MOLECULAR DYNAMICS SIMULATION OF SOLVATED CELL WALL POLYSACCHARIDES: A STRUCTURAL AND DYNAMIC STUDY

A Dissertation in
Chemical Engineering
by
Bingxin Fan

© 2015 Bingxin Fan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2015
The dissertation of Bingxin Fan was reviewed and approved* by the following:

Janna K. Maranas
Professor of Chemical Engineering and Material Science and Engineering
Graduate Program Chair of Chemical Engineering
Dissertation Advisor, Chair of Committee

Seong H. Kim
Professor of Chemical Engineering and Material Science and Engineering

James D. Kubicki
Professor of Geochemistry

Scott T. Milner
Joyce Chair and Professor of Chemical Engineering

*Signatures are on file in the Graduate School
Abstract

We investigate behavior of cell wall polysaccharides using computational approaches. Research on cell wall has become an increasingly emphasized topic due to the lignocellulosic biomass is a great candidate for biofuel production. Understanding the cell wall hierarchical structure is not only beneficial for fundamental understanding. It also provides a scientific basis for developing more effective methods for the conversion process. In this thesis, we focus on both the structural behavior of the polysaccharides and the dynamic properties of their surrounding water.

The load-bearing network of cell wall is the cellulose/xyloglucan network. In order to avoid harsh chemical extraction treatment and to simulate the polysaccharides in the length scale that close to that in plant cell walls, we developed a coarse-grained force field for both cellulose and (XXXG) xyloglucan, and combined these two force fields for studying the network structure. The force field for cellulose is built based on atomistic simulation of a 6x6x40 microfibril. The force sites are defined as the geometric average of the six member glucose ring. The force field is parameterized such that the chain configuration, intermolecular packing, and hydrogen bonding of the two levels of modeling are consistent. To retain the directionality of the interfibril interactions, we define pair-wised interactions between the interchain neighbors, and add the potentials sequentially until the crystal structure of the coarse-grained fibril matches that of the atomistic target. The coarse-grained simulation shows that microfibrils longer than 100nm tend to form kinks along their
longitudinal direction. The kink structure may be linked to the periodic disorder of the microfibril observed based on small angle neutron scattering measurements.

The xyloglucan force field is build based on atomistic simulation of 15 (XXXG)₃ segments, which is the shortest length of xyloglucan that shows significant interaction on cellulose microfibril surfaces. In order to make the two force fields compatible, we also define the xyloglucan force sites as the geometric average of the glucose rings or xylose rings. Thus, there are two types of beads in the xyloglucan chains. The coarse-grained simulation uses the atomistic chain configuration and intermolecular spacings as the target. Upon completion, we performed a simulation combining the two force fields. The simulation box in this case contains one 6x6x200 microfibril surrounded by 20 randomly placed (XXXG)₅₀ xyloglucan chains. The coarse-grained simulation shows three types of xyloglucan based on their interaction with cellulose chains: the bridge chains (that interact with both the microfibril and its periodic image), the single chains (that only interacts with the microfibril), and the isolated chains (that do not directly interact with the microfibril). We also see that some of the isolated chains may bind to the other two types of xyloglucan and participating indirectly in bridging the microfibrils. Therefore, even though some xyloglucans may not directly interact with the microfibril, they may also contribute to the mechanical strength of the network structure. In addition, we observe that the interaction between xyloglucan and cellulose tends to extend along the fibril longitudinal direction. The above observations are very useful for revising the current cell wall models.
Water occupies up to 90% of the cell wall, and it has been shown that water may modify the mechanical properties of cellulose by varying the hydration level. Thus, in order to fully understand the properties of the cell wall network structure, one cannot ignore the role of water. We studied the dynamic properties of water at 5% and 20% hydration levels by fitting self-intermediate scattering function using a stretched exponential model. Atomistic simulation allows us to completely decouple the contributions of translation and rotation in the scattering functions. By applying jump model, we can determine the translational diffusion coefficient, jump length, and the residence time of water proton within the local cage. We observe that multiple types of translational motion exist in 20% hydrated system, but even the faster motion is still slower than bulk water. We further performed simulation of the fibril at these two hydration levels at various temperatures. From each simulation, the rotation and translation is analyzed separately. The activation energy of rotation and translation are obtained using Arrhenius plots of relaxation time of rotation and diffusion coefficient of translation, respectively. The translation activation energy is comparable with bulk water, but the rotational activation energy of the confined water is much higher comparing to the bulk water, indicating a difference in the water rotation mechanism. However, this mechanism seems to be independent on hydration level. By performing anisotropy analysis on water rotation, we determined that the difference in the values of rotation relaxation times of the confined water is due to the difference in their extent of anisotropy.
In summary, we study the structural properties of the cell wall load-bearing network by developing coarse-grained force fields by requiring consistency with atomistic simulation target. The surrounding of the network is examined based on atomistic simulation, which can provide useful insights on the mechanism of the water motions.
# TABLE OF CONTENTS

List of Figures ............................................................................................................. x
List of Tables ................................................................................................................ xiv
Acknowledgements ...................................................................................................... xv

Chapter 1 Introduction ................................................................................................. 1
  1.1 Primary Cell Wall ................................................................................................. 1
    1.1.1 Cellulose Microfibril .................................................................................... 2
    1.1.2 Xyloglucan ................................................................................................. 4
    1.1.3 Water ........................................................................................................... 6
  1.2 Study Primary Cell Wall in silico ........................................................................ 6
    1.2.1 Atomistic Simulation .................................................................................. 6
    1.2.2 Necessity of Coarse-Grained Simulation .................................................... 7
  1.3 Thesis Overview .................................................................................................. 8
  1.4 Reference ............................................................................................................ 10

Chapter 2 Coarse-Grained Simulation of Cellulose Iβ with Application to Long Fibrils .................................................................................................................. 14
  2.1 Introduction ......................................................................................................... 15
  2.2 Methodology ...................................................................................................... 18
    2.2.1 Simulation Details ...................................................................................... 18
    2.2.2 Coarse-Grained Force Field Development .................................................. 20
  2.3 Result and Discussion ......................................................................................... 27
    2.3.1 Force Field Parameters and Cellulose Crystal Structure ............................ 27
    2.3.2 Applying the CG Force Field to Longer Fibrils .......................................... 31
  2.4 Conclusions ....................................................................................................... 36
  2.5 References ......................................................................................................... 37

Chapter 3 Molecular Level Scale Investigation of the Xyloglucan-Cellulose Microfibril Assembly ............................................................................................................. 43
  3.1 Introduction ....................................................................................................... 44
  3.2 Method ............................................................................................................... 47
    3.2.1 Atomistic Simulation .................................................................................. 47
    3.2.2 Coarse-Grained Simulation .......................................................................... 48
  3.3 Result and Discussion ......................................................................................... 53
3.3.1 Xyloglucan Coarse-Grained Force Field ..................................................53
3.3.2 Xyloglucan and Microfibril Assembly .......................................................55
3.4 Conclusion .......................................................................................................60
3.5 Reference .........................................................................................................61

Chapter 4 A Computational Study on the Mutual Effect between Cellulose Iα Microfibril and the Interfacial Water .................................................................64
4.1 Introduction .....................................................................................................65
4.2 Methodology ...................................................................................................67
  4.2.1 Simulation Details .......................................................................................67
  4.2.2 Analyzing Water Dynamic Properties ......................................................68
4.3 Results and Discussion ....................................................................................72
4.4 Conclusion .......................................................................................................79
4.5 References ......................................................................................................80

Chapter 5 Dynamics of Surface Water on Cellulose Iα Microfibrils by Molecular Dynamics Simulation .................................................................84
5.1 Introduction .....................................................................................................85
5.2 Simulation Details ............................................................................................87
5.3 Method of Data Analysis ................................................................................87
5.4 Results and Discussion ...................................................................................90
  5.4.1 Rotational Motion ......................................................................................90
  5.4.2 Translational Motion ................................................................................93
5.5 Conclusion .......................................................................................................102

Chapter 6 Conclusion .............................................................................................107
  6.1 Cell Wall Load-Bearing Network Structure ..................................................107
  6.2 Dynamic Properties .......................................................................................110
  6.3 Summary .......................................................................................................111
  6.4 Recommendations for Future Study .............................................................112
  6.5 Reference ......................................................................................................114

Appendix A LAMMPS Input Script .....................................................................116
  A.1 Cellulose Microfibril Coarse-Grained Simulation Input Script ......................116
  A.2 Microfibril-Xyloglucan Network Coarse-Grained Simulation Input Script ......116
Appendix B Tabulated Potentials of Coarse-Grained Xyloglucan Simulation ..........117
  B.1 Tabulated Angle Potential ............................................................................117
B.2 Tabulated Dihedral Potential

Appendix C Jump Diffusion Model Fitting
List of Figures

Fig 1.1. Cellobiose unit is the repeating unit of cellulose chains ...........................................2
Fig 1.2. The relationship between the unit cells of cellulose Iα and Iβ. Modified from Reference 25 ..................................................................................................................4
Fig 1.3. The basic structure of xyloglucan. The position that can be substituted is labeled as R. Figure modified based on 26 .................................................................5
Fig 2.1. Cellulose microfibril at the atomistic level. In this paper, the dimension of microfibrils is denoted by the number of chains on each side surface and the number of glucose units along the longitudinal direction. In case of atomistic simulation, the microfibril has a diamond shape cross-section consisting of 36 chains with 6 chains on each surface, and each chain contains 40 glucose units. This is referred as the 6x6x40 microfibril .................................................................19
Fig 2.2 Coarse-grained mapping of cellulose chains with beads centered on (a) O4 (the glycosidic oxygen) or (b) ring center ..................................................................................21
Fig 2.3. Intramolecular probability distributions of O4 (dashed line) and ring center (solid line) models: (a) bond length, (b) bending angle, (c) torsional angle. Intermolecular pair distribution functions of (d) O4 model and (e) ring center model. ....22
Fig 2.4. Snapshot of the position of cellulose CG beads after 15 ns. Only van der Waals potential is considered in the nonbonded potential term .................................................25
Fig 2.5. Comparison of pair distribution functions obtained from CG simulation with van der Waals interaction only (dashed line) and the target from atomistic simulation trajectories (solid line) ........................................................................25
Fig 2.6. The target pair distribution function decomposed into specific structural features. (a) Interlayer separation distance distribution in (110) planes; (b) interlayer separation distance distribution in (11-0) planes; (c) intralayer separation distance distribution in (100) planes; (d) interlayer separation distance distribution in (010) planes; (e) CG mapping of the cross section view of the microfibril, in which the structurally significant separation distances are shown; (in the atomistic cross-sectional view, a and b unit cell axes are highlighted in solid lines, and Miller indices of the crystal planes are highlighted in dashed lines) (f) target pair distribution function, in which the structurally significant peaks are labeled as their corresponding planes ........................................................................................................26
Fig 2.7. Demonstration of potentials required to reproduce the pair distribution function peaks. The intermolecular interactions of simulation above are van der Waals interaction and (a) intralayer bonds in (100), (b) intralayer bonds in (100) and interlayer bonds in (110), (c) intralayer bonds in (100) and interlayer bonds in both (110) and (11-0), (d) intralayer bonds in (100) and interlayer bonds in (110), (11-0) and (010). Potentials are not yet tuned..................................................................................29
Fig 2.8. Comparison of CG simulation with atomistic targets: (a) bond length, (b) bending angle, (c) torsional angle distribution, and (d) pair distribution functions. Dashed curves: distributions following simulation with the CG force field for 5ns. Solid curves: atomistic target distributions.

Fig 2.9. Cross sectional and longitudinal view of the coarse-grained cellulose microfibril (MD simulation snapshot).

Fig 2.10. CG simulation snapshot of microfibril 100 and 400 glucose units long.

Fig 2.11. Comparison of bended (lower black points) and non-bended (upper grey points) 200 residue microfibril potential energy.

Fig 2.12. Various potentials against residue number for bended 200-residue microfibril. The type of potential is indicated in the y-axis, and residue number is shown in x-axis. (a)-(c) is the intramolecular potential variation along the fibril. (d)-(h) indicates the intermolecular potential variation. (a)(e)(h) contain a peak at kinked region, indicating those interactions are unfavorable for the kink. Interestingly, these directions correspond to the principle axis of the cellulose microfibril crystal structure. This implies that the formation of the kink disturbs the local crystal structure of the microfibril. (c) and (d) form a well at kinked region, indicating that the long range interactions involving multiple beads are in favor of the kink formation.

Fig 3.1. Schematic representation of xyloglucan structure. (Taken from Reference 7)

Fig 3.2. Snapshot of atomistic simulation of 15 chains of (XXXG)₃.

Fig 3.3. Comparison of atomistic (left) and coarse-grained (right) representation of xyloglucan chains.

Fig 3.4. Coarse-grained bond, angle, and torsion distributions and their corresponding bonded potentials prior iterations. The blue circles are the distributions produced based on the atomistic simulation, and the red dashed curves are the initial bonded potential input into the coarse-grained simulation.

Fig 3.5. Pair distribution of the interchain pairs (top row) and the Morse-potential fits around their first peak (bottom row). The type of interchain pair is labeled as each column.

Fig 3.6. Comparison of coarse-grained simulation (red dots or curves) with atomistic targets (black dots or curves).

Fig 3.7. Coarse-grained simulation snapshot of the cellulose microfibril and xyloglucan assembly. (a) Initial configuration. (b) Snapshot of simulation after 10ns. Each snapshot includes the simulation box itself with its periodic images above and below in y-axis.

Fig 3.8. Possible locations of xyloglucan chains. Note the molecules are represented in their dynamic bonds of 5.4 Å, which means all beads with instant distances longer than 5.4Å will be recognized as “bonds” in the figure. This is why there are some bonds exist between cellulose chains in the snapshots, but these bonds are not physically there in the simulation. This is also true for other simulation snapshots.
Fig 3.9. Example of the bridge structure formed by multiple xyloglucan chains.............58
Fig 3.10. Xyloglucan interacts with cellulose microfibril along the fibril longitudinal
direction. The simulation time is labeled on top of each snapshot.........................60

Fig 4.1: Initial configuration of water solvated microfibrils atomistic system.............68
Fig 4.2. Oxygen distance traveled over 500ps of simulations at various water contents.
In 5% hydration system and bulk water system, the distance distributions contain only
one peak whereas in 20% hydration system, the distribution contains two peaks. This
indicates that there are two types of translational motion in 20% hydration system,
and the translational motion in the other two systems can be described by a single
type of motion. The distance is computed using the summation of displacement
within 1ps...........................................................................................................71

Fig 4.3. Bulk water rotational SISF fitted with KWW equation. Rotational SISFs are
obtained from Q=0.5 to Q=2.0 with increment of 0.1. The hollow points are
calculations, and fitting curves are shown as red solid lines..................................73

Fig 4.4. Relaxation time and stretch factor values in Eq. 5 and Eq. 6 for (a)(b) bulk
water, (c)(d) 20% hydration, and (e)(f) 5% hydration systems. Empty circles represent
rotational motion, and solid points represent translational motion. In case of 20%
hydration system, the slower motion is indicated by triangles.........................74

Fig 4.5. Bulk water relaxation rate vs. Q^2 fitted with jump model..........................75

Fig 4.6. Rotational SISF at Q=0.7 shows that the time independent term of rotational
SISF is independent of water content..............................................................76

Fig 4.7. h=0.20 SISF fitting fitted to jump diffusion model. From the fitting of this
curve, we determine that the diffusion coefficient of faster water (shown in (a)) in this
system is 1.2 × 10-9m2/s, the residence time is 7.7ps, and the jump length is 2.33Å.
In case of slower water (shown in (b)), the diffusion coefficient is 0.14 × 10-9m2/s,
the residence time is 36ps, and the jump length is 1.74Å..............................77

Fig 4.8. Tracking a single oxygen mean square displacement over 200ps to
demonstrate the jump diffusion motion of water molecules. The time lapse image of
the first 75ps (with interval of 1ps) is shown, and the oxygen atoms are color-coated
based on the local cage they belong to..........................................................................77

Fig 4.9. h=0.05 SISF fitting fitted to jump diffusion model. From the fitting of this
curve, we determine that the diffusion coefficient of water in this system is 0.028 ×
10-9m2/s, the residence time is 1813 ps, and the jump length is 5.5Å................79

Fig 5.1. Arrhenius plot of the relaxation rotational time at (a) 5% hydration and (b) 20%
hydration...........................................................................................................92

Fig 5.2. Distribution of average water oxygen travel distance over 500ps at 20%
hydration level. The travel displacement of each water oxygen is calculated every 1ps,
and the summation of this value over 500ps derives the travel distance of one water
oxygen, and the distributions are generated based on the result of all water oxygen
atoms in the system. The circle points denote the data points based on the calculation.
The solid lines denote the fitting to the data points. The simulations were performed at 253K (green), 273K (blue), 283K (red), and 298K (black).

Fig 5.3. Jump lengths do not show temperature dependence trend. Dashed lines are the average jump length calculated based on all data points for each translational motion.

Fig 5.4. Temperature dependence of diffusion coefficients. Diffusion coefficients obtained from jump diffusion model are shown as circles. The dashed lines are fittings according to Arrhenius equation. (a) 5% hydration level. (b) 20% hydration level: the fast water diffusion coefficients are colored in blue, and the slow water data are shown in red.

Fig 5.5. Local coordinate frame of the water molecule. Modified based on Reference 34.

Fig 5.6. Summary of relaxation time in the reorientation axes at both hydration levels.

Fig 5.7. Temperature dependence of relaxation time of each rotational orientation axis. The 5% hydration level data are shown in diamond shape data points, and the 20% hydration level data are shown in circle. τ_{HH}, τ_{μ}, and τ_{⊥} are color coated in red, blue, and orange, respectively.

Fig 6.1. Cellulose coarse-graining: (a) Intramolecular and (b) intermolecular coarse-grained beads with point dipole; (c) intermolecular potential are defined as a function of r and θ.

Fig E.1. Jump diffusion model fitting for water proton translational motion at 20% hydration level.

Fig E.2. Jump diffusion model fitting for water proton translational motion at 5% hydration level.
List of Tables

Table 1.1 Cellulose Iβ coarse-grained force field potential parameters ..........................30
Table 3.1. Xyloglucan (XXXG)_n chain coarse-grained force field potential parameters ..54
Table 5.1 Rotational motion activation energy and Arrhenius prefactor ..........................92
Table 5.2 Summary of fraction of slow and fast water at each temperature ......................95
Table 5.3. Summary of jump diffusion model fitting results ........................................96
Table 5.4. Arrhenius plot fitting results of translational motion ....................................97
Table 5.5 Comparison of activation energy of rotation in the reorientation axes. E_A is the activation energy of overall rotational motion obtained from SISF fittings ............101
Acknowledgements

I would like to thank my advisor, Dr. Janna Maranas, for giving me this opportunity to work with her. Thank you for admitting me to the Chemical Engineering Department at Penn State. Thank you for your great patience and guidance on my thesis project. I appreciate your emphasis on a friendly and interactive research group environment and your care about our general well-being. I am also grateful for your encouragement for attending conferences and meetings, from which I gained extensive practice on writing and presentation skills. Thank you so much for an enriching and memorable graduate school experience.

I would like to thank Dr. Scott Milner, Dr. Seong Kim, and Dr. James Kubicki. I really appreciate you taking the time to serve as my committee members. Thank you for reading my thesis and giving me great advices. I would like to thank Dr. Loukas Petridis (ORNL) for valuable discussion on dynamics of interfacial water on cellulose microfibrils. Thank you for great insights and suggestions. I would like to thank Dr. Linghao Zhong and Dr. Zhen Zhao for their contribution of atomistic simulation coordinates of cellulose microfibril, which provides us the basis for the development of our coarse-grained microfibril force field.

I would like to thank the prior and current members in the Janna Maranas research group for their help and advice along my progress. I would also like to thank my friends at Penn State for their support and for making life here more enjoyable and memorable.

Last but not least, I would like to thank my families for their love and support.
I would like to thank my grandparents for their caring words and making sure that I eat healthy. I thank my parents for taking care of me, encouraging me, and always being there whenever I need you. Thank you for being so understanding, respecting my decisions, and making me your number one priority no matter how busy you are. Finally, I thank my husband, Shih-Chun Huang, for also being a great friend and colleague. I am so grateful for having met you at Penn State and being in the same research group. Thank you for the great times discussing about our research, always being optimistic and sharing your positive energy, looking after me when I am ill or upset, and staying with me through ups and downs.

Thank you all so much! This work could not have been completed without support from all of you.

The work is supported as part of The Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001090.
Chapter 1 Introduction

1.1 Primary Cell Wall

The growing plant cells are surrounded by the primary cell wall which is highly rich in polysaccharides. The primary cell walls is a complex structure between 150 to 200 nm thick and serve essential roles in cell structure, mechanical support, cell growth, and morphogenesis. \(^1,2\) They mainly consist of three polysaccharides which are the basic building blocks of the wall: cellulose, hemicellulose, and pectin. The polysaccharides are assembled in a hierarchical structure. \(^3\) 20-30% of the polysaccharides are cellulose, which can form microfibrils that serve as the main structural component. \(^2\) Hemicelluloses occupy 25-30% of the polysaccharide fraction, which forms a network with cellulose, and this network serves as the load-bearing structure of the primary cell wall. \(^1\) Pectins occupies about 30-35% of cell wall dry weight and can be removed from the cell wall by weak chemical treatments such as EDTA/CDTA solution, and this process do not significantly disrupt the cellulose/xyloglucan network. \(^4\)

Recently, researches on cell wall have gained increasing attention due to the potential to convert lignocellulose material to biofuel. \(^5,6\) Despite its economic and environmental importance, researchers have not reached a final conclusion on the structure of the cell wall, and many aspects still remains unclear. Better understanding of the wall structure is essential for developing improved methods for dissembling methods for conversion to biofuel. In this dissertation, we focus on the study the load-bearing network of the primary cell wall using computational approaches.
1.1.1 Cellulose Microfibril

Cellulose is an unbranched polysaccharide of contiguous β-(1,4)-linked D-glucopyranose residues, and is the primary structural component in plant cell walls. The degree of polymerization of cellulose in primary cell walls are reported in two fractions (250-500 and 2500-4000), and that in secondary cell walls is 10000-15000. \(^2\) The β-(1,4)-linkages of the glucose units requires the residues to be positioned at 180° relative to their neighbor residues. Therefore, the cellobiose unit is considered as the repeating unit of cellulose chains. (Fig 1.1)

![Fig 1.1. Cellobiose unit is the repeating unit of cellulose chains](image)

Among all the cell wall polysaccharides, cellulose is the most stable component. The amount of cellulose in plant cell walls vary significantly depending on cell wall type and species. Typically, secondary cell walls have higher cellulose content (~50%) comparing to primary walls (20-30%). \(^2\) In native plants, individual cellulose chains tend to organize into microfibrils. Diameters of microfibrils in higher plants are usually between 2 to 8nm, while larger microfibrils exist in cellulosic algae. \(^7,8\) The most commonly accepted cellulose microfibril size in plants is 36 chains. \(^9,10\) Smaller numbers
of chains have been recently proposed indirectly based on fibril dimension measurements. The microfibril assembled by individual chains have great mechanical properties. For example, the Young’s modulus of cellulose along the longitudinal direction of the fibril is between 120 to 170 GPa. Therefore, cellulose is also a great candidate for advanced material development.

Based on the crystalline structure, native cellulose microfibrils can be characterized into two allomorphs, namely Iα and Iβ. Cellulose Iα has a triclinic unit cell, while that of Iβ is monoclinic. The unit cell parameters and structure of the two allomorphs are shown in Fig 1.2. The conformation and chain packing in the a-b plane are very similar, but their layering in the fibril longitudinal direction (c direction) is different. It has been shown that the two allomorphs may be converted via high temperature treatment (from Iα to Iβ) or by bending microfibril for 39°. Despite of the highly crystallized structure, cellulose microfibril is believed to contain some disorder. The exact location of the disordered region is still in debate, and two major conclusions are proposed: periodic disorder along the longitudinal direction of the microfibrils, and disordered chains at the surface of the microfibrils.

Cellulose is the most abundant biological material in nature. Its high abundance makes it more likely as a candidate for biofuel production. However, the current conversion process has low yield due to the recalcitrance of cellulose microfibrils towards enzymatic hydrolysis. The high stability of the microfibril makes cellulose insoluble, thus extremely difficult to be extracted from the cell wall network. Most of the experimental characterization requires some extent of extraction in order to characterize
cellulose microfibrils, but cellulose may have been modified during the harsh chemical treatments. Therefore, methods that require no or minimum treatments are essential for understanding the native properties of cellulose microfibrils.

Fig 1.2. The relationship between the unit cells of cellulose Iα and Iβ. Modified from Reference 25

1.1.2 Xyloglucan

Hemicelluloses are a group of polysaccharides that can be solubilized from wall by treatment with aqueous solutions of alkaline after removal of pectic polysaccharides. In cell walls, hemicelluloses may adopt a cellulose-like conformation and cause a tendency to interact with cellulose microfibrils via hydrogen-bonding. Among various types of hemicelluloses, xyloglucan has been proven to be the most abundant hemicellulose in dicot primary cell walls. Similar to cellulose microfibrils, the backbone of xyloglucan also consists of β-1,4-linked glucose units. The basic structure of
xyloglucan is a motif that consists of a backbone of four glucose units, with xylose branch on three of the consecutive glucose substituted at the O6 position. (Fig 1.3) The O2 position of xylose can be substituted with galactose. The O2 position of galactose can then be further substituted with fucose. 27

Fig 1.3. The basic structure of xyloglucan. The position that can be substituted is labeled as R. Figure modified based on 26

The Young’s modulus of hemicellulose is 5-8 GPa, which is 20 times smaller than that of cellulose microfibrils. 3,28 However, it can significantly enhance the strength of the cell wall network structure by tethering between the microfibrils and form biomechanical “hot-spots” which may form tight contact between microfibrils and prevent cell wall creep upon treatment with endoglucanase. 1 Therefore, it is essential to include xyloglucan in the cell walls such that the network structure may be hold, which is impossible to be achieved by microfibrils alone. Understanding the details about how the interactions are formed will provide important insights on the architecture of cell walls.
1.1.3 Water

In the natural condition of cell walls, a large fraction is occupied by water, which fraction ranges from 25% (such as in wheat and barley roots) to 90% (such as in Ulva lactuca).\textsuperscript{29} Despite that research implemented on water-cellulose interface is limited, the importance of water on other various types of surfaces and structures (such as enzymes, metals, polymers, and carbohydrates) is extensively published. Water has been proven to play important roles in metal surface, protein, and polymer characterization. Water can form ordered adsorbed layer on metal surfaces, and the behavior of the ordered layer depends on the structural arrangement of the metal atoms underneath.\textsuperscript{30} In addition, water may affect the enzymatic reactions by serving as a modifier of the solvent and controls the polarity and solubility of the reactants and products.\textsuperscript{31} In the case of cellulose microfibril, it has been shown by atomistic simulation that the configuration of exterior glucose units are different from that of the interior ones, indicating that water may be the cause of the surface disorder of the microfibrils. In addition, water may also affect the mechanical property of the microfibril by varying the hydration level of the surrounding.\textsuperscript{32} All of above observations imply that water is not simply a solvent. Ignoring water may cause difference in the properties of the polysaccharides comparing their native condition.

1.2 Study Primary Cell Wall in silico

1.2.1 Atomistic Simulation

As mentioned previously, many experimental methods for characterizing cell wall polysaccharides requires harsh chemical treatments to extract those components. During
the treatments, the native structure of the wall component may have been modified. Molecular dynamics (MD) simulation offers an alternative to chemical extraction, but is subject to a force field that is simultaneously accurate and computationally efficient. There are several commonly used force field used for atomistic simulation, such as CHARMM35, \(^{33-37}\) GLYCAM06, \(^{36-38}\) GROMOS96 45A4, \(^{37,39-42}\) OPLS, \(^{43-45}\) and PCFF. \(^{46-49}\) Atomistic simulation is a great tool to study detailed local behavior of the cell wall structure such as the xyloglucan segment interaction with cellulose microfibril segments, the interaction between water molecules and cellulose microfibrils, and the effect of varying surface on the interactions. The atomistic level details are useful for providing insights on the local structures of the network, and allow one to vary details of the simulation conveniently.

### 1.2.2 Necessity of Coarse-Grained Simulation

Atomistic simulation has been applied to study cellulose microfibrils and their interaction with xyloglucans. Due to the high computational demand, the dimension of cellulose microfibrils in atomistic simulation is limited to less than 80 residues. In the studies of interactions with xyloglucan, the segment length of fibril is further reduced in order to compensate for the computational cost of xyloglucan, whose length is less than 12 residues long in most simulations. \(^{50-52}\) The cellulose degree of polymerization in primary cell wall is at least 500 residues, and the xyloglucan chains molecular weight is at least on the order of \(10^5\) g/mol. \(^2,53\) Such dimensions are far beyond the current limit of
atomistic simulations. Therefore, methods with lower computational cost are desired to study the cell wall load-bearing network structure.

Coarse-grained model is a mesoscopic model, in which each force site is defined by multiple atoms. By treating each force site as one particle and only defining the interaction between the force sites, the computational cost is significantly reduced. \(^\text{54}\) Coarse-grained simulation has been widely used to study macromolecule behavior such as lipids, proteins, and polymers. By coarse-graining the system, one loses the atomistic level details such as hydrogen bonding, but gain the ability to simulate much larger cellulose microfibril and xyloglucan chains, which better mimics conditions within the wall. Based on the level of details required, one can adjust the number of atoms included in each force site accordingly.

1.3 Thesis Overview

This dissertation focuses on investigating the properties of the load-bearing components of primary cell wall. We emphasize on the structural properties of the network formed by cellulose microfibrils and xyloglucan chains via coarse-grained simulations in Chapters 2 and 3. As water is a large fraction in the native wall, the mutual influences of polysaccharides (particularly cellulose microfibril) and their surrounding water is studied using atomistic simulation in Chapter 4. The dynamic properties of water are obtained at several temperatures, and we summarize its properties to provide insights on the possible mechanisms of water motions in Chapter 5. The following questions/topics are addressed in this dissertation:
- The coarse-grained force field for cellulose microfibrils (Chapter 2)
- What information does the coarse-grained simulation reveal about the structural feature of the full length microfibril?
- The coarse-grained force field for xyloglucan (Chapter 3)
- How do the xyloglucan chains interact with cellulose microfibrils? (Chapter 3)
- How do the xyloglucan chains that do not interact with cellulose microfibrils contribute to the strength of the network structure? (Chapter 3)
- What effects the cellulose microfibril surface have on the motions of water? (Chapter 4)
- Are the mechanisms of water motions different in confined water compared to bulk water? (Chapter 5)
- What cause the change in water motion mechanisms if there is any? (Chapter 5)

We applied both atomistic simulation and coarse-grained simulation to study the structural feature of the cell wall network structure, and the dynamic feature of the polysaccharide surrounding water. This work provides a fundamental understanding of the interactions between the cell wall components. We summarize our results and proposed possible future works in Chapter 6.
1.4 Reference


(25) CERMAV. Crystalline Structure of Native Celluloses


(48) Mazeau, K. \textbf{2005}.


Chapter 2 Coarse-Grained Simulation of Cellulose Iβ with Application to Long Fibrils

This chapter is published as: B. Fan and J. K. Maranas, Cellulose 2014, accepted

Cellulose microfibrils are recalcitrant toward dissolution, thus it is difficult to extract and characterize them without modifying their native state. To study the molecular level behavior of microfibrils over 100 sugar residues, we construct a coarse-grained model of solvated cellulose Iβ microfibril using one bead per sugar residue. We derive the coarse-grained force field from atomistic simulation of a 36 chain, 40-residue microfibril by requiring consistency between the chain configuration, intermolecular packing and hydrogen bonding of the two levels of modeling. Coarse-grained force sites are placed at the geometric center of each glucose ring. Intermolecular van der Waals and hydrogen bonding interactions are added sequentially until the microfibril crystal structure in the atomistic simulation is achieved. This requires hydrogen bond potentials for pairs that hydrogen bond in cellulose Iβ, as well as those that can hydrogen bond in other structures, but not in cellulose Iβ. Microfibrils longer than 100 nm form kinks along their longitudinal direction, with an average periodicity of 70 nm. The behavior of kinked regions is similar with a bending angle of approximately 20°. These kinked regions might be linked to observations of periodic disorder from small angle neutron scattering and acid hydrolysis.
2.1 Introduction

Cellulose is an unbranched polysaccharide of contiguous β-(1,4)-linked D-glucopyranose residues, and is the primary structural component in plant cell walls. As the most abundant biological material on earth, cellulose has gained increasing attention due to its potential for conversion to useful products including biofuels and direct energy.  

Cellulose chains in native plants tend to organize into microfibrils, which sometimes bundle into macrofibrils and larger organized structures. Microfibrils in higher plants have diameters between 2 and 8 nm, while larger microfibrils occur in cellulosic algae. Although no direct proof is available, and smaller numbers of chains have been suggested, it is widely accepted that the cellulose microfibril in plants contains 36 chains. Cellulose has been studied for more than 16 decades, yet we lack agreement on some microfibril properties. Direct imaging suggests that native cellulose microfibrils have some flexibility: they contain bends and turns along the extending direction. Their cause is still unclear: both the fibril-fibril interaction and the interaction between cellulose and other components can contribute to this behavior. Another area of contention is the presence of disorder in the microfibril. The existence of disordered regions is clear from X-ray scattering, which does not reveal its location. Two possibilities have been proposed. The disordered regions may be located on the surface of the microfibril, as supported by the correlation between surface hydroxyl group percentage and cellulose crystallinity, or they may be periodic, as demonstrated combining small-angle neutron scattering and acid hydrolysis on ramie cellulose, and
suggested by the 50-400 nm nanowhiskers that result from acid hydrolysis on native cellulose. In order to progress further on these issues, molecular level information is advantageous.

The insolubility of cellulose makes it difficult to extract from cell walls for experimental characterization of structure. Removing other cell wall components requires chemical treatments (such as strong alkali and heating), which may alter the properties of the extracted cellulose. Molecular dynamics (MD) simulation offers an alternative to chemical extraction, but is subject to a force field that is simultaneously accurate and computationally efficient. Simulation of cellulose microfibrils is often accomplished using atomistic force fields, such as CHARMM35, GLYCAM06, GROMOS96 45A4, OPLS, and PCFF. Cellulose in primary cell walls has been reported in two molecular weight fractions: 500–1000 and 5000–8000 glucopyranose residues. The length of cellulose in secondary cell walls is even higher. In contrast, computational requirements limit atomistic simulations to a single, 36-chains microfibril, in most cases with 40 residues. Thus, methods with lower computational cost are required to simulate microfibrils of native lengths.

In this paper, we report a coarse-grained (CG) force field that is used to simulate 36 chain microfibrils of up to 400 residues. A CG model is a mesoscopic model in which force sites contain multiple atoms. Each force site is treated as one particle, and the interactions are only defined between the CG force sites. This significantly reduces the number of force and the computational cost. CG simulation has been widely applied to simulate biomolecular systems, including lipids, proteins, carbohydrates,
and polymers. Coarse-grained models of cellulose have been developed and used to study processes (such as the enzymatic pathways of degradation process) or systems (such as microfibrils with disordered regions) involving long microfibrils.

Two levels of coarse-graining are reported: three beads per residue and one bead per residue. In the three-bead models, force sites were assigned based on the MARTINI coarse-grained force field (each site contains two carbon atoms with oxygen and hydrogen atoms adjacent to them) or the M3B model (force sites are placed on C1, C4, and C6 atoms). The coarse-grained models used atomistic simulation of a solvated Iβ microfibril as the target system. Several of these models were applied to study the interaction between the carbohydrate-binding domain and cellulose surfaces.

The first single-bead cellulose CG model was developed based on an atomistic simulation of octaose, the 8-ring oligomer of cellulose. The authors placed CG force sites on the glycosidic oxygen atoms, and the model was used to prepare a relaxed system of bulk amorphous cellulose. More recently, a solvent-free single-bead CG model was used to study the transition of cellulose fibrils from crystalline to amorphous structures, and the residue-scale REACH (Realistic Extension Algorithm via Covariance Hessian) coarse-grained force field was used to calculate mechanical properties such as Young’s modulus and persistence length.

Compared to single-bead models, three-bead models have more molecular detail. For example, each bead in a three-bead model contains one hydroxyl group, which dictates the directionality of hydrogen bonds. Single-bead models lose this directionality, but are required to study larger systems. In this report, we present a
solvated single-bead model for cellulose. The bonded potentials are obtained using the iterative Boltzmann Inversion method, and the van der Waals potential is assigned from the pair distribution function of amorphous cellulose chains. We solve the directionality problem by including bonded potentials between pairs of CG beads that either do or can form a hydrogen bond. We use the model to simulate fibrils with 100 – 400 residues. For fibrils longer than 100nm, we observe bends along the longitudinal direction. In the longest fibril we tested (400 residues), multiple bends are present. This behavior has not been previously reported, and may provide insights about the periodicity and the behavior of the disordered regions in cellulose microfibrils in future studies.

2.2 Methodology

2.2.1 Simulation Details

We used an atomistic target simulation provided by collaborators in which they used the CHARMM simulation package with a 36 chain 40-residue microfibril (Fig 1) in a TIP3P water box. A solvation shell (minimum of 10-Å) prevents the microfibril from interacting with its periodic images. The carbohydrates are represented with the C35 force field, and long-range electrostatics are handled using the particle-mesh Ewald algorithm with a 1-Å charge grid size. Non-bonded interactions are truncated at 10 Å. The SHAKE algorithm maintains constant lengths for bonds involving hydrogen atoms. After performing minimization, the system undergoes stepwise heating of 10 ns at 100K, 200K and 300K. A 20-ps equilibration is performed at 300K. The 9-ns production run
was performed in the NVE ensemble using a 1 fs time step. Atomic coordinates form the last 3-ns production were output every 0.5 ps for later analysis.

Fig 2.1. Cellulose microfibril at the atomistic level. In this paper, the dimension of microfibrils is denoted by the number of chains on each side surface and the number of glucose units along the longitudinal direction. In case of atomistic simulation, the microfibril has a diamond shape cross-section consisting of 36 chains with 6 chains on each surface, and each chain contains 40 glucose units. This is referred as the 6x6x40 microfibril.

We use the LAMMPS \(^6^2\) package to perform the CG simulations. The initial system contains a microfibril solvated in a 10Å water shell. In order to more accurately describe interactions between the microfibril and water in the first solvation shell, we describe water molecules using a coarse-grained explicit force field instead of using an implicit solvent model. A maximum of 3000 iterations of conjugate gradient minimization is performed to relieve large strains. All CG simulations are performed in the NVT ensemble using a 1 fs time step. Here we require a 1fs as our time step because some of the potentials are sharp. For example, the bond potential of the coarse-grained beads (Table 1) and that of the atomistic bond potentials reported in C35 force field have the same order of magnitude. In this case, using a small time step will prevent forces from becoming too high during the simulation. The temperature is maintained at 300K using a Nose-Hoover thermostat. We allow the system to equilibrate for at least 5ns before collecting the production trajectories. The cut-off distances for all non-bonded interactions are set to 15Å.
2.2.2 Coarse-Grained Force Field Development

We derive the CG force field based on the atomistic simulation described above. The objectives in deriving this force field are to realize the largest computational gain while maintaining microfibril structure, and to keep the force field general enough that branches may be easily added for simulation of xyloglucan or other polysaccharides. We choose one force site per residue and compensate for the loss of directional hydrogen bonds by introducing interactions for pairs that can or do form hydrogen bonds. These interactions resemble bonded interactions that occur between specific pairs, but are allowed to break and reform during the simulation. We refer to these potentials as hydrogen bonds, although it is important to note that only one of them corresponds to an actual hydrogen bond in the native cellulose Iβ crystal structure. The others are required to reproduce the glucose residue packing pattern. Atomistic simulations suggest that some do form hydrogen bonds at high temperature. It is not problematic to enforce a specific separation between pairs that do not hydrogen bond, as the atomistic simulations clearly show a preferred separation distance without such bonding. To achieve generality for later simulation of branched polysaccharides, we develop our force field starting with CG bond stretching, angle, and torsional potentials, and non-bonded pair interactions. The hydrogen bonds are added consecutively, and their impact on the microfibril structure is noted at each addition.

The force field parameters are obtained by Boltzmann inversion of the corresponding intramolecular and intermolecular probability distributions. In order to obtain the distributions, we first determine the location of the force sites. Two potential
placements of the CG bead center are the glycosidic oxygen (O4 model) or the center of the D-glucose rings (ring center model) (Fig 2). To determine the best choice, we consider the probability distributions of intramolecular bonding (bond lengths, bending angles, and torsional angles) and intermolecular pair interactions (intermolecular pair distribution function) (Fig 3a-c) to look for obvious difficulties in using them as CG potential targets. The two models have similar bond length and bending angle distributions, but the torsional angle distribution is much softer in the ring center model. Choosing a softer torsion potential minimizes the risk of artificially reducing conformational freedom of the glucan chains. More importantly, the pair distribution function of the ring center model (Fig 3e) is less structured than the O4 model (Fig 3d), so that we can decompose the pair distribution function into individual hydrogen bonds and more accurately describe the preferred separations with intermolecular potentials. Therefore, we place the force site at the D-glucose ring center, which benefits computational cost and accuracy.

Fig 2.2 Coarse-grained mapping of cellulose chains with beads centered on (a) O4 (the glycosidic oxygen) or (b) ring center
The CG force field incorporates bonded and non-bonded potentials. Bonded potentials include bond stretching, bending, and torsional transitions. The target intramolecular distributions (Fig 3a-c) are Gaussian, and thus we represent the CG bonded potentials by harmonic springs:

\[ U_{\text{intra}}(x) = k_0 (x - x_0)^2 \]  \hspace{1cm} (1)

where \( x \) is the bond length, bending angle, or torsional angle. We obtain the parameters \( x_0 \) and \( k_0 \) from atomistic intramolecular distributions (Fig 3a-c) using iterative Boltzmann inversion.

The intermolecular potential must describe van der Waals interactions and hydrogen bonding. Because the coarse-grained beads are neutral, electrostatic interactions are not considered. We represent the van der Waals interactions with a 6-12 Lennard-Jones potential:
\[ U_{LJ}(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right] \]  

where \( r \) represents the separation distance. To obtain the preferred separation distance \( \sigma \) and the potential well depth \( \varepsilon \), we extract the data near the first peak of a published amorphous cellulose pair distribution function, apply Boltzmann Inversion, and fit inverted data points with Equation 2. Based on the fitting, we assign \( \sigma \) as 5.4 Å and \( \varepsilon \) as 0.10 kcal/mol.

As with the atomistic simulation, the coarse-grained microfibril is solvated with water, which we also represent in CG form. He et al. (2010) reported a series of CG water models with different levels of coarse-graining, functional forms of the potential energy, and cut-off distances for the non-bonded interactions. We selected a CG water bead that represents two water molecules with a 12-6 Lennard-Jones potential (\( \sigma = 3.779 \) Å, and \( \varepsilon = 1.118 \) kcal/mol).

Our approach is to add specific intermolecular potentials sequentially, thus we performed an initial simulation of the cellulose microfibril with only the bonded and van der Waals potentials. As expected, the structure disorders from the original crystalline form, and water penetrates the microfibril and disturbs the packing (Fig 4). This is reflected in the intermolecular pair distribution function (Fig 5), which is far from the atomistic level target. We thus must add specific intermolecular potentials to represent preferential interchain interactions, including hydrogen bonds. To accomplish this, we first associate the peaks in the target pair distribution function with specific features in the cellulose microfibril structure. As shown in Fig 6, the first four peaks correspond to adjacent bead preferred separation distances in the (11-0), (110), (010), and (100) planes
respectively. We anticipate that the remaining peaks are secondary repeats of these
distances, as they occur at roughly twice the distances of the first four. The atomistic and
CG representation of the microfibril cross section are shown in Fig 6e, in which the
chains are extended into the paper. Miller indices were used to describe the planes and
directions in the structure: the side surfaces of the diamond shape cross section are
labeled as (11-0) and (110) planes; the layers of glucan units are aligned parallel to the
unit cell axis b in (100) plane; the vertical stacking direction is the unit cell axis a, which
is in the (010) plane. The equilibrium values of the directional intermolecular distance
distributions (Fig 6a-d) correspond to the pair distribution peak locations (Fig 6f). The
consistency of the distances suggests the origins of the structurally significant peaks.
Because features in the pair distribution function peaks at separation distances above 9Å
most likely originate from the closer interactions, adding the four new potentials may be
sufficient to reproduce long range features.
Fig 2.4. Snapshot of the position of cellulose CG beads after 15 ns. Only van der Waals potential is considered in the nonbonded potential term.

Fig 2.5. Comparison of pair distribution functions obtained from CG simulation with van der Waals interaction only (dashed line) and the target from atomistic simulation trajectories (solid line)
Fig 2.6. The target pair distribution function decomposed into specific structural features. (a) Interlayer separation distance distribution in (110) planes; (b) interlayer separation distance distribution in (11-0) planes; (c) intralayer separation distance distribution in (100) planes; (d) interlayer separation distance distribution in (010) planes; (e) CG mapping of the cross section view of the microfibril, in which the structurally significant separation distances are shown; (in the atomistic cross-sectional view, a and b unit cell axes are highlighted in solid lines, and Miller indices of the crystal planes are highlighted in dashed lines) (f) target pair distribution function, in which the structurally significant peaks are labeled as their corresponding planes.
Boltzmann Inversion can be used on the intermolecular distance distribution curves (Fig 6a-d) to obtain a harmonic representation of these bonds, but such bonds are unbreakable and unrealistic. To represent the intermolecular constraints more realistically, we use Morse potentials, which weaken as distance between beads increases.

\[ U_{nb} = D_0 \left[ 1 - e^{-\alpha(r-r_0)} \right]^2 \]  

Here \( D_0 \) determines the depth of the energy well, \( r_0 \) is the distance of minimum energy, and the stiffness parameter \( \alpha \) determines the curvature of the potential around \( r_0 \). As the intermolecular bonds in (100) planes is a hydrogen bond, and the others are similar in strength, we assign \( D_0 \) in the Morse potential as 5.0 kcal/mol, which is representative of the O-H⋯O hydrogen bond energy strength. We determine \( \alpha \) by requiring that the Morse potential retain the same curvature as the harmonic potential around \( r_0 \). The difference between harmonic and Morse potentials should not be significant when simulating a 40-residue microfibril, but this choice may be crucial when simulating longer fibrils, in which intermolecular bonds may break and the formation of disordered regions may occur.

2.3 Result and Discussion

2.3.1 Force Field Parameters and Cellulose Crystal Structure

We add pseudo-bonds representing interchain interactions sequentially based on their physical meaning and significance. A glucose residue in the microfibril cross-section forms two intra-layer and six inter-layer bonds, where a layer ((100) plane, see Fig 6e) contains parallel, hydrogen bonded glucose rings. It has been reported that the intralayer
hydrogen bonds are the dominant intermolecular interaction in native cellulose microfibrils, and thus we add this interchain potential first. At higher temperatures (500 K), intralayer hydrogen bonds are weakened, and this enables formation of interlayer bonds. We next add interlayer bonds between adjacent glucose residues in the (110) and (110) planes. As shown in Fig 7, each addition brings the pair distribution function closer to the atomistic target, but the third peak at 7.7Å is not captured. This is also an interlayer bond between glucose residues in the (010) plane. To model this, we add a potential between adjacent beads in this plane in order to fully reproduce the accurate crystal structure of cellulose microfibril.

Following addition of interchain bonds (morse potentials), the intrachain bonds (harmonic potentials) required refinement. We present the full force field in Table 1. As mentioned in the methodology section, the reason that we choose the ring center model is that it introduces softer torsion that minimizes the reduction in conformational freedom. This enables us to more accurately decompose the pair distribution function to individual type of intermolecular interactions. Based on the refined potential parameters (Table 1), the torsional potential is more than 30 times softer than the bending angle potential, which is consistent with our strategy of bead location selection. Additionally, the strength of interchain bonds can be evaluated by α, from which we note that the potential in the intralayer direction (i.e. (100)) is at least two times sharper than the interchain interactions. This agrees with the structure of cellulose microfibril, as the intralayer interaction is contributed mostly by hydrogen bonds, and the interlayer interactions are contributed mostly by packing and Iβ crystal structure. This force field maintains the
target interchain and intrachain distributions (Fig 8) when used in a 20ns simulation of a 6x6x40 microfibril.

Fig 2.7. Demonstration of potentials required to reproduce the pair distribution function peaks. The intermolecular interactions of simulation above are van der Waals interaction and (a) intralayer bonds in (100), (b) intralayer bonds in (100) and interlayer bonds in (110), (c) intralayer bonds in (100) and interlayer bonds in both (110) and (11-0), (d) intralayer bonds in (100) and interlayer bonds in (110), (11-0) and (010). Potentials are not yet tuned.
Table 1.1 Cellulose 1\beta coarse-grained force field potential parameters

<table>
<thead>
<tr>
<th>Intrachain Bonds</th>
<th>12-6 Lennard Jones</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>( x_0 ) [Å]</td>
</tr>
<tr>
<td>stretching</td>
<td>5.237 Å</td>
</tr>
<tr>
<td>bending</td>
<td>175.6°*</td>
</tr>
<tr>
<td>torsion</td>
<td>180°*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interchains Bonds</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>location</td>
<td>( r_o ) [Å]</td>
<td>( D_o ) [kcal/mol]</td>
</tr>
<tr>
<td>(100)</td>
<td>8.32</td>
<td>5.0</td>
</tr>
<tr>
<td>(110)</td>
<td>6.68</td>
<td>5.0</td>
</tr>
<tr>
<td>(11-0)</td>
<td>5.90</td>
<td>5.0</td>
</tr>
<tr>
<td>(010)</td>
<td>7.69</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Fig 2.8. Comparison of CG simulation with atomistic targets: (a) bond length, (b) bending angle, (c) torsional angle distribution, and (d) pair distribution functions. Dashed curves: distributions following simulation with the CG force field for 5ns. Solid curves: atomistic target distributions.
Simulation with the CG force field does not disrupt the cellulose Iβ crystal structure: after 20ns of CG simulation time, the structure is highly ordered with unit cell parameters within 2% of X-ray diffraction results. A snapshot of the final structure of the 6x6x40 simulation is shown in Fig 9, and lattice parameters are presented in Table 2. Note that parameters a and b (Fig 6e) are enforced directly by the Morse potentials. The 40-residue fibril twists with a helical angle of 1.6°/nm, similar to the atomistic simulation (1.4°/nm).

Fig 2.9. Cross sectional and longitudinal view of the coarse-grained cellulose microfibril (MD simulation snapshot)

2.3.2 Applying the CG Force Field to Longer Fibrils

Using the same force field, we simulated microfibrils between 100 and 400 glucose residues. The initial structures of the microfibrils are built based on experimentally determined cellulose Iβ lattice parameters. Each microfibril was placed in a 10-Å solvation shell, and the same minimization strategy was employed. We used 5 ns production runs. As shown in Fig 10, microfibrils longer than 100nm form kinks comprised of two bends. The kinked regions are 15-20 residues (7-10nm) long, and the bending angles are 20°. Multiple kinks form in the 400-residue microfibril, with spacings of 50 and 100 nm. This is a unique observation, made possible by simulating microfibrils longer than 100 residues. We believe that kinks are a reproducible feature in fibrils longer
than 100nm, as this feature is observed in all the long fibril simulations we performed. It is interesting in light of reports of periodic disorder, both directly using SANS on ramie fibers, 9 and indirectly through acid hydrolysis. 11,66 The SANS study estimates that disordered regions are 2-3 nm, and the periodicity based on both methods is 50-150 nm. Thus, it is reasonable to conclude that the kinks are linked to the periodic disorder. Multiple simulations on long fibrils are required to generate adequate statistics on periodicity. Interestingly, the kinks appear to relieve twist. Compared to the 40-residue fibril simulation, the extent of twist in the longer fibrils is significantly weaker. The 40-residue fibril twists uniformly along the longitudinal direction, whereas the longer fibrils do not twist with the exception of one end of the 300 residue fibril. This suggests that kinks may occur to relieve the twist which leads to the formation of periodic disorder.

If kinks occur spontaneously, they must be energetically favorable. In Fig 11, we compare the potential energy of 200-residue kinked and straight microfibrils. The straight configuration is obtained by fixing the positions of beads in the center interior chain, thus forcing the microfibril to remain straight and eliminating twist. As expected, the potential energy of the straight microfibril is higher than that of the kinked microfibril throughout the simulation. This strengthens the previous conclusion that kinks are a reproducible feature of long microfibrils. The difference per bead is 0.4 kcal/mol, about twice the energy of a trans-gauche rotation. As shown in Fig 12, the energy difference is not distributed evenly among residues, but is concentrated at the kink location. In addition to the variation with residue position, we investigated the potentials that stabilize kinks. Energies for the eight different potentials are plotted in Fig 12. It is observed that kinks
cause energy penalty in the bonded potential, (010) potential, and (100) potential in the residues in the kinked region. (Fig 12a,e,h) This suggests that the formation of kinks disturbs the crystal packing in that region, while the regions of the fibril away from kinks retain their unit cell structure. Because the formation of kinks disturbs crystal packing, intra-chain bond stretching/bending, and interchain interactions in the unit cell axis directions a and b (Fig 6e) resists bending. In these potentials, the energetic cost is localized in the 10-20 residues directly involved in the kink. Interestingly, torsional and van der Waals potentials (Fig 12c,d) provide the driving force for kink formation, as they decrease energy in kinked regions. Compared to the 10-20 residues affected in the unit cell potentials, torsional and van der Waals potentials are affected over a wider range: 40 (torsion) and 50 (van der Waals) residues. These potentials involve more force sites per bead (four for torsion and more than 20 for van der Waals) and thus it appears that these interactions initiate kinks, only later disrupting the unit cell structure.
Fig 2.10. CG simulation snapshot of microfibril 100 and 400 glucose units long

Fig 2.11. Comparison of bended (lower black points) and non-bended (upper grey points) 200 residue microfibril potential energy
Fig 2.12. Various potentials against residue number for bended 200-residue microfibril. The type of potential is indicated in the y-axis, and residue number is shown in x-axis. (a)-(c) is the intramolecular potential variation along the fibril. (d)-(h) indicates the intermolecular potential variation. (a)(c)(h) contain a peak at kinked region, indicating those interactions are unfavor of the kink. Interestingly, these directions correspond to the principle axis of the cellulose microfibril crystal structure. This implies that the formation of the kink disturbs the local crystal structure of the microfibril. (c) and (d) form a well at kinked region, indicating that the long range interactions involving multiple beads are in favor of the kink formation.
2.4 Conclusions

We have reported a solvated single-bead coarse-grained force field for the cellulose Iβ microfibril based on atomistic simulation. The force field is constructed such that chain configuration, intermolecular packing and hydrogen bonding of the CG system are consistent with that of the atomistic system, and then used to simulate long (100-400 residues in length) fibrils. The most important feature of these long fibril simulations is the appearance of kinks in the longitudinal direction. The kinks are spaced by 50-150nm, and appear to relieve twist. Microfibril simulations with more than 200 residues do not twist with the exception of one end in the 300-residue case. The periodicity and kink size are on the same order of magnitude as reported for fibrils with periodic disorder. Based on the above observations, we conclude that twist may be an end effect, which can be relieved by increasing the fibril length. Torsion and van der Waals interactions are favored in kink formation, and are responsible for initiating kinks. A primary cell wall model suggested recently that load-bearing structure in primary walls is dominated by biomechanical “hot spots” (less than 2% of the total xyloglucan) in which xyloglucan and cellulose microfibrils are in close proximity. It is not possible to have short stretches of microfibrils in close contact without bending. We suggest that kinks provide preferred locations to form “hot spots” while retaining the interfibril spacing. NMR spectra show that both cellulose Iα (triclinic) and Iβ (monoclinic) exist in higher plants. It has been shown that bending can cause the interconversion between the two allomorphs, and that the Iα:Iβ ratio is sensitive to the bending angle. Our force field is designed to simulate cellulose Iβ, and the extent of bending is not sufficient to allow the microfibril
to fully convert to Iα (which requires bending of 39°). It is likely that bends like those observed here initiate this interconversion. The coarse-grained force field is developed based on Iβ allomorph. Thus, its ability to study cellulose Iα structure and interconversion is limited. However, the natural occurrence of kinks implies that microfibrils can interconvert spontaneously.

2.5 References


(13) Albersheim, P.; Darvill, A.; Roberts, K.; Sederoff, R.; Staehein, A. Plant Cell

(14) Morán, J. I.; Alvarez, V. a.; Cy拉斯, V. P.; Vázquez, A. Cellulose 2007, 15, 149–159.

(15) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R.


(17) Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.;

(18) Matthews, J. F.; Bergensträhle, M.; Beckham, G. T.; Himmel, M. E.; Nimlos, M.

(19) Matthews, J. F.; Beckham, G. T.; Bergenstrahle-Wohlert, M.; Brady, J. W.;

(20) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonza, J.; Daniels, C. R.;


Chapter 3 Molecular Level Scale Investigation of the Xyloglucan-Cellulose Microfibril Assembly

The work in this chapter is adapted from the author's manuscript: B. Fan and J. K. Maranas. In preparation.

The network structure formed by cellulose microfibrils and xyloglucan is the load-bearing component of the primary cell wall. To determine the molecular level behavior of the network structure, we develop a coarse-grained force field of (XXXG) xyloglucan chains by requiring consistency between the chain configuration and intermolecular interactions between the coarse-grained simulation and atomistic simulation. We combine this force field with our previously developed cellulose microfibril force field to simulate an assembly of cellulose and xyloglucan, and use the simulation with its periodic images to mimic the network structure. We observe that the xyloglucan chains can be characterized into bridging chains (that interact with two fibrils), single chains (that interact with one fibril), and isolate chains (that do not interact with fibril directly). The isolated chains can bind to other xyloglucan chains to participate in bridging microfibrils indirectly. In addition, we also observe that the interfibril regions of xyloglucan chains tend to coil. Our observations differ from many of the commonly accepted cell wall models, in which xyloglucan chains are drawn as single extended chains, but this better
explains why xyloglucan contributes significantly to the extensibility and strength of the cell wall structure.

3.1 Introduction

Primary cell wall is a highly complex structure made of cellulose, hemicellulose, pectin, and other proteins and inorganic molecules. Cellulose consists of linear chains of β-(1,4)-linked glucose units, and the individual chains can bundle together to form highly rigid microfibrils. There exist multiple types of hemicellulose. Xyloglucan are considered as the major hemicellulose in the primary cell wall of eudicots and nongraminaceous monocots, where it is believed to play an important role in cell wall structure and function: firstly, it may act as a spacer to prevent microfibrils from aggregating after biosynthesis; secondly, xyloglucan can also act as bridges to cross-link adjacent cellulose microfibrils. \(^1\)-\(^3\) The network formed by cellulose microfibril and hemicellulose polysaccharides is the major load-bearing structure of the primary plant cell wall. \(^4\)-\(^6\)

Similar to cellulose microfibrils, xyloglucan polymers also have a glucan backbone. However, most of the glucose units (G) are substituted at O6 position with α-D-xylose (X), forming the XXXG basic units. O2 of xylose residues can be substituted with β-D-galactose, which may be further substituted with α-L-fucose at its O2. \(^7\) (Fig 3.1)
The existence of xyloglucan cross-links between microfibrils has been proven by nuclear magnetic resonance (NMR) and extraction techniques. NMR study supports the existence of both mobile and immobile xyloglucan fractions. The mobile fraction is thought to be the cross-links, and the immobile fraction is thought to be the xyloglucans entrapped or adsorbed by cellulose microfibrils. However, there is no direct evidence to support this conclusion. Since most of the experimental measurements are measuring sample averages, and the xyloglucan chains are too small to be detected by direct visualization technique, it is difficult to answer details of how xyloglucan interacts with cellulose microfibril using experiments.

Molecular dynamics simulation is a great tool for this purpose. One has control of what to include in the simulation system and their contents. It can also provide molecular level detail for the interactions without applying any harsh chemical pre-treatments. Atomistic simulation has been applied to study effect of xyloglucan side chain variation on the interaction between xyloglucan and cellulose microfibril. Zhang et al. studied the effect of water and side chain variation on the adsorption of xyloglucan on cellulose, and
reported that variation of one side chain results in difference in the equilibrium structure of the adsorbed xyloglucan, but it does not significantly influence the interaction energy and binding affinity. 9 Zhao et al. simulated xyloglucan segments that located between microfibrils to mimic the trapped domain of xyloglucan chains. They observed that the xyloglucan segments have stronger interaction with cellulose at the microfibril hydrophobic surfaces. 10 Hanus and Mazeau suggested that the xyloglucan molecules with short side chains are capable to form flat outstretched conformation with all the sugar units interacting with the cellulose surface, while xyloglucan with longer side chains are unable to adopt this conformation. 6 The published atomistic simulations are restricted to study only the xyloglucan segments whose backbones are 4 to 12 glucose-unit long. However, in reality, the length of xyloglucan backbone is $10^2$-$10^3$ larger than the segment size. 11 To obtain a more realistic description for the xyloglucan cellulose interaction, a more computational efficient model is desired to simulate xyloglucan chains with lengths closer to that in primary cell walls.

We have previously developed a coarse-grained force field for cellulose which was applied to simulation for microfibrils up to 400-glucose residues long. In this paper, we developed a coarse-grained force field for xyloglucan with XXXG repeat units. The coarse-grained beads are defined by each sugar monomers. We simulate a 200-glucose unit long microfibril surrounded by 20 (XXXG)$_{50}$ chains, and observe several functions of xyloglucan in the load-bearing network of cell wall, which contain both consistent and challenging aspects based on the current cell wall models.
3.2 Method

3.2.1 Atomistic Simulation

We used an atomistic simulation of xyloglucan segments as our simulation target. Prior the atomistic simulation, we need to decide specifications for the system. Xyloglucan is a polysaccharide with molecular weight on the order of $10^6$ g/mol. It is impossible to conduct a simulation of such long xyloglucan chains. It is essential for us to determine a reasonable length for the xyloglucan chains for the simulation, while retaining its key properties in the cell wall network. Lopez et al. has been reported that requires at least 3 motifs in order to observe significant interactions between xyloglucan and microfibrils. Therefore, we picked 3 as the length of xyloglucan chains in the atomistic system. We included 15 chains in the simulation box such each xyloglucan chain in the resulting configuration will not interact with its periodic image. Initially, the 15 chains are aligned such that there are 5 parallel chains each in the x, y, and z direction of the simulation box. This is to ensure that the result will not be affected by the anisotropy of the system.

The simulation was performed using LAMMPS simulation package. The carbohydrates are represented with the C35 force field, and the long-range electrostatics are handled using the standard Ewald summation algorithm. The non-bonded interactions are cut-off at 10 Å. The time-step of the simulation is 1fs. After performing minimization, the system undergoes box shrinking process which the initially extended (XXXG)$_3$ segments will coil to form smaller simulation boxes. We first performed simulation in the NVT ensemble at 300K to slowly shrink the box from 70Å
to about 40Å long in 20ns. Then the system undergoes simulation in the NPT ensemble at 1 bar for 20ns such that the box is allowed to shrink further to produce desired density of the xyloglucan segments. The final density of the simulation box after box shrinking is ~1.4g/cm³. (Fig 3.2) The simulation is then switched to NVT at 300K for 2ns production run with trajectories output every 1ps for later analysis. During the production run, SHAKE ¹⁶ algorithm is used to maintain constant lengths for bonds involving hydrogen atoms.

![Fig 3.2. Snapshot of atomistic simulation of 15 chains of (XXXG)₃.](image)

### 3.2.2 Coarse-Grained Simulation

The coarse-grained force field of xyloglucan is developed based on the atomistic simulation described above. In order to achieve the largest computational efficiency in the model while retaining the xyloglucan structure, and to keep consistency with our
previous developed coarse-grained cellulose model, we decided to use one force site per for the xyloglucan chains. The force sites are defined as the geometric average of the six member ring of each glucose or xylose unit. Therefore, the atomistic xyloglucan chains can be simplified as the shown in Fig 3.3.

![Fig 3.3. Comparison of atomistic (left) and coarse-grained (right) representation of xyloglucan chains.](image)

The force field containing two types of potential, i.e. bonded potential and non-bonded potential. The bonded potential $U_b$ accounts for bond stretching, bond angle bending and torsions. The bond stretching potential is described by a harmonic spring as expressed by the following:

$$U_b(r) = k_0(r - r_0)^2$$  \hspace{1cm} (Eq. 1)

where $r$ is the length of the bond, and $r_0$ is the equilibrium bond length. There are two types of bonded potentials in the coarse-grained system, i.e. GG and GX. The equilibrium bond lengths are obtained based on the GG and GX distance distributions generated based on the atomistic simulation trajectories. (Fig 3.4) $r_0(GG) = 4.80\text{Å}$ and $r_0(GX) = 5.27\text{Å}$. Both values are smaller than that of the distance separating two
glycosidic units in the cellulose chains or oligomers. This is reasonable as the sugar units in the xyloglucan chains are not fully extended, instead, they may form coils which results in smaller distances between the sugar units. The data points of the bond length distribution for each type of bond are fit with a Gaussian distribution, from which we can derive the initial value of the potential parameter \( k_o \) by applying Boltzmann Inversion.  

Different from the bond length distribution, the bending angle and the torsional angle distributions shows multiple peaks along their angle ranges, which are impossible to be described by analytical equations. The bending angle potentials consist of GGG and GGX angles, and the torsional angle potentials consist of GGGG, GGGX, and XGGX torsions. These potentials are represented in the tabulated forms. The initial potentials generated by applying Boltzmann inversion to the bending angle and torsion angle distributions are shown in Fig 3.4.
The nonbonded potentials $U_{nb}$ account for the interchain interactions between the GG, GX, and XX beads. We use pair distribution functions of those interchain groups to determine the initial input for the nonbonded potential. The initial inputs are generated based on fitting of the first peak of the pair distribution functions (Fig 3.5) to Morse potentials (Eq. 2).

$$U_{nb} = D_0\left[1 - e^{-\alpha(d-d_0)}\right]^2$$ (2)

where $d$ is the distance between the non-bonded beads, $d_0$ represents the distance of minimum energy, the stiffness parameter $\alpha$ determines the curvature of the potential around $d_0$, and $D_0$ determines the depth of the energy well. The cut-off distances for all three non-bonded interactions are set to 15Å.
After determined the initial inputs for both bonded and non-bonded interactions, the coarse-grained simulation of the xyloglucan chains are performed using the following specifications. The coarse-grained simulation is performed by LAMMPS simulation package. We used the final equilibrated atomistic simulation trajectory of the xyloglucan chains as the starting system. The atomistic trajectory undergoes coarse-grained mapping. The system is then minimized by a maximum of 3000 iterations of conjugate gradient minimization to relieve any possible large strains. The simulation is then performed in NVT ensemble at 300K using a Nose-Hoover thermostat. The time-step set to 1fs such that it is consistent with our previous coarse-grained simulation of cellulose microfibril, which makes the two force fields compatible for simulation of microfibril xyloglucan network.
3.3 Result and Discussion

3.3.1 Xyloglucan Coarse-Grained Force Field

For the bonded potentials, the initial input of the coarse-grained force field of (XXXG)₃ xyloglucan chains are optimized using iterative Boltzmann Inversion ²⁰ which modifies the force field parameter based on the previous simulation results. For the non-bonded potential, the fit parameters are adjusted manually based on the previous simulation results after we complete the iterations for the bonded-potentials, as changing the non-bonded potentials (which is much weaker comparing to the bonded potentials) have very small effect on the bond, angle, or torsion distributions. We adjust D₀ to optimize the peak heights and adjust α to match the peak curvatures. The complete coarse-grained force field parameters are listed in Table 3.1, and the comparison of the bonded and nonbonded distributions with their atomistic targets are shown in Fig 3.6. The final k₀ for the bond stretching potentials are 9.5 kcal mol⁻¹ Å⁻² for GG bonds and 2.1 kcal mol⁻¹ Å⁻² for GX bonds. Comparing to the GG bonds in the coarse-grained cellulose microfibril force field, ¹⁹ the GG bonds are almost 10 times smaller, indicating that the crystallinity plays an important role in strengthen the interactions between the sugar units. This strength of interaction is reflected by the recalcitrance and strong mechanical properties of the microfibril. The coarse-grained force field well describes the xyloglucan chains as indicated by the good agreements between the coarse-grained distributions and their atomistic targets. (Fig. 3.6)
Table 3.1. Xyloglucan (XXXG)\textsubscript{n} chain coarse-grained force field potential parameters

<table>
<thead>
<tr>
<th>Intrachain Bonds</th>
<th>Interchain Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>x\textsubscript{0}</td>
</tr>
<tr>
<td>bond GG</td>
<td>4.80 Å</td>
</tr>
<tr>
<td>bond GX</td>
<td>5.27 Å</td>
</tr>
<tr>
<td>bending</td>
<td>Tabulated (See Appendix B.1)</td>
</tr>
<tr>
<td>torsion</td>
<td>Tabulated (See Appendix B.2)</td>
</tr>
</tbody>
</table>

Fig 3.6. Comparison of coarse-grained simulation (red dots or curves) with atomistic targets (black dots or curves)
3.3.2 Xyloglucan and Microfibril Assembly

We combine our previously developed cellulose force field with the xyloglucan force field to conduct a simulation on the microfibril and xyloglucan assembly. The simulation box contains: a cellulose Iβ microfibril that consists of 36 chains, each chain is 200-glucose units long; 20 xyloglucan (XXXG)_{50} chains. The amount of xyloglucan and the amount of cellulose in this system are approximately equal, which is consistent with the typical condition in the cell walls of *Arabidopsis thaliana*. The simulation is run *in vacuo*, but to ensure the proper interfibril distance between the microfibril and its periodic images. The microfibril contains about 6Å margin on each side, resulting in the interfibril distance to be approximately 15Å. This is consistent with the observations made using small angle neutron scattering experiments. The xyloglucan chains are generated randomly configured and randomly placed in the simulation box such that the beads on the generated chain do not overlap with the cellulose microfibril particles or other xyloglucan chains.
Fig 3.7. Coarse-grained simulation snapshot of the cellulose microfibril and xyloglucan assembly. (a) Initial configuration. (b) Snapshot of simulation after 10ns. Each snapshot includes the simulation box itself with its periodic images above and below in y-axis.

The simulation is performed at 300K for ns. The intermolecular interactions between the beads in xyloglucan chains and the beads in cellulose microfibrils are described by the interchain interactions shown in the coarse-grained xyloglucan force field. After minimization, the simulation was performed for 10ns using 1fs as the timestep. The initial and final snapshots of the simulation box are shown in Fig 3.7. We observed that one cellulose chain in the microfibril becomes disordered comparing to the rest of the cellulose chains indicating that the interaction between cellulose microfibril and xyloglucan can disturb the crystallinity of the microfibril, but overall, the rest of the cellulose chains still hold the crystalline structure of the microfibril.
By observing the simulation trajectory and its periodic images, we find that the xyloglucan can be characterized into three different groups based on their interaction with cellulose microfibril: xyloglucan that interacts with both the microfibril and the fibril’s periodic image, xyloglucan that only interacts with only one microfibril, and xyloglucan that is not in contact with the microfibril. (Fig 3.8)

Fig 3.8. Possible locations of xyloglucan chains. Note the molecules are represented in their dynamic bonds of 5.4 Å, which means all beads with instant distances longer than 5.4Å will be recognized as “bonds” in the figure. This is why there are some bonds exist between cellulose chains in the snapshots, but these bonds are not physically there in the simulation. This is also true for other simulation snapshots.

In the case that the xyloglucan is interacting with both the microfibril and its imaging, the xyloglucan chain also acts as bridge between microfibrils and enhances the overall mechanical property of the network structure. Experimental measurements by NMR has characterized this bridging xyloglucan chains as the immobile fraction, $^{6,8}$ and the common current cell wall models often view the bridging xyloglucan as one long chain that connecting more than one microfibrils. Does this mean the non-bridging xyloglucans
are only space fillers that do not contribute to the strength of the network? Our simulation shows that this is unlikely to be true. The xyloglucan chains that only interacts with one fibril or do not interact with any fibril may also bind with other chains to participate in the bridge. For example, in Fig 3.9, there are 3 xyloglucan chains in this region, and each is shown in different color. The green chain is the only one that binds to the microfibril, and the purple and pink chains are not incorporated with the fibril at all. However, the interfibril space is filled by the purple and pink chains, while both chains are coiled with the green chain. Therefore, when pulling the network structure in the direction between the two fibrils shown in the figure, the bridge structure formed by the multiple xyloglucan chains may also contribute to the strength and prevent the network structure from collapsing.

Fig 3.9. Example of the bridge structure formed by multiple xyloglucan chains.
In addition to the xyloglucan location, the coarse-grained simulation also brings useful information about the interaction between fibril and xyloglucan chains. It is suggested in various cell wall models that xyloglucan chains forms a coating on cellulose microfibril, which we did not observe in our model. Instead, we observe that xyloglucan chains tend to align with the microfibril at the interaction sites. Fig 3.10 shows the instant snapshots of one xyloglucan chain on the microfibril along the 10ns simulation time. Along the simulation progress, the xyloglucan first interacts with cellulose microfibril in a small region. The interaction region then becomes more extended along the microfibril within 7 ns and stay stabilized. The direction of xyloglucan chain extension is consistent with the results based on atomistic simulation, which demonstrated that xyloglucans have higher preference for interaction in the longitudinal direction than interaction perpendicular to the cellulose chain direction. Atomistic simulation is capable to describe interactions of short xyloglucan segments, but simulating long xyloglucan chains is beyond its limit. Experimental results such as NMR have suggested xyloglucan may coated along the microfibril, but it direct proof and detail about the coating region cannot be provided by these measurements. Our simulation provides derives a possible formation process of the xyloglucan interaction sites on the microfibril.
3.4 Conclusion

We develop a coarse-grained force field for XXXG based xyloglucan chains based on the atomistic simulation of 15 (XXXG)$_3$ segments. The force field is constructed such that the chain configuration and intermolecular interaction of the coarse-grained system matches that of the atomistic simulation. The coarse-grained force field is then applied to simulate xyloglucan and cellulose microfibril assembly. The simulation box contains one microfibril with 36 chains, and each chain is 200 glucose units long. The xyloglucan chains also have a 200 glucose units long backbone, and we include 20 chains in the box. In this way, the weight ratio of cellulose and xyloglucan in the simulation box is approximately equal, which is similar to that in Arabidopsis thaliana.\textsuperscript{21} We use the coarse-grained simulation system with its periodic image to mimic the network structure. We observe that the xyloglucan chains can be characterized into three types based on their location: the bridging xyloglucan (that interacts with two microfibrils), the non-bridging xyloglucan (that interacts with only one microfibril), and the isolated xyloglucan (that does not directly interacting with the microfibril). Isolated xyloglucan is thought to
serve as a space-filler of the network structure to maintain proper interfibril distances. However, our simulation shows that the isolated xyloglucan can interact with other xyloglucan chains and become part of the bridge structure, which may also potentially contribute to the strength of the network structure. In addition, the bridge structure, formed by a xyloglucan chain or formed by multiple xyloglucan chains, prefers a coiled structure in the interfibril space instead of being fully extended. It has been suggested that xyloglucan/cellulose networks provide a balance between the mechanical strength and extensibility of the cell wall. The coiled bridge provide the network more flexibility comparing to the fully extended structure, thus can serve as a greatly contribute to the wall extensibility.

### 3.5 Reference


(19) Fan, B.; Maranas, J. K. *Cellulose* **2014**.


(22) Huang, S.-C.; Park, Y. B.; Cosgrove, D.; Maranas, J. K. *in prep.* **2015**.


Chapter 4 A Computational Study on the Mutual Effect between Cellulose Iα Microfibril and the Interfacial Water

The work in this chapter is adapted from the author’s manuscripts:


Cellulose microfibrils are the load-bearing structure in the plant cell wall, and it has been shown that microfibril mechanical property is influenced by the hydration level in the system. In this paper, we demonstrate water at the microfibril surface behaves significantly different frombulk water, which proves that the influence between water and cellulose microfibril is mutual. We performed atomistic simulation of four aligned cellulose Iα fibrils solvated in water. The hydration levels of the system are varied between 5% and 20%. The water dynamics properties in these systems are studied by fitting the self-intermediate scattering function using a stretched exponential model. One of the unique features of molecular dynamics simulation is that it enables us to study translational motion and rotational motion separately. By applying jump model to the fitting results, we determine the translational diffusion coefficient, jump length, and residence time within the local cage. We observe a great decrease in both translational and rotational motion as the hydration level becomes lower. Multiple types of
translational motion exist in 20% hydrated system. However, the faster motion is still much slower than bulk water, indicating that all water in 20% system is confined. Since 20% hydration has reached the level of water content in secondary cell walls, ignoring the effect of water when analyzing cell walls may cause severe difference from reality.

4.1 Introduction

Cellulose consists of linear $\beta$-(1,4)-linked glucose chain. As the most abundant biological material on earth, it has been studied for more than 17 decades due to its potential contribution to biofuel production. Despite of the high demand, the current cellulose processing cannot achieve commercial viability. This problem is mainly caused by biomass recalcitrance, namely the slow kinetics of decomposing cellulose into monosaccharides. In plant cell walls, individual cellulose chains bundle together to form microfibrils after been synthesized. These microfibrils have very organized structure of diameter 2-8 nm in higher plants. Depending on the source and type of cell wall, the degree of polymerization of the primary cell wall microfibrils ranging from 500 to 8000. Microfibrils in secondary cell wall are even much longer. To develop an efficient process to decomposing cellulose for energy applications, it is essential to gain the knowledge of how cell wall components interact with cellulose microfibrils. The interaction between cellulose microfibrils and other cell wall polysaccharides has been studied deeply, and the contribution of those polysaccharide components to cell wall properties has been widely reported. Studies on mutants provided information about types of polysaccharides that are not essential in plant growth. Experimental and
computational methods are applied to study the interaction of individual cellulose microfibril, mainly focusing on its hydrogen bonding network. Fundamental knowledge of these interactions (between fibril and polysaccharides, or among microfibrils) are essential for both building cell wall models and provide a scientific basis for decomposing these components. However, the role of water, one of the major cell wall components, is often undervalued or even neglected in previous studies.

Water occupies a large fraction of the cell wall, and its fraction ranges from 25% (such as in wheat and barley roots) to 90% (such as in *Ulva lactuca*). Despite that research implemented on water-cellulose interface is limited, the importance of water on other various types of surfaces and structures (such as enzymes, metals, polymers, and carbohydrates) is extensively published. Water content can significantly influence the performance of enzymes, which lose their full functionality and cannot maintain their natural conformation without a certain level of water in their structures. Water serves as a modifier of the solvent and controls the polarity as well as the solubility of the reactants and products of enzymatic reactions. In the case where water is consumed or produced in the enzymatic reaction, it also affects the turnover rate. On metal electrodes, the arrangement, dielectric and dynamic properties of water can be altered by the charged surface. Those aspects all significantly influence the surface reactions and device efficiency. As for filled polymers, exposure to water nearly always changes their mechanical properties by serving as a plasticizer, swelling the polymer, destroying interface adhesion, or causing phase separation. Consequently, water absorption results in reduction of tensile strength and modulus and enlargement in damping. In addition,
previous publications have shown that water content affects the mechanical properties of cellulose microfibrils. Compared to dried system, cellulose microfibrils can be stiffened by adding a small amount of water (20% w/w) into the parallel aligned fibrils. Therefore, understanding the behavior of water can contribute to better controlling of the material properties, and to optimize materials for application. All information above implies that the fundamental significance of water on surfaces cannot be neglected, and one must examine the interface behavior in order to fully understand the chemical reactivity of cellulose.

In this chapter, we focuses on study of water at water-cellulose Iα interface using atomistic simulation. Rotational and translational motions are decoupled by computing self-intermediate scattering function (SISF) of each motion. The motions are characterized by determine relaxation time and stretch factor of each motion. Jump diffusion model is applied to calculate the diffusion coefficient of water molecules, as well as the proton residence time in the local cages and jump length between the cages.

4.2 Methodology

4.2.1 Simulation Details

The simulated system contains four aligned cellulose Iα fibers, each consisting of 36 chains with a chain length of 80. The fibers were solvated at two different degrees of hydration to match the experiments: $h=0.05$, similar to the dry sample, and $h=0.20$. (Fig 4.1Fig) The CHARMM C36 carbohydrate force field and the Tip4P-EW water model were employed. The simulations were performed with the program GROMACS.
4.5.5\textsuperscript{20} using a time step of 2 fs. Bonds involving hydrogen atoms were constrained using the LINCS\textsuperscript{21} algorithm (fourth order with one iteration), and for water the Settle algorithm was used.\textsuperscript{22} Neighbor searching was performed every 10 steps. Periodic boundary conditions were employed, and the PME algorithm\textsuperscript{23,24} was used for electrostatic interactions. A cutoff of 12 Å was used for the neighbor searching and real-space electrostatics. For the van der Waals interactions the switch function was used for distances 9-10 Å. Temperature coupling was performed with the V-rescale algorithm\textsuperscript{25} ($\tau = 0.1$ fs) and pressure coupling with the Parrinello-Rahman algorithm\textsuperscript{26} ($\tau = 1$ fs). Each system was simulated at a pressure of 10\textsuperscript{5} Pa and at 298 K. Simulations were 11 ns long at each temperature. Data from the last 10 ns of simulation at each temperature were analyzed.

\begin{center}
\includegraphics[width=\textwidth]{Fig4_1.png}
\end{center}

Fig 4.1: Initial configuration of water solvated microfibrils atomistic system

\subsection*{4.2.2 Analyzing Water Dynamic Properties}

The dynamics of water can be characterized by analyzing the self-intermediate scattering function (SISF), which is defined as

\[ S(Q,t) = \frac{1}{N} \sum_{n=1}^{N} \exp(i \cdot Q \cdot |r_n(t) - r_n(0)|) \]  \hspace{1cm} (Eq. 1)
where $Q$ is the scattering vector length, and $\mathbf{r}_n(t)$ is the position vector of the $n^{th}$ hydrogen atom. In order to ensure that our results agree with the experimental measurements (such as quasi-elastic neutron scattering), we only consider hydrogens on water molecules when performing the calculations as the scattering cross-section of hydrogen is much larger than that of oxygen, the contribution of oxygen to the scattering function is negligible. This expression takes into account all possible motions, and is not feasible to analyze all motions simultaneously. To simplify the analysis, the motions are generally decoupled (based on their time-scale) into translational motion of the molecules, rotational motion about their center of mass, and vibrational motion of individual bonds. The independence of those motions is generally assumed for liquid samples, thus the scattering function can be written as

$$S(Q,t) = S_T(Q,t) \cdot S_R(Q,t) \cdot S_V(Q,t)$$  
(Eq. 2)

where $S_T$, $S_R$, and $S_V$ corresponds to the translational, rotational, and vibrational contribution, respectively. Since SHAKE algorithms was turned on when performing the simulations, bond lengths are restricted, which implies that $S_V(Q,t)$ can be dropped from the above expression, i.e.

$$S(Q,t) = S_T(Q,t) \cdot S_R(Q,t)$$  
(Eq. 3)

One of the unique features of molecular dynamics is that this method enables us to analyze contribution of translational motion and rotational motion separately. Rotational scattering function, contributed from the re-orientation of water molecules, is obtained by

$$S_R(Q,t) = \frac{1}{N} \langle \sum_{n=1}^{N} \text{Exp}[(i \cdot Q \cdot |\mathbf{u}_n(t) - \mathbf{u}_n(0)|)] \rangle$$  
(Eq. 4)
where the rotational vector \( \mathbf{u}_n \) is derived by removing the center of mass (to which the \( n^{th} \) hydrogen belong) from \( \mathbf{r}_n \). The decay of \( S_R(Q,t) \) is fitted to the first three terms of Sears expansion rotational model, \(^{29,30}\)

\[
S_R(Q, t) = j_0^2(QR) + 3j_1^2(QR)Exp \left[ -\frac{1}{3} \left( \frac{t}{\tau_R} \right)^{\beta_R} \right] + 5j_2^2(QR)Exp \left[ -\left( \frac{t}{\tau_R} \right)^{\beta_R} \right]
\] (Eq. 5)

where \( \tau_R \) is the rotational relaxation time, and \( \beta_R \) (ranging between 0 and 1) is the stretch factor of rotation which is a description of distribution of \( \tau_R \), i.e. smaller \( \beta_R \) implies more widely distributed \( \tau_R \). The expression above contains a time-independent part and a time-dependent part. \( j_0(QR) \), the first spherical Bessel function denotes the rotational elastic incoherent structure factor (EISF), i.e. the time-independent component in the rotational structure function. This term arises by expanding the classical rotational diffusion model of a molecule, \(^{31,32}\) and it depends on the rotational radius, \( R \), which is the distance between hydrogen atom and water molecule center of mass. Based on the TIP4P-Ew model \(^{19}\) parameters, \( R \) equals to 0.9186 Å.

Translational motion can be decomposed from \( S(Q,t) \) based on Eq. 3, and it also can be described using KWW function in the following form \(^{33}\)

\[
S_T(Q, t) = A_0 \cdot Exp \left[ -\left( \frac{t}{\tau_T} \right)^{\beta_T} \right]
\] (Eq. 6)

where \( \tau_T \) and \( \beta_T \) represent the relaxation time and stretched factor of translation, respectively. \( A_0 \) is the non-ergodicity parameter, which equals to unity in most of our fittings. In the case that there are multiple types of translational motions, the overall translational scattering function can be described as a summation of multiple translational KWW functions,
\[ S_T(Q, t) = \sum_{i=1}^{n} x_i \cdot A_{0i} \cdot \exp \left[ -\left( \frac{t}{\tau_{TI}} \right)^{\beta_{TI}} \right] \]  
(Eq. 7)

in which \( x_i \) represents the fraction water protons following the \( i^{th} \) type of translational motion. We computed the distribution of the distance water oxygen atoms have traveled over a 500ps time frame (Fig 4.2), from which we determined that 5% hydration and bulk water systems can be described by a single type of motion, whereas 20% hydration system requires two types of motion to describe it.

![Fig 4.2. Oxygen distance traveled over 500ps of simulations at various water contents. In 5% hydration system and bulk water system, the distance distributions contain only one peak whereas in 20% hydration system, the distribution contains two peaks. This indicates that there are two types of translational motion in 20% hydration system, and the translational motion in the other two systems can be described by a single type of motion. The distance is computed using the summation of displacement within 1ps.](image)

The relaxation time and stretch factor obtained from fitting can be used to compute the average relaxation time \( \tau_\beta \) of each type of rotational and translational motion.\(^{34}\)

\[ \tau_\beta = \int_0^\infty \exp \left[ -\left( \frac{t}{\beta} \right)^{\beta} \right] dt \]  
(Eq. 8)
In the case of translational motion, the relaxation rate $\Gamma_T$ is calculated from the relaxation time (Eq. 9), and fitted to jump diffusion model $^{15}$ (Eq. 10) as a function of $Q$ to obtain diffusion coefficient of the water protons and the mean residence time $\tau_0$ within a cage.

$$\Gamma_T = \frac{\hbar}{\tau_0} \quad \text{(Eq. 9)}$$

$$\Gamma_T = \frac{hD_0Q^2}{1 + D_0Q^2\tau_0} \quad \text{(Eq. 10)}$$

where $\hbar$ represents Planck’s constant, $D_0$ denotes the diffusion coefficient of water molecules, and $\tau_0$ is the residence time of protons, which indicates the time scale that protons remain in the local cage before jump to another. The jump length of proton translational motion, $L$, can be determined as

$$L = \sqrt{6D_0\tau_0} \quad \text{(Eq. 11)}$$

### 4.3 Results and Discussion

To validate our results and compare the difference between bulk and restricted water motions, we first analyzed bulk water simulation results. As mentioned above, one of the unique features of molecular dynamics is that it enables us to analyze contribution of translational motion and rotational motion separately. We compute the rotational SISF of water protons, $S_R(Q,t)$, by removing the translation of water molecule center of mass from the translation of water protons. (Eq. 4) Different from translational motion whose SISF decays very fast, rotational motion always shows a plateau which indicates the value of elastic incoherent scattering factor (EISF). As shown in Fig 4.3, the rotational SISF of bulk water can be accurately fitted with a single KWW equation, $j_0$. 

72
(QR) exactly matches the plateau region of all SISFs. In addition, we could extract the rotational relaxation time and stretch factor. Since the motion of water molecule center of mass has been removed when calculating the rotational SISF, rotational motion should not depend on the length scale of the observation window. Therefore, the relaxation time and stretch factor value should not vary significantly with Q. This is also consistent with our fitting results as shown in (Fig 4.4). The average relaxation time is 0.93±0.05ps, and the average stretch factor is 0.83±0.02, and we could determine that the average rotational relaxation time of bulk water is 0.99±0.05ps. It has been shown that water rotational relaxation follows an Arrhenius behavior, based on which the average rotational relaxation time of bulk water is 1.14 ps, 29 and our simulation result is in good agreement with this experimental value.

Fig 4.3. Bulk water rotational SISF fitted with KWW equation. Rotational SISFs are obtained from Q=0.5 to Q=2.0 with increment of 0.1. The hollow points are calculations, and fitting curves are shown as red solid lines.
Translational motion can be calculated based on Eq. 3. Since bulk water only contains one type of translational motion (Fig 4.2 Fig), a single KWW equation was used to fit the translational SISFs, from which the relaxation time $\tau_T$ and $\beta_T$ was extracted. Average relaxation time, $\tau_{\beta'}$, was obtained using Eq. 8, and it is used to compute relaxation rate $\Gamma_T$ as shown in Eq. 9. Based on the Q dependence of the relaxation rate (Fig 4.5), the diffusion coefficient of water is determined as $3.3 \times 10^{-9} m^2/s$, which
agrees with the experimental measured value of $2.3 \times 10^{-9} \text{m}^2/\text{s}$\textsuperscript{35}. The residence time of water molecules in a local cage is 0.07 ps. Proton jump length is determined based on Eq. 11, which equals to 0.38Å. Both the residence time and jump length indicates that the motion of bulk water is not restricted by local cage. The results are consistent with our expectation, which indicates that we could use SISF fittings and parameters to characterize the rotational and translational motion of water.

![Fig 4.5. Bulk water relaxation rate vs. Q^2 fitted with jump model.](image)

To confirm that the rotational SISF fitting equation is validate for all water content, we compared their SISF time independent parts and observed that, at the same Q value all SISFs decays to the same plateau. (Fig 4.6) As the rotational geometry determines the time independent parts, this also implies that the rotation geometry of water is independent of water content and presence of cellulose microfibrils. Therefore, we conclude that only one type of rotational motion is required when fitting the rotational SISFs. The decay of SISFs becomes faster with increasing water content, which indicates faster and less stretched rotational motion. As in $h=0.20$, the average relaxation time and

75
stretch factor of the rotational motion is $1.44 \pm 0.10$ ps and $0.62 \pm 0.01$, respectively. (Fig 4.4)

Fig 4.6. Rotational SISF at $Q=0.7$ shows that the time independent term of rotational SISF is independent of water content.

Since there are two translational motion in $h=0.20$ system, it requires two stretched exponential terms in the translation KWW fitting equation, which requires the fraction of each type. This was done by fitting the oxygen travelling distance distribution of $h=0.20$ system using two Gaussian functions. Based on the integral of the two Gaussian functions over distance from 0 to 800 Å$^2$, we determine that the fractions of slow water and fast water are 40% and 60%, respectively. Eq. 7 is applied to determine the relaxation time and stretch factor of each translational motion at various $Q$. As shown in Fig 4.7, we plot relaxation rate calculated based on the fitting results vs. $Q^2$. Both types of motions are precisely described by the jump diffusion model. In addition, we use time series of mean square displacement (MSD) and time lapse image of single oxygen atoms (sample plots shown in Fig 4.8) to check whether our fitting agrees with the simulation. The sudden jumps in the MSD time series indicates occurrence of jumps from
a local cage, and it is usually followed by a flatter region which implies that oxygen atom remains in the cage during that time. We generated the time lapse image of the first two cages and color coated the oxygen based on the time frame it belongs to show that the two cages are well separated from each other.

Fig 4.7. $h=0.20$ SISF fitting fitted to jump diffusion model. From the fitting of this curve, we determine that the diffusion coefficient of faster water (shown in (a)) in this system is $1.2 \times 10^{-9} m^2/s$, the residence time is 7.7ps, and the jump length is 2.33Å. In case of slower water (shown in (b)), the diffusion coefficient is $0.14 \times 10^{-9} m^2/s$, the residence time is 36ps, and the jump length is 1.74Å.

Fig 4.8. Tracking a single oxygen mean square displacement over 200ps to demonstrate the jump diffusion motion of water molecules. The time lapse image of the first 75ps (with interval of 1ps) is shown, and the oxygen atoms are color-coated based on the local cage they belong to.
The average rotation relaxation time of $h=0.05$ system is $3.7\pm0.4$ ps with a stretch factor of $0.41\pm0.01$. (Fig 4.4) Since the interfibril distance of the 5% hydration system is much less than that of the 20% hydration system, water molecules are more likely to be in contact with both surfaces or lying flat on the cellulose surface. This can explain the large difference in the rotational relaxation time between the two levels of hydration. The translational motion of water in 5% hydration system is much slower than water in 20% hydration system. (Fig 4.9) Since diffusion coefficient is inversely proportional to viscosity (based on Stokes-Einstein equation), the decrease in diffusion coefficient of water implies greater water viscosity at lower water content. Due to this increase in friction between neighboring particles, it is much more difficult to slide the microfibrils if the hydration level is low. One of the major differences between primary and secondary cell walls is their hydration level. Secondary cell walls are much lower in water content, and they reach saturation at only about 30% water. 36 It has also been shown that, in spruce wood (secondary cell wall), the adjacent cellulose microfibrils are laterally packed over part of their lengths to form loose bundles, which is thought to be in absence of lignin and hemicelluloses 36–38 and water accessible. 39 The highly viscous water between the loosely bundled fibrils can prevent microfibrils from sliding apart from each other. This indicates that water may strengthen the mechanical property of secondary cell walls.
Fig 4.9. $h=0.05$ SISF fitting fitted to jump diffusion model. From the fitting of this curve, we determine that the diffusion coefficient of water in this system is $0.028 \times 10^{-9} m^2/s$, the residence time is 1813 ps, and the jump length is 5.5Å.

4.4 Conclusion

In this paper, we use molecular dynamics simulation to prove that water in plant cell walls behaves very differently than bulk water. The simulations of four laterally aligned microfibril solvated in water were performed at 5% and 20% hydration. We showed that the dynamic property of water is significantly related to the water content inside the system. We generated water proton self-intermediate scattering function, and use jump diffusion model \(^{15}\) to extract the diffusion coefficient, residence time, and jump length of water protons at 5%, 20% hydration, and compared our results with bulk water. Compared to bulk water, the rotational motions in 20% hydrated system and 5% hydrated system are slowed by 2 folds and 16 folds respectively. The much slower rotational motion in 5% hydrated system is an indication that water rotation is restricted either by more than one surface, or the water is lying flat on the cellulose surface. Furthermore, the translational motion also decays dramatically as water content decreases. Since diffusion coefficient is inversely proportional to viscosity of water, we believe that the slow water can result in a greater friction between adjacent particles and cause difficulty to slide the
microfibrils apart. This is similar to the secondary cell wall, whose hydration level is quite low. Therefore, water may contribute to strengthen the loosely bundled microfibril aggregates in the secondary walls.

Although our current study only focuses on the influence of cellulose microfibril surface on water property, there exist other types of polysaccharides, such as hemicellulose, pectin, and lignin, within cell walls. It has been shown that the presence of these polysaccharides has influences on cell wall hydrophilicity (which is an indication of water accessibility) and water diffusivity. For example, previous publication demonstrated that removal of hemicellulose induces hydrophilicity variation which affects the bioconversion efficiency of cellulose.\textsuperscript{40} The understanding of water dynamics in cell wall is not complete. A more comprehensive story requires study of a system that containing other cell wall polysaccharides, which not only contributes to a more complete knowledge on plant cell wall structure, but also can potentially provide insights on more effective bioconversion methods.

4.5 References


80


(19) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. 


Chapter 5 Dynamics of Surface Water on Cellulose I\(\alpha\) Microfibrils by Molecular Dynamics Simulation

The work in this chapter is adapted from the author’s manuscript: B. Fan, L. Petridis, and J. K. Maranas. In preparation.

We conduct molecular dynamics simulations of \(I\alpha\) solvated by water at hydration level of 5\% and 20\%. The simulations are conducted at temperatures ranging between 213K-298K. The translational motion and rotational motion are decoupled and analyzed by fitting the self-intermediate scattering functions with a stretched exponential function. The resulting rotational relaxation time and translation diffusion coefficient are fitted using Arrhenius equation to obtain the activation energy of the motions. We observe that, at both hydration levels, the water motions are significantly slower than bulk water, and the motion becomes slower at lower hydration level. The activation energies of translation of confined water are similar to that of bulk water indicating that the translation mechanism may remain unchanged. The activation energy of rotation of the confined water is also independent of hydration level, but is two times higher than that of the bulk water, indicating that confinement modifies the rotation mechanism. We performed anisotropy analysis of the rotational motion, and determined that the rotational activation energies in all rotation principle axes are identical, but the extent of anisotropy
increases with decreasing hydration level. Therefore, the slowed rotation upon decreasing hydration level is induced by anisotropy.

5.1 Introduction

Cellulose is the most abundant biological material on earth, and it consists of linear chains of β-(1,4)-linked glucose. The linear cellulose chains can crystallize to form microfibrils, which serves as the load-bearing structure in the plant cell wall. Native cellulosics are mostly found in two allomorphs, namely cellulose Iα and cellulose Iβ. In primary cell walls, cellulose microfibril diameters are generally between 2-8 nm in higher plants, while much larger bundles are found in secondary cell walls, bacteria, and algae. Cellulose has been discovered for more than 17 decades and continues gaining increasing attentions for its great potential for applications. The most long-standing application of cellulose is paper manufacturing and textiles, which cannot be established without the mechanical and physical properties of cellulose and its abundance. The focus of cellulose application now is to convert it to sustainable environmental-friendly biofuel.

Native cellulosics are nearly always found in hydrated state with water fraction ranging from 25% to 90%. Yet, the research implemented on water-cellulose interface is very limited. It is widely accepted that the property of water in the vicinity of a surface or interface are significantly different from that of bulk water at the same temperature. Water has been shown to play important roles in characterizing metal surfaces, proteins, and polymers. On metal surfaces, the arrangement, dielectric, and dynamic properties of surface water are altered. For example, molecular dynamics studies of water on Pt
surfaces show that the order of the adsorbed layer of water depends on the structural arrangement of the Pt surface underneath the water layer. These properties all significantly affect the surface reactions and energy device efficiency.\(^3,4\) Not only can water have specific structure arrangement on surface, it also may induce structure arrangement to amphiphatic molecules upon its presence. One of such example is phospholipids, which consists of a hydrophilic head and hydrophobic tails. Phospholipids in water can form lipid bilayers, which is the major component of the cell membrane.

In the native state of cellulose, the presence of water is mandatory for wall-degrading enzymes and for polymer creep during primary wall growth.\(^5\) In industrial applications, cellulose is hydrolyzed to soluble sugars via chemical or enzymatic processes. Enzymatic processes are more desirable due to milder reaction conditions, high specificity, and better control of the processes.\(^6\) Both processes requires aqueous environment, in which water is in close interaction with cellulose and can affect enzyme activities. Despite the lack of emphasizing on water-cellulose interface, the current available publications have demonstrated the significance of water on cellulose properties and applications.\(^5,7,8\)

In this chapter, the rotation and translation motion at various temperature and hydration levels are analyzed by fitting the self-intermediate scattering function to a stretched exponential function. The relaxation time and diffusion coefficient of rotational motion and translational motion, respectively, are fitted with an Arrhenius equation, from which we obtained the activation energy of each motion, which are useful for understanding the mechanisms of the water motions.
5.2 Simulation Details

The simulated system contains four aligned cellulose Iα microfibrils, which consists of 36 chains with a chain length of 80 glucose unit each. The simulation was performed under two levels of hydration, i.e. \( h=0.05 \) (or 5% hydration) and \( h=0.20 \) (or 20% hydration), where \( h \) is defined as the ratio of water mass and fibril mass. We employed the CHARMM C36 carbohydrate force field\(^9,10\) and the TIP4P-EW\(^11\) water model to describe the system. The simulations were performed with GROMACS 4.5.5 program\(^12\) using a time step of 2fs. LINCS algorithm\(^13\) (fourth order with one iteration) was used to constrain all bonds involving hydrogen atoms, and SETTLE algorithm\(^14\) was used for water molecules. The cutoff distance for neighbor searching and electrostatic interactions was set to 12 Å. Neighbor searching was performed for every 10 timesteps, and the long-range electrostatic interactions were computed using the PME algorithm.\(^15,16\) For the van der Waals interactions the switch function was used for distances 9-10 Å. We applied V-rescale algorithm\(^17\) for temperature coupling and Parrinello-Rahman algorithm\(^18\) for pressure coupling. The simulation was equilibrated for 1ns following a production run for 10ns.

5.3 Method of Data Analysis

The rotational and translational motions of water are characterized by computing the self-intermediate scattering function (SISF). The overall SISF is a description of both rotational and translational motions of water protons, and it is generally assumed that the
translational motion of the water molecules and the rotational motion about their center of mass are decoupled for liquid samples,\textsuperscript{19} i.e.

\[ S(Q, t) = S_T(Q, t) \cdot S_R(Q, t) \]  

(Eq. 1)

where \( Q \) is the scattering vector length, and \( S_T \) and \( S_R \) denotes the rotational and translational SISFs, respectively. When calculating SISFs, only the hydrogen atoms in water molecules were considered such that our result can be comparable with experimental measurements such as quasi-elastic neutron scattering, in which the scattering cross-section of oxygen is negligible compared to that of hydrogen.\textsuperscript{20} The SISF can be calculated from the trajectories of the atomistic simulation as

\[ S(Q, t) = \frac{1}{N} \langle \sum_{n=1}^{N} \exp(i \cdot |r_n(t) - r_n(0)|) \rangle \]  

(Eq. 2)

where \( r_n(t) \) corresponds to the position vector of the \( n^{th} \) proton. Note that we do not take vibrational motion into account as SHAKE algorithms was applied when performing the simulations, and it results in restricted bond length (i.e. restricted bond vibration).

Molecular dynamics simulation allows us to easily separate the contributions of rotational and translational motion. The main contribution of rotational scattering function is from the re-orientation of water molecules. By removing the water molecule (to which the \( n^{th} \) proton belong) center of mass from \( r_n \), we can obtain the rotational vector \( u_n \). The rotational scattering function is then obtained by

\[ S_R(Q, t) = \frac{1}{N} \langle \sum_{n=1}^{N} \exp[(i \cdot |u_n(t) - u_n(0)|)] \rangle \]  

(Eq. 3)

We fit the decay of \( S_R(Q, t) \) to the first three terms of the Sears expansion rotational model,\textsuperscript{21,22}
\[ S_R(Q,t) = j_0^2(QR) + 3j_1^2(QR) \exp \left[ -\frac{1}{3} \left( \frac{t}{\tau_R} \right)^{\beta_R} \right] + 5j_2^2(QR) \exp \left[ - \left( \frac{t}{\tau_R} \right)^{\beta_R} \right] \] (Eq. 4)

where \( j_i(QR) \) is the \( i^{th} \) spherical Bessel function. We extract the rotational relaxation time \( \tau_R \) and the stretch factor \( \beta_R \) from the fitting curves. The time independent part of this equation, \( j_0(QR) \), denotes the rotational elastic incoherent structure factor (EISF), which depends on the geometry of the rotational motion. This term arises by expanding the classical rotational diffusion model of a molecule. \(^{21,22}\) The rotational radius, \( R \), is set to 0.9186 Å, based on the TIP4P-Ew water force field parameters. \(^{11}\)

Translational SISF can be decoupled from rotational motion SISF based on Eq. 1, and it is well-described by a Kohlrausch-Williams-Watts (KWW) function. \(^{23}\) Multiple types of translational motion can exist simultaneously in a system, and the SISF in this case can be fitted by the summation of multiple KWW functions, with each KWW function describes one type of translational motion. (Eq. 5)

\[ S_T(Q,t) = \sum_{i=1}^{n} x_i \cdot A_{0i} \cdot \exp \left[ - \left( \frac{t}{\tau_{Ti}} \right)^{\beta_{Ti}} \right] \] (Eq. 5)

where \( \tau_T \) and \( \beta_T \) represent the relaxation time and stretched factor of translation, respectively. \( A_0 \) is the non-ergodicity parameter, which equals to unity in most of our fittings, and \( x_i \) is the fraction of water protons following the \( i^{th} \) type of translational motion. In the simplest case that there is only one type of translational motion, the above fitting equation becomes

\[ S_T(Q,t) = A_0 \cdot \exp \left[ - \left( \frac{t}{\tau_T} \right)^{\beta_T} \right] \] (Eq. 6)

We can determine the average relaxation time of the motion, \( \tau_{\beta} \), based on the relaxation time and stretch factor obtained from the SISF fittings based on Eq. 4-6. \(^{24}\)
\[ \tau_\beta = \int_0^\infty \exp \left[ -\left( \frac{t}{\tau} \right)^\beta \right] dt \] (Eq. 7)

When analyzing translational motion, we also calculated the relaxation rate \( \Gamma_T \) from the relaxation time (Eq. 8), and fitted to jump diffusion model. \(^{19}\) (Eq. 9)

\[ \Gamma_T = \frac{\hbar}{\tau_\beta} \] (Eq. 8)

\[ \Gamma_T = \frac{\hbar D_0 Q^2}{1 + D_0 Q^2 \tau_0} \] (Eq. 9)

where \( \hbar \) represents Planck’s constant. The jump diffusion model fitting will derive the diffusion coefficient of the water protons \( D_0 \) and the mean residence time \( \tau_0 \) within a local cage, which indicates the time scale that protons remain in the local cage before jump to another. The jump length of proton translational motion, \( L \), can be determined as

\[ L = \sqrt{6D_0 \tau_0} \] (Eq. 10)

5.4 Results and Discussion

5.4.1 Rotational Motion

We have shown that the motion of water is slowed due to confinement on cellulose microfibril surface. We also showed that both rotational and translational motion become faster with increasing hydration levels. However, it was not clear that whether the increase in dynamics is due to change in mechanism. The rotational SISFs of water protons, \( S_R(Q, t) \), at various temperatures are obtained by removing the translation of water molecule center of mass from the translation of water protons. (Eq. 3) At each temperature, each scattering vector length, and each hydration level (5% or 20%), we obtain fitting parameters \( \tau_R \) and \( \beta_R \). The values of \( \tau_R \) and \( \beta_R \) almost stays constant at
each temperature and water content, i.e. it is relatively independent of scattering vector length. This validates our fitting results as rotational motion should behave the same way when varying the size of the observation window unless it is too small to visualizing rotation of individual water molecules. The average rotational relaxation, $\tau_\beta$, was obtained based on Eq. 7. Rotational motion of water molecules is caused by breaking individual hydrogen bonds between water and water, or between water and cellulose microfibril. $\tau_\beta$ is the evaluation of rotational motion. Both neutron scattering experiments and molecular dynamics simulation have shown that the temperature dependence of the relaxation time of water rotational motion has an Arrhenius behavior in polymer/water solutions\textsuperscript{25} and in water only samples.\textsuperscript{26} We plot the rotational relaxation time $\tau_\beta$ vs. the temperatures, and found that the confined water on cellulose Iα surface also follows Arrhenius behavior (Fig 5.) and is well described by

$$\tau_\beta = A_0 \cdot Exp\left(\frac{E_A}{k_B T}\right)$$

(Eq. 11)

where $A_0$ is a prefactor constant that depend on the hydration level of the system, $E_A$ is the activation energy of the rotational motion at that hydration level, and $k_B$ is the Boltzmann constant.
Fig 5.1. Arrhenius plot of the relaxation rotational time at (a) 5% hydration and (b) 20% hydration.

Based on the fitting equations shown in (Fig 5.1), we can derive the activation energy of the rotational motion. (Table 5.1)

<table>
<thead>
<tr>
<th>Hydration Level</th>
<th>$E_A$ [kJ/mol]</th>
<th>$A_0 \times 10^{-3}$ ps</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>23.5</td>
<td>13.2</td>
</tr>
<tr>
<td>20%</td>
<td>24.4</td>
<td>0.553</td>
</tr>
</tbody>
</table>

The rotation activation energy for bulk water is $7.74 \text{kJ/mol}$, which is more than two fold smaller than that of the confined water on. This implies that the mechanism of bulk water rotation is definitely different from that of the microfibril interfacial water molecules. Suprisingly, even though there is a huge difference between the values of rotational relaxation times at the two hydration levels, the energy barriers for water rotation are only differed by less than 4%. This suggests that the rotation mechanisms of the confined water on cellulose Iα microfibril surface may be independent of the hydration level.
5.4.2 Translational Motion

Similar to rotational motion, the translational SISFs can also be described using KWW functions. (Eq. 6) One of the challenges in analyzing the translational SISFs is that multiple types of translational motion may exist simultaneously in a system, which may form multiple solvation shell on surfaces. In this case, the SISFs are fitted using the summation of multiple KWW equations, each multiplied by a prefactor indicating the fraction of water following that type of translation. (Eq. 5) To determine the existence of multiple translations and calculate the fraction of each type, we generate the distribution of the average travel distance of individual water oxygens in 500ps. We observe that the distribution at 5% hydration level contains a single peak at all temperatures, while the distributions at 20% hydration level are double peaked (Fig 5.2). The double peaked distributions indicate that there are two types of water translational motions in the 20% hydration systems, and are used to accurately determine the fraction of water undergoing each motion.
Fig 5.2. Distribution of average water oxygen travel distance over 500ps at 20% hydration level. The travel displacement of each water oxygen is calculated every 1ps, and the summation of this value over 500ps derives the travel distance of one water oxygen, and the distributions are generated based on the result of all water oxygen atoms in the system. The circle points denote the data points based on the calculation. The solid lines denote the fitting to the data points. The simulations were performed at 253K (green), 273K (blue), 283K (red), and 298K (black).

The travel distance distribution data are fitted as the summation of two Gaussian distributions,

\[
dist = A_1 \exp\left(-\frac{(d-\mu_1)^2}{2\sigma_1^2}\right) + A_2 \exp\left(-\frac{(d-\mu_2)^2}{2\sigma_2^2}\right)
\]  
(Eq. 12)

and we take the integral of the first term and the second term in the above expression over the entire travel distance range, which in turn gives us the fraction of the slow and fast water at each temperature. (Table 5.2) The fitting result shows that the fraction of each translational type is not constant; instead, the fraction of slow water generally increases as temperature decreases.
Table 5.2 Summary of fraction of slow and fast water at each temperature.

<table>
<thead>
<tr>
<th>T[K]</th>
<th>Slow fraction</th>
<th>Fast fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>298</td>
<td>0.42</td>
<td>0.58</td>
</tr>
<tr>
<td>283</td>
<td>0.43</td>
<td>0.57</td>
</tr>
<tr>
<td>273</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>253</td>
<td>0.54</td>
<td>0.46</td>
</tr>
</tbody>
</table>

The fractions are plugged into Eq. 5 to obtain more accurate fitting for relaxation times and stretch factor for both types of translational motions. The average translational motion relaxation time is obtained by Eq. 7, and we in turn calculate the translational relaxation rate by Eq. 8. Fitting the translational relaxation rate as a function of scattering length vector will derive the diffusion coefficient of water protons and the residence time of water protons in their local cages (Eq. 9), and both of these parameters are used when calculating the jump length of water protons between local cages. (Eq. 10) The jump diffusion model fitting curves are reported in Appendix E, and the fitting parameters are summarized in Table 5.3.
Table 5.3. Summary of jump diffusion model fitting results.

<table>
<thead>
<tr>
<th>Hydration Level</th>
<th>T [K]</th>
<th>Fast Water</th>
<th>Slow Water</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$D_0$ [Å²/ps]</td>
<td>$\tau_0$ [ps]</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>298</td>
<td>0.112</td>
<td>6.69</td>
<td>2.12</td>
</tr>
<tr>
<td>283</td>
<td>0.0900</td>
<td>16.1</td>
<td>2.95</td>
</tr>
<tr>
<td>273</td>
<td>0.0716</td>
<td>26.9</td>
<td>3.40</td>
</tr>
<tr>
<td>253</td>
<td>0.0249</td>
<td>40.4</td>
<td>2.46</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>298</td>
<td>0.002829</td>
<td>1813</td>
<td>5.55</td>
</tr>
<tr>
<td>283</td>
<td>0.001576</td>
<td>1678</td>
<td>3.98</td>
</tr>
<tr>
<td>273</td>
<td>0.000934</td>
<td>2960</td>
<td>4.07</td>
</tr>
<tr>
<td>253</td>
<td>0.000593</td>
<td>7623</td>
<td>5.21</td>
</tr>
<tr>
<td>233</td>
<td>0.000234</td>
<td>9093</td>
<td>3.57</td>
</tr>
<tr>
<td>213</td>
<td>0.000123</td>
<td>11087</td>
<td>2.87</td>
</tr>
</tbody>
</table>

Although the trends of temperature dependence of diffusion coefficient and residence time are obvious, we do not see a significant dependence of jump length on temperature. Nonetheless, we do observe that the jump lengths at 5% hydration level are in general greater than that at 20% hydration level. (Fig 5.3) It has been shown that greater jump length is often associated with less frequently occurred jump events and higher energy barriers. This implies that the jump diffusion mechanism may be altered by changing hydration level. We generate the plots of temperature dependence of the diffusion coefficients for each translational motion, and obtain the activation energy of water proton translation by fitting the data to Eq. 11 (Fig 5.4), and the results are summarized in Table 5.4.
Fig 5.3. Jump lengths do not show temperature dependence trend. Dashed lines are the average jump length calculated based on all data points for each translational motion.

Fig 5.4. Temperature dependence of diffusion coefficients. Diffusion coefficients obtained from jump diffusion model are shown as circles. The dashed lines are fittings according to Arrhenius equation. (a) 5% hydration level. (b) 20% hydration level: the fast water diffusion coefficients are colored in blue, and the slow water data are shown in red.

<table>
<thead>
<tr>
<th>Hydration Level</th>
<th>$E_A$ [kJ/mol]</th>
<th>$A_0$ [$\text{Å}^2$/ps]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>19.1</td>
<td>5.36</td>
</tr>
<tr>
<td>20% slow</td>
<td>21.7</td>
<td>90.0</td>
</tr>
<tr>
<td>20% fast</td>
<td>21.4</td>
<td>735</td>
</tr>
</tbody>
</table>
As shown in the table above, the activation energy of different types of translational motions are only differed by less than 6% between each other. The published translation activation energy for bulk water is between 19 to 21 kJ/mol, \(^{29-32}\) which is exceptionally similar to the activation energies of the confined water on cellulose microfibrils. This implies that the mechanism of translation of the confined water may be the same as that of bulk water.

### 5.4.3 Anisotropy Analysis

The both translational motion and rotational motion of the confined water share the same activation energy at difference water contents. We would like to investigate the possible origin that causes this difference in relaxation time and diffusion coefficient values while maintaining the mechanism unchanged. A previous publication by Suzuki et al. \(^{33}\) studied the effective diffusion coefficient and activation energy for deuterated water in compact sodium bentonite. They reported the effective diffusion coefficient in the parallel direction and perpendicular direction to the preferred orientation of the montmorillonite. They observed that the effective diffusion coefficient in the parallel direction is more than 50% greater than that in the normal direction, while the activation energy of translation in both directions are almost equal. \(^{33}\) In other words, it is possible to have anisotropic translational motion while keeping the translation activation energy isotropic.

In our simulation, it is challenging to track the isotropy of the translational motion: it is difficult to define parallel and normal contribution of the translation as there are surfaces exists in all x, y, z directions. We performed this analysis for our rotational
motions to see whether the motion is anisotropic, and whether the anisotropy affects the activation energy of rotation.

The water molecule has three axes of reorientation, i.e. vector HH that connects the two hydrogen atoms, vector $\mu$ that directs along the bisector of the H-O-H bend (corresponding to the direction of water dipole moment), and vector $\perp$ that is perpendicular to the plane of the water molecule. $^{34}$ (Fig 5.5) We can define a unit vector $v$ that parallel to one of the axes and calculate its autocorrelation function as

$$C(t) = \langle v(0) \cdot v(t) \rangle$$  \hspace{1cm} (Eq. 13)

Fig 5.5. Local coordinate frame of the water molecule. Modified based on Reference 34.

Similar to the SISFs, the autocorrelation function can be fitted with a KWW stretched exponential function. By integrating the exponential function, we could obtain the relaxation time of reorientation of water molecules in each direction, which is summarized in Fig 5.6.
Based on the rotational SISF fitting, we observed that the rotational motion of water at both 5% and 20% hydration levels are confined (significantly slower than bulk water). The autocorrelation function fitting shows that the relaxation times of reorientation are not equal in the three axes, and this indicates that the water reorientation is anisotropic at both 5% and 20% hydration levels. As expected, the extent of anisotropy increases with decreasing hydration level. In both cases, $\tau_\mu$ is the largest among the relaxation times at the same temperature. At 20% hydration level, relaxation time of dipole direction is about twice as much as that of other two axes. At 5% hydration level, this difference increased to almost ten folds. Similar to previous analysis, we fit the temperature dependence of
relaxation times to Arrhenius equation (Fig 5.7), and the activation energy data are summarized in Table 5.5.

![Figure 5.7](image)

**Fig 5.7.** Temperature dependence of relaxation time of each rotational orientation axis. The 5% hydration level data are shown in diamond shape data points, and the 20% hydration level data are shown in circle. $\tau_{HH}$, $\tau_\mu$, and $\tau_\perp$ are color coated in red, blue, and orange, respectively.

**Table 5.5** Comparison of activation energy of rotation in the reorientation axes. $E_A$ is the activation energy of overall rotational motion obtained from SISF fittings.

<table>
<thead>
<tr>
<th>Hydration Level</th>
<th>$E_{A,HH}$ [kJ/mol]</th>
<th>$E_{A,\mu}$ [kJ/mol]</th>
<th>$E_{A,\perp}$ [kJ/mol]</th>
<th>$E_A$ [kJ/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>27.3</td>
<td>24.9</td>
<td>28.5</td>
<td>23.5</td>
</tr>
<tr>
<td>20%</td>
<td>23.5</td>
<td>25.8</td>
<td>24.1</td>
<td>24.4</td>
</tr>
</tbody>
</table>

Although the rotational motion is anisotropic, and the extent of anisotropy differs in the reorientation axes, the activation energies of rotation in the different reorientation axes are almost identical, and agree well with the previous value obtained from the rotational SISF fitting. This observation is consistent with the previous publication, 33
from which we could concluded that the large differences in the relaxation times and
diffusion coefficients may be contributed from the anisotropy extend of the motion
without modifying its mechanism.

5.5 Conclusion

We studied the interfacial water molecules among cellulose Iα microfibril surfaces at two
hydration levels and various temperatures. The dynamics of water molecules is
significantly confined upon presence of the microfibril surfaces, resulting in much lower
rotation relaxation times, greater translation local cage residence times, and smaller
diffusion coefficients comparing to bulk water properties. As the hydration level
decreases, both rotation and translation motion are slowed further. To determine whether
the mechanisms of the motions are changed due to confinement, we calculated the
activation energies of translational motion and rotational motion. We did not observe
dependence of activation energy on hydration level for both types of motions. However,
our results indicates that the activation energy of translation is very similar to that of bulk
water, while the activation energy of rotation is about three times smaller for confined
water. This implies that the mechanism of translation of confined water may be the same
as bulk water, while their rotation mechanisms are surely different.

The motions of bulk water are viewed as isotropic motions, while we found that
this is not the case in confined water. To determine why we observe great difference in
the rotation relaxation time at 5% and 20% hydration levels and only see minimal
differences in the rotation activation energy, we calculated the rotation relaxation times at
three rotational axes, namely HH axes, dipole direction axes, and OH axes. At both hydration levels, the relaxation time in the dipole direction axes is much greater than those in the other two axes. Therefore, the rotational motion of confined water is anisotropic, and the extent of anisotropy increases with decreasing hydration level. Nonetheless, the activation energies of rotation at all three axes are in great agreement with that of the overall rotation activation energy. From this, we concluded that the higher relaxation times of rotational motion at low hydration levels are caused by the anisotropy of the motions. It has been reported that water molecules (~1wt%) in a cellulose-PVA blend can undergo rapid anisotropic flips around the dipole axis. 35 This flip motion also results in a much greater relaxation in dipole direction compared to HH and OH directions. Therefore, the confined water at the presence of cellulose surface may share the same rotation mechanism.

The above results show that the dynamics of confined water greatly differ from those of bulk liquid water. Therefore, one cannot simply ignore water molecules when studying biochemical processes. Especially in the case of low hydration levels, the properties of water are strongly altered.
5.6 Reference


(32) Khakimov, a. M.; Rudakova, M. a.; Doroginitskii, M. M.; Filippov, a. V.

   *Biophysics (Oxf).* **2008**, *53*, 147–152.


Chapter 6 Conclusion

In this dissertation, the structural and dynamic properties of cell wall components are investigated using computational approaches. We present coarse-grained force fields for cellulose and for (XXXG) xyloglucan, and focus on the structural properties of cell wall load-bearing components in Chapter 2 and 3. The coarse-grained force fields are applied for simulation full length fibrils and xyloglucan chains which provides molecular level detail of the interactions and structure features of these two components. As the primary cell wall is largely occupied by water, the dynamic properties of water at the cellulose microfibril interface are studied and we show that the water and microfibril mutually influence each other in the cell wall. (Chapter 4) Additionally, we examine whether the mechanism of water motion is influenced by confinement or hydration level. (Chapter 5) We find that the rotation motion mechanism is altered upon hydration while this may not be true for translational motion.

6.1 Cell Wall Load-Bearing Network Structure

The network formed by cellulose microfibril and xyloglucan has been identified as the load-bearing network of the cell wall. ¹ To understand the molecular level detail of the interaction of the two components and investigate their individual structural features, we developed a coarse-grained simulation for each polysaccharide. Different from atomistic simulation, the force sites in the coarse-grained simulation are defined a group of atoms.
The interactions are only defined between the force sites, thus, the computational cost of coarse-grained simulation is much less than that of atomistic simulation. We used atomistic simulation of shorter polysaccharide segment as the target system. The force field parameters are generated based on iterative Boltzmann Inversion \(^2\) by requiring consistency of the intrachain and interchain configurations between the atomistic model and the coarse-grained model. Note that in cellulose microfibril, the glucose units may hydrogen bond with their interchain neighbors as a result of the crystalline structure, to maintain the directionality of the interchain bonds, we defined pairwise interchain potentials between glucose beads and its neighbors in the \(\mathbf{a}\)-\(\mathbf{b}\) plane, and added the potentials sequentially to the force field until the pair distribution function of the interchain force sites is matched with the target. We conduct a simulation of microfibrils with lengths up to 200nm, and observe that microfibrils longer than 100nm tends to form kinks along their longitudinal direction. We think that this kink may be related to the periodic disorder feature of the microfibril as suggested by small angle neutron scattering measurements on ramie. \(^3\)

The coarse-grained force field of xyloglucan chains is used in combination with the cellulose force field. We simulate one 6x6x200 microfibril with 20 chains of \((XXXG)_{50}\) xyloglucan chains randomly placed in the simulation box. The simulation is performed \textit{in vacuo} with 6Å spacing between the side of microfibril and the box, which corresponding to a 15Å interfibril distance between the cellulose microfibril and its periodic images. The resulting xyloglucan chains can be characterized into three fractions based on their interaction with cellulose microfibril: the bridging chains, which interact...
with two microfibrils (the cellulose microfibril in the simulation box and its neighboring periodic image); the single chains, which interact with only one microfibril; the isolated chains, which do not directly interact with the cellulose microfibril or its periodic image. In many cell wall models, the bridging xyloglucan chains are draw as extended chains along two fibrils. Based on our simulation results, we can conclude that this is not an accurate description for the bridging regions. In fact, most of the xyloglucan chains in the bridging regions we observe form coil-like structures. In addition, the single chains and the isolated chains may also contribute to the strength of the cell wall structural network. We observe that the bridge regions may be formed by multiple xyloglucan chains coiled together. Therefore, the chains that are not in directly contact with cellulose microfibril may partially serve as the bridge, which in turn enhances the mechanical property of the cell wall.

The coarse-grained simulation also provides significant insights about the formation of the xyloglucan cellulose interaction. The simulation shows several bridging chains or single chains that do not initially interacting with the microfibril. We monitored the behavior of the xyloglucan chains along the progress of the simulation, and observed that upon contact of xyloglucan and cellulose microfibril, the interacting part between the two polysaccharides extends along the longitudinal direction of the microfibril. This agrees with the atomistic simulation result which shows that interaction between microfibril and xyloglucan parallel to longitudinal direction is more energetically favorable comparing to interacting perpendicular to the longitudinal direction. We think that this feature of interaction sites extending along the microfibril makes the
conformation of cellulose microfibril and xyloglucan alike, and provide a potential interacting site between microfibrils to form biological “hot spots”.

6.2 Dynamic Properties

Water can influence the mechanical properties of cellulose microfibril by varying the hydration level. Thus, in order to fully understand the properties of the cell wall network structure, the role of water should be considered carefully. We studied the dynamic properties of water at 5% and 20% hydration levels by fitting self-intermediate scattering function using a stretched exponential model. A unique feature of molecular dynamics simulation is that it enables one to decompose the motion of translation completely from the self-intermediate scattering function by fixing the center of mass of each water molecule. In this way, the contribution of rotation and translation to the scattering function can be decoupled and analyzed separately. We determine the translational diffusion coefficient, jump length, and the residence time of water proton within the local cage. Multiple types of translational motion are observed in 20% hydrated system. The motions of water on cellulose microfibril is significantly slowed comparing to bulk water properties indicating that water cannot be omitted when studying the cell wall properties. The analysis on scattering function is then performed on simulation at various temperatures. The activation energy of rotation and translation are obtained using Arrhenius plots of relaxation time of rotation and diffusion coefficient of translation, respectively. Although the translation activation energies of the bulk water and the confined water on cellulose microfibril are comparable, the rotational activation
energy of the confined water is much higher, indicating that confinement induces change in rotational motion mechanism. Even though the activation energy of the confined water rotation is independent of hydration level, the rotation motion at 5% hydration is significantly smaller than that of 20% hydration. To determine the origin of the dependence of water rotation on hydration level, we analyzed the autocorrelation function of the relaxation at each rotation principle axis. We conclude that the difference in the values of rotation relaxation times of the confined water is due to the difference in their extent of anisotropy.

6.3 Summary

The major conclusions and achievements of this dissertation are:

- We derive a coarse-grained force field for cellulose microfibril and xyloglucan that can be used for simulating a piece of cell wall network structure. (Chapter 2 and 3)

- Cellulose microfibrils longer than 100nm tends to form kinks along their longitudinal directions. (Chapter 2)

- Xyloglucan in the network can be characterized into bridging chains, single chains, and isolated chains, based on their interaction with cellulose microfibrils. (Chapter 3)

- Xyloglucan that does not directly bridging two microfibrils may form bridge-like structure with other xyloglucan chains, and may also contribute to the mechanical strength of the wall. (Chapter 3)
• The interaction between cellulose and xyloglucan extends along the longitudinal direction of the microfibril. (Chapter 3)

• Microfibrils and its surrounding water in plant cell mutually influence each other. (Chapter 4)

• Water is slowed by the presence of cellulose microfibril surface, and the motion becomes slower at lower water content. (Chapter 4 and 5)

• The slowed motion upon altering hydration level is due to change induced in anisotropy of the water molecules. (Chapter 5)

• The rotational mechanism in the confined water is altered comparing to bulk water, while this may not be true for translation. (Chapter 5)

6.4 Recommendations for Future Study

The current coarse-grained cellulose force field requires pre-defined intermolecular potential between pairs of glucose beads. This definition retains the directionality of the hydrogen bonds, but it cannot recognize hydrogen bond reformation once the bond is broken due to chain slippage. Therefore, improving the coarse-grained force field using a non-pair-specific intermolecular bond is desired. One method to define non-pairwise intermolecular bond is to add directionality to each coarse-grained bead. In this way, the bead in individual chains can align in alternating fashion as a result of the β-(1,4)-linkage (Fig 6.1a), and the intermolecular point dipoles will be aligned parallel to each other. (Fig 6.1b) In this case, the strength of the intermolecular bond is evaluated based on the angle between the point dipole and the position vector and the distance between the
neighboring interchain beads as described by $\theta$ and $r$ respectively in Fig 6.1c. This definition of intermolecular potential in the cellulose microfibril allows the force field to be applied to study the self-assembly of the cellulose chains themselves, and the microfibril/xyloglucan assemble process.

![Fig 6.1. Cellulose coarse-graining](image)

The current xyloglucan model is developed based on (XXXG) motifs. The xylose branch in xyloglucan chains can be further substituted by galactose and fucose in plant cell walls. Atomistic simulation on xyloglucan segments with variation in side chains have shown that, comparing to chains built by XXXG motifs, it is more difficult for the segments with longer side chains to adopt flat conformation on cellulose microfibril. The equilibrium structure of the adsorbed xyloglucan is also related to the xyloglucan side chain. Coarse-grained simulation on long xyloglucan chains with varying side group is essential for identifying the preferred interaction sites between cellulose and xyloglucan. Furthermore, the current simulation is conducted in vacuo. As water and cell
wall polysaccharides can mutually influences each other, it is desired to conduct simulations with solvated polysaccharides.

As shown in Chapter 4 and 5, water molecules surrounded in the interfacial region between cellulose microfibrils are highly confined. The current water model developed by He et al. \(^8\) is based on bulk properties of water, which may not provide an accurate enough description for the interfacial water presented in the simulation of cell wall polysaccharides. Therefore, it is desired to obtain a water coarse-grained force field that can induce the influence of water on the microfibril. For example, inducing change in mechanical properties upon addition of water, and show slowed dynamics near cellulose microfibril surfaces. This can provide significant insights on whether the slowed interfacial water molecules can serve as a shield to prevent interaction between cellulose and other macromolecules such as protein.

6.5 Reference


Appendix A LAMMPS Input Script

A.1 Cellulose Microfibril Coarse-Grained Simulation Input Script

```plaintext
units real
dimension 3
neigh_modify delay 0 every 1 check yes
atom_style full
bond_style hybrid harmonic morse
angle_style harmonic
improper_style harmonic
pair_style lj/cut 15.0
time_step 1
special_bonds lj/cut 0.0 1.0 1.0
```

A.2 Microfibril-Xyloglucan Network Coarse-Grained Simulation Input Script

```plaintext
units real
dimension 3
neigh_modify delay 0 every 1 check yes
atom_style full
bond_style hybrid harmonic morse
angle_style harmonic table linear 87
dihedral_style table linear 120
improper_style harmonic
pair_style hybrid lj/cut 15.0 lj/cut 15.0
time_step 1
special_bonds lj/cut 0.0 1.0 1.0
```
Appendix B Tabulated Potentials of Coarse-Grained Xyloglucan Simulation

B.1 Tabulated Angle Potential

<table>
<thead>
<tr>
<th>GGGAngle</th>
<th>GGXangle</th>
</tr>
</thead>
<tbody>
<tr>
<td>N 87 FP 0 0</td>
<td>N 87 FP 0 0</td>
</tr>
<tr>
<td>1 0</td>
<td>6.0039</td>
</tr>
<tr>
<td>2 51</td>
<td>0.72116</td>
</tr>
<tr>
<td>3 52.5</td>
<td>0.099787</td>
</tr>
<tr>
<td>4 54</td>
<td>0.39885</td>
</tr>
<tr>
<td>5 55.5</td>
<td>0.10132</td>
</tr>
<tr>
<td>6 57</td>
<td>0.50515</td>
</tr>
<tr>
<td>7 58.5</td>
<td>0.83793</td>
</tr>
<tr>
<td>8 60</td>
<td>0.5486</td>
</tr>
<tr>
<td>9 61.5</td>
<td>0.91876</td>
</tr>
<tr>
<td>10 63</td>
<td>1.3548</td>
</tr>
<tr>
<td>11 64.5</td>
<td>1.6767</td>
</tr>
<tr>
<td>12 66</td>
<td>2.195</td>
</tr>
<tr>
<td>13 67.5</td>
<td>2.4852</td>
</tr>
<tr>
<td>14 69</td>
<td>1.8512</td>
</tr>
<tr>
<td>15 70.5</td>
<td>1.5884</td>
</tr>
<tr>
<td>16 72</td>
<td>1.2898</td>
</tr>
<tr>
<td>17 73.5</td>
<td>1.3121</td>
</tr>
<tr>
<td>18 75</td>
<td>1.248</td>
</tr>
<tr>
<td>19 76.5</td>
<td>1.1195</td>
</tr>
<tr>
<td>20 78</td>
<td>0.99115</td>
</tr>
<tr>
<td>21 79.5</td>
<td>0.94894</td>
</tr>
<tr>
<td>22 81</td>
<td>1.3713</td>
</tr>
<tr>
<td>23 82.5</td>
<td>1.3377</td>
</tr>
<tr>
<td>24 84</td>
<td>1.4152</td>
</tr>
<tr>
<td>25 85.5</td>
<td>1.1979</td>
</tr>
<tr>
<td>26 87</td>
<td>1.0453</td>
</tr>
<tr>
<td>27 88.5</td>
<td>1.0079</td>
</tr>
<tr>
<td>28 90</td>
<td>1.0614</td>
</tr>
<tr>
<td>29 91.5</td>
<td>0.9955</td>
</tr>
<tr>
<td>30 93</td>
<td>1.0889</td>
</tr>
<tr>
<td>31 94.5</td>
<td>1.0515</td>
</tr>
<tr>
<td>32 96</td>
<td>0.84514</td>
</tr>
<tr>
<td>33 97.5</td>
<td>0.6905</td>
</tr>
<tr>
<td>34 99</td>
<td>0.5802</td>
</tr>
<tr>
<td>35 100.5</td>
<td>0.59515</td>
</tr>
<tr>
<td></td>
<td>36</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
</tr>
<tr>
<td>37</td>
<td>103.5</td>
</tr>
<tr>
<td>38</td>
<td>105</td>
</tr>
<tr>
<td>39</td>
<td>106.5</td>
</tr>
<tr>
<td>40</td>
<td>108</td>
</tr>
<tr>
<td>41</td>
<td>109.5</td>
</tr>
<tr>
<td>42</td>
<td>111</td>
</tr>
<tr>
<td>43</td>
<td>112.5</td>
</tr>
<tr>
<td>44</td>
<td>114</td>
</tr>
<tr>
<td>45</td>
<td>115.5</td>
</tr>
<tr>
<td>46</td>
<td>117</td>
</tr>
<tr>
<td>47</td>
<td>118.5</td>
</tr>
<tr>
<td>48</td>
<td>120</td>
</tr>
<tr>
<td>49</td>
<td>121.5</td>
</tr>
<tr>
<td>50</td>
<td>123</td>
</tr>
<tr>
<td>51</td>
<td>124.5</td>
</tr>
<tr>
<td>52</td>
<td>126</td>
</tr>
<tr>
<td>53</td>
<td>127.5</td>
</tr>
<tr>
<td>54</td>
<td>129</td>
</tr>
<tr>
<td>55</td>
<td>130.5</td>
</tr>
<tr>
<td>56</td>
<td>132</td>
</tr>
<tr>
<td>57</td>
<td>133.5</td>
</tr>
<tr>
<td>58</td>
<td>135</td>
</tr>
<tr>
<td>59</td>
<td>136.5</td>
</tr>
<tr>
<td>60</td>
<td>138</td>
</tr>
<tr>
<td>61</td>
<td>139.5</td>
</tr>
<tr>
<td>62</td>
<td>141</td>
</tr>
<tr>
<td>63</td>
<td>142.5</td>
</tr>
<tr>
<td>64</td>
<td>144</td>
</tr>
<tr>
<td>65</td>
<td>145.5</td>
</tr>
<tr>
<td>66</td>
<td>147</td>
</tr>
<tr>
<td>67</td>
<td>148.5</td>
</tr>
<tr>
<td>68</td>
<td>150</td>
</tr>
<tr>
<td>69</td>
<td>151.5</td>
</tr>
<tr>
<td>70</td>
<td>153</td>
</tr>
<tr>
<td>71</td>
<td>154.5</td>
</tr>
<tr>
<td>72</td>
<td>156</td>
</tr>
<tr>
<td>73</td>
<td>157.5</td>
</tr>
<tr>
<td>74</td>
<td>159</td>
</tr>
<tr>
<td>75</td>
<td>160.5</td>
</tr>
<tr>
<td>76</td>
<td>162</td>
</tr>
<tr>
<td>77</td>
<td>163.5</td>
</tr>
<tr>
<td>78</td>
<td>165</td>
</tr>
<tr>
<td>79</td>
<td>166.5</td>
</tr>
<tr>
<td>80</td>
<td>168</td>
</tr>
<tr>
<td>81</td>
<td>169.5</td>
</tr>
<tr>
<td>82</td>
<td>171</td>
</tr>
<tr>
<td>83</td>
<td>172.5</td>
</tr>
<tr>
<td>84</td>
<td>174</td>
</tr>
<tr>
<td>85</td>
<td>175.5</td>
</tr>
<tr>
<td>86</td>
<td>177</td>
</tr>
<tr>
<td>87</td>
<td>180</td>
</tr>
</tbody>
</table>
B.2 Tabulated Dihedral Potential

<table>
<thead>
<tr>
<th>#</th>
<th>GGGGtorsion</th>
<th>GGGXtorsion</th>
<th>XGGXtorsion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N 120 DEGREES NOF</td>
<td>N 120 DEGREES NOF</td>
<td>N 120 DEGREES NOF</td>
</tr>
<tr>
<td>1</td>
<td>-180 1.7098</td>
<td>-180 0.78629</td>
<td>-180 -0.56707</td>
</tr>
<tr>
<td>2</td>
<td>-177 2.5878</td>
<td>-177 0.61742</td>
<td>-177 1.1098</td>
</tr>
<tr>
<td>3</td>
<td>-174 2.6428</td>
<td>-174 1.0028</td>
<td>-174 2.4353</td>
</tr>
<tr>
<td>4</td>
<td>-171 2.2264</td>
<td>-171 1.0191</td>
<td>-171 1.5465</td>
</tr>
<tr>
<td>5</td>
<td>-168 1.9404</td>
<td>-168 1.1007</td>
<td>-168 1.2795</td>
</tr>
<tr>
<td>6</td>
<td>-165 2.2511</td>
<td>-165 1.1768</td>
<td>-165 1.5367</td>
</tr>
<tr>
<td>7</td>
<td>-162 2.2382</td>
<td>-162 1.4759</td>
<td>-162 0.40622</td>
</tr>
<tr>
<td>8</td>
<td>-159 2.2742</td>
<td>-159 1.4227</td>
<td>-159 0.3113</td>
</tr>
<tr>
<td>9</td>
<td>-156 2.3628</td>
<td>-156 1.9447</td>
<td>-156 1.0049</td>
</tr>
<tr>
<td>10</td>
<td>-153 2.4386</td>
<td>-153 1.5543</td>
<td>-153 1.3073</td>
</tr>
<tr>
<td>11</td>
<td>-150 2.4808</td>
<td>-150 1.3125</td>
<td>-150 0.84667</td>
</tr>
<tr>
<td>12</td>
<td>-147 2.269</td>
<td>-147 1.7592</td>
<td>-147 0.82805</td>
</tr>
<tr>
<td>13</td>
<td>-144 2.1318</td>
<td>-144 1.608</td>
<td>-144 0.68498</td>
</tr>
<tr>
<td>14</td>
<td>-141 1.6929</td>
<td>-141 1.5248</td>
<td>-141 1.1507</td>
</tr>
<tr>
<td>15</td>
<td>-138 2.0477</td>
<td>-138 1.7646</td>
<td>-138 1.2681</td>
</tr>
<tr>
<td>16</td>
<td>-135 2.1729</td>
<td>-135 1.9416</td>
<td>-135 0.8501</td>
</tr>
<tr>
<td>17</td>
<td>-132 2.1865</td>
<td>-132 1.6201</td>
<td>-132 1.1697</td>
</tr>
<tr>
<td>18</td>
<td>-129 2.4586</td>
<td>-129 1.7717</td>
<td>-129 0.9316</td>
</tr>
<tr>
<td>19</td>
<td>-126 2.2167</td>
<td>-126 2.0977</td>
<td>-126 0.77495</td>
</tr>
<tr>
<td>20</td>
<td>-123 2.4391</td>
<td>-123 2.0202</td>
<td>-123 0.71992</td>
</tr>
<tr>
<td>21</td>
<td>-120 2.2643</td>
<td>-120 2.0271</td>
<td>-120 0.67794</td>
</tr>
<tr>
<td>22</td>
<td>-117 2.2238</td>
<td>-117 2.2842</td>
<td>-117 0.65123</td>
</tr>
<tr>
<td>23</td>
<td>-114 1.8488</td>
<td>-114 2.1759</td>
<td>-114 0.41883</td>
</tr>
<tr>
<td>24</td>
<td>-111 2.1435</td>
<td>-111 2.0726</td>
<td>-111 0.84194</td>
</tr>
<tr>
<td>25</td>
<td>-108 1.9356</td>
<td>-108 2.3997</td>
<td>-108 0.98178</td>
</tr>
<tr>
<td>26</td>
<td>-105 1.9688</td>
<td>-105 2.2888</td>
<td>-105 0.95113</td>
</tr>
<tr>
<td>27</td>
<td>-102 1.9704</td>
<td>-102 2.2692</td>
<td>-102 0.94933</td>
</tr>
<tr>
<td>28</td>
<td>-99 1.8774</td>
<td>-99 2.357</td>
<td>-99 0.84198</td>
</tr>
<tr>
<td>29</td>
<td>-96 1.9629</td>
<td>-96 2.2389</td>
<td>-96 1.1202</td>
</tr>
<tr>
<td>30</td>
<td>-93 1.8311</td>
<td>-93 2.3463</td>
<td>-93 1.4975</td>
</tr>
<tr>
<td>31</td>
<td>-90 1.7244</td>
<td>-90 2.2803</td>
<td>-90 1.5945</td>
</tr>
<tr>
<td>32</td>
<td>-87 1.8316</td>
<td>-87 1.9981</td>
<td>-87 1.2985</td>
</tr>
<tr>
<td>33</td>
<td>-84 1.771</td>
<td>-84 1.7682</td>
<td>-84 1.2537</td>
</tr>
<tr>
<td>34</td>
<td>-81 1.8785</td>
<td>-81 1.7373</td>
<td>-81 1.2344</td>
</tr>
<tr>
<td>35</td>
<td>-78 1.7319</td>
<td>-78 1.4853</td>
<td>-78 1.1582</td>
</tr>
<tr>
<td>36</td>
<td>-75 1.7362</td>
<td>-75 1.418</td>
<td>-75 1.4091</td>
</tr>
<tr>
<td>37</td>
<td>-72 1.7463</td>
<td>-72 1.3009</td>
<td>-72 1.0915</td>
</tr>
<tr>
<td>38</td>
<td>-69 1.7136</td>
<td>-69 1.202</td>
<td>-69 1.638</td>
</tr>
<tr>
<td>39</td>
<td>-66 1.7568</td>
<td>-66 1.2588</td>
<td>-66 1.3698</td>
</tr>
<tr>
<td>40</td>
<td>-63 1.6002</td>
<td>-63 1.2775</td>
<td>-63 2.0102</td>
</tr>
<tr>
<td>41</td>
<td>-60 1.5457</td>
<td>-60 1.3998</td>
<td>-60 1.9058</td>
</tr>
<tr>
<td>42</td>
<td>-57 1.6067</td>
<td>-57 1.3601</td>
<td>-57 2.3398</td>
</tr>
<tr>
<td>43</td>
<td>-54 1.3272</td>
<td>-54 1.2338</td>
<td>-54 2.0793</td>
</tr>
<tr>
<td>44</td>
<td>-51 1.2819</td>
<td>-51 1.2724</td>
<td>-51 1.8144</td>
</tr>
<tr>
<td>45</td>
<td>-48 1.2735</td>
<td>-48 1.2134</td>
<td>-48 1.5954</td>
</tr>
<tr>
<td>46</td>
<td>-45 1.2719</td>
<td>-45 1.0545</td>
<td>-45 1.2241</td>
</tr>
<tr>
<td>47</td>
<td>-42 1.3497</td>
<td>-42 1.0053</td>
<td>-42 1.1188</td>
</tr>
<tr>
<td>48</td>
<td>-59</td>
<td>1.3582</td>
<td>48</td>
</tr>
<tr>
<td>49</td>
<td>-36</td>
<td>1.4388</td>
<td>49</td>
</tr>
<tr>
<td>50</td>
<td>-33</td>
<td>0.86164</td>
<td>50</td>
</tr>
<tr>
<td>51</td>
<td>-30</td>
<td>1.0986</td>
<td>51</td>
</tr>
<tr>
<td>52</td>
<td>-27</td>
<td>1.1038</td>
<td>52</td>
</tr>
<tr>
<td>53</td>
<td>-24</td>
<td>0.48703</td>
<td>53</td>
</tr>
<tr>
<td>54</td>
<td>-21</td>
<td>0.35716</td>
<td>54</td>
</tr>
<tr>
<td>55</td>
<td>-18</td>
<td>0.39346</td>
<td>55</td>
</tr>
<tr>
<td>56</td>
<td>-15</td>
<td>0.97113</td>
<td>56</td>
</tr>
<tr>
<td>57</td>
<td>-12</td>
<td>0.92468</td>
<td>57</td>
</tr>
<tr>
<td>58</td>
<td>-9</td>
<td>0.92848</td>
<td>58</td>
</tr>
<tr>
<td>59</td>
<td>-6</td>
<td>0.48127</td>
<td>59</td>
</tr>
<tr>
<td>60</td>
<td>-3</td>
<td>1.0938</td>
<td>60</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>-0.60963</td>
<td>61</td>
</tr>
<tr>
<td>62</td>
<td>3</td>
<td>0.0062316</td>
<td>62</td>
</tr>
<tr>
<td>63</td>
<td>6</td>
<td>0.54626</td>
<td>63</td>
</tr>
<tr>
<td>64</td>
<td>9</td>
<td>0.52597</td>
<td>64</td>
</tr>
<tr>
<td>65</td>
<td>12</td>
<td>0.52721</td>
<td>65</td>
</tr>
<tr>
<td>66</td>
<td>15</td>
<td>0.7327</td>
<td>66</td>
</tr>
<tr>
<td>67</td>
<td>18</td>
<td>0.84815</td>
<td>67</td>
</tr>
<tr>
<td>68</td>
<td>21</td>
<td>0.90041</td>
<td>68</td>
</tr>
<tr>
<td>69</td>
<td>24</td>
<td>0.7488</td>
<td>69</td>
</tr>
<tr>
<td>70</td>
<td>27</td>
<td>0.70217</td>
<td>70</td>
</tr>
<tr>
<td>71</td>
<td>30</td>
<td>0.4787</td>
<td>71</td>
</tr>
<tr>
<td>72</td>
<td>33</td>
<td>0.52121</td>
<td>72</td>
</tr>
<tr>
<td>73</td>
<td>36</td>
<td>0.73826</td>
<td>73</td>
</tr>
<tr>
<td>74</td>
<td>39</td>
<td>0.98369</td>
<td>74</td>
</tr>
<tr>
<td>75</td>
<td>42</td>
<td>1.0626</td>
<td>75</td>
</tr>
<tr>
<td>76</td>
<td>45</td>
<td>1.2382</td>
<td>76</td>
</tr>
<tr>
<td>77</td>
<td>48</td>
<td>1.4848</td>
<td>77</td>
</tr>
<tr>
<td>78</td>
<td>51</td>
<td>1.3411</td>
<td>78</td>
</tr>
<tr>
<td>79</td>
<td>54</td>
<td>1.2637</td>
<td>79</td>
</tr>
<tr>
<td>80</td>
<td>57</td>
<td>1.3198</td>
<td>80</td>
</tr>
<tr>
<td>81</td>
<td>60</td>
<td>1.3646</td>
<td>81</td>
</tr>
<tr>
<td>82</td>
<td>63</td>
<td>1.4003</td>
<td>82</td>
</tr>
<tr>
<td>83</td>
<td>66</td>
<td>1.3186</td>
<td>83</td>
</tr>
<tr>
<td>84</td>
<td>69</td>
<td>1.5154</td>
<td>84</td>
</tr>
<tr>
<td>85</td>
<td>72</td>
<td>1.3795</td>
<td>85</td>
</tr>
<tr>
<td>86</td>
<td>75</td>
<td>1.6958</td>
<td>86</td>
</tr>
<tr>
<td>87</td>
<td>78</td>
<td>1.6779</td>
<td>87</td>
</tr>
<tr>
<td>88</td>
<td>81</td>
<td>1.605</td>
<td>88</td>
</tr>
<tr>
<td>89</td>
<td>84</td>
<td>1.9691</td>
<td>89</td>
</tr>
<tr>
<td>90</td>
<td>87</td>
<td>1.8618</td>
<td>90</td>
</tr>
<tr>
<td>91</td>
<td>90</td>
<td>1.9619</td>
<td>91</td>
</tr>
<tr>
<td>92</td>
<td>93</td>
<td>1.9083</td>
<td>92</td>
</tr>
<tr>
<td>93</td>
<td>96</td>
<td>2.0882</td>
<td>93</td>
</tr>
<tr>
<td>94</td>
<td>99</td>
<td>2.0679</td>
<td>94</td>
</tr>
<tr>
<td>95</td>
<td>102</td>
<td>2.2996</td>
<td>95</td>
</tr>
<tr>
<td>96</td>
<td>105</td>
<td>2.1458</td>
<td>96</td>
</tr>
<tr>
<td>97</td>
<td>108</td>
<td>2.5344</td>
<td>97</td>
</tr>
<tr>
<td>98</td>
<td>111</td>
<td>2.3655</td>
<td>98</td>
</tr>
<tr>
<td>99</td>
<td>114</td>
<td>2.2325</td>
<td>99</td>
</tr>
<tr>
<td>100</td>
<td>117</td>
<td>2.5201</td>
<td>100</td>
</tr>
<tr>
<td>101</td>
<td>120</td>
<td>2.3506</td>
<td>101</td>
</tr>
<tr>
<td>102</td>
<td>123</td>
<td>2.6338</td>
<td>102</td>
</tr>
<tr>
<td>103</td>
<td>126</td>
<td>2.572</td>
<td>103</td>
</tr>
</tbody>
</table>

120
<table>
<thead>
<tr>
<th>104</th>
<th>129</th>
<th>2.8367</th>
<th>104</th>
<th>129</th>
<th>1.7519</th>
<th>104</th>
<th>129</th>
<th>2.4071</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>132</td>
<td>2.6067</td>
<td>105</td>
<td>132</td>
<td>1.7279</td>
<td>105</td>
<td>132</td>
<td>1.9757</td>
</tr>
<tr>
<td>106</td>
<td>135</td>
<td>2.8293</td>
<td>106</td>
<td>135</td>
<td>2.0116</td>
<td>106</td>
<td>135</td>
<td>1.5796</td>
</tr>
<tr>
<td>107</td>
<td>138</td>
<td>2.9861</td>
<td>107</td>
<td>138</td>
<td>2.105</td>
<td>107</td>
<td>138</td>
<td>1.3022</td>
</tr>
<tr>
<td>108</td>
<td>141</td>
<td>2.8528</td>
<td>108</td>
<td>141</td>
<td>1.7767</td>
<td>108</td>
<td>141</td>
<td>1.0298</td>
</tr>
<tr>
<td>109</td>
<td>144</td>
<td>3.0352</td>
<td>109</td>
<td>144</td>
<td>1.2731</td>
<td>109</td>
<td>144</td>
<td>1.4908</td>
</tr>
<tr>
<td>110</td>
<td>147</td>
<td>2.8385</td>
<td>110</td>
<td>147</td>
<td>1.9304</td>
<td>110</td>
<td>147</td>
<td>1.8237</td>
</tr>
<tr>
<td>111</td>
<td>150</td>
<td>3.1588</td>
<td>111</td>
<td>150</td>
<td>2.3381</td>
<td>111</td>
<td>150</td>
<td>1.4138</td>
</tr>
<tr>
<td>112</td>
<td>153</td>
<td>2.8013</td>
<td>112</td>
<td>153</td>
<td>2.763</td>
<td>112</td>
<td>153</td>
<td>1.1091</td>
</tr>
<tr>
<td>113</td>
<td>156</td>
<td>3.0279</td>
<td>113</td>
<td>156</td>
<td>1.9539</td>
<td>113</td>
<td>156</td>
<td>0.65183</td>
</tr>
<tr>
<td>114</td>
<td>159</td>
<td>1.8611</td>
<td>114</td>
<td>159</td>
<td>1.1976</td>
<td>114</td>
<td>159</td>
<td>0.54585</td>
</tr>
<tr>
<td>115</td>
<td>162</td>
<td>1.6378</td>
<td>115</td>
<td>162</td>
<td>2.118</td>
<td>115</td>
<td>162</td>
<td>0.47345</td>
</tr>
<tr>
<td>116</td>
<td>165</td>
<td>2.8963</td>
<td>116</td>
<td>165</td>
<td>2.0756</td>
<td>116</td>
<td>165</td>
<td>1.2052</td>
</tr>
<tr>
<td>117</td>
<td>168</td>
<td>2.8622</td>
<td>117</td>
<td>168</td>
<td>1.1528</td>
<td>117</td>
<td>168</td>
<td>0.45246</td>
</tr>
<tr>
<td>118</td>
<td>171</td>
<td>2.2998</td>
<td>118</td>
<td>171</td>
<td>0.94183</td>
<td>118</td>
<td>171</td>
<td>0.34495</td>
</tr>
<tr>
<td>119</td>
<td>174</td>
<td>1.4254</td>
<td>119</td>
<td>174</td>
<td>1.7333</td>
<td>119</td>
<td>174</td>
<td>1.2066</td>
</tr>
<tr>
<td>120</td>
<td>177</td>
<td>2.4018</td>
<td>120</td>
<td>177</td>
<td>1.5071</td>
<td>120</td>
<td>177</td>
<td>-0.087184</td>
</tr>
</tbody>
</table>
Appendix C Jump Diffusion Model Fitting

Fig E.1. Jump diffusion model fitting for water proton translational motion at 20% hydration level.
Fig E.2. Jump diffusion model fitting for water proton translational motion at 5% hydration level
VITA

BINGXIN FAN

EDUCATION

Pennsylvania State University (University Park)  Ph.D. in Chemical Engineering  Aug 2010 - May 2015

University of Minnesota (Minneapolis, MN)  B.S. in Chemical Engineering  Aug 2008 - May 2010

Michigan Technological University (Houghton, MI)  Undergraduate student in Chemical Engineering and Mathematics  Aug 2006 - May 2008

PUBLICATIONS


SELECTED PRESENTATIONS

- XIIIth International Cell Wall Meeting July 2013 (Nantes, France)
  Coarse-grained simulation of solvated Iβ cellulose microfibril
  Session: Advanced Understanding of CW Structure, Biosynthesis & Function

- American Physical Society March Meeting 2013 (Baltimore, MD)
  Coarse-grained simulation of solvated Iβ cellulose microfibril
  Session: Multiscale Modeling--Coarse-graining in Space and Time

- 243rd American Chemical Society National Meeting March 2012 (San Diego, CA)
  Coarse-grained simulation of cellulose microfibrils
  Session: Biosynthesis, Structure, and Interactions of Native Cellulose