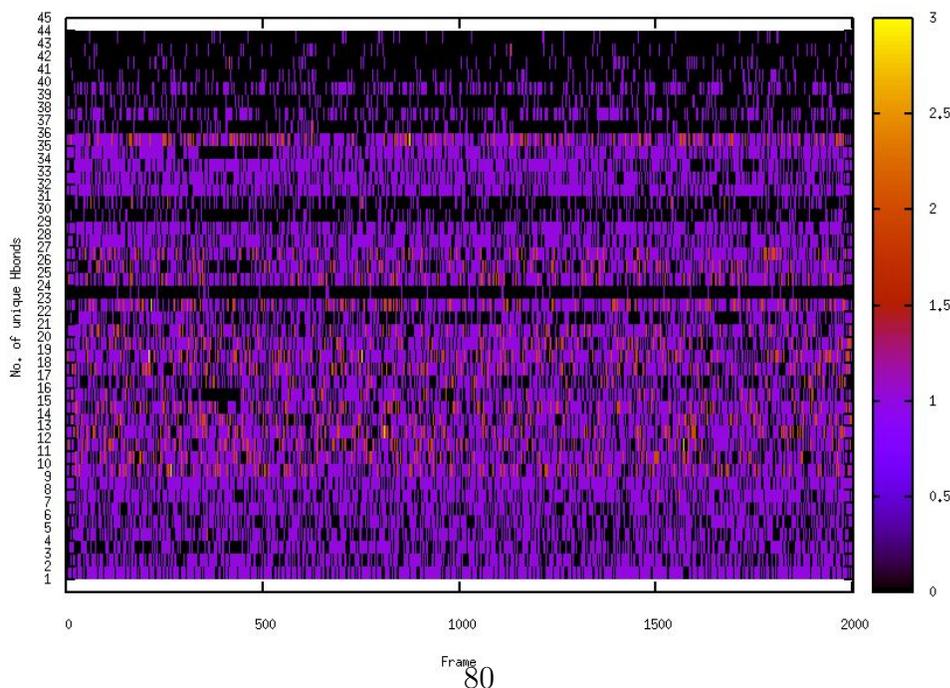
Figures 3.10 and 3.11 show the number of unique hydrogen bonds formed between oligomer and water molecules in the (100)pes system as a function of time, represented in terms of frames. Each frame corresponds to 0.002 ns i.e. the whole timescale is for 4 ns. In fig. 3.10 the interaction between pentaose and water in presence of salt, near and far from the (100) surface is shown. Similarly in fig. 3.11 the interaction between MC and water in the absence of salt, near and far from the (010) surface is shown.

At first look, it can be deduced that the number of hydrogen bonds formed between oligomer and water molecules is larger when the oligomer is far from the surface. It can also be seen that the number of unique hydrogen bonds between oligomer and water is higher in the case of pentaose, this of course, is due to the presence of more hydrogen bond acceptors in case of the pentaose. There are no specific patterns to follow in these interactions, but if specific hydrogen bonds were found to influence the overall interactions, then they will be studied individually to infer from them.

FIGURE 3.10: Hydrogen bond between Oligomer and water - I. The number of unique hydrogen bonds between oligomer and water in the (100)pes system as a function of time, in terms of frames (1 frame = 0.002 ns) is shown, the number of hydrogen bonds in each unique type per frame is denoted along the right-hand side y-axis

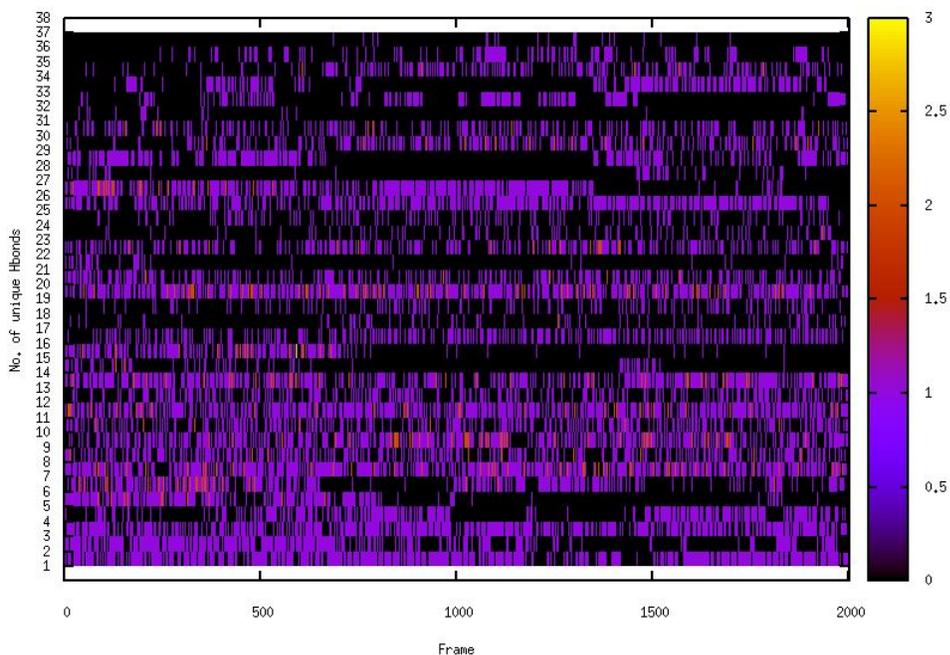


(a) (100)pes - closer to the surface

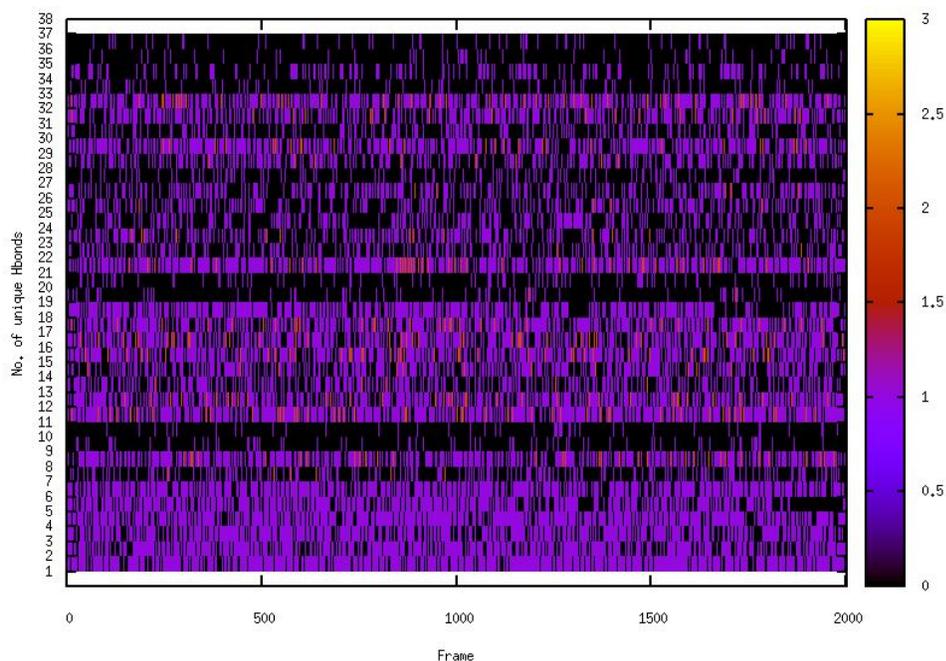


(b) (100)pes - farther from the surface

FIGURE 3.11: Hydrogen bond between Oligomer and water - II. The number of unique hydrogen bonds between oligomer and water in the (010)mc system as a function of time, in terms of frames (1 frame = 0.002 ns) is shown, the number of hydrogen bonds in each unique type per frame is denoted along the right-hand side y-axis



(a) (010)mc - closer to the surface



(b) (010)mc - further from the surface

### 3.3 Free energy calculation

The most important tool used by us to understand the CNC-oligomer interaction was the free energy calculation. The free energy values were obtained in terms of PMF via umbrella sampling. The free energy values calculated in this work are the free energies of desorption. One important factor to remember is that the calculated values are not absolute free energies but free energy differences, which makes these values a relative quantity i.e. when interpreting, the discussion must be placed in terms of two/more states. These values give the estimate of the likelihood of occurrence between two macrostates. For example, questions such as, what is the likelihood of desorption/adsorption occurring at a given distance from the substrate? what happens under different physical cases? and to what degree does it happen? can be addressed with these measurements.

#### 3.3.1 PMF profile

The PMF value is equivalent to the free energy. The relationship between these two entities was discussed in subsection 2.4.2.8. Figure 3.12 shows the PMF profile of the (100)pes system as the function of  $\xi$ . This system was chosen to illustrate the observations as it carries the pentaose oligomer, which is a short cellulose chain and it exhibits maximum affinity towards CNC naturally. Also, salt is present in this system and all zones are visible distinctly.

Hence this system serves as a good overview of all the PMF profiles. It can be seen in fig. 3.12, that the profile has repeating step-like structures. The steps in the PMF curve

are constituted by different zones along the reaction coordinate and these zones are discussed in the next section. Different positions of interest along the reaction coordinate are marked in red and these positions are where the oligomer configurations were captured from normal MD simulation for analysis. The discussion on configurations is made in subsection 3.3.2.

### 3.3.1.1 Zones in the PMF profile

The PMF profiles obtained from our simulations had similar characteristics of the free energy curves obtained previously by Payne et al. for decrystallization of cellulose-I $_{\beta}$ . Also, a similar stepped structure was observed in the PMF profiles obtained by Muthukumar and Khare for desorption of a celohexaose chain from the CNC surface. This shows the consistency in the free energy observations. Like it was mentioned in the previous section, the stepped structure of PMF curve is the composed of three different zones along the  $\xi$ . We refer to these zones as the favourable zone, slopes and plateaus.

- **The favourable zone:** This zone or to be more precise, this position is the point along the PMF profile with the lowest possible PMF value. i.e. it occurs in the deepest point of the PMF curve's well. This zone corresponds to the position with the highest likelihood of oligomer occurrence near the CNC surface. For each curve, this position is the most favourable zone for a given oligomer with respect to the CNC surface under the given physical condition (with/without salt). In fig. 3.12 it is marked as "1".

The desorption of the oligomers occurs in steps where one monomer unit at a time desorbs from the surface. As they desorb, the intermolecular and intramolecular

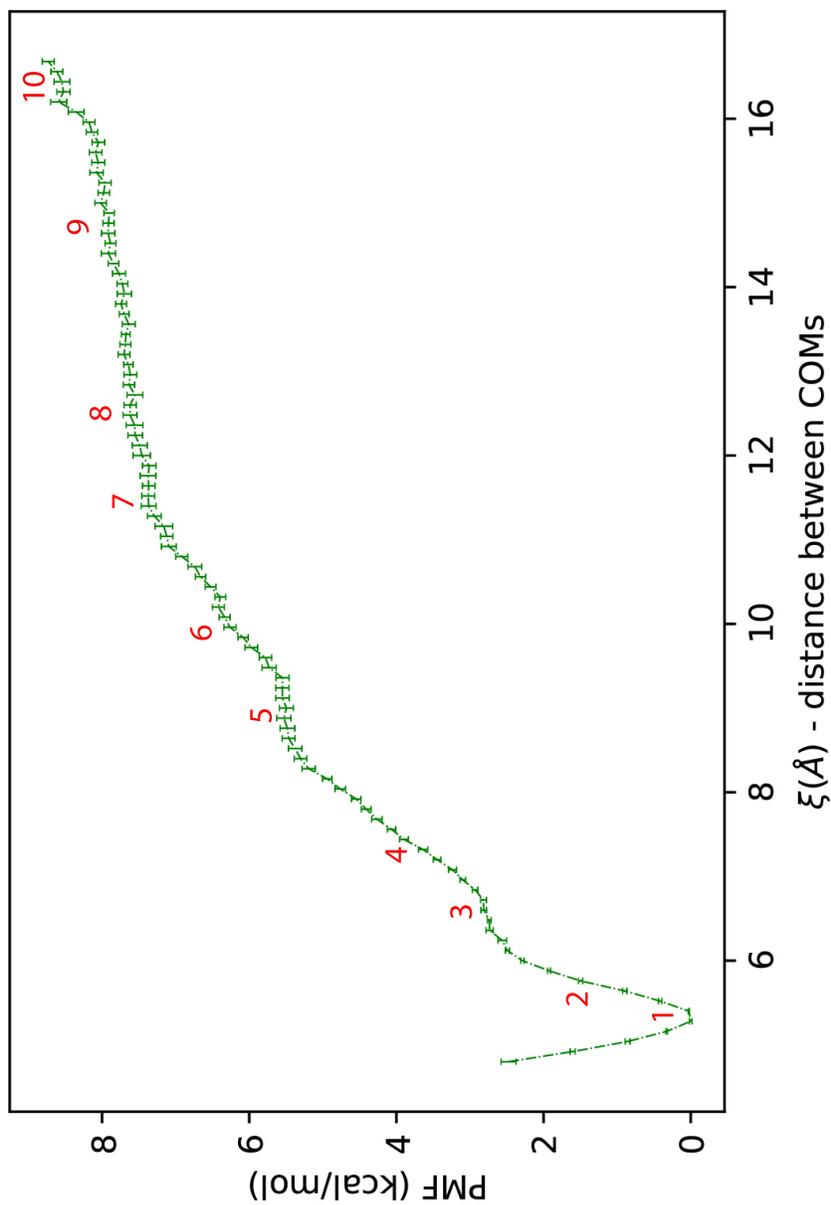


FIGURE 3.12: PMF as the function of  $\xi$  for the (100)pes system with 25 windows. The curve has three different zones namely, the favourable zone, marked by 1; the slopes, marked by 2, 4, 6 and the plateaus, marked by 3, 5, 7. Positions of interest where the oligomer configurations were captured are marked in red.

interactions start to change. The significant interactions at the interface are the hydrogen bonding interactions. Two major types of hydrogen bond groups are the intermolecular hydrogen bondings between the oligomer-CNC surface pair and the intramolecular hydrogen bondings observed between the units of the oligomer. The slopes and plateaus are resultants of favourable and unfavourable events which result in increase or decrease of free energies.

- **Unfavourable event:** When an oligomer unit desorbs from the surface, it loses its hydrogen bonds with the surface and the hydrophobic unit is exposed to the water molecules. This is a highly unfavourable event and it results in the increase of free energy. So, as the number of desorbed units increases, the free energy value gradually increases and this is the reason for the monotonic increase in the function value.
- **Favourable event:** The favourable event is the formation of intramolecular hydrogen bondings within the oligomer chain. As more and more monomer units are desorbed, they form hydrogen bonds with the adjacent units, this is a highly favourable event and it results in the decrease of free energy.
- **Slopes:** All oligomer units in this study carry 5 monomer units. The slopes are associated with a sequential increase in the free energy and this increase in free energy is due to the unfavourable events i.e. desorption of oligomer units from the surface and exposing itself to water. In general, every time an odd-numbered/unbalanced oligomer unit desorbs from the surface it results in an increase in free energy and gives rise to the gradual increase in slope we see on the PMF curve. The zones which exhibit this behaviour are collectively called 'slopes'.

The two of the slopes are marked by “2, 4” in fig. 3.12.

- **Plateaus:** The plateaus are formed when there is a balance between the unfavourable and favourable events i.e. when the free energy increase due to the desorption of oligomer unit is balanced by the free energy decrease due to the formation of intramolecular hydrogen bonds within the oligomer unit. These intramolecular hydrogen bonds are possible due to the cellobiosyl repeat mentioned in subsection 1.1.2. This balance in free energy increase and decrease results in a flat region along the PMF curve where there is no difference in free energy between two immediate reference states.

The zones which exhibit these characteristics are collectively referred to as 'plateaus'. When two adjacent monomer units are desorbed, they are coplanar and they form intramolecular hydrogen bonds with ease. Hence the plateaus set in immediately after the desorption of an even-numbered unit. Some of the plateaus are marked by “3, 5, 7” in fig. 3.12. Similar observations and arguments have been made previously<sup>[49]</sup> based on hydrogen bonding networks observed in cellulose<sup>[17,52]</sup>. The PMF curve propagates along the reaction coordinate and eventually, a constant free energy of desorption value is achieved.

### 3.3.2 Configurations along the PMF curve

To confirm our theory provided in the above section, which can be summarized as "The oligomer takes different configurations along the reaction coordinate. The slopes are due to the increase in free energy associated with odd-numbered unit's desorption and

plateaus are due to the balance between free energy increase and decrease. The free energy decrease is due to the formation of intramolecular hydrogen bonding resulting from the desorption of even-numbered units".

The configurations of the pentaose unit from the (100)pes system were recorded along the reaction coordinate at the different position marked in fig. 3.12.

- Figure 3.13(a) shows the pentaose (pe) oligomer adsorbed to the surface and all 5 units are seen aligned along the z-direction. This is the configuration observed at the favourable zone and all of these snapshots are from unbiased MD simulations. Statistically more reasonable configurations like the ones where all rings are perfectly parallel to the  $YZ$  plane are possible too.
- Figure 3.13(b) and (d) show the configuration observed in the slopes. In case of (b), the configuration is only slightly different from (a) with one of the rings slightly rotated along the chain length but the desorption of three rings is clearly visible in (d). This is consistent with our initial hypothesis regarding the slopes, which is, with the desorption of odd-numbered units, the free energy increases.
- Figure 3.13(c) and (e) show the configurations observed in the plateau zones. In case of (c) two desorbed units and in case of (e) four desorbed units can be seen. This too is consistent with our definitions of plateau zone. These even-numbered desorbed oligomer units form balancing intramolecular hydrogen bonds resulting in the decrease of free energy. This is followed by the balance between free energy increase and decrease which results in the observed plateaus.

As we move along the reaction coordinate, it can be seen that the oligomer gradually

desorbs from the surface. In fig. 3.13(f) we can see the partial desorption of the last unit. This position, marked as '6' in fig. 3.12 corresponds to a region with two slopes connected by a short plateau. It indicates that with an increase in the distance between the surface and the oligomer, desorption becomes easier.

The subsequent zones can be seen with shorter or no plateaus and this is due to the reason that there are no more adsorbed units which can desorb and form new intramolecular hydrogen bonds. Hence the PMF value gradually increases without any free energy decreases to balance it. There were no significant configuration changes in (g) and (h) except the fact that the chain has attained more degrees of freedom. In the final two configurations (i) and (j) which are about 14 Å away from CNC's first layer's COM (i.e. beyond the short-range interaction cut-off of 12 Å), it can be seen that the chain has higher degrees of freedom. It can be seen by the orientation changes of the chain with respect to the previous configurations. Overall, the configurations observed from MD simulations along the  $\xi$  are consistent with the hypothesis made based on the PMF curves.

### 3.3.3 Overall comparison

The PMF curve and its characteristics in accordance with the molecular interactions were discussed in the previous section. In addition to previously mentioned systems, one additional system, which is a modified version of (100)mcs was simulated. The new system is referred to as (100)mcs2 and in this system, a salt concentration of 100 mM as opposed to other salted systems (50 mM) was used. This system was studied as a

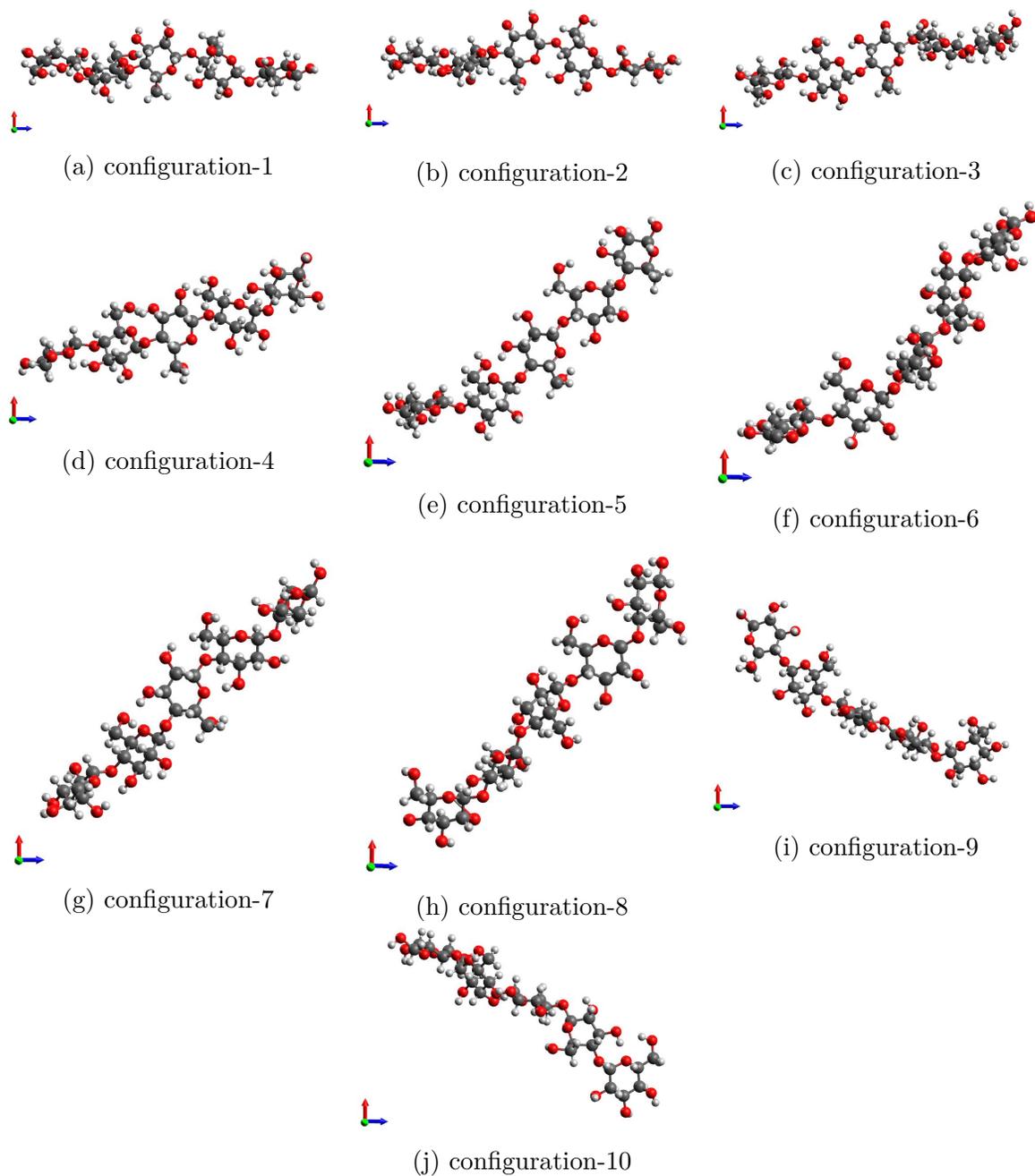


FIGURE 3.13: Configurations of pentaose oligomer at different positions along  $\xi$ . The x, y and z-directions are given in red, green and blue colour respectively.

special case in hopes to find the effects of higher salt concentrations on the system properties and the PMF profile. Figure 3.14 shows the PMF profiles of all the systems. The PMF as the function of  $\xi$  is given and at first glance, it can be seen that there are two major groups of curves belonging to the two different surfaces. Within these two major groups, a subset of curves belonging to each oligomer can be seen. The systems with salt exhibit slightly different PMF profiles. This shows that the PMF profile varies based on the surface, polymer and presence of salt. Hence a break down and interpretation of the curves in terms of these physical factors are necessary and it is done in the subsequent sections.

In case of the (100) surface curves, it can be seen that there is a clear distinction between MC and pentaose and this is due to the higher affinity of pentaose towards CNC when compared to MC. The favourable zone is almost same in all the cases, and the small differences in the position are within the window width used for PMF calculation (0.5 Å). The salt effects will be dealt with in subsection 3.3.3.1. It can be seen in fig. 3.14 that there are three groups closer to the surface which slowly diverge. Some of these curves have sections with small or no plateaus and these zones correspond to a scenario where the balance between the favourable-unfavourable events (discussed in subsection 3.3.1.1) is short and this is quite evident in the cases of pentaose where there is more hydrogen bond (HB) donor-acceptor pairs compared to MC. Hence the number of steps in case of pentaose is higher than that of MC.

In case of the (010) surface curves, it can be seen that there is no clear distinction between MC and pentaose even near the surface. The curves diverge slowly as we move away from the surface. Also, there are not many steps in the curves when compared to

the ones observed in the (100) surface cases. This can be explained by the existence of zones with lower slopes and longer plateaus i.e. both the free energy increase due to desorption and the balance between the event which gives the plateaus are persistent when compared to the other surface.

We can break down the curves further and look into each window further but it is highly unlikely that we would observe distinctions between windows as the recorded COLVAR value is prone to fluctuations. The (010) surface curves have no clear distinction between them closer to the surface and it indicates that there is not much difference in the affinity of the oligomers towards CNC along this surface. Also, this surface has persistent events (free energy increase/decrease and their balance). Overall differences observed in the major group of curves can be explained by the differences in the surfaces. The oligomer characteristics seem to affect individual PMFs (number of steps, plateau length etc.). Hence the interactions clearly differ between surfaces with specific characteristics depending on the oligomer.

### **3.3.3.1 Salt effects**

In fig. 3.14 it can be seen clearly that the systems with salt stand out. The presence of salt in the bulk even at a very low concentration of 50 mM showed significant changes in the oligomer interactions. The presence of salt makes the steps more visible. For example, in case of (100)pes, it can be seen that the steps are very distinct and they correlate with the number of monomer units. This is a good indication of bulk system's influence on the interactions within the oligomer unit i.e. longer plateaus are formed when even-numbered oligomer units form intramolecular hydrogen bondings and the

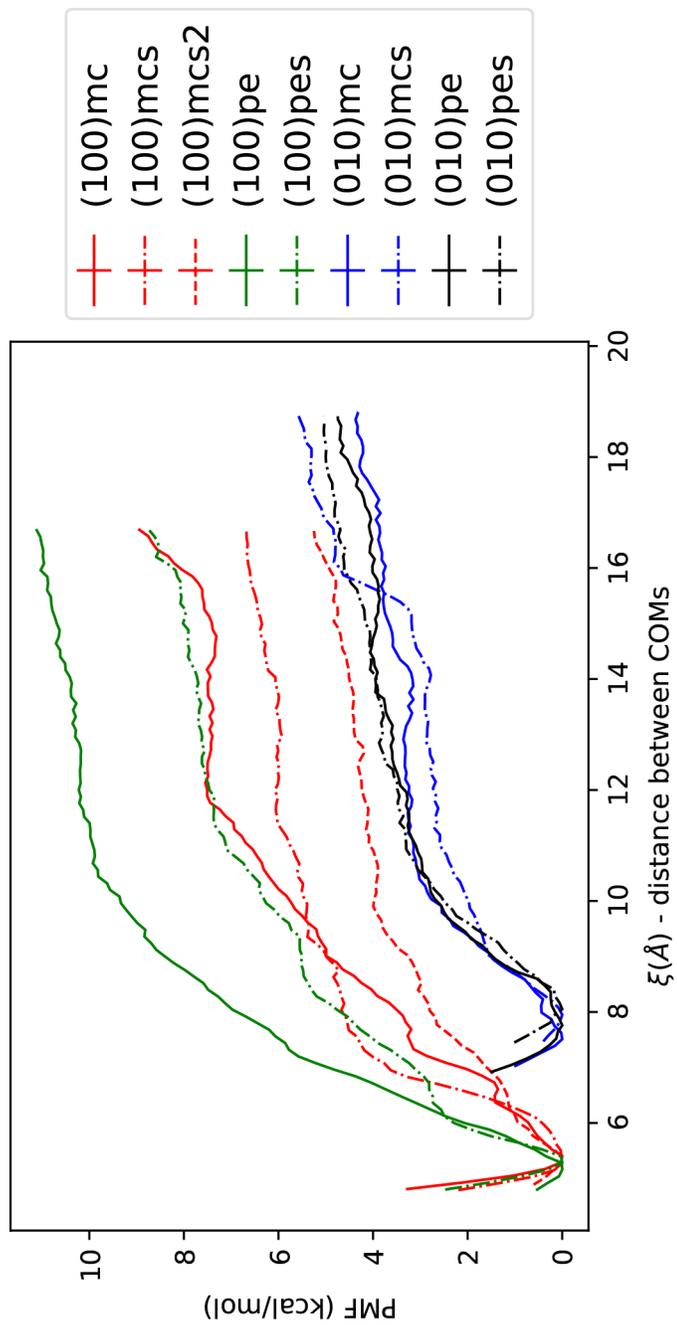


FIGURE 3.14: PMF curves of the systems as the function of  $\xi$ . Two major groups belonging to two different surfaces can be seen.

presence of salt seems to favour this event. Similarly steeper slopes denote the increase in free energy resulting from the oligomer unit's exposure to water (bulk) and the presence of salt makes this existing unfavourable event even more unfavourable.

To check if the increase in salt concentration enlarges the observed effects, the (100)mcs2 with 100 mM salt concentration was simulated and its effect is clearly significant compared to the case of (100)mcs with 50 mM salt concentration. In case of the (010) surface, we observed that there is no clear distinction between different systems. So if the effect of salt was significant we should be able to observe it more clearly on this surface than in the (100) surface and we were able to witness it. In fig. 3.14 we can see that at and above a distance of 14 Å between the COMs of CNC and oligomer, the effect of salt becomes significant. This is accompanied by the sudden and steep increase in the PMF values immediately after this point.

It should be noted that this distance is outside our short-range interaction cut-off of 12 Å, where all non-bonded interactions are uniform. We do not expect any specific interactions between CNC and the oligomer except for the expected uniform interactions existing in the whole system. So it can be interpreted as the driving force changing from intermolecular interactions to the bulk effect. One more important observation on the (010) surface is the presence of two different favourable zones for systems with and without salt. This can be explained by the position of ions in the systems.

The (010) surface carry sulphate groups and the surface had a net charge of -40 Coulomb in all the cases. Compared to  $\text{Cl}^-$ ,  $\text{Na}^+$  is smaller in size and it is positively charged. Hence it can easily diffuse towards the (010) surface. The position of salt ions

in all salted systems is shown in figs. 3.15 and 3.16. It can be noted that the ions are diffusing well in the bulk water for the (100) systems, while there is a clear segregation of ions in the cases of (010) systems.

Hence the ionic positions do have some effects on the oligomer occurrence on the surface but this needs further study to establish the actual mechanism. The PMF value is higher for systems with salt and this makes the oligomer in these systems to adsorb readily on to the CNC surface. It was experimentally observed that the polymer adsorption increases with increase in salt concentration. Hence our observation that presence of salt strongly influences the oligomer interaction is consistent with that of the experimental one.

### **3.3.3.2 Bulk effects**

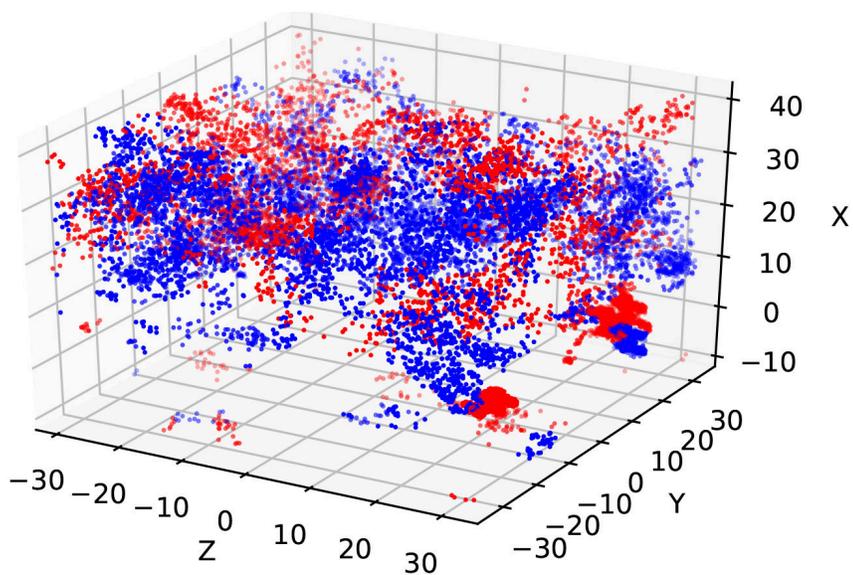
It should be mentioned that the bulk effects other than the presence of salt might significantly affect the oligomer interactions. For example,

- (1) Gelation is driven by phase separation caused by hydrophobic bondings above the lower critical solution temperature (LCST)
- (2) Based on the hydrophilic lipophilic balance (HLB) values of MC (10-11.25) and cellulose (12.45), cellulose is more hydrophilic compared to MC<sup>[63]</sup>

Effects like these are significant when the whole system is considered. So bulk properties do have some effects on the interactions. Since we were interested more in the intermolecular interactions we did not explore them.

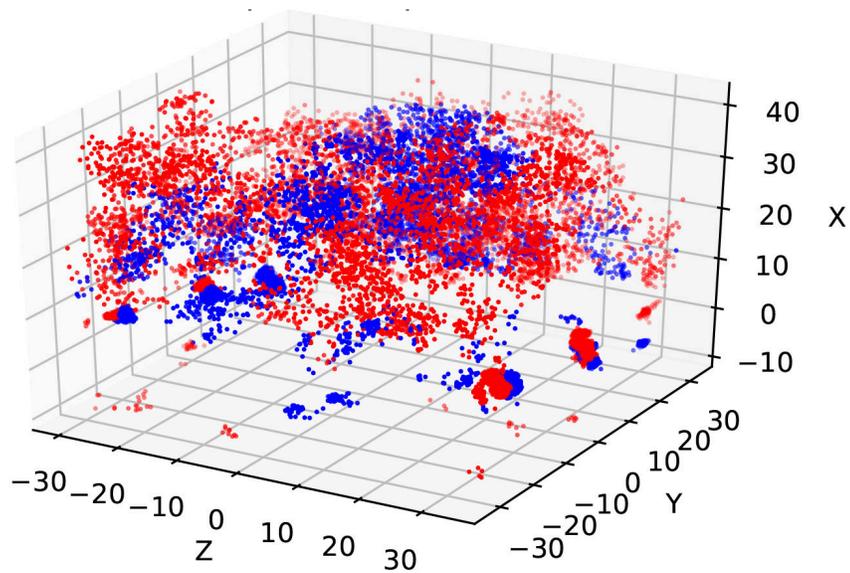
FIGURE 3.15: The position of ions on (100) surface, the Sodium ions are given in blue and the Chloride ions are given in red.

■ Na+  
■ Cl-



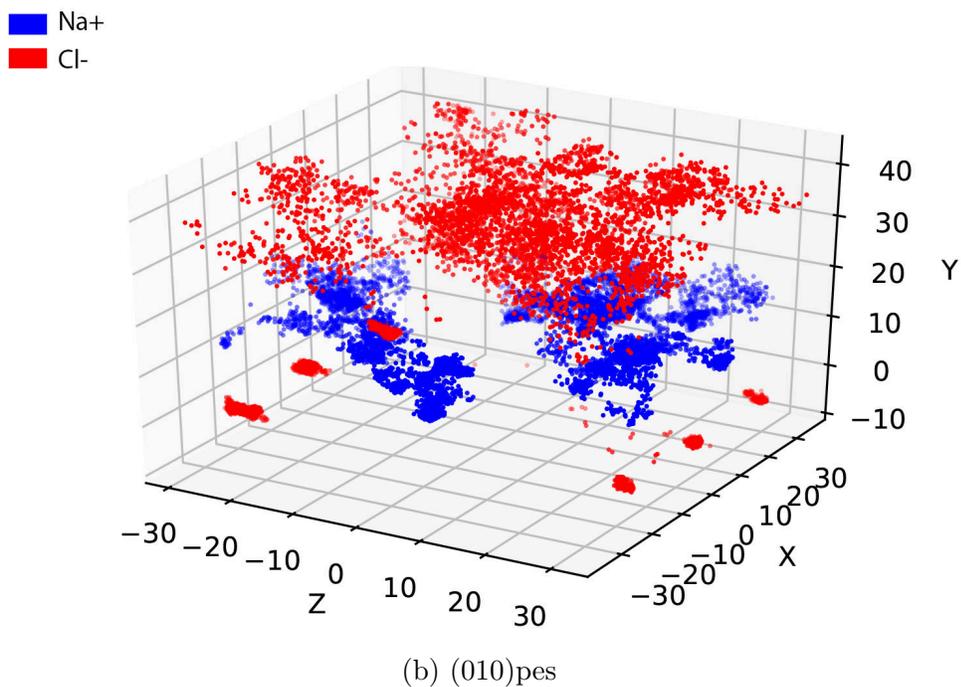
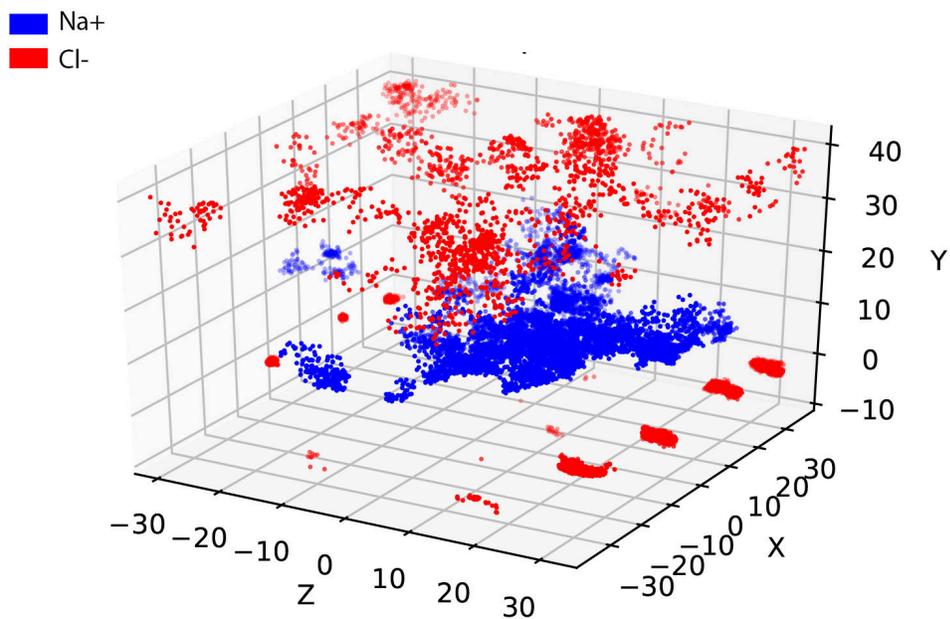
(a) (100)mcs

■ Na+  
■ Cl-



(b) (100)pes

FIGURE 3.16: The position of ions on (010) surface, the Sodium ions are given in blue and the Chloride ions are given in red.



## 3.4 Other analyses

In general, MD simulations hand us data with numerous degrees of freedom and it is often important to understand parameter-property relationships from these data. It can be done by dimensional reduction and data from it can be interpreted into information using methods like cluster analysis. In our case, for each system, the distance of the oligomer from the surface seem to influence the observations. Hence a set of properties exhibited by the oligomer were recorded as a function of oligomer position along the normal direction and these data were used to find clusters in observed property values.

### 3.4.1 Oligomer Properties

A series of properties exhibited by the oligomers such as oligomer's end-to-end distance, radius of gyration ( $R_g$ ), root-mean-square deviation (RMSD), surface accessible surface area (SASA) and bulk properties like water shells binding the oligomer were recorded as a function of oligomer's position along the normal axis ( $x$  for (100) surface and  $y$  for (010) surface). These properties will be a good estimate of oligomer's dynamics along the reaction coordinate. In figs. 3.17 to 3.21 showing the oligomer properties, the CNC surface is marked in red.

- **End-to-end distance:** The first oligomer property which was recorded is the end-to-end distance. All oligomer units were made up of 5 monomer units ( $\approx 2.6$  nm in length). The persistence length ( $l_p$ ) of MC as measured by Bodvik et al. in dilute solutions is 5.8 nm. Though the oligomer's length is well within its persistence length, the end-to-end distance does fluctuate to a very small degree due to its

interactions with the substrate and bulk. Figure 3.17 shows the end-to-end distance of oligomers in different systems. It can be seen that irrespective of the system the values are concentrated near the surface. This indicates that the oligomer once adsorbed on to the surface does not bend much. In cases of systems with salt, we can see multiple clusters along the reaction coordinate. This indicates that the oligomer takes specific configuration with almost constant end-to-end distance at each cluster.

- **RMSD:** From the RMSD values shown in fig. 3.18, it can be seen that similar to end-to-end distance and other properties, the data points are concentrated closer to the surface showing that the oligomer takes a constant/similar configuration near the surface. RMSD values calculated here are in reference to the initial state. One interesting observation in the cases of the (010)mcs and (010)pe showed in fig. 3.18(f) and (g) is the existence of two clusters of data points closer to the surface. It indicates that the oligomer takes two entirely different configurations which vary from the initial configuration.
- **The Radius of Gyration ( $R_g$ ):** The  $R_g$  is a more reasonable measure of chains linearity when compared to the end-to-end because end-to-end distance gives the lowest separation distance irrespective of the direction.  $R_g$  gives a sense of the size of the polymer coil and it is more intuitive than end-to-end distance. Figure 3.19 shows the  $R_g$  values of the oligomers observed at different positions along the normal direction. Same as the cases of other oligomer properties, concentrated  $R_g$  values are found closer to the surface and the values are relatively higher, this denotes that the oligomer is linear and it takes similar configurations near the surface. In cases of systems with salt, a band of  $R_g$  values are observed which can be interpreted as

that the oligomer maintains its configuration after desorbing from the surface.

- **SASA:** The next property which was calculated for the oligomer chains was the SASA, using the linear combinations of pairwise overlaps (LCPO) method. In AMBER the LCPO method is used to find the implicit solvent effects, though the model used in the actual simulation is an explicit solvent model, this value gives us the change in solvent accessibility to the oligomer chain as they have hydrophobic residues along the chain length.

This value is often used to determine the transfer free energy for biomolecules. Figure 3.20 shows the results of this calculation for each system. As a general trend, it can be seen that irrespective of the systems all oligomers along all surfaces exhibit lower SASA value near the surface. This is the measure of the surface area of the oligomer accessible to the water molecules. As expected the values are slightly lower near the hydrophobic (100) surface compared to the hydrophilic (010) surface.

- **Water shell:** The oligomers in the systems have good hydrogen bonding capacities, hence the bulk system around them need to be considered when interpreting their properties. Figure 3.21 shows the water shells around the oligomer with lower and upper bound, 3.8 and 5 Å respectively. Like it was mentioned earlier, the systems are non-homogeneous and hence measures like RDF of water molecules would not make much sense due to the lack of rotational symmetry. So the water molecule count was calculated with respect to the oligomer position. The number of water molecules around the oligomers was almost constant in all the cases. The slight differences between the cases with and without salt can be attributed to the interactions of

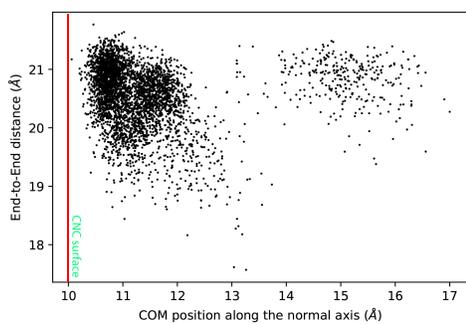
ions with bulk water.

Overall the observed oligomer properties help us to draw the following conclusions, (1) The oligomers exhibit similar configurations closer to the surface, (2) The different clusters of data along the normal axis show existence of different oligomer configurational states along it and (3) The cluster density gives the importance of each of these states. Though this is a good qualitative measure, quantitative measures will help to get further insight into the oligomer interactions. For that, numerical cluster analysis like K-means clustering, hierarchical agglomeration etc. needs to be done systematically, and that will help to classify the observations better.

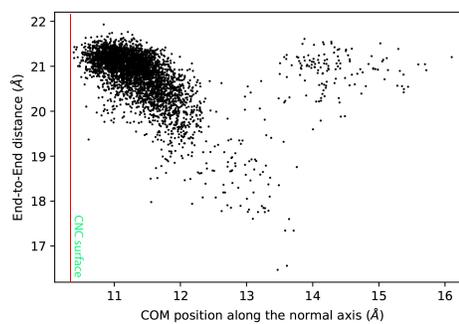
### 3.4.2 Intramolecular Hydrogen bond within the oligomer

Based on the PMF curves and other analyses made with respect to the oligomers, it is very clear that the oligomers proximity to the surface, the oligomer intermolecular and intramolecular interactions are very crucial. In case of the oligomer adsorption to the surface, it was explained in subsection 3.3.1.1 that the intramolecular hydrogen bond formation is the favourable event which drives the free-energy lower and results in constant free energy zones/plateaus in the PMF profiles. Hence as we move away from the surface, the total number of intramolecular hydrogen bonding formed within the oligomer chain should increase and eventually reach a constant value. To test this hypothesis, 25 different section from the (100)pes system, which was used to explain the PMF profile and the configurations in it (subsection 3.3.1) were studied for hydrogen bonds. Each of these 25 sections corresponds to each of the windows used to compute the PMF values of the system. Figures 3.22 and 3.23 summarizes the observations made

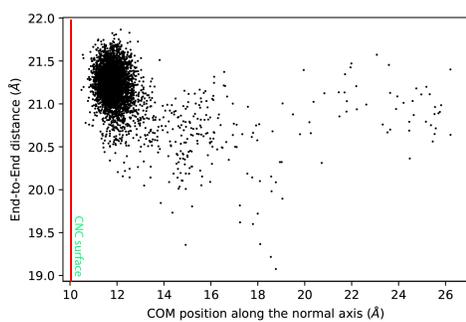
FIGURE 3.17: End-to-end distance of the oligomer chains.



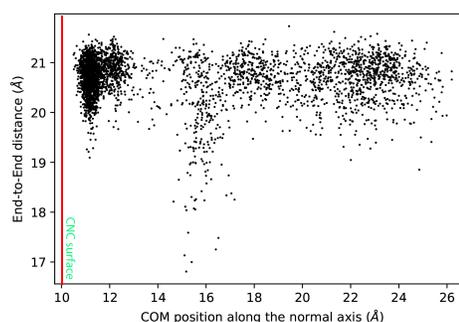
(a) (100)mc



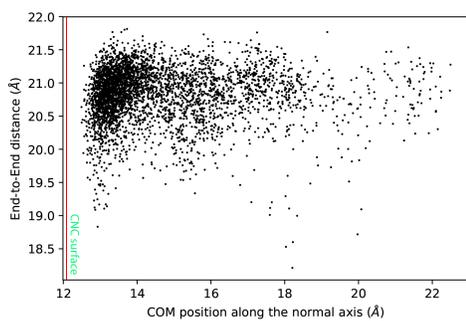
(b) (100)mcs



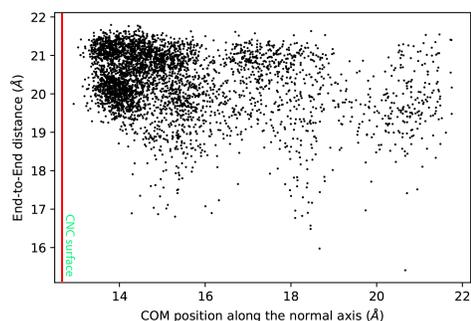
(c) (100)pe



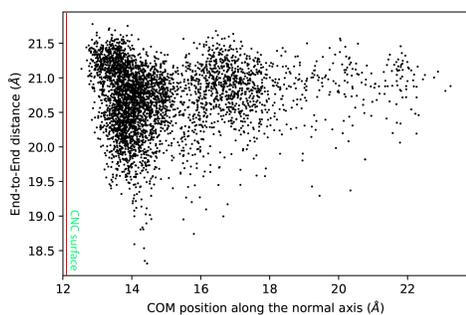
(d) (100)pes



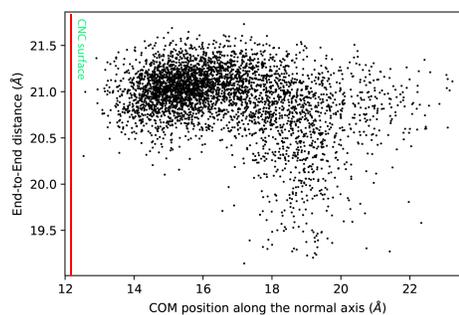
(e) (010)mc



(f) (010)mcs

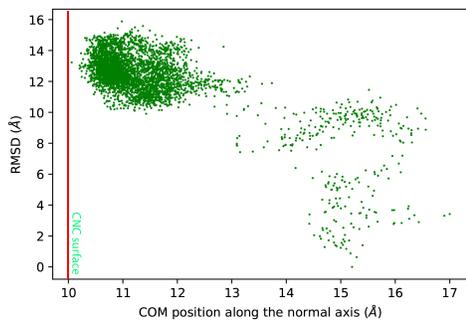


(g) (010)pe

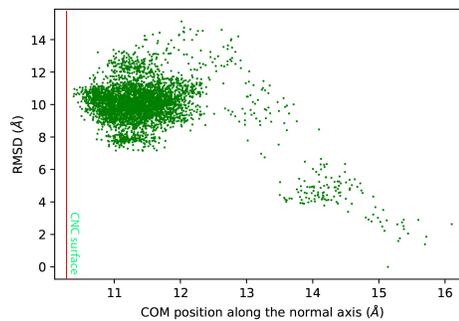


(h) (010)pes

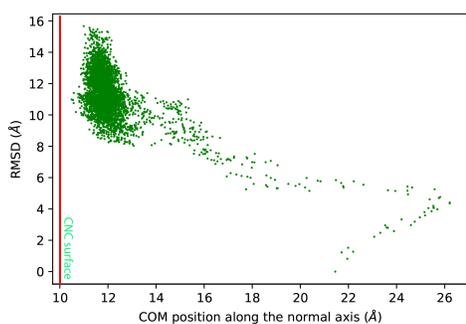
FIGURE 3.18: RMSD values of the oligomer chains.



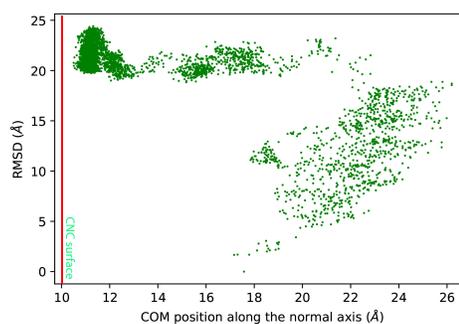
(a) (100)mc



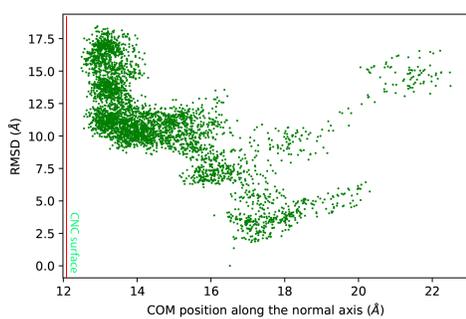
(b) (100)mcs



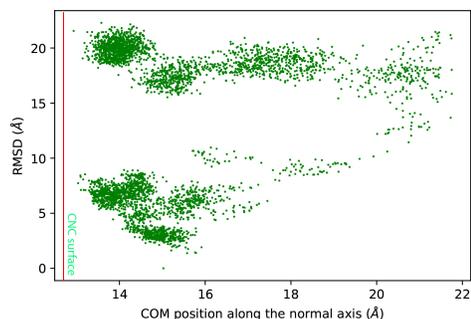
(c) (100)pe



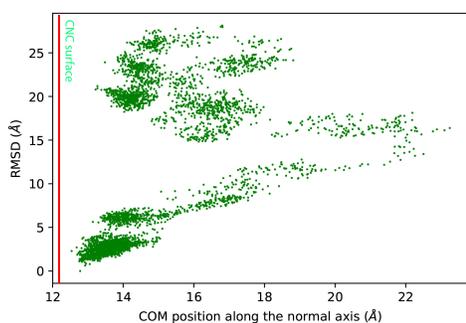
(d) (100)pes



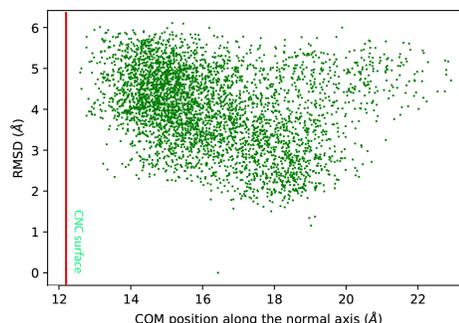
(e) (010)mc



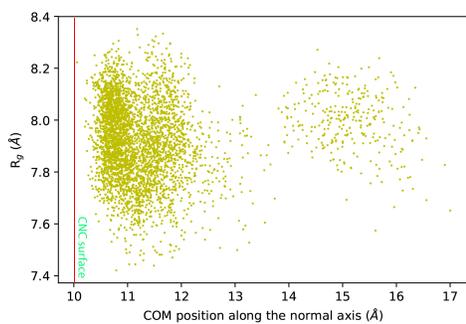
(f) (010)mcs



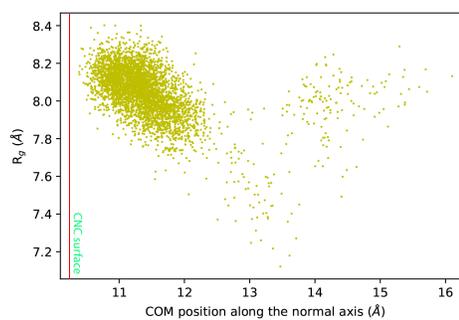
(g) (010)pe



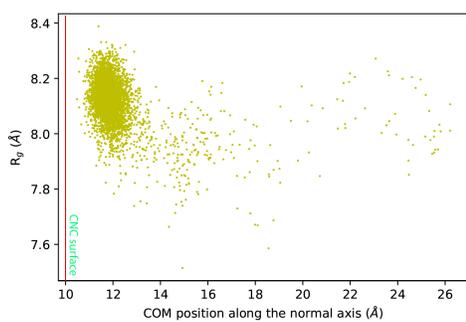
(h) (010)pes

FIGURE 3.19:  $R_g$  of the oligomer chains.

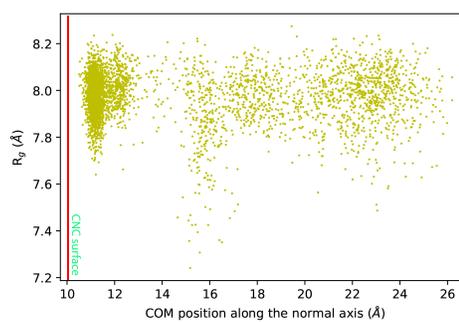
(a) (100)mc



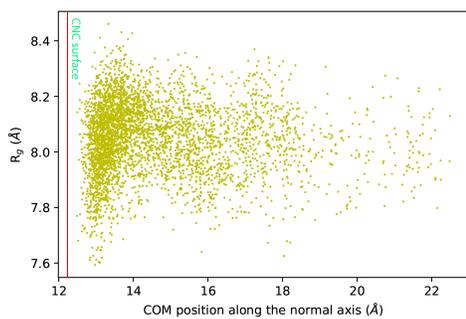
(b) (100)mcs



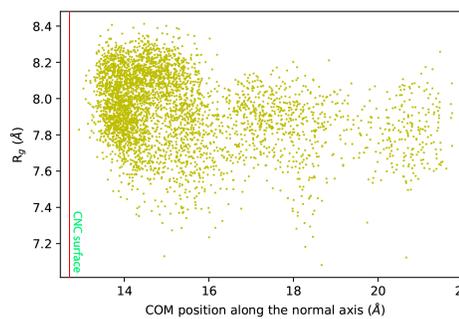
(c) (100)pe



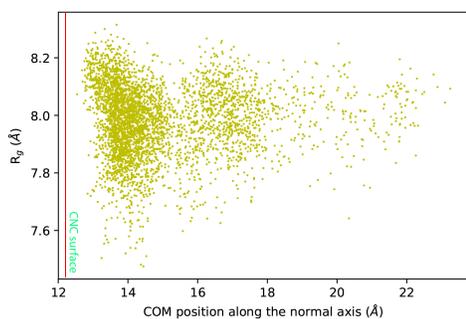
(d) (100)pes



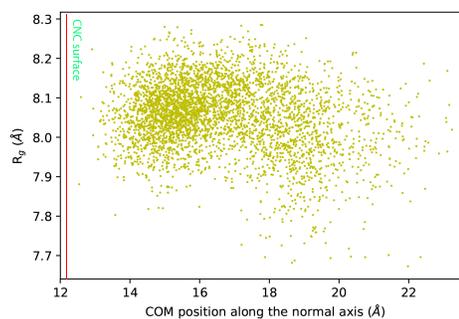
(e) (010)mc



(f) (010)mcs

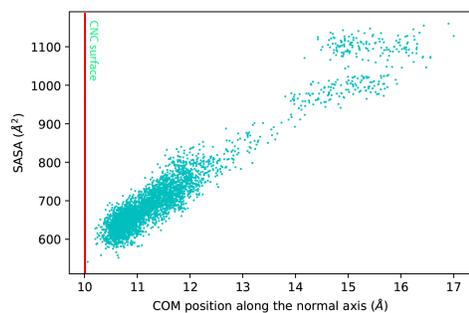


(g) (010)pe

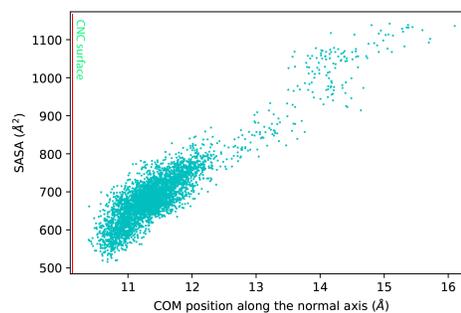


(h) (010)pes

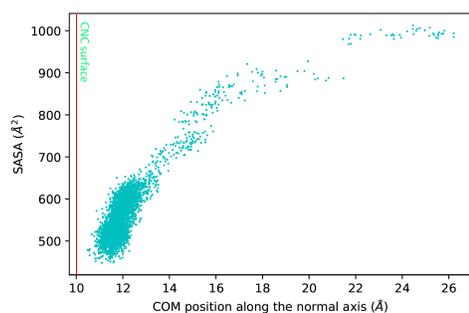
FIGURE 3.20: SASA of the oligomer chains.



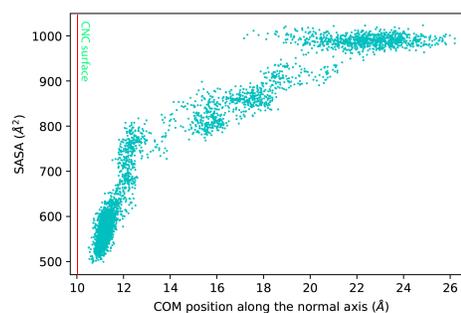
(a) (100)mc



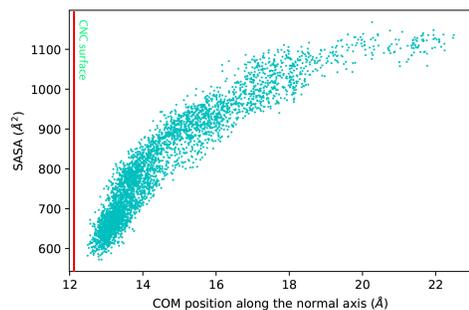
(b) (100)mcs



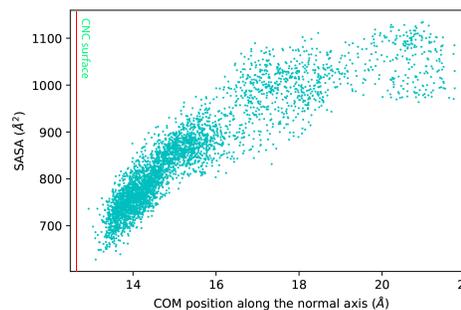
(c) (100)pe



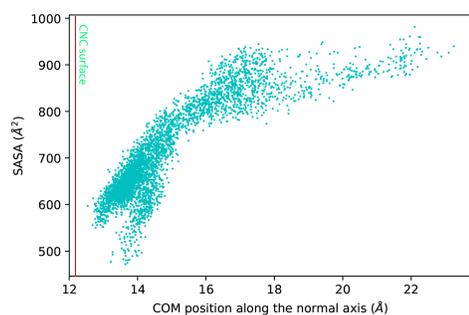
(d) (100)pes



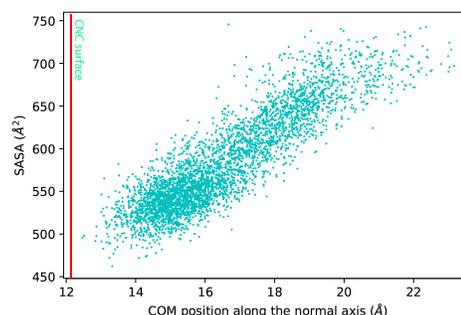
(e) (010)mc



(f) (010)mcs

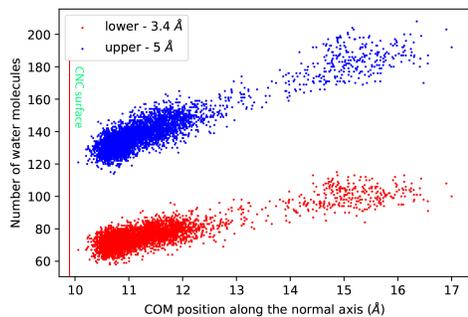


(g) (010)pe

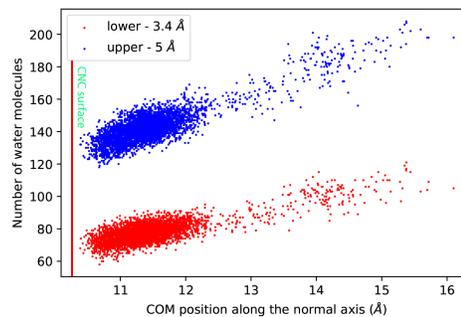


(h) (010)pes

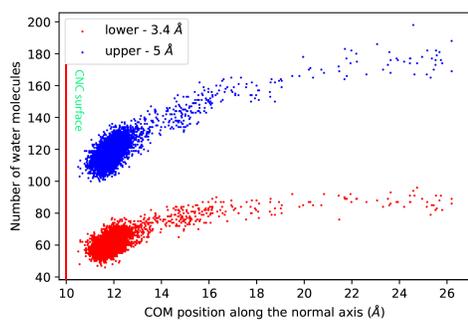
FIGURE 3.21: First and second water shells.



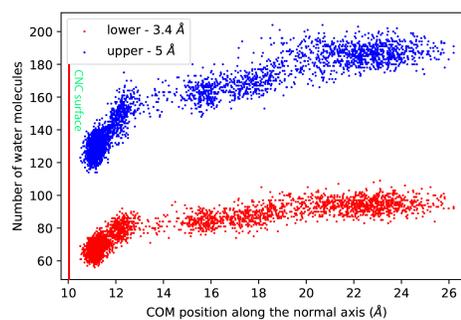
(a) (100)mc



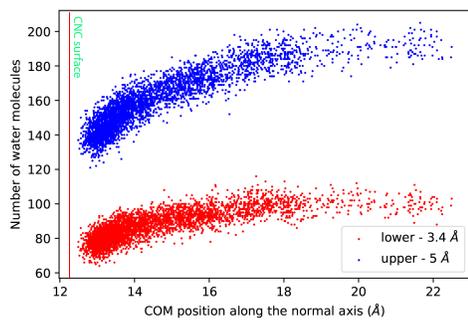
(b) (100)mcs



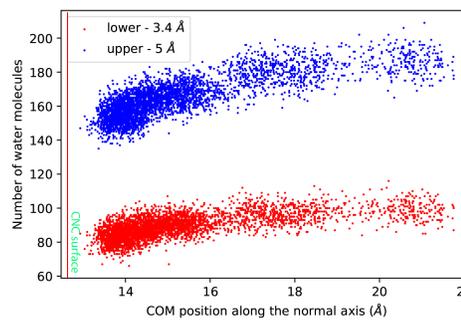
(c) (100)pe



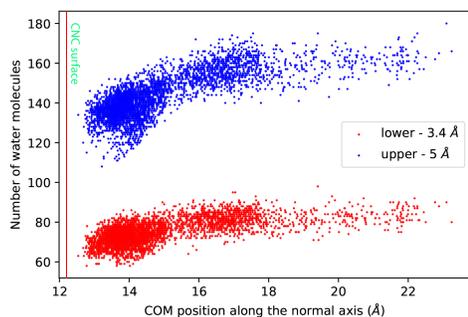
(d) (100)pes



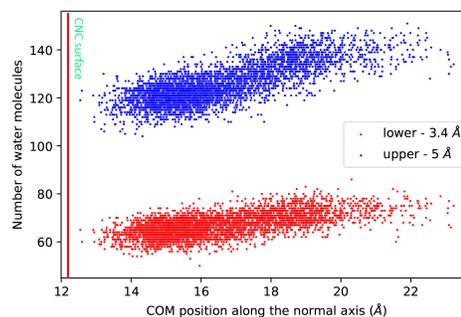
(e) (010)mc



(f) (010)mcs



(g) (010)pe



(h) (010)pes

from these systems.

The number of intramolecular hydrogen bonds are given as a function of time, it can be seen that the number of hydrogen bonds increases gradually as the distance from surface increases. Increasing window number denotes increasing distance. It can be seen that just like we hypothesized, the number of intramolecular hydrogen bonds gradually increases. Interestingly the number does not drop as we move away from the surface. This agrees with our other hypothesis about the reason for the plateaus becoming shorter and slopes becoming less steeper for higher values of the reaction coordinate.

The hypothesis was as follows, “Eventually all oligomer units desorb and form intramolecular hydrogen bonds and no additional free-energy reduction can take place after a certain point, but the unfavourable hydrophobic interactions still remain and it will gradually increase the free energy”. Now if we observe the hydrogen bond plots, it is clear that the values reach a constant value after a certain point. Some slight changes in the overall count of the hydrogen bonds can be attributed to the presence or absence of unique hydrogen bonds which occur infrequently.

### **3.5 Summary**

The result section was grouped into model validation, direct MD observations, free energy calculations and other related observations. As a part of the model validation, the bulk water simulation results were presented. From those results, it was observed that the water diffusivity is equal to previously observed values and justification for

FIGURE 3.22: Intramolecular Hydrogen bond within the oligomer in windows 0-11 of the (100)pes. The plots show the number of intramolecular hydrogen bonds found in the oligomer as the function of time.

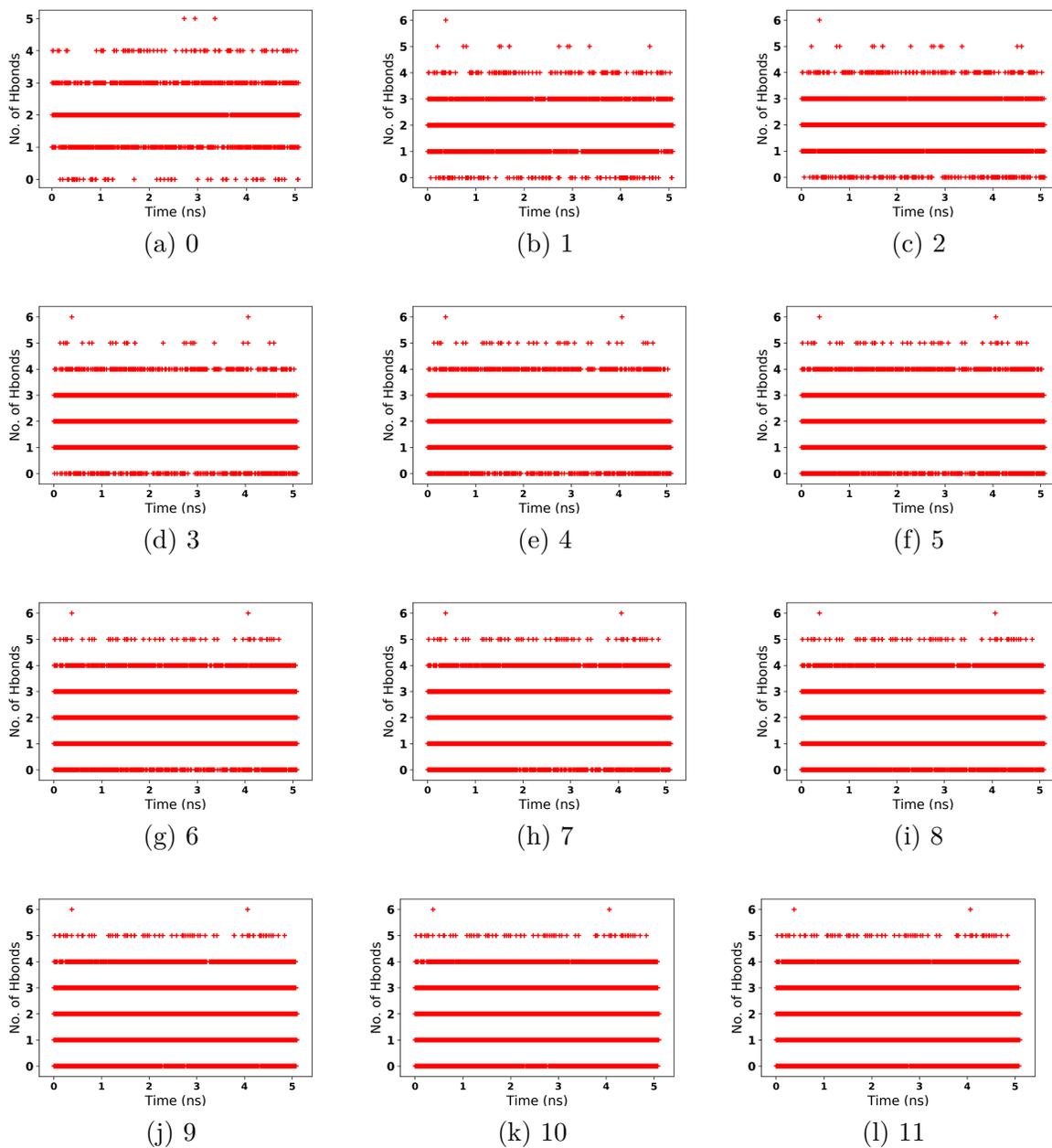
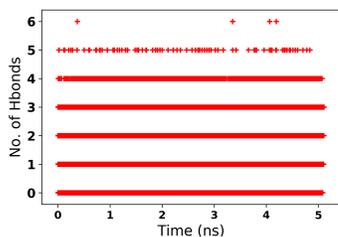
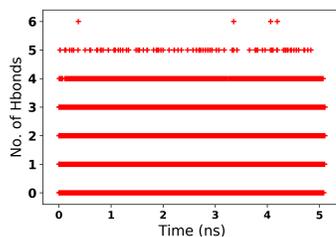


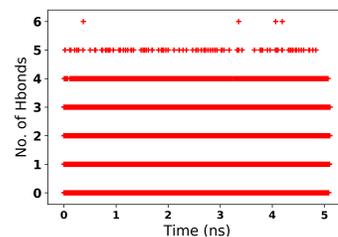
FIGURE 3.23: Intramolecular Hydrogen bond within the oligomer in windows 13-24 of the (100)pes system. The plots show the number of intramolecular hydrogen bonds found in the oligomer as the function of time.



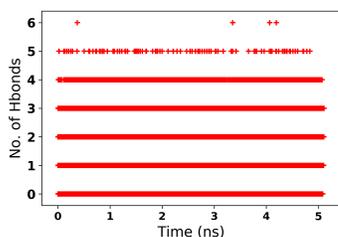
(a) 13



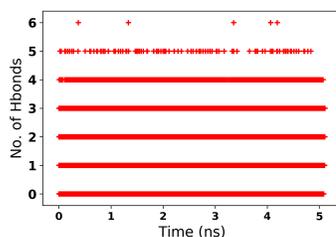
(b) 14



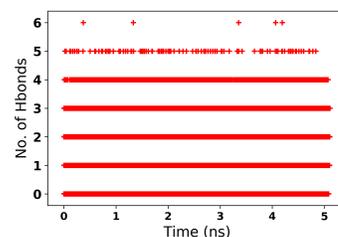
(c) 15



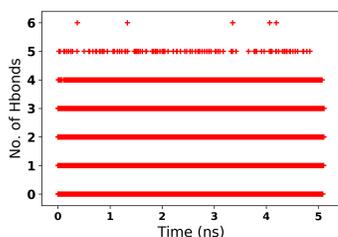
(d) 16



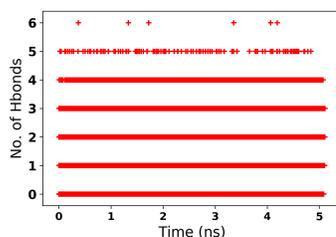
(e) 17



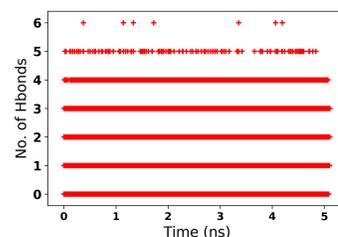
(f) 18



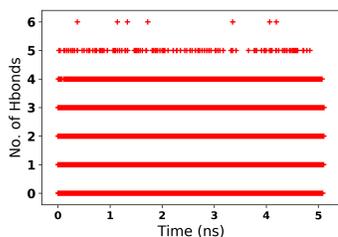
(g) 19



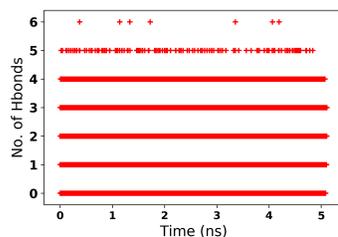
(h) 20



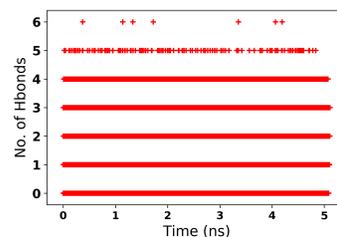
(i) 21



(j) 22



(k) 23



(l) 24

selecting the TIP3P model over other models due to its ability to predict thermodynamic properties was discussed.

It was also shown that the average HB distance is closer to the first  $g_{OO}$  peak which was near 2.7 Å. Complementing this, uniform hydrogen bonds were observed within the bulk water. The next section included the direct MD observations. This section dealt with the interactions of different components and the overall system observations such as bulk water properties. To show that the adsorption and desorption take place as the simulation proceeds, the COM values of the oligomers as a function of time was given and no unique, system specific observations were made in this analysis. Next, to incorporate the observed water properties, the water density profile along the simulation box and its self-diffusion coefficient observed in each system was shown.

The density profile was almost uniform in all cases, some minor fluctuations were attributed to the interaction of water with oligomers and salt. With respect to the oligomer, the oligomer movements in terms of MSD was discussed. The regions in MSD which were closer to being linear are the ones corresponding to oligomer's occurrence in the bulk, while the fluctuating zones were explained to be belonging to the period of adsorption/desorption of oligomer. To incorporate the oligomer-water interactions, which was uniform in most cases (depends mostly on the proximity of the oligomer), two different systems in two scenarios were presented and their hydrogen bonding interactions were explained.

It was found that the number of unique hydrogen bonds between the oligomer and water

depended on its distance from the surface. After these initial observations and results, the Free energy calculation (free energy differences) were presented. The free energy calculations in terms of PMF for one system (100)pes was examined to present the different zones in the PMF profile. A total of three zones (the favourable zone, slope and plateaus) were explained based on the two favourable and unfavourable events. These favourable and unfavourable events, the intramolecular hydrogen bonds and hydrophobic interaction between oligomer and water result in free energy decrease and increase respectively.

The slopes and plateaus were correlated with the desorption of odd and even numbered oligomer units from the CNC surface and in order to test this theory, ten different oligomer configurations along the PMF profile was recorded and the results were found to be in accordance with the proposed theory. As an overall comparison between all the systems, all PMF profiles were presented together. From this two major PMF groups, oligomer and salt dependent subgroups were found. The presence of salt was found to induce two favourable zones in case of the (010) surface. Also, the salt seemed to make the slope zones less steep and the plateaus longer. We also proposed that there are two driving forces for adsorption (1) The intermolecular interactions and (2) The bulk interactions, based on the observation that the PMF values of the salted system in case of the (010) system shoots up after the minimum interaction cut-off distance.

From all these observations it was clear that oligomer interactions do vary for each system and its proximity to the surface does have some implications on the observed properties. Hence a series of oligomer properties were recorded with respect to the oligomer's position along the normal axis and it was discussed that the proximity indeed

has a clear influence on the observed properties. Finally, the intramolecular hydrogen bond interactions for one system (100)pes was discussed and the observations from it made us conclude that the intramolecular interactions are indeed occurring similar to our proposed hypothesis.

# Chapter 4

## Conclusions

Cellulose nanocrystals exhibit a range of characteristics due to their mechanical and physical interactive properties. CNC interactions with polymers show some interesting properties such as strong gelation, emulsion stabilization etc. in gels, emulsions and films etc. Taking the fact that there is no clear idea when it comes to the molecular interactions between CNC and polymers as motivation, we simulated a set of eight systems with different physical properties through all-atom MD simulation method. Our primary goals were mainly focused on finding,

- The molecular interactions between different components within a system.
- The driving force for adsorption/desorption.
- The interactions which had maximum implications on the observed properties.
- Whether the proximity of oligomer from the surface have any effects on the interactions? etc.

To answer these question we conducted a series of extensive MD analyses. Our primary results were in the form of PMF curves which are equivalent to the free energy differences observed between different macrostates of the system. We observed that the extent of interaction between CNC surface and oligomer varied depending on the physical parameters, namely, the CNC surface, oligomer type and presence/absence of salt.

The interactions were stronger between pentaose and CNC compared to MC and this is attributed to the presence of more hydrogen bond acceptors in the former. There were distinction in the interactions between CNC and oligomer with respect to oligomer's proximity from the surface. On the (100) surface the degree of oligomer and CNC interaction was diverse for each physical case, oligomer type and presence of salt etc. On the (010) surface there were no such differences but differences were observed as we moved away from the surface. This made us conclude that on the (010) surface there is not much difference in the CNC-oligomer interactions.

We hypothesized that the differences in observations are due to the hydrogen bonds between the components and the bulk interactions. Our results from the PMF curves, hydrogen bond and configuration analysis correspond well with this. From the PMF curves we proposed that the increase in free energy value is due to the hydrophobic interactions between oligomer and water and also that the intermediate constant free energy zones are formed due to the balance in free energy decrease and increase. These free energy decrease and increase result from the desorption of even and odd numbered oligomer residues from the CNC surface. Our observations from configuration analysis shows that this is exactly the case.

Our observations from the intramolecular hydrogen bonding interaction within the oligomer chain correspond well with our theory that the intramolecular hydrogen bonds increases as the oligomer desorbs from the surface. This strongly supports our theory. The presence of salt induced two different favourable zones on the (010) surface and the bulk effects seem to take effect as we cross the non-bonded interaction cut-off distance. This is a clear indications that presence of salt and bulk properties on the whole does influence the CNC-oligomer interactions.

A series of results on the properties exhibited by the oligomer as a function of distance along normal axis made us conclude that the interaction distance does have strong implications on the properties exhibited by the oligomer. Another interesting proposal that there are two driving forces for adsorption was made based on the PMF profiles of systems with and without salt was made. The difference in PMF values and curve characteristics suggest that the interactions are indeed different in the presence of salt. The presence of salt makes the favourable and unfavourable events presistent, making the slopes less steep and plateaus longer. How these factors affect the interactions is still unclear though. From a physical point of view, all the components which were variables when varied had clear implications on the CNC-oligomer interaction. In addition to this, the proximity of different oligomers from the CNC surfaces had clear implications on their interaction. Hence the oligomer's interaction with CNC depends on the type of the oligomer, the surface properties of CNC and also the bulk properties such as the presence of salt and hydrophobicity and/or hydrophilicity of different components etc. The two driving forces, molecular interaction and bulk interaction which take effect as we move towards and away from the surface respectively, were influenced by the

components which were varied systematically in this study. For example, the affinity of MC and pentaose to CNC surfaces varied due to the fact that they carried different number of hydrogen bond acceptors i.e. their ability to form hydrogen bonds varied which resulted in the differences in their interactions closer to the surface. Similarly, MC is more hydrophobic than pentaose and hence their interactions with water in bulk varies resulting in different affinities at longer distance scales. The adsorption and desorption phenomenon is driven by two factors which take effect at different distances from the surface.

Overall, we observed that the event of oligomer adsorption on CNC had two underlying hands which drive the mechanism, (1) intermolecular and intramolecular interactions, mainly in terms of hydrogen bonds between CNC and oligomers, and (2) The interactions of oligomers with bulk, water and salt. We have found quantitative evidence in terms of PMF for these interactions and some qualitative observations for some interactions like intramolecular interactions but we still need to stratify the interactions more systematically for which we are in need of a larger system or potentially a coarse-grained model. We set out to understand the different energetic interactions and driving forces involved in polymer adsorption on CNC which we plan on using to understand the system dynamics such entropy changes etc. And they can explain the observations made in different systems involving CNC and polymers. Our results helped us in understanding the molecular interactions and gave us some clear ideas as to what drives the adsorption phenomenon in CNC-polymer systems. Our observations and results correspond to the adsorption and related phenomenon at a very small length scale. The size scales used in this work is ideal for understanding the atomic level interactions but the CNC-polymer systems are typically larger in size. In

order to answer and understand system properties in actual length scales a larger model is required. This includes the determination of different CNC-polymer configurations these systems take as a result of different driving factors and interactions. The information which we need to model a realistic version of these systems include the interactions potential between different groups etc. The presented work will serve as a good backbone in coarse graining these systems which will subsequently help us to determine the configuration, driving forces and system properties more precisely. The present work will serve as the foundation for our future explorations in this course.

# Appendix A

## Post-processing

This section details the external tools used to perform the analysis on the system. Trajectory data outputted by LAMMPS in \*.lammpstrj and \*.dcd formats were used to perform most of the analysis. System specific and certain numerical analyses were carried out using custom codes.

### MDAnalysis

MDAnalysis<sup>[16,47]</sup> is an object-oriented Python library used to analyze the trajectories obtained from MD simulations from various simulation software in a variety of formats. It has very powerful atom selection and manipulations tools. It enables the user to decode the particle trajectories in the form of NumPy arrays. This is a very flexible and fast framework for many complex analysis tasks. MDAnalysis was used to prepare density profiles, RMSD, RDF,  $R_g$  etc. This package can be imported into any python framework and is very easy to call and utilize.

## **MDTraj**

MDTraj is another trajectory analysis tool<sup>[46]</sup> used to process MD simulation particle trajectories, but this tool was mainly used for interconversion between different trajectories format for the ease of employment with other tools.

## **CPPTRAJ**

All analysis related to the oligomer residues like the SASA,  $R_g$  analyses were carried out using the CPPTRAJ module<sup>[58]</sup> in AMBER. CPPTRAJ is a complimentary program to the already existing ptraj program. CPPTRAJ can handle trajectories with different topology files within the same run. CPPTRAJ is capable of analyzing multiple structures and proving multiple output files.

## **VMD**

VMD is another visualization tool<sup>[28]</sup>, which has a very robust atom selection feature, VMD along with topotools<sup>[36]</sup> is highly compatible with LAMMPS. Although not many analyses were carried out using VMD, it was an intermediate step to verify and build systems based on the observations.

# Appendix B

## Input files

In this section of the thesis, samples of the input files used in this study are shown.

### A1 Cellulose-builder

#### Input

```
PHASE=I-BETA # Accepted values are: I-ALPHA , I-BETA , II , III_I . PBC=ALL #  
Accepted values are: NONE (default), A , B , ALL . PCB_c=TRUE # Accepted values  
are: FALSE , TRUE .
```

#### Command

```
$ ./cellulose-builder.sh 2 8 6
```

## A2 Packmol

### Input

```
tolerance 2.0  
filetype pdb  
output 1026sulpe.pdb  
structure 1026sul.pdb  
number 1  
center  
fixed 0. 0. 0. 0. 0. 0.  
end structure  
structure pentase.pdb  
number 1  
inside box -16 12.2 -8 16 20 8  
end structure  
avoid_overlap yes  
movebadrandom  
randominitialpoint  
nloop 1000  
seed -1
```

## Command

```
$Packmol < input.inp
```

## A3 AMBER

### **tleap**

```
source leaprc.GLYCAM_06j-1
source leaprc.gaff
286mcs = loadmol2 286mc.mol2
addIons2 286mcs Na+ 5
addIons2 286mcs Cl- 5
saveamberparm 286mcs 286mcs.top 286mcs.crd
quit
```

### **CPPTRAJ**

```
parm 1026pesnew.top
trajin ps.dcd 0 2000
hbond All out hb.dat avgout avg.dat
rms rmsd1 :502-507 out rmsd.agr
distance etoe :503@C1 :507@C1 out etoe.agr
radgyr rogl :502-507 out rogl.dat
```

surf :502-507 out surf.dat

watershell :502-507 out watershell.dat :WAT

run

runanalysis lifetime All[solutehb] out sol.lt.dat

# Appendix C

## Calculations

In this section of the thesis, some of the model related calculations are shown.

### A1 Water molecules count

The number water molecules to pack in each simulation box was calculated using a density value of  $1\text{g}/\text{cm}^3$ . A general calculation set is shown below.

$$\begin{aligned} N_w &= (V - V_{CNC}) / V_{Water} \\ &= ([X*Y*Z] - (N_u * V_u)) / (29.9 \text{ \AA}^3) \\ &= ([X*Y*Z] - (N_u * 658.026 \text{ \AA}^3)) / (29.9 \text{ \AA}^3). \end{aligned}$$

Where,

$N_w$  = Number of water molecules

$V_{Water}$  = Volume occupied by one water molecule.

$V$  = Total Volume of the simulation cell.

$V_{CNC}$  = Volume of the CNC.

X, Y, Z = Simulation cell dimensions along x, y, z directions.

$N_u$  = Total number of unit cells in CNC.

$V_u$  = Volume of one unit cell (658.026 Å<sup>3</sup> for  $I_\beta$ ).

## A2 Salt ions count

$$N_i = C_s * \{[X*Y*Z] - (N_u * V_u)\} * N_A$$

Where,

$N_i$  = Number of anion/cation.

$C_s$  = Salt concentration (converted from mM to M/Å<sup>3</sup>).

X, Y, Z = Simulation cell dimensions along x, y, z directions.

$N_u$  = Total number of unit cells in CNC.

$V_u$  = Volume of one unit cell (658.026 Å<sup>3</sup> for  $I_\beta$ ).

Avogadro's number (6.022 X 10<sup>23</sup> mol<sup>-1</sup>) ( $N_A$ ).

# Bibliography

- [1] Susanna Ahola, Petri Myllytie, Monika Österberg, Tuija Teerinen, and Janne Laine. Effect of polymer adsorption on cellulose nanofibril water binding capacity and aggregation. *BioResources*, 3(4):1315–1328, 2008.
- [2] Gregg T Beckham and Michael F Crowley. Examination of the  $\alpha$ -chitin structure and decrystallization thermodynamics at the nanoscale. *The Journal of Physical Chemistry B*, 115(15):4516–4522, 2011.
- [3] Kaouther Ben Azouz, Elaine C Ramires, Winke Van den Fonteyne, Nadia El Kissi, and Alain Dufresne. Simple method for the melt extrusion of a cellulose nanocrystal reinforced hydrophobic polymer. *ACS Macro Letters*, 1(1):236–240, 2011.
- [4] Rasmus Bodvik, Andra Dedinaite, Leif Karlson, Magnus Bergström, Petra Bäverbäck, Jan Skov Pedersen, Katarina Edwards, Göran Karlsson, Imre Varga, and Per M Claesson. Aggregation and network formation of aqueous methylcellulose and hydroxypropylmethylcellulose solutions. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 354(1-3):162–171, 2010.
- [5] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan,

- A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, and P.A. Kollman(2016). *AMBER 2016*. University of California, San Francisco, 2016.
- [6] ED Cranston and DG Gray. Formation of cellulose-based electrostatic layer-by-layer films in a magnetic field. *Science and Technology of Advanced Materials*, 7(4):319–321, 2006.
- [7] Emily D Cranston and Derek G Gray. Morphological and optical characterization of polyelectrolyte multilayers incorporating nanocrystalline cellulose. *Biomacromolecules*, 7(9):2522–2530, 2006.
- [8] Emily D Cranston and Derek G Gray. Birefringence in spin-coated films containing cellulose nanocrystals. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 325(1-2):44–51, 2008.
- [9] Maria H Deinema and LPTM Zevenhuizen. Formation of cellulose fibrils by gram-negative bacteria and their role in bacterial flocculation. *Archives of Microbiology*, 78(1):42–57, 1971.
- [10] Shuping Dong and Maren Roman. Fluorescently labeled cellulose nanocrystals for bioimaging applications. *Journal of the American Chemical Society*, 129(45):13810–13811, 2007.
- [11] Samuel Eyley and Wim Thielemans. Imidazolium grafted cellulose nanocrystals for ion exchange applications. *Chemical Communications*, 47(14):4177–4179, 2011.
- [12] Anwasha N Fernandes, Lynne H Thomas, Clemens M Altaner, Philip Callow, V Trevor Forsyth, David C Apperley, Craig J Kennedy, and Michael C Jarvis.

- Nanostructure of cellulose microfibrils in spruce wood. *Proceedings of the National Academy of Sciences*, 108(47):E1195–E1203, 2011.
- [13] G. Fiorin, M.L. Klein, and J. Hénin. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.*, 2013. doi: 10.1080/00268976.2013.813594.
- [14] B Lachele Foley, Matthew B Tessier, and Robert J Woods. Carbohydrate force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):652–697, 2012.
- [15] Thiago CF Gomes and Munir S Skaf. Cellulose-builder: A toolkit for building crystalline structures of cellulose. *Journal of computational chemistry*, 33(14):1338–1346, 2012.
- [16] Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler JE Reddy, Manuel N Melo, Sean L Seyler, David L Dotson, Jan Domanski, Sébastien Buchoux, Ian M Kenney, et al. Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. In *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016.
- [17] Adam S Gross and Jhih-Wei Chu. On the molecular origins of biomass recalcitrance: the interaction network and solvation structures of cellulose microfibrils. *The Journal of Physical Chemistry B*, 114(42):13333–13341, 2010.
- [18] Alan Grossfield. Wham: the weighted histogram analysis method. *version*, 2(9):06, 2012. URL <http://membrane.urmc.rochester.edu/content/wham>.
- [19] Youssef Habibi, Lucian A Lucia, and Orlando J Rojas. Cellulose nanocrystals: chemistry, self-assembly, and applications. *Chemical reviews*, 110(6):3479–3500, 2010.

- [20] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1):17, 2012.
- [21] Walter Norman Haworth. The structure of carbohydrates. *Helvetica Chimica Acta*, 11(1):534–548, 1928. ISSN 1522-2675.
- [22] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.
- [23] Zhen Hu. *Tailoring Cellulose Nanocrystal, Polymer and Surfactant Interactions for Gels, Emulsions, and Foams*. PhD thesis, McMaster University, 2015.
- [24] Zhen Hu, Emily D Cranston, Robin Ng, and Robert Pelton. Tuning cellulose nanocrystal gelation with polysaccharides and surfactants. *Langmuir*, 30(10):2684–2692, 2014.
- [25] Zhen Hu, Tyler Patten, Robert Pelton, and Emily D Cranston. Synergistic stabilization of emulsions and emulsion gels with water-soluble polymers and cellulose nanocrystals. *ACS Sustainable Chemistry & Engineering*, 3(5):1023–1031, 2015.
- [26] Zhen Hu, Heera S Marway, Hesham Kasem, Robert Pelton, and Emily D Cranston. Dried and redispersible cellulose nanocrystal pickering emulsions. *ACS Macro Letters*, 5(2):185–189, 2016.
- [27] Zhen Hu, Richard Xu, Emily D Cranston, and Robert H Pelton. Stable aqueous foams from cellulose nanocrystals and methyl cellulose. *Biomacromolecules*, 17(12):4095–4099, 2016.

- [28] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [29] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [30] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [31] Sridhar Kumar Kannam, Daniel P Oehme, Monika S Doblin, Michael J Gidley, Antony Bacic, and Matthew T Downton. Hydrogen bonds and twist in cellulose microfibrils. *Carbohydrate polymers*, 175:433–439, 2017.
- [32] Ali Chami Khazraji and Sylvain Robert. Self-assembly and intermolecular forces when cellulose and water interact using molecular modeling. *Journal of Nanomaterials*, 2013:48, 2013.
- [33] Karl N Kirschner and Robert J Woods. Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences*, 98(19):10541–10545, 2001.
- [34] Karl N Kirschner, Austin B Yongye, Sarah M Tschampel, Jorge González-Outeiriño, Charlisa R Daniels, B Lachele Foley, and Robert J Woods. Glycam06: a generalizable biomolecular force field. carbohydrates. *Journal of computational chemistry*, 29(4): 622–655, 2008.

- [35] Dieter Klemm, Brigitte Heublein, Hans-Peter Fink, and Andreas Bohn. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition*, 44(22):3358–3393, 2005.
- [36] Axel Kohlmeyer. Topotools: Release 1.7, 2016. URL <http://doi.org/10.5281/zenodo.50249>.
- [37] Tetsuo Kondo and Chie Sawatari. Intermolecular hydrogen bonding in cellulose/poly (ethylene oxide) blends: thermodynamic examination using 2, 3-di-o-and 6-o-methylcelluloses as cellulose model compounds. *Polymer*, 35(20):4423–4428, 1994.
- [38] William R Krigbaum and Frederick T Wall. Viscosities of binary polymeric mixtures. *Journal of Polymer Science Part A: Polymer Chemistry*, 5(4):505–514, 1950.
- [39] Shankar Kumar, John M Rosenberg, Djamel Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.
- [40] Koon-Yang Lee, Yvonne Aitomäki, Lars A Berglund, Kristiina Oksman, and Alexander Bismarck. On the use of nanocellulose as reinforcement in polymer matrix composites. *Composites Science and Technology*, 105:15–27, 2014.
- [41] Yijin Mao and Yuwen Zhang. Thermal conductivity, shear viscosity and specific heat of rigid water models. *Chemical Physics Letters*, 542:37–41, 2012.
- [42] Pekka Mark and Lennart Nilsson. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, 105(43):9954–9960, 2001.

- [43] Pekka Mark and Lennart Nilsson. Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations. *Journal of computational chemistry*, 23(13):1211–1219, 2002.
- [44] José Mario Martínez and Leandro Martínez. Packing optimization for automated generation of complex system’s initial configurations for molecular dynamics and docking. *Journal of computational chemistry*, 24(7):819–825, 2003.
- [45] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez. Packmol: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.
- [46] Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee-Ping Wang, Thomas J. Lane, and Vijay S. Pande. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528 – 1532, 2015. doi: 10.1016/j.bpj.2015.08.015.
- [47] Naveen Michaud-Agrawal, Elizabeth J Denning, Thomas B Woolf, and Oliver Beckstein. Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, 32(10):2319–2327, 2011.
- [48] R Mills. Self-diffusion in normal and heavy water in the range 1-45. deg. *The Journal of Physical Chemistry*, 77(5):685–688, 1973.
- [49] Lakshmi Muthukumar and Rajesh Khare. Molecular dynamics simulation of free energy of desorption of cellobiose from a cellulose crystal surface. In *Applications*

- of Molecular Modeling to Challenges in Clean Energy*, pages 1–17. ACS Publications, 2013.
- [50] Yoshiharu Nishiyama, Paul Langan, and Henri Chanzy. Crystal structure and hydrogen-bonding system in cellulose  $i\beta$  from synchrotron x-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 124(31):9074–9082, 2002.
- [51] Yoshiharu Nishiyama, Junji Sugiyama, Henri Chanzy, and Paul Langan. Crystal structure and hydrogen bonding system in cellulose  $i\alpha$  from synchrotron x-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 125(47):14300–14306, 2003.
- [52] Christina M Payne, Michael E Himmel, Michael F Crowley, and Gregg T Beckham. Decrystallization of oligosaccharides from the cellulose  $i\beta$  surface with molecular simulation. *The Journal of Physical Chemistry Letters*, 2(13):1546–1550, 2011.
- [53] B Peng, N Dhar, HL Liu, and KC Tam. Chemistry and applications of nanocrystalline cellulose and its derivatives: a nanotechnology perspective. *The Canadian Journal of Chemical Engineering*, 89(5):1191–1206, 2011.
- [54] Suma Peri, Lakshmi Muthukumar, M Nazmul Karim, and Rajesh Khare. Dynamics of cello-oligosaccharides on a cellulose crystal surface. *Cellulose*, 19(6):1791–1806, 2012.
- [55] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995. URL <http://lammps.sandia.gov>.
- [56] Bengt G Ranby. Aqueous colloidal solutions of cellulose micelles, 1949.

- [57] Michael S Reid, Marco Villalobos, and Emily D Cranston. Benchmarking cellulose nanocrystals: from the laboratory to industrial production. *Langmuir*, 33(7):1583–1598, 2016.
- [58] Daniel R Roe and Thomas E Cheatham III. Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7):3084–3095, 2013.
- [59] Ute Römling. Molecular biology of cellulose production in bacteria. *Research in microbiology*, 153(4):205–212, 2002.
- [60] Peter Ross, Raphael Mayer, and Moshe Benziman. Cellulose biosynthesis and function in bacteria. *Microbiological reviews*, 55(1):35–58, 1991.
- [61] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer physics communications*, 91(1-3):275–282, 1995.
- [62] Salah Mnaour Salah. Application of nano-cellulose in textile. *J Textile Sci Eng*, 3:142, 2013.
- [63] N Sarkar and LC Walker. Hydration—dehydration properties of methylcellulose and hydroxypropylmethylcellulose. *Carbohydrate polymers*, 27(3):177–185, 1995.
- [64] Herrmann Staudinger. Die chemie der hochmolekularen organischen stoffe im sinne der kekuléschen strukturlehre. *European Journal of Inorganic Chemistry*, 59(12):3019–3043, 1926.
- [65] von Hermann Staudinger, Karl-Heinz In den Birken, and Magda Staudinger. Über den micellaren oder makromolekularen bau der cellulosen. *Macromolecular Chemistry and Physics*, 9(1):148–187, 1953.

- [66] Richard P Swatloski, Scott K Spear, John D Holbrey, and Robin D Rogers. Dissolution of cellulose with ionic liquids. *Journal of the American chemical society*, 124(18):4974–4975, 2002.
- [67] Tuula T Teeri, Harry Brumer, Geoff Daniel, and Paul Gatenholm. Biomimetic engineering of cellulose-based materials. *TRENDS in Biotechnology*, 25(7):299–306, 2007.
- [68] Albin F Turbak, Fred W Snyder, and KAREN R Sandberg. Microfibrillated cellulose, a new cellulose product: properties, uses, and commercial potential. In *J. Appl. Polym. Sci.: Appl. Polym. Symp.:(United States)*, volume 37. ITT Rayonier Inc., Shelton, WA, 1983.
- [69] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.
- [70] Masahisa Wada, Yoshiharu Nishiyama, Henri Chanzy, Trevor Forsyth, and Paul Langan. The structure of celluloses. *Powder Diffraction*, 23(2):92–95, 2008.
- [71] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.
- [72] Theodore H Wegner and Philip E Jones. Advancing cellulose-based nanotechnology. *Cellulose*, 13(2):115–118, 2006.
- [73] R Woods. Glycam web. complex carbohydrate research center, university of georgia, athens, 2014.

- [74] Akira Yamakawa, Shiho Suzuki, Takeshi Oku, Kenta Enomoto, Motohide Ikeda, Joseph Rodrigue, Keita Tateiwa, Yoshinobu Terada, Hiroyuki Yano, and Shinichi Kitamura. Nanostructure and physical properties of cellulose nanofiber-carbon nanotube composite films. *Carbohydrate Polymers*, 171:129–135, 2017.
- [75] S Yamanaka, K Watanabe, N Kitamura, M Iguchi, S Mitsuhashi, Y Nishi, and M Uryu. The structure and mechanical properties of sheets prepared from bacterial cellulose. *Journal of Materials Science*, 24(9):3141–3145, 1989.
- [76] Zhen Zhang. Surface modification of cellulose nanocrystal for advanced applications. 2017.
- [77] Zhen Zhang. *Surface modification of cellulose nanocrystals by esterification and ATRP reactions for advanced applications*. Theses, Université de Bordeaux, September 2017. URL <https://tel.archives-ouvertes.fr/tel-01685016>.
- [78] Shengdong Zhu, Yuanxin Wu, Qiming Chen, Ziniu Yu, Cunwen Wang, Shiwei Jin, Yigang Ding, and Gang Wu. Dissolution of cellulose with ionic liquids and its application: a mini-review. *Green Chemistry*, 8(4):325–327, 2006.